Financial Misreporting:
Hiding in the Shadows or in Plain Sight?

Delphine Samuels
dsamuels@mit.edu
Sloan School of Management
MIT

Daniel J. Taylor
dtayl@wharton.upenn.edu
The Wharton School
University of Pennsylvania

Robert E. Verrecchia
verrecch@wharton.upenn.edu
The Wharton School
University of Pennsylvania

Draft: April 2018

Abstract:

This paper examines how the quality of a firm's information environment influences a manager's subsequent decision to misreport. The conventional wisdom is that high-quality information facilitates monitoring and increases the cost of misreporting, suggesting a negative relation between the quality of the information environment and misreporting. However, high-quality information also increases the weight that investors place on earnings in valuing the firm. This in turn increases the benefit of misreporting, suggesting a positive relation. We formalize these two countervailing forces—"monitoring" and "valuation"—in the context of a parsimonious model of misreporting. We show that the combination of these two forces leads to a unimodal relation. Specifically, as the quality of the information environment improves, misreporting first increases, reaches an inflection point, and then decreases. We find evidence of such a relation in multiple empirical measures of misreporting. Misreporting is greatest in a medium-quality environment and least in *both* high- and low-quality environments.

We thank John Core, Vivian Fang, Paul Fischer, Mirko Heinle, Rick Lambert, Rodrigo Verdi, and seminar participants at MIT and Wharton for helpful comments. We thank our respective schools for financial support.

1. Introduction

Understanding the motives for misreporting is of paramount interest to investors, regulators, and practitioners. Perhaps for this reason, a vast academic literature examines the causes of misreporting and the various mechanisms that can mitigate it. Many studies in this literature explicitly link the misreporting decision to the cost-benefit tradeoff facing the manager, and seek to provide insight on settings where managers will be more (or less) likely to misreport. In this paper, we examine how the "valuation" and "monitoring" roles of information alter the cost-benefit tradeoff facing the manager, and influence the manager's subsequent decision to misreport.

The conventional wisdom is that a "high-quality" information environment reduces the incidence of misreporting. For example, in their review of the corporate governance literature, Armstrong, Guay, and Weber (2010) discuss how a high-quality information environment facilitates monitoring and increases the manager's expected cost of misreporting. Indeed, the notion that high-quality information reduces misreporting features prominently in SEC rulemaking (e.g., Clayton, 2017). Broadly speaking, the conventional wisdom arises from the "monitoring role" of information: high-quality information increases the *expected cost* of misreporting. If the expected cost is higher, intuition suggests that the incidence of misreporting should be *lower*.

However, this intuition is incomplete. High-quality information also increases the *expected benefit* of misreporting. It is widely accepted that the price response to each dollar of reported earnings, or ERC, is larger in a high-quality information environment (e.g., Collins and Kothari, 1989; Teoh and Wong, 1993). The larger the ERC, the greater the valuation benefit from a dollar of inflated earnings. As a result, the "valuation role" of information suggests that a high-quality information environment increases the manager's expected benefit from misreporting. If the expected benefit is higher, intuition suggests that the incidence of misreporting should be *higher* (e.g., Ferri, Zheng, and Zou, 2017). We show that these two roles introduce countervailing effects that result in misreporting being greatest in a medium-quality information environment, and least in both high-quality *and* low-quality environments.

We illustrate the intuition that underlies our empirical predictions using a parsimonious model of misreporting based on Fischer and Verrecchia (2000). As is standard in the literature, we assume a privately informed manager can misreport, or “bias,” reported earnings, and that there are a variety of different “types” of managers, where each type of manager has a different preference for bias. In deciding whether to bias earnings, the manager trades off the benefit and cost, and investors have rational expectations about the extent of bias and adjust price accordingly (e.g., Guttman, Kadan, and Kandel, 2006; Bertomeu, Darrough, and Xue, 2017).

We operationalize the notion of the “information environment” by assuming that investors are uncertain about the manager’s type, and that as the quality of the environment improves, investors know more about the manager and can better anticipate bias (e.g., Fischer and Verrecchia, 2000; Marinovic, Liang, and Varas, 2017; Bertomeu, Cheynel, Li, and Liang, 2018; Frankel and Kartik, 2018). While researchers have used the term “information environment” in a variety of different contexts, in our model it refers specifically to the precision of investors’ prior beliefs about the manager’s type. In our model, a high (low) quality information environment is one in which investors have more (less) information about the manager. We incorporate the valuation role of information by assuming that the manager cares about stock price; and incorporate the monitoring role of information by assuming that the cost of bias increases with the quality of investors’ information about the manager’s type. In effect, we assume that the cost of bias increases with investors' ability to anticipate bias. The latter assumption is the distinguishing feature of our model relative to prior work (e.g., Fischer and Verrecchia, 2000).

In the presence of these countervailing forces, we show that the relation between bias and the information environment is *unimodal*: there is a unique inflection point, below which the relation is positive, and above which the relation is negative (see e.g., Figure 1).[1] The intuition for this result is that starting in a low-quality environment, a lower expected benefit results in lower bias. However, as the information environment improves and investors know more about the manager, bias increases because the expected benefit increases. Eventually, bias peaks at a unique inflection point and then begins to decline, because in a high-quality environment the expected cost is higher. Consequently, the relation is neither unambiguously positive nor unambiguously negative: bias is greatest in a medium-quality environment and least in *both* high- and low-quality environments. We use the insights from this analysis to inform the design of our empirical tests.

A distinguishing feature of our analysis is that we predict that the relation between misreporting and the information environment takes a specific, non-linear functional form. While prior work often frames empirical predictions in terms of linear relations, our predictions focus on the shape of the relation. Our focus on the shape of the relation should mitigate concerns about omitted variables and reverse causality. For example, an alternative explanation for a unimodal relation would need to suggest an alternative theoretical construct that not only explains the relation between misreporting and the information environment, but also why the relation flips sign at approximately the same inflection point. While we cannot rule out such a possibility, it seems unlikely. In the economics literature, this empirical approach is known as "identification by functional form" (e.g., Lewbel, 2018).

[1] Formally, a function, $y = f(x)$, is unimodal if for some threshold value t, it is monotonically increasing for $x \leq t$ and monotonically decreasing for $x \geq t$.

We estimate the shape of the relation using multiple measures of the theoretical constructs and multiple distinct sets of tests. Following prior research, we focus on restatements due to intentional misrepresentation as a measure of misreporting (e.g., Hennes, Leone, and Miller, 2008; Armstrong, Larcker, Ormazabal, and Taylor, 2013; Fang, Huang, and Wang, 2017). Consistent with the extant theory, we measure "bias" in reported earnings using the difference between earnings as originally reported and earnings as restated. In subsequent analyses, we assess the robustness of our results to alternative measures of misreporting, including a binary indicator variable for whether there was a restatement due to intentional misrepresentation, and a binary indicator variable for whether the SEC issued an Accounting and Auditing Enforcement Release.

In principle, there are a variety of proxies we could use to measure the quality of the information environment. The theory is very broad, and is silent on the channel through which investors might learn about the manager. Consequently, we focus on three broad proxies common in the literature: analyst following, institutional investor following, and media coverage. While these proxies are broad, each of them provides some insight on the ability of investors to anticipate the manager's reporting decision (e.g., Dyck, Morse, and Zingales, 2010). Consistent with how we model the information environment, a large body of prior research indicates that each of these proxies is associated with both heightened ERCs and heightened monitoring (e.g., Bushee, 1998; Jiambalvo, Rajgopal, and Venkatachalam, 2002; Bushman, Piotroski, and Smith 2004; Piotroski and Roulstone, 2004; Miller, 2006; Bowen, Rajgopal, and Venkatachalam, 2008. These findings provide construct validity, and suggest that our proxies capture the countervailing forces that motivate our analysis. To ensure that we measure investors' information prior to the potentially manipulated earnings report, we lag our measures by one year relative to the restatement period.

To reduce the effect of measurement error, we use principal component analysis to extract the common component to all three proxies, and use the common component as our primary measure.

We begin our empirical analysis by graphically presenting the shape of the relation. In particular, we sort firms into quintiles based on measures of the information environment and present average values of misreporting for each quintile. We find that misreporting is monotonically increasing in the first three quintiles, peaks in the fourth quintile, and then declines sharply in the fifth quintile. We observe a similar pattern across all measures of the information environment (including each of analyst following, institutional following, and media coverage) and across all measures of misreporting (including the amount of bias, and indicator variables for restatements and AAERs).

Next, we conduct two sets of empirical tests. First, we estimate polynomial regression models that include both linear and 2^{nd}-order polynomial terms. If the relation is unimodal (i.e., first increases and then decreases), we expect the 2^{nd}-order polynomial term on our measure of the information environment to load incremental to the linear term, and for the coefficient to be negative. In estimating these specifications, we control for *ad hoc* unimodal relations between misreporting and our control variables (e.g., we control for the possibility of an atheoretic unimodal relation between misreporting and growth options). Consistent with our predictions, across all specifications, we find robust evidence of a negative coefficient on the 2^{nd}-order polynomial term.

Second, we estimate spline regression models that treat the shape of the relation as piecewise linear. Specifically, we estimate linear regressions on either side of a threshold and test whether the slope coefficient below (above) the threshold is positive (negative). One estimation challenge with these models is that, while the theory suggests that a threshold exists, it is silent on the location of the threshold within our sample. Accordingly, we use two different approaches to estimate these models. In the first approach, we use the quality of the average firm's information environment as a rough approximation of the threshold, and estimate different slope coefficients for observations below and above the threshold. Consistent with a unimodal relation, we find a positive (negative) relation for firms with below- (above-) average quality.

In the second approach, we use the multivariate adaptive regression spline method (MARS) to estimate simultaneously both the threshold that minimizes the mean-squared error, and the sign of the relation on either side of the threshold (e.g., Friedman, 1991). The advantage of this approach is that MARS can reliably identify not only the slope coefficients of the piecewise-linear function around the threshold, but also the point at which the function is piecewise linear. We find that the threshold is at the 61st percentile and, consistent with a unimodal relation, the slope to the left (right) of the threshold is positive (negative). Thus, for observations in the top two quintiles, our results suggest that small improvements in the information environment will *decrease* misreporting. However, for observations in the bottom three quintiles, our results suggest that small improvements will *increase* misreporting.

Our theoretical and empirical findings should be of interest to both academics and policy makers. A growing theory literature analyzes the motivations for misreporting. For example, Peng and Roell (2014), Marinovic and Varas (2016), Marinovic and Povel (2017) examine how misreporting relates to the structure of compensation contract. Guttman and Marinovic (2018) examine how misreporting relates to debt contracts. Beyer, Guttman, and Marinovic (2016) use a dynamic structural model to estimate the unobserved bias and noise in reported earnings. Fang, Huang, and Wang (2017) model the bias in reported earnings as a function of the noise in reported earnings. Bertomeu, Cheynel, Li, and Liang (2018) use a structural model to estimate the unobserved cost of misreporting. In contrast to these papers, we focus on how the "valuation" and "monitoring" roles of information alter the cost-benefit tradeoff facing the manager, and influence the manager's subsequent decision to misreport.

Moving beyond the theory literature, our findings have wide-ranging implications for the large empirical literature that examines the causes and consequences of misreporting, especially as it relates to information dissemination, investor attention, and information intermediaries (e.g., Bloomfield, 2002; Miller and Skinner, 2015; Blankespoor, deHaan, and Zhu, 2018). For example, our results suggest a unique and novel channel through which a variety of empirical constructs could presumably influence misreporting. If a given empirical construct (e.g., information dissemination or investor attention) increases the quality of the information environment, *ceteris paribus*, our results suggest that this construct will have a unimodal relation with misreporting. Insofar as a linear specification masks this heterogeneity (and potentially result in the absence of an average effect), we recommend future empirical research consider non-linear specifications. In this regard, our findings are also relevant to the literature on predicting accounting fraud (see Dechow, Ge, Larson, and Sloan, 2011 for a review). Our results suggests future research in this area may wish to consider non-linear predictions models and/or focus on testing prediction models in medium-quality information environments—a setting where theory and evidence suggests such behavior will be most pronounced and prediction models will have more power.

