



How should we think about earnings quality? A discussion of “Earnings quality: Evidence from the field”[☆]



Mark W. Nelson^a, Douglas J. Skinner^{b,*}

^a S.C. Johnson Graduate School of Management, Cornell University, NY 14850, USA

^b University of Chicago Booth School of Business, IL 60637, USA

ARTICLE INFO

Available online 11 November 2013

JEL classification:

G30

M14

M41

Keywords:

Earnings quality

Earnings management

FASB

Survey methods

CFOs

ABSTRACT

Dichev, Graham, Harvey and Rajgopal (DGHR, in this issue) survey chief financial officers (CFOs) to elicit their views on earnings quality, broader trends in financial reporting, and the prevalence of earnings management. They provide some interesting insights on these issues. We discuss how CFOs' incentives in the financial reporting process are likely to affect what we can learn from them about earnings quality. We also discuss how DGHR's methodological choices regarding survey sample and question design affect their inferences, including what we can infer about the prevalence and magnitude of earnings management.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Dichev, Graham, Harvey and Rajgopal (hereafter DGHR, in this issue) continues a series of papers in accounting and finance that draw inferences about corporate financial policies from surveys of corporate chief financial officers (CFOs).¹ These papers provide many novel insights in a variety of areas. The goal of DGHR is to provide survey evidence on CFOs' views about earnings quality. To supplement and enrich these results, the authors also present results from one-on-one interviews from a small sample of CFOs and two accounting standard-setters.

We summarize and discuss a number of aspects of DGHR, including motivation, method, results, and inferences. We agree with the authors that the findings potentially inform policy and research discussions on earnings quality issues. Nevertheless, we also raise questions about the appropriateness of applying this approach to the earnings quality literature as well as about some of the inferences drawn by the authors.

[☆] This is an invited discussion of the paper by Dichev et al. (2013) based on our invited discussions of the paper at the 2012 JAE conference. We thank Rob Bloomfield, Scott Emmett, Michelle Hanlon, and Bob Libby for comments on prior versions.

* Corresponding author. Tel.: +1 7737027137.

E-mail addresses: mwn2@cornell.edu (M.W. Nelson), dskinner@chicagobooth.edu (D.J. Skinner).

¹ See Graham and Harvey (2001, 2002) on corporate finance, Graham et al. (2005, 2006) on financial reporting, Brav et al. (2005) on payout policy, and Ben-David et al. (2013) on managerial miscalibration. In accounting, many surveys have also elicited data from auditors; see, e.g., Bell et al. (2001), Gibbins and Newton (1994), Gibbins et al. (2001), Nelson et al. (2002), and Gibbins et al. (2005).

2. Contribution

2.1. Is there a single definition of earnings quality?

To open the paper, the authors state that “there are broad disagreements about how to define and measure it (earnings quality) (p. 2).” To support this statement, they list several candidate measures, including earnings persistence, predictability, asymmetric loss recognition, income increasing accruals, and the absolute value of accruals. Although these are certainly all earnings-related measures, it is not clear that the use of these measures reflects disagreement in the literature about what earnings quality means.

Like art, how one thinks about earnings quality is to some degree in the eye of the beholder: this term is primarily used in financial statement analysis—the decisions being made and the decision contexts are likely to vary. In a recent review on this topic, Dechow et al. (2010, p. 344) indicate that “(h)igher quality earnings provide more information about the features of a firm’s financial performance that are relevant to a *specific decision made by a specific decision maker* (emph. added)” and “(w)e reach no single conclusion on what earnings quality is because “quality” is contingent on the decision context.” Given the inherently context-specific nature of earnings quality, it is not surprising that earnings quality is not defined or measured in a uniform way in our literature. Therefore, DGHR should not be viewed as resolving a disagreement in the literature as much as contributing insights from CFOs with respect to this topic.

2.2. What can CFOs tell us about earnings quality?

Previous survey work in this genre produces novel insights beyond what is possible from traditional empirical archival research. For example, a great deal of empirical research in corporate finance attempts to provide a better understanding of the determinants of capital structure and payout policy. However, that research has made limited progress in answering some open questions in this literature, such as why industrial firms are not more heavily levered given the apparent tax advantages of debt and why dividends are such a persistent phenomenon in spite of the apparent advantages of stock repurchases. Evidence from large-sample surveys of CFOs provides researchers with fresh ideas about what CFOs actually do in practice and so helps to calibrate empirical research. Brav et al. (2005) provide evidence that managers often set dividend policy based on changes in nominal dividends-per-share, as opposed to following more conventional ways of thinking about dividend policy (such as payout ratios based on the Lintner model). This evidence is useful because it helps researchers develop and test alternative theories of payout policy—survey evidence can be used to inform and influence other methods of investigation, including large sample empirical work.