With respect to policymakers, our results suggest that policy interventions that improve investors' information will not necessarily reduce the incidence of misreporting. Instead, the effect of the policy intervention will depend on the firm's (or country's) pre-existing environment. In low-quality information environments (e.g., markets with weak disclosure regulations), small improvements in information quality increase the valuation benefit of misreporting more than the cost, resulting in greater misreporting. However, in high-quality environments (e.g., markets with strong disclosure regulations), small improvements increase the cost of misreporting more than the benefit, resulting in lower misreporting. This suggests that regulations that improve transparency can sometimes have unintended adverse consequences: the association between transparency and misreporting is not necessarily negative.

The remainder of our paper proceeds as follows. Section 2 formally develops our empirical predictions. Section 3 describes the sample and measurement choices used in our empirical tests. Section 4 describes the research design, results and robustness tests. Section 5 provides concluding remarks.

2. Hypothesis development

2.1 Overview

In this section, we illustrate the intuition that underlies our empirical predictions using a parsimonious model of misreporting based on Fischer and Verrecchia (2000), hereafter FV. As in FV, we consider a setting where a manager privately observes true earnings, but then exercises discretion over the extent to which reported earnings diverge from true earnings. We refer to the divergence between reported earnings and true earnings as “bias.” In the context of our model, the divergence between reported earnings and true earnings constitutes “misreporting” because the manager intentionally introduces bias to obfuscate the report.

As in FV, we assume that there is a continuum of different types of managers, and each type of manager has a different preference for bias. We operationalize the notion of “information environment” by assuming that investors are uncertain about the manager’s type. This uncertainty precludes shareholders from perfectly backing out the bias. As the quality of the information environment improves, investors know more about the manager, and can better anticipate the extent of bias.[2]

We begin by discussing the level of expected bias in the setting considered in FV: where the quality of information about the manager affects the benefit of bias but not the cost. FV show that expected bias is largest in high-quality environments, where investors know more about the manager. The intuition for this result is that when investors know more about the manager, the price response per-dollar of reported earnings (or ERC) is higher, and a higher ERC increases the manager's benefit from inflating earnings. We refer to this as the "valuation role" of information.

Ferri, Zheng, and Zou (2017) test the valuation role of information in the context of managerial compensation disclosures. They test whether compensation disclosure increases the ERC, and by virtue of increasing the ERC, also increases the level of accrual-based earnings management. They find strong evidence of the former, and no evidence of the latter. One potential explanation for these results is that compensation disclosures not only affect the benefit of bias through the valuation role (as in FV), but also the cost of bias through the monitoring role: the extension of FV we consider below.

We extend FV to a setting where the cost of bias is greater when investors know more about the manager. We motivate this assumption by appealing to a body of empirical work that suggests heightened legal penalties and enforcement in high-quality information environments (e.g., Bushman and Smith, 2001; Armstrong, Guay, and Weber, 2010; Leuz and Wysocki, 2016). We refer to this as the "monitoring role" of information. The notion that the cost of bias is larger when investors know more about the manager introduces a countervailing force that militates against misreporting in high-quality information environments.

---

[2] FV is but one example of a class of signaling model in which agents are heterogeneous along two dimensions: "natural action" and "gaming ability." See Frankel and Kartik (2018) for a broad discussion of this type of model.

2.2. Setup

Unless otherwise noted, the basic setup and notation follows FV, and all random variables are mean-zero and independently normally distributed. Henceforth, we use a ~ to denote a random variable. To simplify the exposition, we state formal results in the Appendix.

At the beginning of the first period, the manager privately observes a noisy signal of terminal cash flow: $\tilde{e} = \tilde{v} + \tilde{n}$. One can think of $\tilde{e}$ as true earnings, $\tilde{v}$ as terminal cash flow, and $\tilde{n}$ as the noise in earnings that results from a non-opportunistic application of accounting rules. Having observed the realization of true earnings, the manager then reports earnings to the capital market. We represent reported earnings by $r = e + b$ where b is the bias in reported earnings: b is a choice variable of the manager. After observing reported earnings, risk neutral investors trade the firm's shares in a perfectly competitive market and price is equal to expected future cash flow, $P = E[\tilde{v} \mid r]$.

In choosing the optimal bias, a risk neutral manager trades off the cost of bias with the benefit of bias. We represent the manager's preferences with the objective function:

$$\max_{b}\left(\tilde{x}P - \frac{1}{2}C(\pi_x)b^2\right). \tag{1}$$

The first term, $\tilde{x}P$, represents the benefit to bias through the share price, where $\tilde{x}$ is a random variable. The realization of $\tilde{x} = x$ is determined by nature prior to the first period, and is known only by the manager. One way to interpret $\tilde{x}$ is to suggest that there is a continuum of different manager "types," with each type having a different preference for bias. Investors are uncertain of the managers type, through $\tilde{x}$, but nonetheless anticipate that the average manager personally benefits from bias being positive (i.e., $\mu_x > 0$). Uncertainty about the manager's type prevents shareholders from perfectly backing out the bias. Henceforth, we use the notation π_x to represent the precision of $\tilde{x}$, i.e., $\pi_x = \frac{1}{\sigma_x^2}$. As such, π_x represents investors' information about the manager's type. Note that it is not relevant for our analysis whether investors are uncertain about the first term of the manager's objective function or the second term: Bertomeu, Cheynel, Li, and Liang (2018) show an equivalent representation is to scale the cost function by $\tilde{x}$.

The second term represents the cost of bias. Henceforth we refer to $C(\pi_x)$ simply as the "cost function" and assume it is strictly positive. One can think of the cost of bias as representing an amalgamation of the effort cost of bias, the probability of detection, and the pecuniary and non-pecuniary penalties to the manager in the event of detection. FV consider a circumstance where the cost function is a constant, i.e., $C(\pi_x) = c$. The distinguishing feature of our analysis is that we consider a setting where the cost function is increasing in π_x, i.e., $\frac{dC(\pi_x)}{d\pi_x} > 0$.

2.3 The benefit of bias

The Appendix shows that there is a unique linear equilibrium with price given by $P = \alpha + \beta(e+b)$. Here, β represents the earnings response coefficient, or ERC. The greater the ERC, the greater the price response to reported earnings $(e+b)$, and hence the greater the benefit of bias. Consistent with extant empirical work (e.g., Collins and Kothari, 1989; Teoh and Wong, 1993; Ferri, Zhang, and Zou, 2017), the Appendix shows that $\frac{\partial\beta}{d\pi_x} > 0$. This establishes the mechanism through which π_x increases the benefit of bias—by increasing the value-relevance of reported earnings. The greater the value-relevance of reported earnings, the greater the benefit to inflating reported earnings.

In a setting where π_x affects the benefit of bias *but not the cost*, i.e., $C(\pi_x) = c$, the model collapses to that in FV and the expression for expected bias is $E[b] = \frac{\beta}{c}\mu_x$. The numerator of this expression encapsulates the benefit of bias through the ERC, and the denominator encapsulates the cost of bias. Because only the numerator is increasing in π_x, $\frac{\partial E[b]}{d\pi_x} > 0$. This is the counterintuitive result of FV: expected bias is largest when investors know more about the manager.

2.4 The cost of bias

Next, we extend FV to incorporate the possibility that $\frac{dC(\pi_x)}{d\pi_x} > 0$. As before, expected bias is given by the ratio of benefit to cost, $E[b] = \frac{\beta}{C(\pi_x)}\mu_x$. However, in this setting, both the numerator *and* the denominator are increasing in π_x. It is straightforward to show that the sign of $\frac{\partial E[b]}{d\pi_x}$ is given by the sign of the expression:

$$\frac{\partial \beta}{d\pi_x}\frac{1}{\beta} - \frac{\partial C(\pi_x)}{d\pi_x}\frac{1}{C(\pi_x)}. \tag{2}$$

Eqn (2) has a simple economic interpretation. The first (second) term represents the growth rate in the benefit (cost) of bias for a one-unit increase in π_x. When the first term is greater than the second term, the growth rate in the benefit exceeds the growth rate in the cost and $\frac{\partial E[b]}{d\pi_x} > 0$. However, when the first term is less than the second term, the growth rate in the cost exceeds the growth rate in the benefit and $\frac{\partial E[b]}{d\pi_x} < 0$.

We show that the relation is unimodal for a wide variety of cost functions. That is, there exists a unique inflection point, denoted π_x^*, such that: (1) for all values of π_x below the inflection point ($\pi_x < \pi_x^*$), the growth rate in benefits exceeds the growth rate in costs and $\frac{\partial E[b]}{d\pi_x} > 0$; and (2) for all values of π_x above the inflection point ($\pi_x > \pi_x^*$), the growth rate in costs exceeds the growth rate in benefits and $\frac{\partial E[b]}{d\pi_x} < 0$. The Appendix derives two sufficiency conditions for $C(\pi_x)$ that guarantee unimodality, and shows that a variety of standard cost functions, including linear, convex, and concave functions, satisfy these conditions.

Figure 1 illustrates the relation between $E[b]$ and π_x for standard linear, convex, and concave cost functions. The dashed, red line plots the relation for a standard convex cost function: $C(\pi_x) = \omega + \rho\pi_x^2$, $\omega > 0$, $\rho > 0$. The solid, black line plots the relation for a standard linear cost function: $C(\pi_x) = \omega + \rho\pi_x$, $\omega > 0$, $\rho > 0$. The dotted, green line plots the shape of the relation for a standard concave cost function: $C(\pi_x) = \omega + \rho\sqrt{\pi_x}$, $\omega > 0$, $\rho > 0$. In each case, the relation is unimodal: $E[b]$ initially increases in π_x, reaches a unique inflection point, and then decreases. Having discussed the economic theory that motivates our predictions, next we examine whether our predictions are empirically descriptive.

3. Sample construction and variable measurement

3.1 Sample

We construct our sample using data from CRSP, Compustat, I/B/E/S, Thomson-Reuters, RavenPack, and Audit Analytics from 2004 to 2012. Our sample begins in 2004, when data on our measures of the information environment and misreporting first become available, and ends in

2012 when our measure of misreporting ends. Column (1) of Panel A of Table 1 shows that there are a total of 46,148 firm-year observations over this period on the merged CRSP/Compustat universe with non-missing income, total assets, and market value. After requiring additional information needed to construct the control variables used in our analysis (e.g., plant, property, and equipment, sales growth, debt, buy-and-hold returns over the fiscal year), column (2) shows that the sample drops to 41,831 firm-years. However, this sample still represents 90% of the total available Compustat/CRSP population, suggesting our sample is comprised of a wide variety of information environments. For example, column (3) suggests 24% of firm-years in our sample do not have any analyst following (1 – (31,677/41,831) = 24%).

A salient feature of our sample is the presence of substantial cross-sectional variation in our measures of the information environment. The greater the variation, the greater the power of our tests. Indeed, one concern with empirically testing the theoretical shape of a functional form is that the empirical sampling variation may not be sufficiently large to replicate the entirety of the theoretical shape. Consider how this concern might affect our empirical analysis. Figure 1 shows the shape of the relation predicted by theory. However, e*x ante*, we do not know where firms in our sample will fall on the *x*-axis. It could be that, within our sample, investors' have sufficiently high-quality information such that all firms fall to the right of the inflection point. In such a circumstance, we would only observe a portion of the theoretical shape. For example, if all of the firms in our sample fall to the right (left) of the inflection point, we would not observe an inflection point, and would only observe a negative (positive) relation. Thus, for our tests of the shape of the functional form to be meaningful, we need to maximize sampling variation in our measures of the information environment: we need to observe firms on both sides of the inflection point. This requires having as broad a sample as possible. In this regard, our empirical tests are joint tests of a unimodal shape and a sufficiently broad sample that we observe firms on both sides of the inflection point.

Panel B of Table 1 presents descriptive statistics for the variables used in our analysis. Variable definitions follow Armstrong, Larcker, Ormazabal, and Taylor (2013). *$Assets* is the dollar value of total assets (in millions). *$Sales* is the dollar value of sales (in millions). *Acquisition* is an indicator variable for whether an acquisition accounts for 20% or more of total sales. *Capital* is net plant, property, and equipment scaled by total assets. *Financing* is the total debt and equity issuance scaled by total assets. *FirmAge* is the number of years the firm appears on Compustat. *Intangibles* is the ratio of research and development and advertising expense to sales. *InterestCov* is the ratio of interest expense to net income. *Leverage* is long term debt plus short term debt, scaled by total assets. *MB* is market value of equity plus book value of liabilities divided by book value of assets. *NAnalyst* is the number of analysts with one-year ahead earnings forecasts on I/B/E/S as of the end of the fiscal year. *NInstit* is the number of institutional owners listed on Thomson Reuters as of the end of the fiscal year. *NMedia* is the number of news releases about the firm over the fiscal year on RavenPack Analytics.[3] *Returns* is the buy-and-hold return over the fiscal year. *ROA* is income before extraordinary items scaled by total assets. *SalesGrowth* is the change in sales scaled by prior-period sales. *Size* is the natural logarithm of total assets. These variables are calculated net of any restatements and winsorized at the 1st/99th percentiles.

Consistent with a wide range of firms and information environments, Panel B suggests the interquartile range of total assets (sales) is $146 million to $2.67 billion ($68 million to $1.6 billion). The interquartile range is 1 to 8 analysts, 23 to 157 institutions, and 9 to 27 media articles.

[3] To ensure news articles pertain to the firm, we include only articles with RavenPack relevance score of 100 (out of 100). Results are robust to including articles with relevance scores above 75.

Panel B also suggests the average firm has double-digit sales growth (*SalesGrowth*) of 15%, earned a loss (average *ROA* of –0.02), and has an annual return (*Return*) of 9.1%.

3.2 Measurement of key variables

3.2.1 Measures of the information environment

Our tests require an empirical measure of the quality of the information environment. In principle, there are a variety of proxies that could be used: while our model provides specific guidance on how the quality of investors' information about the manager might manifest in the manager's cost-benefit tradeoff, it is silent on the source of investors' information. If anything, we argue that prior empirical work tends to interpret the parameter π_x in Fischer and Verrecchia (2000) too narrowly; as speaking to the quality of information about pecuniary compensation rather than the quality of information about the manager more broadly (e.g., Fang, Huang, and Wang, 2017; Ferri, Zhang, and Zhu, 2017). While managerial compensation contracts are undoubtedly one source of information about the manager's type, they represent only a small fraction of the total available information. For example, investors might glean information about the manager from an analysis of previous disclosures, media articles, analyst reports, or acquire information from other intermediaries. In this regard, the theory applies broadly and is not specific to a certain information channel. In effect, the theory only requires that the information is present prior to observing the realization of current period earnings.

Given that the theory applies broadly, we consider several popular proxies in the literature: analyst following, institutional investor following, and media coverage. Importantly, prior research indicates that each of these proxies is a noisy measure of the extent to which shareholders can anticipate managerial actions (and thus better understand the manager's type). One of the key advantages of these proxies is that prior research suggests each proxy is related both to greater value-relevance of reported earnings and to greater monitoring, and thus captures both the "valuation" and "monitoring" roles of information that motivate our analysis. For example, the literature consistently finds that analysts not only provide valuable information to investors and enhance the value-relevance of earnings, but also perform an important monitoring role that can help deter misreporting (e.g., Dyck, Morse, and Zingales, 2006; Yu, 2008; Bradshaw, Ertimur, and O'Brien). With fiduciary responsibilities towards their clients, institutional investors play important roles in impounding earnings information into prices (Jiambalvo, Rajgopal, and Venkatachalam, 2002; Piotroski and Roulstone, 2004) and in monitoring managers (e.g., Bushee, 1998; Bowen, Rajgopal, and Venkatachalam, 2008). Similar to analysts and institutions, the recent literature also suggests that media coverage can play a dual role: it improves the value-relevance of financial information either through dissemination or content creation, and facilitates monitoring and deterrence (e.g., Bloomfield, 2002; Miller and Skinner, 2015; Blankespoor, deHaan, and Zhu, 2018). The findings in prior work establish construct validity.

To reduce the effect of measurement error, we use principal component analysis to extract the common component to analyst following, institutional following, and media coverage. We use the common component to measure the quality of the information environment (*InfoQual*). In addition, to be faithful to the underlying theory, we lag our empirical measures one year relative to the potential period of misreporting. This ensures we measure the quality of information prior to current period earnings.

Table 2 presents the results from our principal component analysis. Variables are normalized to be mean 0 with standard deviation 1 prior to this analysis. Panel A shows that the first principal component explains 66.3% of the variation in the three variables and has an eigenvalue near two. The first principal component loads positively on all three variables (loadings of 0.413, 0.445, and 0.366 on *NAnalyst*, *NInstit*, and *NMedia* respectively) and is the only component with an eigenvalue greater than one.

Panel B reports descriptive statistics for our measure. The mean and median values of *InfoQual* are 0.00 and –0.280 respectively, suggesting a right skew toward high-quality information environments.[4] Panel C shows that *InfoQual* is highly correlated with each of the underlying components. *InfoQual* has a spearman (pearson) correlation of 0.84 (0.87) with analyst following, 0.91 (0.90) with institutional following, and 0.74 (0.76) with media coverage.

Our model explicitly predicts that the earnings response coefficient (ERC) is larger in high-quality information environments. Indeed, our theoretical predictions hinge on this result (see Section 2.3). Accordingly, we validate our measure by estimating its relation to the earnings response coefficient (ERC). If we do not find a positive relation, this would suggest that the empirical measure does not adequately capture the theoretical construct of interest. Although prior work has effectively shown that the ERC is increasing in the three underlying components of our measure (e.g., analyst following, institutional following, and media coverage), and Table 2 suggests our measure is highly correlated with these components, we nonetheless conduct the validation test for the sake of completeness.

Table 3 presents results. Panel A presents descriptive statistics for variables used in our analysis. *BHAR* is the market-adjusted buy and hold return over the twelve months ending three months after the fiscal year end. *Surprise* is the forecast error from a random walk model of annual earnings scaled by beginning of period price.[5] *Beta* is the slope coefficient from a single factor market model. *Ln(MV)* is the natural log of market value. *BM* is the book-to-market ratio. To

[4] The mean of a linear combination of normalized variables is zero by construction.
[5] We use a random walk model as opposed to the analyst consensus forecast because a substantial portion of our sample (24%) is missing analyst forecasts.

ensure that these variables reflect the information available to investors at the time, we compute all measures using unrestated values (i.e., values as initially reported). This research design choice is consistent with the ERC featured in our model, which is measured relative to reported earnings that potentially include bias. Variables are winsorized at the $1^{st}/99^{th}$ percentiles. Panel A suggests the average market-adjusted return is 3.1%, and the average earnings surprise is 1.7% of beginning of period price.

Panel B presents regression results. Column (1) presents the results from estimating the relation between the ERC and the quality of the information environment. We find a strong positive relation (*Surprise*InfoQual*, *t*-stat of 2.45). Column (2) presents results after controlling for the relation between the ERC and firm risk (*Beta*), size (*LnMV*), and growth (*BM*). Consistent with prior research (e.g., Collins and Kothari, 1989), we find riskier firms have a larger ERC (*Surprise*Beta, t*-stat 4.67), larger firms have a lower ERC (*Surprise*Ln(MV), t*-stat –4.08), and low growth firms have a lower ERC (*Surprise*BM*, *t*-stat –4.08). We continue to find a positive and incremental relation between the ERC and the information environment (*Surprise*InfoQual*, *t*-stat of 2.91).

Column (3) presents results from a similar specification as column (2), except that all of the independent variables are transformed into scaled quintile ranks that range from 0 to 1. Using ranks ensures that our independent variables are of similar scale, minimizes the effects of outliers, and allows for a meaningful comparison of the relative economic significance of each variable (i.e., each coefficient measures the difference in returns between the top and bottom quintile, *ceteris paribus*). We continue to find a significantly positive coefficient on *Surprise*InfoQual* (coeff. 0.312, *t*-stat of 4.47). In addition, the relation is similar in economic magnitude to firm risk (*Surprise*Beta* coeff. 0.326) and twice that of firm growth (*Surprise*BM* coeff. –0.184). We conclude that our empirical measure adequately captures the quality of the information environment.

3.2.2 Measures of financial misreporting

In our model, financial misreporting or "bias" refers to the difference between reported earnings and true earnings. In an effort to adhere closely to the theory, we measure misreporting using the difference between earnings as originally reported and earnings as restated, and focus on restatements due to intentional misrepresentation. We consider restatements classified by Audit Analytics as resulting from fraud or an SEC investigation to be intentional misrepresentations (e.g., Hennes, Leone and Miller, 2008; Armstrong, Larcker, Ormazabal, and Taylor, 2013).[6] By focusing on restatements due to misrepresentation, we can precisely measure bias in a circumstance where the cost of bias (e.g., probability of detection) is larger in high quality information environments.

We obtain from Audit Analytics a dataset of restatements classified as resulting from fraud or an SEC investigation. The database tracks Non-Reliance Restatements announced on Form 8-K between 2004 and 2016, and provides information on the date of the announcement, the period to which the restatement applies, and the effect of the restatement on net income.[7] The database contains information on 1,018 restatements due to intentional misrepresentations between 2004 and 2016.[8]

Panel A of Table 4 presents descriptive statistics on this database. Panel A shows that the restatements cover a total of 2,620 affected years and account for $41 billion in overstated net income. Notably, Panel A suggests the number of restatement announcements is heightened in the

[6] The Audit Analytics database provides us with a broader sample than the GAO database used in prior work (see Karpoff, Koester, Lee, and Martin, 2017 for a review of the different measures and databases).
[7] Firms begin disclosing 8-K item 4.02 "Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review" in August 2004.
[8] We remove from our sample intentional restatements missing data on the announcement date, the period for which the restatement applies, or the effect of the restatement on net income. By removing these observations from the sample we ensure we do not erroneously classify the observation as having zero misreporting.

period immediately following Sarbanes-Oxley and begins to approach zero by 2016. One reason for this trend is that it takes time for reporting issues to be detected and corrected (especially in the case of intentional misrepresentation). We take two steps to mitigate concerns that this time lag affects our results. First, we end our sample in 2012. In other words, we use data on restatement announcements through 2016 to measure misreporting through 2012. In doing so, we allow for a minimum of four years between the year of restatement and the year of the announcement. Second, we include year fixed-effects in our regressions to control for common time trends. As such, our estimated regression coefficients are within-year estimates.

Next, we match the Audit Analytics database to the sample in Table 1 based on the period that was restated. We define *RestateHLM* as an indicator variable equal to one if financial results in a given firm-year were restated, and *Bias* as the amount of restated earnings scaled by beginning total assets and expressed in basis points (i.e., 5.23% of assets is expressed 523). If earnings are not restated due to intentional misrepresentation, *Bias* is zero. In this manner, *Bias* measures the amount of the misreporting, not just the probability of misreporting. Because our theory speaks to the amount of misreporting, we use *Bias* as our primary measure of misreporting. In subsequent analyses we report results from repeating our tests measuring misreporting using both *RestateHLM* and an indicator variable for whether the firm-year was the subject of an SEC Accounting and Auditing Enforcement Release (*AAER*). We find similar results across all three measures (see Section 4.3 for more details).

Panel B of Table 4 presents the number of observations restated each year from 2004 to 2012 due to intentional misrepresentation, the probability of restatement (average *RestateHLM*), and the average values of *Bias*. Consistent with prior literature, Panel B shows that the probability of restatement is approximately equal to 0.01 in every year of our sample, and that average bias is consistently around 0.13 basis points of assets. The reason the average *Bias* is so small is that 99% of observations do not misreport, and thus have zero bias. Among those who do misreport (i.e., among firms with *RestateHLM* = 1) average *Bias* is 21 basis points.