The reason these studies have been so informative is largely because CFOs (managers) are directly involved in the decisions of interest—we learn from managers about how and why they make the decisions they do. It is less obvious that managers’ views about earnings quality are of interest because it is not clear why they have any particular comparative advantage in determining a universal definition of earnings quality that can be used in the literature. Earnings quality is an amorphous idea that is used in a variety of financial statement analysis contexts, so perhaps the views of key users, such as analysts or investors, would be of more interest.

DGHR’s perspective is that managers’ views on earnings quality are of interest because managers make the financial reporting decisions that generate the earnings numbers that are of interest. But it is not clear that earnings quality is a construct that is of much import to them or that they have earnings quality in mind when they make financial reporting decisions. It is also possible and perhaps likely that managers’ particular views about earnings quality are affected by their financial reporting incentives, a point we discuss further in Section 3.

To the extent there is disagreement in the literature, it seems to reflect differences of opinion about whether earnings quality has mostly to do with earnings persistence (higher quality earnings are more persistent) or the extent to which earnings reflect underlying economic reality. For example, artificially smoothed earnings are likely to be more persistent (and so of high quality in one sense) but may not reflect the underlying volatility of the firm’s business (and so of low quality in another). In this situation, the more volatile time series could be seen to be of higher quality because it more faithfully reflects the underlying economics. However, if the artificially smoothed series better reflects the firm’s ability to generate long run economic value (that is, the smoothed earnings trend signals the firm’s long run ability to deliver increases in earnings in spite of shorter-run volatility in its business), it is less clear that the artificially smoothed series is of lower quality. A lot depends on management intent and the decision context of the user.

To the extent managers use their discretion in the financial reporting process to improve the signal value of earnings (that is, to increase informativeness about the underlying economics), management discretion can improve earnings quality. Moreover, the normal operation of the accruals process is such that it smooths out the volatility of cash flows and so results in a smoother and more informative measure of economic performance (e.g., Dechow, 1994; Dechow and Skinner, 2000). Conversely, to the extent that managers use their financial reporting discretion to distort the signal value of earnings (that is, they engage in opportunistic earnings management to obscure the economics), earnings quality is reduced. All of this means that it will be hard to reach definitive conclusions about how different earnings measures (such as accruals, discretionary accruals, the use of unusual items, the absolute value of accruals) relate to earnings quality without specifying something about the user and the decision context.

Put differently, DGHR view CFOs as a lens through which to identify the characteristics of high-quality earnings. However, like any lens, this one can add distortion. Without knowing the particular incentives of the CFO, it is difficult to

know whether her preferences for various characteristics of earnings are likely to increase or decrease earnings quality. Because we cannot back out that distortion, it is hard to interpret DGHR's results as revealing some "true" dimensions of earnings quality. We discuss this concern further in [Section 3](#).

2.3. Key insights from DGHR

With that background in mind, we believe there are some interesting and useful findings in the paper. There are three main sets of results.

First, CFOs indicate that high quality earnings are sustainable and persistent. This result is largely consistent with the literature, in the sense that financial statement analysis texts typically define earnings quality in terms of persistence (e.g., [Penman, 2013, Ch. 18](#)). It is comforting that CFOs also describe earnings quality in these terms. Of course, this result may also have something to do with managers' incentives to produce predictable earnings that meet earnings targets on an ongoing basis, so their responses may naturally be slanted in this direction.

The second set of results relates to the relation between accounting standards and earnings quality. These results mostly reflect CFOs' views about trends in financial reporting. The authors find that managers are concerned about certain trends they perceive in financial reporting, including the increasing number and complexity of rules, the gradual decline of the matching principle, and the increasing use of fair value accounting. Managers would prefer rules that reflected and evolved from practice as opposed to being mandated from "above," an idea that is also well-established in the academic literature ([Watts, 2006](#); [Ball, 2009](#)). The general thrust of the argument here seems to be that if managers generally use their financial reporting discretion to improve the signal value of earnings numbers (that is, to communicate their private information in an unbiased way), giving them more scope to do so through the financial reporting process will enhance earnings quality. The other side of this argument, and one typically heard from standard-setters, is that giving management more discretion also opens the door for distortive earnings management, a point most managers are unlikely to make.