4. Empirical tests and results

4.1 Univariate sorts

We begin our analysis by graphically presenting the shape of the relation between misreporting and the quality of the information environment. In Panel A of Table 5, we sort firms into quintiles based on the quality of the environment ($InfoQual_t$). For each quintile, we report the average value of misreporting in the subsequent year (i.e., $Bias_{t+1}$). Consistent with a unimodal relation, we find that misreporting is monotonically increasing in the first four quintiles, and declines sharply in the fifth quintile. Panel A also presents results from repeating this procedure separately for analyst following ($NAnalyst_t$), institutional investor following ($NInstit_t$), and media coverage ($NMedia_t$). We find a similar pattern across these measures as well. We present these results graphically in Figure 2. The shaded areas represent 95% confidence intervals. In Panel B, we repeat the procedure and present average values of an indicator variable for misreporting ($RestateHLM_{t+1}$) and find similar results. Finding a similar pattern across multiple measures of misreporting and multiple measures of the information environment suggests the evidence of a unimodal relation is not an artefact of the specification of our subsequent statistical tests: one can observe it in the raw data. Next, we conduct multiple formal statistical tests.

4.2 Polynomial regressions

We formally test for a unimodal relation between misreporting and the quality of the information environment by estimating polynomial regression models that include both linear and

2^{nd}-order polynomial terms (e.g., *InfoQual* and $InfoQual^2$). If the relation is unimodal, we expect the 2^{nd}-order polynomial term to be statistically significant and negative.

We estimate the following base model:

$$Bias_{t+1} = \alpha + \beta_1\, InfoQual^2_t + \beta_2\, InfoQual_t + \gamma\, Controls_t + \varepsilon_t. \qquad (3)$$

where $Bias_{t+1}$ is our measure of misreporting, *InfoQual* is our measure of the quality of the information environment, and *Controls* is a vector of control variables. All independent variables are normalized to be mean zero and standard deviation of one, are lagged one period relative to the measure of misreporting, and are net of any restatements. Standard errors are clustered by firm.[9]

Based on the prior literature examining restatements (e.g., Burns and Kedia, 2006; Efendi, Srivastava and Swanson, 2007; Armstrong, Larcker, Ormazabal and Taylor, 2013), we include the following control variables in our tests: firm size (*Size*), growth opportunities (*MB*), leverage (*Leverage*), past accounting performance (*ROA*), past stock performance (*Returns*), capital intensity (*Capital*), research and development and advertising expense (*Intangibles*), firm age (*FirmAge*), sales growth (*SalesGrowth*), debt and equity financing during the year (*Financing*), the size of acquisitions made over the year (*Acquisition*), and the firm's interest coverage (*InterestCov*). All variables are as defined in Table 1.

Table 6 presents results from estimation eqn. (3). Column (1) estimates a univariate version of eqn. (3) that does not include control variables. Column (2) estimates results including control variables and year fixed-effects, which control for common time trends. Column (3) presents results from additionally including industry fixed-effects that control for persistent differences

[9] Inferences are robust to clustering by industry and to two-way clustering by firm and year using bootstrapping to correct for few clusters (only 9 years of data).

across industries.[10] Consistent with a unimodal relation, we find a negative and significant coefficient on *InfoQual*^{2} across all our specifications (*t*-stat ranges between –3.72 and –4.57). In addition, the signs and significance of our controls (e.g., positive coefficients on *ROA*, *InterestCov*, *SalesGrowth*, and negative coefficients on *Capital* and *Financing*) are generally consistent with those in prior research (e.g., Armstrong, Larcker, Ormazabal, and Taylor, 2013).

We next conduct two sets of tests designed to assess the robustness of our results to controlling for other potential non-linearities in the data. First, we follow Fang, Huang, and Wang (2017) and control for a non-linear relation between intentional restatements and two measures of the industry rate of restatements due to error (*PctError* and *ErrorAmount*). *PctError* is the percentage of observations in the respective industry-year that were restated due to unintentional error (where unintentional errors are those restatements not classified as intentional manipulation). *ErrorAmount* is the average difference between net income and restated net income due to error in the respective industry-year, scaled by the standard deviation of net income among non-restating firms in the industry-year. These variables control for industry earnings quality.

Table 7 presents results. Columns (1) and (2) present results from estimating eqn. (3) including *PctError* and *PctError*^{2}, and columns (3) and (4) present results from estimation eqn. (3) including *ErrorAmount* and *ErrorAmount*^{2}. Consistent with the results in Fang, Huang, and Wang (2017), we find a negative and significant coefficient on *ErrorAmount*^{2} (*t*-stats –3.33 and –2.69), suggesting a non-linear relation between the number of restatements due to error in the industry and misreporting. More importantly, across all specification, we continue to find that the coefficient on *InfoQual*^{2} remains negative and strongly significant (*t*-stats range from –4.02 to –

[10] Industry classifications are based on Fama-French twelve industry groups.

3.72). Moreover, the coefficient magnitudes in Table 7 are virtually unchanged from those presented in Table 6. This suggests our empirical findings are distinct.

Second, we consider the possibility of heretofore undocumented, non-linear relations between misreporting and our control variables. While we are not aware of any theory that would predict a unimodal relation between misreporting and our control variables (e.g., a unimodal relation between misreporting and growth options), we nonetheless control for this possibility. Specifically, we estimate eqn. (3) after including both 1st- and 2nd-order polynomials of all our control variables: for a total of 24 linear and quadratic controls. Results appear in Table 8. For parsimony, we present only the specification with both industry and year fixed effects.

We note two findings. First, the coefficients on the non-linear control variables are generally insignificant. Only 2 of the 11 coefficients on the non-linear control variables are statistically significant at the 10% level or lower. The exceptions are $Size^2$, which is statistically negative (t-stat –2.40), and $Intangibles^2$, which is statistically positive (t-stat 2.45). Interestingly, if one views *Size* as a crude measure of the quality of the information environment and *Intangibles* as an inverse measure of quality of the information environment, then these results are completely consistent with our predictions. Second, the coefficient on $InfoQual^2$ remains negative and strongly significant (t-stat –2.74). These findings suggest that our results are not attributable to non-linear relations between misreporting and our control variables.

4.3 Alternative measures of misreporting

In this section, we assess the robustness of our results to two alternative measures of misreporting. First, we use a binary indicator variable for whether the respective firm-year was the subject of an SEC Accounting and Auditing Enforcement Release (*AAER*). We obtain data on

AAERs from the Center for Financial Reporting and Management. Using this database reduces our sample period to 2004-2011, and our sample size to 37,672 observations.

In Panel A of Table 9, we sort firms into quintiles based on $InfoQual_t$ and report the average likelihood of an AAER in the subsequent year ($AAER_{t+1}$). Consistent with our earlier results in Table 5, we find that the probability of an AAER is monotonically increasing in the first three quintiles, peaks in the fourth quintile, and then declines sharply in the fifth quintile. Next, we estimate eqn. (3) using $AAER_{t+1}$ as the dependent variable. Columns (1) and (2) of Panel B present results. Consistent with our predictions, and with our earlier findings, the coefficient on $InfoQual^2$ is negative and strongly significant (*t*-stats –3.20 and –3.28). This suggests a unimodal relation between the probability of an AAER and the quality of the information environment.

Second, we use a binary indicator variable for whether there was a restatement due to intentional misrepresentation (*RestateHLM*). This measure is directly related to *Bias*, except that it contains no information about the magnitude of the restatement. Nonetheless, it is perhaps the most common measure of misreporting used in the literature. Columns (3) and (4) of Panel B present results from estimating eqn. (3) using $RestateHLM_{t+1}$ as the dependent variable. Given the similarities between *RestateHLM* and *Bias* documented in Table 5, we expect to find similar results. Consistent with our predictions, and with our earlier findings, the coefficient on $InfoQual^2$ is negative and significant (*t*-stats –3.02 and –2.12). This suggests a unimodal relation between the probability of a restatement due to intentional misrepresentation and the quality of the information environment. Collectively, we interpret these findings as suggesting that our results generalize to other measures of misreporting—our results do not appear to be an artefact of measurement choices.

4.4 Alternative research design: Spline regressions

In this section, we aim to provide further evidence on the shape of the relation using spline regressions that treat the relation as piecewise linear. Specifically, we estimate the relation between misreporting (*Bias*) and the quality of the information environment (*InfoQual*) on either side of a threshold τ, and test whether the relation below (above) the threshold is positive (negative):

$$Bias_{t+1} = \alpha + \beta_1 \, (InfoQual - \tau < 0) + \beta_2 \, (InfoQual - \tau \geq 0) + \gamma \; Controls_t + \varepsilon_t, \qquad (4)$$

where all variables are as previously defined. We predict $\beta_1 < 0$ and $\beta_2 > 0$.

One challenge with estimating a piecewise linear form is that the theory does not specify the value of τ in our sample. Accordingly, we use two different approaches to estimate these models. In the first approach, we use the mean level of *InfoQual* as a rough approximation of the threshold (τ = 0), and estimate the relation for observations below and above the mean. Because the mean of *InfoQual* is zero (see Table 2), this approach is equivalent to estimating separate slopes for positive and negative values of *InfoQual*.

In the second approach, we use the multivariate adaptive regression spline method (MARS) to simultaneously estimate both the threshold that minimizes the mean-squared error (denoted τ^*), and the sign of the relation on either side of the threshold (e.g., Friedman, 1991). The advantage of this approach is that MARS can reliably identify not only the slope coefficients of the piecewise linear function around the threshold, but also the point at which the function is piecewise linear.

Table 10 presents our results. Columns (1) and (2) present results from estimating regressions where the threshold is defined at the mean of *InfoQual* (i.e., τ = 0). In these specifications, β_1 measures the relation for observations with below-mean values, and β_2 measures the relation for observations with above-mean values. Column (1) excludes control variables, and column (2) includes controls variables, year fixed effects and industry fixed effects. Consistent with a unimodal relation, in both specifications we find a positive slope below the mean (*InfoQual*

$- \tau < 0$, *t*-stat 3.46 and 3.26) and a negative slope above the mean (*InfoQual* $- \tau \geq 0$, *t*-stat –3.84 and –2.37). Notably the positive slope is between approximately three to four times steeper that the negative slope (e.g., 0.189 v. –0.046). This is consistent with the general shape of unimodal relation presented in Figure 1.

Columns (3) and (4) present results from using MARS to simultaneously estimate both τ^* and the corresponding slope coefficients. There are three noteworthy findings. First, the threshold that minimizes the mean-squared error is $\tau^* = 0.069$, which is at the 61st percentile of *InfoQual*. The location of τ^* at the 61st percentile is generally consistent with the univariate plots in Figure 2, which suggests the relation changes sign in the top two quintiles. Second, the location of τ^* is not affected by the inclusion of control variables or fixed effects. This is consistent with the notion that because we focus on estimating a specific non-linear functional form (i.e., "identification by functional form"), omitted variables are an unlikely explanation for our findings. Third, consistent with a unimodal relation, we continue to find a positive slope below the threshold (*InfoQual* $- \tau < 0$, *t*-stat 3.51 and 3.35) and a negative the slope above the threshold (*InfoQual* $- \tau \geq 0$, *t*-stat –3.75 and –2.23). Thus, for observations in the top two quintiles, our results suggest small improvements in the information environment will *decrease* misreporting. However, for observations in the bottom three quintiles, our results suggest small improvements in information environment will *increase* misreporting. Collectively, our tests provide robust evidence of a unimodal relation between misreporting and the quality of the information environment.

5. Conclusion

In this paper we examine how the quality of the firm's information environment influences the manager's subsequent decision to misreport. The conventional wisdom in the literature is that a high-quality information environment facilitates monitoring and increases the cost of misreporting, suggesting a negative relation. However, a high-quality information environment also increases the weight that investors place on earnings in valuing the firm. This in turn increases the valuation benefit of misreporting, suggesting a positive relation. We formalize these two countervailing forces—"monitoring" and "valuation"—in the context of a parsimonious model of misreporting, and use the insights from the model to guide our subsequent empirical tests.

We show that the theoretical relation misreporting and the quality of the information environment is *unimodal*: there is a unique inflection point, below which the relation is positive, and above which the relation is negative (see e.g., Figure 1). Consequently, the relation is neither unambiguously positive nor unambiguously negative: misreporting is greatest in a medium-quality environment and least in *both* high- and low-quality environments.

We test the novel prediction of our model—that the shape of relation is unimodal—using multiple empirical measures and two distinct research designs. In the first research design, we estimate polynomial regressions that include both linear and 2^{nd}-order polynomial terms. Consistent with a unimodal relation, we find robust evidence of a negative relation between misreporting and the 2^{nd}-order polynomial term for the quality of the information environment. In the second research design, we estimate spline regression models that treat the shape of the relation as piecewise linear around a threshold. Consistent with a unimodal relation, we find a positive relation below the threshold, and a negative relation above the threshold.

Our findings have wide-ranging implications for future researchers and policymakers. With respect to future researchers, our results suggest a novel channel through which a variety of empirical constructs could presumably influence misreporting. If a given empirical construct alters the quality of the information environment, *ceteris paribus*, our results suggest that this construct will have a non-linear, unimodal relation with misreporting. Insofar as a linear specification masks this heterogeneity (and potentially result in the absence of an average effect), we recommend future empirical research consider non-linear specifications. Our findings are also relevant to the empirical literature on predicting accounting fraud. Our results suggests future research in this area may wish to consider non-linear predictions models and/or focus on testing prediction models in medium-quality information environments—a setting where theory and evidence suggests such behavior will be most pronounced and prediction models will have more power.

With respect to policymakers, our results suggest that the effect of a policy intervention that improves investors' information will depend on the firm's pre-existing information environment. In low-quality environments, small improvements in information quality increase the valuation benefit of misreporting more than the cost, and result in greater misreporting. However, in high-quality environments, small improvements in information quality increase the cost of misreporting more than the benefit, and result in lower misreporting. This suggests that regulations that improve transparency can sometimes have unintended adverse consequences: the association between transparency and misreporting is not necessarily negative.

References

Armstrong, C., Guay, W., Weber, J., 2010. The role of information and financial reporting in corporate governance and debt contracting. Journal of Accounting and Economics 50, 179–234.

Armstrong, C., Larcker, D., Ormazabal, G., Taylor, D., 2013. The relation between equity incentives and misreporting: The role of risk-taking incentives. Journal of Financial Economics, 109(2), 327-350.

Bertomeu, J., Darrough, M., Xue, W., 2017. Optimal conservatism with earnings manipulation. Contemporary Accounting Research, 34(1), 252-284.

Bertomeu, J., Cheynel, E., Li, E. Liang, Y., 2018. Estimating the marginal cost of manipulation. Working paper.

Beyer, A., Guttman, I., Marinovic, I., 2016. Earnings management and earnings quality: Theory and evidence. Working paper.

Blankespoor, E., deHaan, E., Zhu, C., 2018. Capital market effects of media synthesis and dissemination: Evidence from robo-journalism. Review of Accounting Studies 23(1), 1-36.

Bloomfield, R., 2002. The incomplete revelation hypothesis and financial reporting. Accounting Horizons 16, 233–243.

Bowen, R., Rajgopal, S., Venkatachalam, M., 2008. Accounting discretion, corporate governance, and firm performance. Contemporary Accounting Research, 25(2), 351-405.

Bradshaw, M., Ertimur, Y., O'Brien, P., 2017. Financial analysts and their contribution to well-functioning capital markets. Foundation and Trends in Accounting 11, 119-191.

Burns, N., Kedia, S., 2006. The impact of performance-based compensation on misreporting. Journal of Financial Economics 79, 35–67.

Bushee, B., 1998. The influence of institutional investors on myopic R&D investment behavior. The Accounting Review 73(3), 305-333.

Bushee, B., Noe, C., 2000. Corporate disclosure practices, institutional investors, and stock return volatility. Journal of Accounting Research 38, 171-202.

Bushman, R., Smith, A., 2001. Financial Accounting Information and Corporate Governance. Journal of Accounting and Economics 32, 237-333.

Bushman, R., Piotroski, J., Smith, A., 2004. What determines corporate transparency? Journal of Accounting Research, 42(2), 207-252.

Clayton J., 2017. Governance and transparency at the commission and in our markets. Remarks at the PLI 49th Annual Institute on Securities Regulation - New York, N.Y.

Collins, D., Kothari, S. P., 1989. An analysis of intertemporal and cross-sectional determinants of earnings response coefficients. Journal of Accounting and Economics, 11(2-3), 143-181.

Dechow, P., Ge, W., Larson, C., Sloan, R., 2011. Predicting material accounting misstatements. Contemporary Accounting Research 28, 17-82.

Dyck, A., Morse, A., Zingales, L., 2010. Who blows the whistle on corporate fraud? The Journal of Finance, 65(6), 2213-2253.

Efendi, J., Srivastava, A., Swanson, E., 2007. Why do corporate managers misstate financial statements? The Role of option compensation and other factors. Journal of Financial Economics 85, 667–708.

Fang, V., Huang, A., Wang, W., 2017. Imperfect accounting and reporting bias. Journal of Accounting Research 55(4), 919-962.

Ferri, F., Zheng, R., Zou, Y., 2017. Uncertainty in managers' reporting objectives and investors' response to earnings reports: Evidence from the 2006 executive compensation disclosures. Journal of Accounting and Economics, forthcoming.

Fischer, P., Verrecchia, R., 2000. Reporting bias. The Accounting Review 75(2), 229-245.

Frankel, A., Kartik, N., 2017. Muddled Information. Journal of Political Economy, forthcoming.

Friedman, J., 1991. Adaptive spline networks. Advances in Neural Information Processing Systems, 675-683.

Guttman, I., Kadan, O., Kandel, E., 2006. A rational expectations theory of kinks in financial reporting. The Accounting Review 81(4), 811-848.

Guttman, I., Marinovic, I., 2018. Debt Contracts in the Presence of Performance Manipulation. Review of Accounting Studies, forthcoming.

Hennes, K., Leone, A., Miller, B., 2008. The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. The Accounting Review 83(6), 1487-1519.

Jiambalvo, J., Rajgopal, S., Venkatachalam, M., 2002. Institutional ownership and the extent to which stock prices reflect future earnings. Contemporary Accounting Research 19(1), 117-145.

Karpoff, J., Koester, A., Lee, S., Martin, G., 2017. Proxies and Databases in Financial Misconduct Research. The Accounting Review 92, 129–163.

Leuz, C., Wysocki, P., 2016. The economics of disclosure and financial reporting regulation: evidence and suggestions for future research. Journal of Accounting Research 54, 525-622.

Lewbel, A., 2018. The identification zoo: Meanings of identification in econometrics. Journal of Economic Literature, forthcoming.

Marinovic, I., Varas, F., 2016. CEO horizon, optimal pay duration, and the escalation of short-termism. working paper.

Marinovic, I., Povel, P., 2017. Competition for talent under performance manipulation. Journal of Accounting and Economics 64(1), 1-14.

Marinovic, I., Liang, Y., Vara, F., 2017. The credibility of financial reporting: A reputation-based approach. The Accounting Review, forthcoming.

Miller, G., 2006. The press as a watchdog for accounting fraud. Journal of Accounting Research 44(5), 1001-1033.

Miller, G. Skinner, D., 2015. The evolving disclosure landscape: How changes in technology, the media, and capital markets are affecting disclosure. Journal of Accounting Research, 53(2), 221-239.

Peng, L., Roell, A., 2014. Managerial incentives and stock price manipulation. Journal of Finance 69, 487-526.

Piotroski, J., Roulstone, D., 2004. The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. The Accounting Review 79(4), 1119-1151.

Teoh, S., Wong, T.J., 1993. Perceived auditor quality and the earnings response coefficient. The Accounting Review 68(2), 346-366.

Yu, F., 2008. Analyst coverage and earnings management. Journal of Financial Economics 88(2), 245-271.

Appendix

This Appendix presents the details underlying our model and derives the results.

A.1 Setup

The timeline of the model is as follows: (i) a risk-neutral manager privately observes true earnings, (ii) the manager issues a potentially biased report of earnings to the capital market, and (iii) investors price the firm's shares based on the report. To facilitate the discussion, henceforth we use a tilde (i.e., ~) to denote a random variable.

True earnings are given by $\tilde{e} = \tilde{v} + \tilde{n}$, where $\tilde{v}$ and $\tilde{n}$ are independently normally distributed with mean zero and variances σ_v^2 and σ_n^2, respectively. We assume these distributions are common knowledge. Having observed the realization of true earnings, $\tilde{e} = e$, the manager's report is given by $r = e + b$, where b represents the bias in reported earnings chosen by the manager. A risk-neutral, perfectly competitive market observes reported earnings and prices the firm's shares. Formally, price is the rational expectation of terminal value conditional on reported earnings, $P = E[\tilde{v} \mid r]$.

The manager chooses the bias in reported earnings in anticipation of share price. The manager's objective function is given by:

$$\max_b \left\{ \tilde{x}P - \frac{1}{2} C(\pi_x) b^2 \right\} \qquad \text{(A1)}$$

where $\tilde{x}$ is independently normally distributed with positive mean, $\mu_x > 0$, and variance σ_x^2. We use the notation π_x to represent the precision of $\tilde{x}$, i.e., $\pi_x = \frac{1}{\sigma_x^2}$. We assume the realization of $\tilde{x} = x$ is known only to the manager at the beginning of the period. We assume $C(\pi_x)$ is continuous, strictly positive, and increasing in π_x, such that $C(\pi_x) > 0$ and $\frac{\partial C(\pi_x)}{\partial \pi_x} > 0$ for all values of π_x. Higher values of $C(\pi_x)$ imply a higher marginal cost of bias. This is the only difference between FV and our model. In FV, $C(\pi_x) = c$, and the marginal cost of bias does not depend on π_x. For expositional convenience, henceforth we refer to $C(\pi_x)$ as "the cost function."

A.2 Equilibrium

An equilibrium to the model is described by a bias function, $b(e,x)$, and a pricing function $P(r)$ that satisfy three conditions. (1) The manager's choice of bias must solve his optimization problem given his conjecture about the pricing function. (2) The market price must equal expected firm value condition on the report and the market's conjecture about bias. (3) Both the conjectured pricing function and conjectured bias must be sustained in equilibrium. We show there is a unique linear equilibrium, where the bias function and price function are given by $b(e,x) = \lambda_e e + \lambda_x x + \delta$ and $P(r) = \alpha + \beta r$. Here, β represents the earnings response coefficient (ERC).

Equilibrium Bias Function. Combining the pricing function and the manager's objective function, the first-order condition implies that the optimal bias function is given by $b(e,x) = \frac{\hat{\beta}}{C(\pi_x)} x$, such that $\lambda_e = 0$, $\delta = 0$, and $\lambda_x = \frac{\hat{\beta}}{C(\pi_x)}$, where we a caret (i.e., ^) to denote a conjecture.

Equilibrium Pricing Function. Assuming the conjectured bias function of the form described above, market price of the firm is given by $P = E[\tilde{v} \mid r] = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_n^2 + \hat{\lambda}_x^2 \sigma_x^2}(r - \hat{\lambda}_x \mu_x)$, such that $\beta = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_n^2 + \hat{\lambda}_x^2 \sigma_x^2}$ and $\alpha = -\beta \hat{\lambda}_x \mu_x$.

Uniqueness. Assume all conjectures in the bias and pricing functions are correct. It is easy to verify the solutions for λ_e, δ, and α are unique functions of the pair $\{\beta, \lambda_x\}$. Thus, to prove the existence of a unique equilibrium we need to show that there exists a unique $\{\beta, \lambda_x\}$ that satisfies

$\beta = \frac{\sigma_v^2}{\sigma_v^2+\sigma_n^2+\lambda_x^2\sigma_x^2}$ and $\lambda_x = \frac{\beta}{C(\pi_x)}$. Substituting the latter into the former, we have $\beta = \frac{\sigma_v^2}{\sigma_v^2+\sigma_n^2+\left(\frac{\beta}{C(\pi_x)}\right)^2\sigma_x^2}$.