In addition to managers' concerns about certain trends in financial reporting, they express concerns about what they see as adverse trends in auditing. Specifically, they express concerns about the quality of audit personnel as well as the extent to which responsibility for audit decisions are increasingly made by the national offices of the Big 4, as opposed to local partners working directly with CFOs. Once again, all of these results must be viewed as coming from CFOs who would prefer more latitude in financial reporting and closer relationships with auditors.

Another interesting finding from the paper relates to CFOs' use of external financial statement numbers for internal decision-making purposes. In general, we know relatively little about the interaction between firms' internal decision-making processes and external financial statements. The fact that managers say they use the external numbers for internal decision-making is therefore of some interest. Relatedly, the authors' findings as to managers' use of competitors' financial statements is also of interest in light of related empirical findings on this point (e.g., [Durnev and Mangan, 2009](#)).

The authors' third set of results has to do with earnings management, which is often linked to earnings quality. This is an area where we have significant concerns. The authors use point estimates based on survey responses to conclude that, in any given period, 20% of managers manage earnings and that the typical misrepresentation accounts for 10% of reported EPS. They further conclude that 60% of earnings management is income increasing while 40% is income decreasing. The authors argue that these results "perhaps for the first time in the literature" give us point estimates of the extent and magnitude of earnings management.

We view these claims as overly strong for a number of reasons. First, the results are based on a sample of 169 managers of public companies, so the methodological issues discussed in detail below come into play, particularly sample representativeness and precision of the estimates. Second, in framing the earnings management questions, the authors are careful to ask managers about earnings management in firms other than their own. This leads one to wonder how managers would know, in general, about earnings management in other firms and leads us back to our initial point about the survey method: survey results make most sense when managers are asked about decisions they actually make. It is not clear why managers have a comparative advantage over other capital market participants (auditors, analysts, independent board members) in knowing about earnings management—how, exactly, do managers know about earnings management at other firms? Perhaps they have direct knowledge of this through industry channels. Perhaps they are projecting their own behavior onto the population of other managers but making some unknown adjustments according to how they perceive themselves as differing from other managers. Perhaps they are just guessing.

We next delve more deeply into the specifics of DGHR's methods, including the strengths and weaknesses of the survey method, the appropriateness of DGHR's survey participants for earnings quality questions, and the specific questions that DGHR use to elicit data.

3. Method

3.1. Overview of survey methods

From a methodological standpoint it is useful to compare surveys to the archival and experimental methods that are more prevalent in empirical accounting research. Archival and experimental studies have complementary strengths and weaknesses ([Libby et al., 2002](#)), and surveys share some of the strengths and weaknesses of each of these other approaches.

Like archival studies, surveys provide a wealth of descriptive data that can provide useful insights. However, also like archival studies, surveys base inferences on measured (rather than manipulated) independent variables, so causal inferences must be made cautiously. Given that DGHR focus on description and exploration rather than on hypothesis testing, this disadvantage is not as important as it otherwise might be.

Like experiments, surveys do not require historical events to generate the data, and can elicit beliefs and opinions that are difficult or impossible to glean from archival analysis. However, like experimental studies, surveys are vulnerable to the potential for the data *collection* process to influence the data *generation* process, because the wording and topic of survey questions can influence responses. Also, the characteristics and representativeness of participants can influence the validity and generality of results. Finally, both surveys and experiments suffer from low post-data-collection flexibility. Once the data is gathered, it is very difficult to adjust the focus of the study without performing a substantially new study.

We first discuss the appropriateness and representativeness of DGHR's sample of survey participants. We then discuss the particular questions that DGHR pose to participants and the analyses they use to generate their key findings.

3.2. Survey participants

3.2.1. Sample appropriateness

One important consideration is whether a sample of CFOs is appropriate to answer the questions posed in the survey. For some of DGHR's survey questions, CFOs seem like a good match. For example, the data reported in Table 2 indicate CFOs' ratings of the importance of earnings for various uses, many of which are internal to the company (e.g., use in executive compensation contracts, use in negotiations with labor, use by the company's own managers). For those questions, CFOs are reporting what they do or what they observe others doing within their organizations, and so are well positioned to make that assessment.

CFOs are less appropriate respondents for other questions, as we discuss above. For example, Table 4 reports CFOs' views about various attributes of earnings quality and Table 8 reports CFOs' ratings of whether various GAAP policies are likely to produce high quality earnings. It is not clear that CFOs have a more valid perspective on these questions than do investors, creditors, standard setters or academics.

Moreover, CFOs' role in the financial reporting process could provide them with incentives that predictably bias their views of the dimensions and determinants of earnings quality. CFOs have their own objectives, which at times may work at cross-purposes to the interests of investors, creditors, etc. Consider a new hypothetical earnings measure which we label CFO-Preferred Earnings ("CFOPE"):

"Predictable and consistent earnings I can control with minimal hassle to placate my investors and creditors, without business shocks or one-time events getting everybody all stirred up and requiring lots of explanation, and without auditors interfering with the message I want to send."

While somewhat facetious, CFOPE captures some attributes of earnings that are likely to be more preferred by CFOs than investors. For example, CFOs might be more likely than investors to prefer earnings that contain fewer long-term estimates and that require fewer explanations (Table 4). CFOs also may be more likely to prefer that standard-setters emphasize matching, issue fewer rules and allow reporting choices to evolve from practice (Table 8), because these attributes give CFOs greater ability to manage the numbers.

Concerns about sample appropriateness also affect the authors' interpretation of interview responses which indicate that some CFOs chafe at detailed accounting guidance and at auditors who refer contentious issues to their firms' central technical specialists. DGHR report one interviewee as stating that "They (auditors) now are much more into the exact wording of something and the interpretation of it versus what's logical. Earlier you could work with your local accounting firm, your local partner and accomplish things. Now, pretty much everything goes up to their think tank at national" (p. 33). Prior research indicates that auditors are better able to resist earnings management pressure when they have a backstop provided by clear accounting standards that preclude management's preferred approach (Nelson et al., 2002; Ng and Tan, 2003). Perhaps such anecdotes illustrate that CFOs don't like to see CFOPE sacrificed in ways that, at least in some circumstances, could improve earnings quality by deterring earnings management. This might also explain why the two standard setters interviewed by the authors have views about trends in financial reporting that differ from those of the CFOs—one of the roles of accounting standards is to limit managers' flexibility in the financial reporting process and so to limit opportunities for earnings management, something that managers generally are unlikely to view positively whether they actively manage earnings or not.

3.2.2. Sample representativeness

Assuming that CFOs are the appropriate population from which to obtain participants to meet the needs of the survey, we next consider whether DGHR obtained a representative sample of CFOs.

DGHR obtain their sample by sending an invitation to approximately 10,300 email addresses that were either supplied by CFO magazine or obtained from a list maintained by the Fuqua School of Business at Duke University. They received 558 responses, for a response rate of 5.4%. Of the 558 responses, 169 indicate they are associated with a public company, and that is the sample that is the focus of the paper. DGHR provide descriptive statistics which indicate that surveyed firms are larger

than typical, with a higher growth rate, more leverage, and higher credit ratings. Thus, these are relatively large and secure public companies. DGHR do not indicate whether the 5.4% response rate is similar for public, private and non-profit entities.

DGHR acknowledge that their response rate is lower than some prior CFO surveys but indicate that it is higher than the 4.5% rate obtained in the quarterly Duke CFO survey. They argue that their response rate could be viewed as higher if one considers that two-thirds of the names on the Duke email list had not opened an emailed survey invitation from Duke in the prior three years of surveys.

DGHR's low response rate is troubling. Sample representativeness becomes more of a concern as response rates decrease, because low response rates increase the likelihood that sample results are a noisy or biased representation of the population. In DGHR's case, approximately 94% of prospective survey participants did not respond, and the fact that two-thirds of participants have not responded to a Duke survey in the past three years does not preclude a systematic difference between the CFOs who respond and those who do not. Prior research provides evidence that willingness to respond to surveys is correlated with various attributes, including opinions about the subject of the survey (Fowler, 2009). For example, prospective participants in DGHR may have been more likely to respond if they are irritated about the current state of GAAP, and less likely to respond if they are satisfied with the status quo or are concerned about divulging information about their firms' financial reporting practices.

The potential lack of representativeness due to DGHR's low response rate is an inherent limitation of the study. It likely matters more for some types of questions than for others. Given a low response rate it is very unlikely that numerical estimates provided by a sample—such as the earnings management estimates discussed above—are reliably indicative of the true numbers in the population (Fowler, 2009). It is more likely that directional relationships generalize, because they are not dependent on interpreting the level of an estimate. We return to this issue again below with respect to specific results.