Thus, we need only show that this expression has a unique solution for β. Rearranging terms, yields the following third-order polynomial:

$$\beta^3 + \beta\left(\sigma_v^2 + \sigma_n^2\right)\pi_x C(\pi_x)^2 - \sigma_v^2 \pi_x C(\pi_x)^2 = 0. \tag{A2}$$

The interested reader will note that this expression mirrors the equilibrium condition in FV (eqn. (15) on p. 236 in FV), except that we have characterized the equilibrium condition in terms of π_x and the more general cost function, $C(\pi_x)$.

As in FV, the solution to eqn. (A2) is strictly positive, because $\beta \leq 0$ implies the expression is negative. Also, the solution to eqn. (A2) will be unique: for positive β, the expression is monotonically increasing in β and tends to infinity as β tends to infinity. Finally, the solution to eqn. (A2) is strictly less than $\frac{\sigma_v^2}{\sigma_v^2+\sigma_n^2}$, because $\beta \geq \frac{\sigma_v^2}{\sigma_v^2+\sigma_n^2}$ implies the expression is positive. This observations stem from the fact that $\beta = \frac{\sigma_v^2}{\sigma_v^2+\sigma_n^2} < 1$ in a circumstance where the manager is forced to report truthfully. Thus, when managers can bias reported earnings, the ERC is lower than in a circumstance where they are forced to report truthfully. This proves the conjectured linear equilibrium is unique, the conjectured bias function and price function are sustained, and that $\beta \in (0, \overline{\beta})$, where $\overline{\beta} = \frac{\sigma_v^2}{\sigma_v^2+\sigma_n^2}$. This latter feature of the equilibrium is intuitive (insofar as β is the weight that investors place on the manager's report) and will be important to establishing unimodal relation between $E[b]$ and π_x.

A.3 Comparative Static

We next examine the sign of the relation between expected bias, $E[b] = \frac{\beta}{C(\pi_x)}\mu_x$, and π_x.

To do so, we first need to derive the implicit solution for $\frac{\partial\beta}{\partial\pi_x}$ from eqn. (A2). Allowing for the fact that this requires a considerable amount of algebra, it is nonetheless straightforward to show

$$\frac{\partial\beta}{\partial\pi_x}=\frac{\beta^3\left(C(\pi_x)+2\pi_x\frac{\partial C(\pi_x)}{\partial\pi_x}\right)}{\left(3\beta^2+\left(\sigma_v^2+\sigma_n^2\right)\pi_x C(\pi_x)^2\right)\cdot\pi_x C(\pi_x)}>0. \quad \text{(A3)}$$

We next we solve for $\frac{\partial E[b]}{\partial\pi_x}$ and evaluate its sign,

$$\frac{\partial E[b]}{\partial\pi_x}=-\beta\frac{1}{C(\pi_x)^2}\frac{\partial C(\pi_x)}{\partial\pi_x}\mu_x+\frac{1}{C(\pi_x)}\frac{\partial\beta}{\partial\pi_x}\mu_x, \quad \text{(A4)}$$

We can further simplify the expression for the sign of the comparative static by substituting eqn. (A3) into equation eqn. (A4), and multiplying by β and $C(\pi_x)$. This yields

$$-\beta C'(\pi_x)+C(\pi_x)\frac{\partial\beta}{\partial\pi_x}=\frac{C(\pi_x)\cdot\left(\beta^3-\sigma_v^2\pi_x^2 C(\pi_x)\frac{\partial C(\pi_x)}{\partial\pi_x}\right)}{\pi_x\cdot\left(3\beta^2+\left(\sigma_v^2+\sigma_n^2\right)\pi_x C(\pi_x)^2\right)}. \quad \text{(A6)}$$

As both $C(\pi_x)$ and the denominator of eqn. (A6) are always positive, we need only consider the sign of the second term in the numerator:

$$\beta^3-\sigma_v^2\pi_x^2 C(\pi_x)\frac{\partial C(\pi_x)}{\partial\pi_x}. \quad \text{(A7)}$$

Using the equilibrium condition in eqn. (A2), one can show that eqn. (A7) reduces to

$$\sigma_v^2\pi_x C(\pi_x)^2-\beta\left(\sigma_v^2+\sigma_n^2\right)\pi_x C(\pi_x)^2-\sigma_v^2\pi_x^2 C(\pi_x)\frac{\partial C(\pi_x)}{\partial\pi_x} \quad \text{(A8)}$$

which is proportional to

$$\propto 1-\beta\left(\frac{\sigma_v^2+\sigma_n^2}{\sigma_v^2}\right)-\pi_x\frac{\frac{\partial C(\pi_x)}{\partial\pi_x}}{C(\pi_x)} \quad \text{(A9)}$$

Hence, the sign of $\frac{\partial E[b]}{\partial\pi_x}$, is given by eqn. (A9).

A.4 Unimodality

Unimodality implies that there exists an inflection point π_x^*, below which $\frac{\partial E[b]}{\partial \pi_x}$ is positive and above which $\frac{\partial E[b]}{\partial \pi_x}$ is negative. Now define $f(\pi_x)$ such that

$$f(\pi_x) = 1 - \beta\left(\frac{\sigma_v^2 + \sigma_n^2}{\sigma_v^2}\right) - \pi_x \frac{\frac{\partial C(\pi_x)}{\partial \pi_x}}{C(\pi_x)}. \tag{A10}$$

To establish conditions that ensure that $E[b]$ is unimodal as π_x increases, one needs to show initially $f(\pi_x) > 0$, then there exists some π_x^* such that $f(\pi_x^*) = 0$ (this is the inflection point), and then for every $\pi_x > \pi_x^*$ $f(\pi_x) < 0$. So note the following. First, $0 \le 1 - \beta\left(\frac{\sigma_v^2+\sigma_n^2}{\sigma_v^2}\right) \le 1$. Second, $1 - \beta\left(\frac{\sigma_v^2+\sigma_n^2}{\sigma_v^2}\right)$ is strictly decreasing in π_x, because β is strictly increasing in π_x (see eqn. (A3)). Third, $\lim_{\pi_x \to 0}\left(1 - \beta\left(\frac{\sigma_v^2+\sigma_n^2}{\sigma_v^2}\right)\right) \to 1$, because $\lim_{\pi_x \to 0} \beta \to 0$.

Thus, a sufficient condition that initially $f(\pi_x) > 0$ is

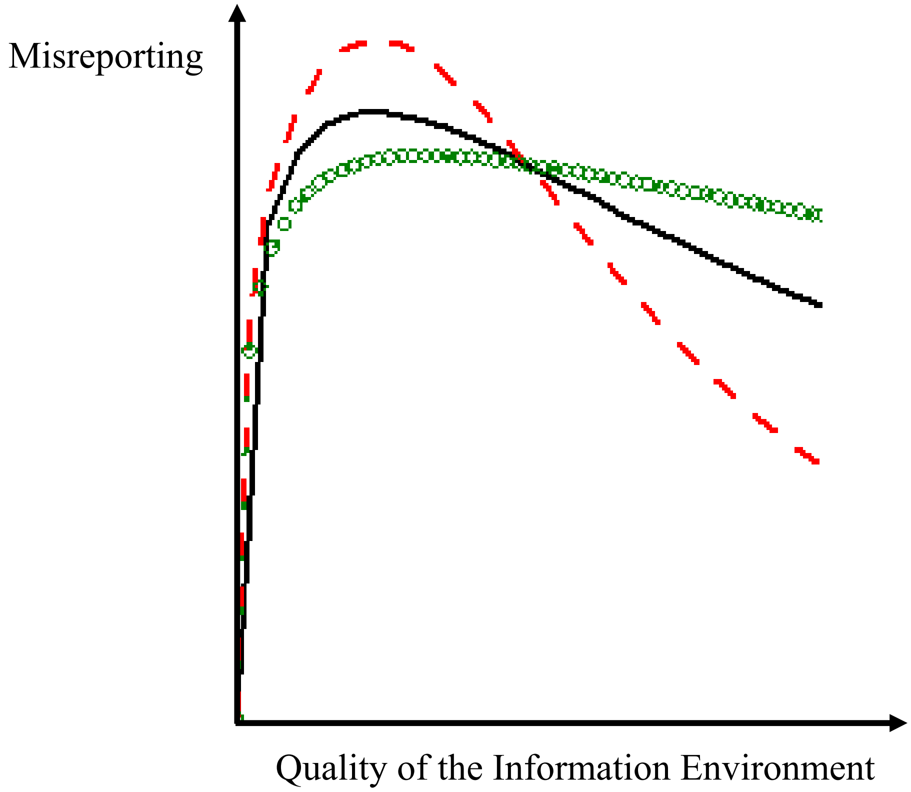
$$\lim_{\pi_x \to 0} \pi_x \frac{\frac{\partial C(\pi_x)}{\partial \pi_x}}{C(\pi_x)} < 1, \tag{A11}$$

and a sufficient condition that there exists some π_x^* such that $f(\pi_x^*) = 0$, and for every $\pi_x > \pi_x^*$ $f(\pi_x) < 0$ is

$$\frac{\partial}{\partial \pi_x}\left(\pi_x \frac{\frac{\partial C(\pi_x)}{\partial \pi_x}}{C(\pi_x)}\right) > 0. \tag{A12}$$

Eqns. (A11) and (A12) represent the two sufficient conditions that guarantee a unimodal relation between $E[b]$ and π_x. An interested reader can easily verify that linear cost functions of the form $C(\pi_x) = \omega + \rho\pi_x$, convex cost functions of the form $C(\pi_x) = \omega + \rho\pi_x^2$, and concave cost functions of the form $C(\pi_x) = \omega + \rho\sqrt{\pi_x}$, where $\omega > 0$ and $\rho > 0$, satisfy these two conditions.

Figure 1. Unimodal Relation

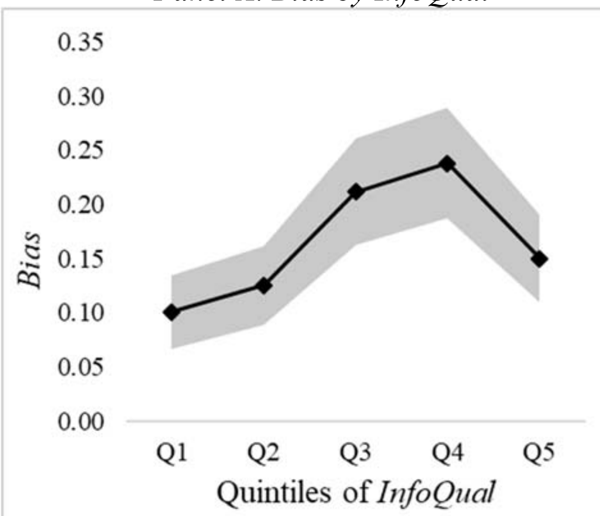


This figure plots E[*b*] as a function of π_x. E[*b*] appears on the *y*-axis. π_x appears on the x-axis. The dashed, red line illustrates the shape of the relation with a convex cost function (i.e., $C(\pi_x) = \omega + \rho\pi_x^2$). The solid, black line illustrates the shape of the relation with a linear cost function (i.e., $C(\pi_x) = \omega + \rho\pi_x$). The dotted, green line illustrates the shape of the relation with a concave cost function (i.e., $C(\pi_x) = \omega + \rho\sqrt{\pi_x}$). For the purposes of this figure, and without loss of generality, we set $\mu_x = 1$, $\sigma_v^2 = ½$, $\sigma_n^2 = ½$, $\omega = ½$, and $\rho = ½$.

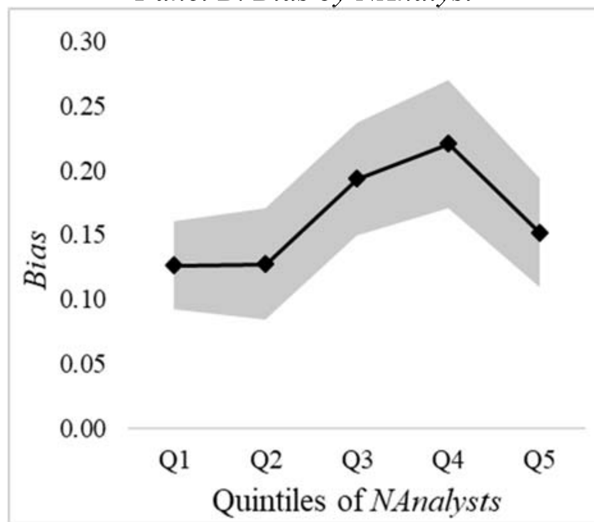
Figure 2. Misreporting and the Information Environment

This table presents the average values of $Bias_{t+1}$ by quintile of $InfoQual_t$ and its components. Panel A presents results for quintiles of *InfoQual*, Panel B for quintiles of *NAnalyst*, Panel C for quintiles of *NInstit*, and Panel D for quintiles of *NMedia*. The shaded areas represent 95% confidence intervals.

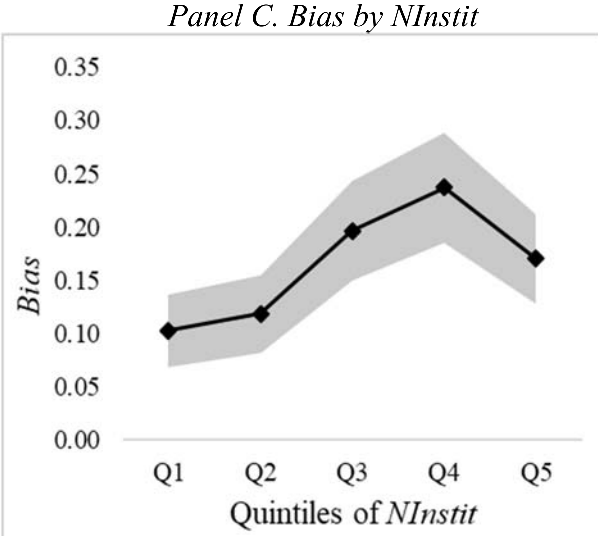
Panel A. Bias by InfoQual



Panel B. Bias by NAnalyst



Panel C. Bias by NInstit



Panel D. Bias by NMedia

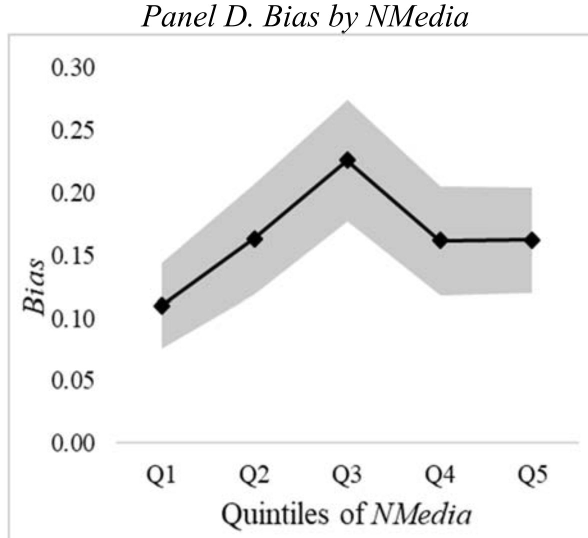


Table 1. Sample

This table describes the construction of our sample. Panel A presents the number of firms appearing in the CRSP/Compustat universe, the number of firms appearing in our sample, and the number of firms in our sample with non-zero analyst following, institutional investor following, and media coverage. Panel B presents descriptive statistics for the variables used in our primary analysis. *Acquisition* is an indicator variable for whether an acquisition accounts for 20% or more of total sales. *Capital* is net plant, property, and equipment scaled by total assets. *Financing* is the total debt and equity issuance scaled by total assets. *FirmAge* is the number of years the firm appears on Compustat. *Intangibles* is the ratio of research and development and advertising expense to sales. *InterestCov* is the ratio of interest expense to net income. *Leverage* is long term debt plus short term debt, scaled by total assets. *MB* is market value of equity plus book value of liabilities divided by book value of assets. *NInstit* is the number of institutional owners listed on Thomson Reuters as of the end of the fiscal year. *NMedia* is the the number of news releases about the firm over the fiscal year on RavenPack Analytics. *Returns* is the buy and hold return over the fiscal year. *ROA* is income before extraordinary items scaled by total assets. *SalesGrowth* is the change in sales scaled by prior-period sales. *Size* is the natural logarithm of total assets. All variables are winsorized at the 1st/99th percentiles and are net of any restatements. Our sample spans fiscal years 2004–2012 and contains 41,831 firm-years.

Panel A. Sample Composition by Year

	Number of firms				
Year	CRSP/Compustat Universe	Sample (requiring controls)	Sample w/ Analyst Coverage > 0	Sample w/ Institutional Coverage > 0	Sample w/ Media Coverage > 0
2004	5,654	5,105	3,560	5,091	4,391
2005	5,592	5,067	3,653	5,058	4,479
2006	5,512	4,969	3,665	4,964	4,478
2007	5,399	4,805	3,657	4,803	4,466
2008	5,135	4,695	3,573	4,685	4,394
2009	4,860	4,486	3,534	4,446	4,184
2010	4,750	4,313	3,434	4,239	4,023
2011	4,664	4,232	3,309	4,127	3,934
2012	4,582	4,159	3,292	4,029	3,877
Total	46,148	41,831	31,677	41,442	38,226

Table 1. Sample (cont'd)

Panel B. Descriptive Statistics

Variable	Mean	Std	25th	Median	75th
$*Assets*	7,002.831	25,269.653	146.518	643.245	2,677.122
$*Sales*	3,321.100	10,067.715	68.333	327.378	1,615.013
Acquisition	0.030	0.171	0.000	0.000	0.000
Capital	0.220	0.242	0.031	0.122	0.332
Financing	0.128	0.231	0.003	0.028	0.139
FirmAge	19.634	14.464	9.000	15.000	25.000
Intangibles	0.210	0.990	0.000	0.012	0.072
InterestCov	0.818	0.884	0.017	0.295	2.000
Leverage	0.204	0.209	0.021	0.153	0.316
MB	1.803	1.350	1.034	1.328	2.011
NAnalyst	5.139	6.131	1.000	3.000	8.000
NInstit	125.399	156.087	23.000	78.000	157.000
NMedia	21.545	20.854	9.000	17.000	27.000
Returns	0.091	0.541	–0.233	0.036	0.306
ROA	–0.020	0.207	–0.014	0.023	0.069
SalesGrowth	0.150	0.440	–0.030	0.079	0.218
Size	6.502	2.173	4.987	6.467	7.892

Table 2. Measuring Quality of the Information Environment

This table describes how we measure the quality of the information environment (*InfoQual*). *InfoQual* is the first principal component from a factor analysis of analyst following (*NAnalyst*), institutional investor following (*NInstit*), and media coverage (*NMedia*). All variables are standardized prior to the factor analysis. Panel A presents the principal component output. Panel B presents descriptive statistics for *InfoQual* (*InfoQual* = 0.413 * *NAnalyst* + 0.445 * *NInstit* + 0.366 *NMedia*). Panel C presents the correlation coefficients between *InfoQual* and its components. Spearman (Pearson) correlations appear above (below) the diagonal.

Panel A. Principal Component Output

Factor	Eigenvalue	Proportion of the variation explained	Cumulative Proportion of the variation explained	First Principal Component Weights	Variables
1st	1.989	66.3%	66.3%	0.413	*NAnalyst*
2nd	0.669	22.3%	88.6%	0.445	*NInstit*
3rd	0.341	11.4%	100.0%	0.366	*NMedia*

Panel B. Descriptive Statistics

Variable	Mean	Std	25th	Median	75th
InfoQual	0.00	1.000	–0.646	–0.280	0.276

Panel C. Correlation Matrix

	InfoQual	*NAnalyst*	*NInstit*	*NMedia*
InfoQual	1.00	0.84	0.91	0.74
NAnalyst	0.87	1.00	0.72	0.37
NInstit	0.90	0.69	1.00	0.53
NMedia	0.76	0.50	0.58	1.00

Table 3. Measuring Quality of the Information Environment—Validation

The table presents results from estimating the earnings response coefficient as a function of the quality of the information environment. Panel A presents descriptive statistics for the variables used in this test. *BHAR* is the market-adjusted buy and hold return over the twelve months ending three months after the fiscal year end. *Surprise* is the forecast error from a random walk model of annual earnings, scaled by price. *Beta* is the slope coefficient from a single factor market model. *Ln(MV)* is the natural log of market value, and *BM* is the book-to-market ratio. All variables are based on unrestated values and winsorized at the $1^{st}/99^{th}$ percentiles. Panel B presents results from estimating the earnings response coefficient. Columns (1) and (2) present results from OLS regressions. Column (3) presents results from OLS regressions using the quintile ranks of the independent variables scaled between zero and one. *t*–statistics appear in parentheses and are based on standard errors clustered by firm. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Sample of 41,379 observations from 2004 to 2012.

Panel A. Descriptive Statistics

Variable	Mean	Std	25^{th}	Median	75^{th}
BHAR	0.031	0.680	–0.276	–0.052	0.190
Surprise	0.017	0.234	–0.024	0.006	0.033
Beta	0.989	0.588	0.562	0.989	1.384
Ln(MV)	12.925	1.972	11.485	12.847	14.267
BM	0.671	0.645	0.301	0.531	0.858

Panel B. Earnings Response Coefficients

	Dependent Variable: *BHAR*		
			Ranks
Variable	(1)	(2)	(3)
Surprise	0.809^{***}	2.699^{***}	0.558^{***}
	(14.37)	(4.94)	(17.86)
*Surprise*InfoQual*	0.160^{**}	0.324^{***}	0.312^{***}
	(2.45)	(2.91)	(4.47)
*Surprise*Beta*	.	0.467^{***}	0.326^{***}
	.	(4.67)	(7.20)
*Surprise*Ln(MV)*	.	$–0.197^{***}$	$–0.807^{***}$
	.	(–4.08)	(–10.01)
*Surprise*BM*	.	$–0.105^{***}$	$–0.184^{***}$
	.	(–4.08)	(–5.02)
InfoQual	0.078^{***}	0.072^{***}	0.046^{*}
	(11.77)	(11.83)	(1.65)
Beta	0.118^{***}	0.102^{***}	$–0.047^{**}$
	(11.82)	(10.96)	(–2.19)
Ln(MV)	$–0.055^{***}$	$–0.049^{***}$	0.106^{***}
	(–11.91)	(–11.53)	(3.24)
BM	$–0.134^{***}$	$–0.138^{***}$	$–0.154^{***}$
	(–24.61)	(–23.12)	(–9.86)
F	215.85	220.02	388.18

Table 4. Measuring Misreporting

This table describes the construction of our measure of misreporting (*Bias*). Panel A describes the Audit Analytics Database used to contrast our measure. Audit Analytics begins tracking Non-Reliance Restatements filed on Form 8-K beginning in 2004. We focus on the subset of restatements related to fraud or SEC investigation. Column (1) of Panel A presents the number of restatements announced each year as a result of intentional misrepresentation. Column (2) presents the total number of firm-years that were restated. Column (3) presents the total effect on net income of those restatements (in millions), multiplied by negative one so that a positive number indicates a downward revision. Panel B presents descriptive statistics for our measures of misreporting after matching restatements to affected firm-years in our sample. Column (1) presents the total number of firm-years in our sample. Column (2) presents the number of firm-years that are subsequently restated. Columns (3) and (4) present average values of our measures of misreporting. *RestateHLM* is a binary measure of misreporting that equals one if the respective firm-year was eventually restated due to intentional misrepresentation, and zero otherwise, and *Bias* is a continuous measure of misreporting, and is calculated as the amount of restated earnings scaled by beginning total assets and expressed in basis points. The sample ends in 2012 to allow a minimum lag of four years between period of restatement and the restatement announcement (e.g., our measures of misreporting are calculated using restatements announced through 2016).