DGHR use two other samples that also suffer from representativeness concerns. First, they report anecdotes from 11 public-company CFOs. Descriptive data indicate that these CFOs are associated with companies that are larger than typical firms in the survey sample (mean (median) sales of \$24B (\$10B)). DGHR indicate that some of these interviews were conducted prior to obtaining the survey data during survey design, and others were conducted afterward to clarify findings, but DGHR do not indicate the number of pre- vs. post-survey interviews or the process by which interviewees were selected. It is unclear whether interviewees are a good reflection of either the survey sample or the more general population of CFOs.

DGHR also interview two former standard setters. The role of these interviews in the paper is not clear. To the extent that the standard setters' views are included to provide a contrast with the views of CFOs, current and former standard setters should have been surveyed systematically to obtain a more representative view of standard setters' opinions. The views of these standard setters are presented as if they represent a unified viewpoint when in fact standard setters sometimes disagree vehemently with each other, as can be observed at FASB meetings and as indicated in the dissents that follow new standards. Certainly a convenience sample of two cannot be viewed as representative of the broader viewpoint of accounting standard setters.

3.3. Survey question format and data analysis

3.3.1. General considerations

Different survey types draw upon different types of participant knowledge. At one extreme is *polling*, in which participants are asked for opinions. Polling is a direct and efficient way to elicit views about just about anything, including preferences, behavior in hypothetical circumstances, etc. A disadvantage of polling is that respondents' views may not reflect actual behavior because of self-presentation biases and lack of self-insight. Also, to the extent that polling is intended to produce evidence about a particular level or proportion of some behavior (e.g., "What percent of Americans support the death penalty?"), sample appropriateness, sample representativeness and question format are very important. DGHR is primarily a polling survey, as participants are asked their opinions about various aspects of earnings quality.

A very different survey type is *episode reporting*, in which participants are asked about particular experiences. An advantage of episode reporting is that participants provide data about a concrete event for which various aspects can be coded and analyzed, which should reduce some of the noise inherent in polling. A disadvantage is that participants may bias their descriptions due to memory limitations and self-presentation concerns, and cannot report data about hypotheticals since those events have not occurred. DGHR conduct some episode reporting through the eleven interviews conducted to supplement their survey evidence.

Regarding question format, the data that are obtained from a survey can be influenced by the way questions are constructed (Fowler, 2009). Aspects of question construction that affect responses include:

- The order in which ratings are made, both within and between questions.
- The structure of rating scales, for example:
 - People tend to avoid scale endpoints
 - Relative ratings are easier than absolute ratings
 - Visual distance on a scale matters
 - People may express uncertainty by moving to the midpoint of qualitative scales, particularly if not given a chance to express their views otherwise

- Question complexity (e.g., requiring mathematical transformations to answer a question).
- Self-presentation concerns (i.e., people do not like to admit they do bad things).

We refer to these aspects as we consider DGHR, first considering questions that elicit relative ratings, and then considering questions that elicit absolute ratings.

3.3.2. Data concerning relative ratings of importance of factors or agreement with statements

Many of DGHR's questions ask participants to rate the extent to which they agree with various statements about different aspects of earnings quality. The ratings are made on 1–5 scales in a list that presents multiple statements simultaneously. As seen in Table 2, DGHR consider the extent to which participants rated each use as important (i.e., giving the use a rating of 4 or 5). DGHR use similar questions to elicit data for Tables 4 (features of high quality earnings), 5 (factors influencing earnings quality), 8 (GAAP policies producing high quality earnings), 9 (changes in standard setting that would produce higher quality earnings), and 13 (motivations for companies managing earnings). Two considerations are important when considering the data and analyses reported in these tables.

First, order of elicitation could matter. For example, a participant who rated the first few uses of earning as being "important" might be less inclined to use that classification for later uses, reasoning that all uses presented in the question are not likely to be important. To control for order effects, DGHR follow the correct practice of randomizing the order in which "uses" appear in this table, thus distributing order effects randomly across levels. DGHR similarly control for order effects for questions that provide data for Tables 4, 8 and 13.