Panel A. Audit Analytics Database

All Restatements Due to Fraud or SEC Investigation (*Audit Analytics database*)			
	Number of Restatements Announced	Total Number of Affected Firm-Years	Total Effect on Net Income (in millions)
Announcement Year	(1)	(2)	(3)
2004	71	200	4,390.873
2005	170	508	16,870.516
2006	181	545	2,363.034
2007	107	257	1,169.040
2008	71	147	311.115
2009	77	169	1,629.941
2010	62	121	742.352
2011	75	159	10,366.056
2012	61	150	886.179
2013	37	97	419.571
2014	50	135	946.829
2015	38	90	1,079.932
2016	18	42	288.381
Total	1,018	2,620	41,463.820

Table 4. Measuring Misreporting (cont'd)

Panel B. Sample Measures of Misreporting by Year

Year	Sample observations (Table 1)	Restated observations due to misrepresentation	Avg *RestateHLM*	Avg *Bias*
2004	5,105	68	0.013	0.295
2005	5,067	52	0.010	0.225
2006	4,969	33	0.007	0.134
2007	4,805	31	0.007	0.138
2008	4,695	32	0.007	0.138
2009	4,486	29	0.007	0.124
2010	4,313	30	0.007	0.138
2011	4,232	29	0.007	0.142
2012	4,159	26	0.006	0.124
Total	41,831	330	0.008	0.165

Table 5. Misreporting and the Information Environment

This table presents average values of our measures of misreporting by quintile of the quality of the information environment. Panel A presents results for *Bias*. Panel B presents results for *RestateHLM*. Misreporting variables are measured in the subsequent year (i.e., in *t*+1).

Panel A. Bias by Quintile of the Quality of the Information Environment

Variable		Quintile 1	2	3	4	5
Quintiles of $InfoQual_t$	$Bias_{t+1}$	0.100	0.125	0.211	0.238	0.150
	Nobs	8,367	8,365	8,367	8,366	8,366
Quintiles of $NAnalyst_t$	$Bias_{t+1}$	0.126	0.127	0.193	0.220	0.152
	Nobs	10,154	6,107	9,184	8,345	8,041
Quintiles of $NInstit_t$	$Bias_{t+1}$	0.103	0.118	0.196	0.237	0.170
	Nobs	8,425	8,320	8,426	8,312	8,348
Quintiles of $NMedia_t$	$Bias_{t+1}$	0.110	0.163	0.226	0.161	0.162
	Nobs	8,695	7,896	8,880	7,877	8,482

Panel B. RestateHLM by Quintile of the Quality of the Information Environment

Variable		Quintile 1	2	3	4	5
Quintiles of $InfoQual_t$	$RestateHLM_{t+1}$	0.004	0.007	0.009	0.011	0.008
	Nobs	8,367	8,365	8,367	8,366	8,366
Quintiles of $NAnalyst_t$	$RestateHLM_{t+1}$	0.006	0.006	0.009	0.010	0.008
	Nobs	10,154	6,107	9,184	8,345	8,041
Quintiles of $NInstit_t$	$RestateHLM_{t+1}$	0.005	0.005	0.009	0.012	0.009
	Nobs	8,425	8,320	8,426	8,312	8,348
Quintiles of $NMedia_t$	$RestateHLM_{t+1}$	0.006	0.007	0.011	0.007	0.008
	Nobs	8,695	7,896	8,880	7,877	8,482

Table 6. Polynomial Regressions

This table presents results from estimating the relation between misreporting and the quality of the information environment using polynomial regressions. Column (1) presents results from estimating the relation between misreporting ($Bias_{t+1}$) and 1st- and 2nd- order polynomials of quality of the information environment ($InfoQual_t$). Column (2) presents results after controlling for firm characteristics and year fixed effects. Column (3) presents results after controlling for firm characteristics, year and industry fixed effects. All variables are as defined in Table 1. *t*–statistics appear in parentheses and are based on standard errors clustered by firm. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Sample of 41,831 firm-years from 2004 to 2012.

	Dependent variable: $Bias_{t+1}$		
Variable	(1)	(2)	(3)
$InfoQual^2$	–0.030***	–0.035***	–0.029***
	(–3.99)	(–4.57)	(–3.72)
InfoQual	0.059**	0.098***	0.064**
	(2.49)	(3.51)	(2.13)
Control variables			
Size	.	–0.039*	–0.007
	.	(–1.83)	(–0.27)
MB	.	0.004	0.009
	.	(0.24)	(0.56)
Leverage	.	0.010	0.013
	.	(0.50)	(0.71)
ROA	.	0.038**	0.033**
	.	(2.50)	(2.19)
Returns	.	0.010	0.006
	.	(0.70)	(0.42)
Capital	.	–0.023*	–0.056***
	.	(–1.70)	(–2.86)
Intangibles	.	–0.012	–0.007
	.	(–1.04)	(–0.54)
Financing	.	–0.019*	–0.021*
	.	(–1.77)	(–1.85)
Acquisition	.	–0.002	–0.006
	.	(–0.23)	(–0.58)
InterestCov	.	0.047***	0.041**
	.	(2.66)	(2.30)
FirmAge	.	–0.001	–0.011
	.	(–0.05)	(–0.54)
SalesGrowth	.	0.028**	0.028**
	.	(2.17)	(2.13)
Year effects	No	Yes	Yes
Industry effects	No	No	Yes
F	15.21	2.87	2.72

Table 7. Controlling for Industry Error Rates

This table presents results from estimating polynomial regression that control for the industry rate of restatements due to error (*IndustryError*). *PctError* is the percentage of observations restated due to error within the industry-year. *ErrorAmount* is the average difference between net income and restated net income due to error within the industry-year, scaled by the standard deviation of net income among non-restating firms in the industry-year (*ErrorAmount*). All other variables are as defined in Table 1. *t*–statistics appear in parentheses and are based on standard errors clustered by firm. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Sample of 41,831 firm-years from 2004 to 2012.

	Dependent Variable: $Bias_{t+1}$			
	IndustryError =*PctError*		*IndustryError* =*ErrorAmount*	
Variable	(1)	(2)	(3)	(4)
$InfoQual^2$	-0.029^{***}	-0.029^{***}	-0.030^{***}	-0.028^{***}
	(–3.85)	(–3.74)	(–4.02)	(–3.72)
InfoQual	0.053^{**}	0.065^{**}	0.060^{**}	0.064^{**}
	(2.27)	(2.15)	(2.56)	(2.12)
$IndustryError^2$	0.004	–0.000	-0.009^{***}	-0.006^{***}
	(0.52)	(–0.05)	(–3.33)	(–2.69)
IndustryError	0.065^{***}	0.076^{***}	0.121^{***}	0.079^{**}
	(5.27)	(3.35)	(2.98)	(2.20)
Controls	No	Yes	No	Yes
Year effects	No	Yes	No	Yes
Industry effects	No	Yes	No	Yes
F	12.27	2.70	9.74	2.70

Table 8. Controlling for Arbitrary Non-Linearities

This table presents results from estimating our polynomial regression specification including controls for arbitrary non-linearities in each of our control variables. The specification includes both linear control variables and 2nd-order polynomial transformations of all controls. *Acquisition* is an indicator variable and is collinear with the 2nd-order transformation. All variables are as defined in Table 1. *t*–statistics appear in parentheses and are based on standard errors clustered by firm. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Sample of 41,831 firm-years from 2004 to 2012.

	Dependent Variable: $Bias_{t+1}$		
$InfoQual^2$	–0.019***		
	(–2.74)		
InfoQual	0.046		
	(1.54)		
Non-Linear Controls		Linear Controls	
$Size^2$	–0.025**	*Size*	0.005
	(–2.40)		(0.20)
MB^2	–0.007	*MB*	0.045
	(–0.94)		(1.38)
$Leverage^2$	–0.005	*Leverage*	0.010
	(–0.41)		(0.32)
ROA^2	–0.000	*ROA*	0.014
	(–0.01)		(0.34)
$Returns^2$	–0.003	*Returns*	0.008
	(–0.44)		(0.44)
$Capital^2$	0.004	*Capital*	–0.066*
	(0.21)		(–1.84)
$Intangibles^2$	0.012**	*Intangibles*	–0.101**
	(2.45)		(–2.13)
$Financing^2$	–0.002	*Financing*	–0.016
	(–0.38)		(–0.60)
$Acquisition^2$	.	*Acquisition*	–0.001
	.		(–0.68)
$InterestCov^2$	–0.049	*InterestCov*	0.070**
	(–1.19)		(2.37)
$FirmAge^2$	–0.011	*FirmAge*	0.004
	(–0.75)		(0.15)
$SalesGrowth^2$	0.004	*SalesGrowth*	0.017
	(0.75)		(0.94)
Year effects	Yes		
Industry effects	Yes		
F	2.17		

Table 9. Alternative Measures of Misreporting

This table presents results from repeating our analysis using two alternative measures of misreporting. *AAER* is an indicator variable for whether the firm-year is subject to an SEC Accounting and Auditing Enforcement Release. *RestateHLM* is an indicator variable for whether the firm-year is restated due to fraud or SEC investigation. Panel A presents average $AAER_{t+1}$ by quintile of $InfoQual_t$. Panel B presents results from estimating polynomial regressions. In columns (1) and (2), the dependent variable is AAER_{t+1}. In columns (3) and (4), the dependent variable is $RestateHLM_{t+1}$. *t*–statistics appear in parentheses and are based on standard errors clustered by firm. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively. Data on AAERs are from the Center for Financial Reporting and Management and spans 2004 to 2011, and 37,672 firm-years. Data on *RestateHLM* is from Audit Analytics and spans 2002 to 2012, and 41,831 firm-years.

Panel A. AAER by Quintile of the Quality of the Information Environment

		Quintile				
Variable		1	2	3	4	5
Quintiles of $InfoQual_t$	$AAER_{t+1}$	0.002	0.003	0.006	0.006	0.004
	Nobs	7,750	7,621	7,564	7,450	7,287

Panel B. Polynomial Regressions

	Dependent variable:			
	$AAER_{t+1}$		$RestateHLM_{t+1}$	
	N = 37,672 from 2004 – 2011		N = 41,831 from 2004 – 2012	
Variable	(1)	(2)	(3)	(4)
$InfoQual^2$	–0.001***	–0.001***	–0.001***	–0.001**
	(–3.20)	(–3.28)	(–3.02)	(–2.12)
InfoQual	0.002**	0.002*	0.003***	0.002
	(2.47)	(1.93)	(2.68)	(1.06)
Controls	No	Yes	No	Yes
Year effects	No	Yes	No	Yes
Industry effects	No	Yes	No	Yes
F	5.11	2.32	4.62	2.32

Table 10. Alternative Design: Spline Regressions

This table presents results from estimating the relation between *Bias* and *InfoQual* using a spline regression with threshold τ:

$$Bias_t = \alpha + \beta_1\,(InfoQual - \tau < 0\,) + \beta_2\,(InfoQual - \tau \geq 0) + \gamma\; Controls_t + \varepsilon_t.$$

Columns (1) and (2) present results from estimating regressions where the threshold is defined at the mean of *InfoQual* (i.e., $\tau = 0$). In these specifications, β_1 estimates the relation for observations below the mean and β_2 estimates the relation for observations above the mean value. Columns (3) and (4) present results from using the multivariate adaptive regression spline (MARS) method to simultaneously estimate both the regression coefficients and the optimal threshold that minimizes the mean-squared error (i.e., $\tau = \tau^*$). All variables are as defined in Table 1. *t*–statistics appear in parentheses and are based on standard errors clustered by firm. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail), respectively.

	OLS Researcher specified threshold (τ)		MARS Optimal threshold selection algorithm (τ^*)	
	(1)	(2)	(3)	(4)
Threshold	$\tau = 0$	$\tau = 0$	$\tau^* = –0.069$	$\tau^* = –0.069$
InfoQual – τ < 0	0.179***	0.189***	0.195***	0.204***
	(3.46)	(3.26)	(3.51)	(3.35)
InfoQual – $\tau \geq 0$	–0.061***	–0.046**	–0.058***	–0.043**
	(–3.84)	(–2.37)	(–3.75)	(–2.23)
p-value: $\beta_1 - \beta_2 = 0$	<0.001	<0.001	<0.001	<0.001
Controls	No	Yes	No	Yes
Year Effects	No	Yes	No	Yes
Industry Effects	No	Yes	No	Yes
F	8.28	2.06	8.20	2.06