However, for questions relevant to Tables 5 and 9, DGHR did not randomize order. All participants saw the lines of the question in the same order as shown in DGHR's Table 5. The authors state that "while order-of-presentation effects cannot be ruled out for questions 3a and 7 (presented in Tables 5 and 9), we believe that they are unlikely to unduly influence the results." (footnote 1). In fact, the order of appearance in the questions presented to respondents and the order of proportions listed in Table 5 have a correlation of 0.60 ($p=0.022$), indicating a significant order effect. The same concern applies to Table 9 – after dropping redundant questions, the correlation between presentation order and importance rating is .64 ($p=0.048$). This significant order effect makes it difficult to reliably interpret the data reported in Tables 5 and 9. DGHR indicate that they did not randomize order for these questions because some of the subparts of these questions fall in meaningful clusters (p. 3), but that rationale does not change the fact that a significant order effect is apparent in the data.

A second consideration is that, given the many questions asked of participants, differences in ratings within these tables could arise by chance. To assess the significance of differences, DGHR perform *t*-tests of the null hypothesis that the mean rating for a given row equals the mean ratings for each of the other rows appearing in each table, and then indicate in the table those differences for which the *p*-value is less than 0.05. For example, in Table 13, the entry of "2-3, 5-12" in the first row of the column labeled "Significant differences in average Rating vs. rows" indicates that the mean rating of "to influence stock price" (row 1) differs at $p < 0.05$ from the mean ratings of other questions reported in rows 2, 3, and 5-12.

This approach suffers from two problems. First, DGHR report tests of means in their tables, but in the paper they base their conclusions on the proportion of ratings for which participants indicated 4 or 5. To align their statistical tests with their interpretations, DGHR should perform McNemar tests of the equality of multiple proportions drawn from a single sample (Wild and Seber, 1993). Second, because DGHR had no *a priori* hypotheses regarding differences in proportions, their *p*-values should be Bonferroni adjusted (regardless of whether DGHR choose to focus on *t*-tests of means or McNemar tests of proportions). Bonferroni adjustment would increase the *p*-values associated with these tests and reduce the extent to which individual proportions appear to differ significantly from each other. This concern applies to Tables 2, 4, 5, 8, 9, and 13.

In combination, the order effects apparent in Tables 5 and 9 and the misaligned statistical tests and understated *p*-values reported in Tables 2, 4, 5, 8, 9, and 13 undermine some assertions made by DGHR regarding differences in the relative proportions indicated in these tables. For example, relevant to Table 5, DGHR state that "...by far the most important factor affecting earnings quality is the firm's business model (74%) ... followed by accounting standards (60.4%). The other three determinants that garner majority opinion are the company's industry (56.8%), macro-economic conditions (55%) and the firm's internal controls (50%). The board of directors (48%), reporting choices (43.2%) and the operating cycle (40.2%) are also thought to influence earnings quality, though to a lesser extent" (p. 17). It is not clear which of these proportions actually differ significantly from each other.

3.3.3. Data concerning absolute ratings on response scales

DGHR also interpret some numbers in absolute terms, rather than as relative differences in response proportions. These interpretations are problematic given the potential for distortion of absolute levels because of particular aspects of their survey construction and sample representativeness.

3.3.3.1. Interpreting levels vs. expressions of uncertainty. Table 6 of DGHR reports analyses of ratings concerning the extent to which earnings quality is innate. A problem with the question used to elicit this data is that it asks for a rating of a concept that is both ambiguous and abstract ("earnings quality innateness") yet provides no way for participants to express uncertainty about their responses. Because the response scale is anchored by "no influence of innate factors" and "entirely determined by innate factors," one way to express uncertainty is to move towards the center of the scale. Therefore, a mean

and median response of approximately 50% is unsurprising, and these concerns undermine the authors' assertion that "Our key findings include ... (ii) about 50% of earnings quality is driven by non-discretionary factors."

3.3.3.2. Interpreting relations between linked questions. Table 7 of DGHR reports analyses of ratings of the amount of discretion in financial reporting. Panel A reports ratings drawn from responses to the following question: "How much discretion in financial reporting does the current accounting standard-setting regime in the United States allow?" and Panel B reports ratings of "Relative to 20 years ago, ... indicate the extent to which you believe companies today have more or less discretion in financial reporting." DGHR interpret their results as follows: "Overall, we find significant evidence that reporting discretion has been reduced over time, and some evidence that discretion today is somewhat less than desirable," which appears to imply that discretion previously was at the right level and now is not. However, a potential concern with interpreting the relation between these two questions is that the first question refers to "standard-setting regime" while the second refers to "discretion in financial reporting" more generally. One could argue that financial reporting offers less discretion due to such factors as increased audit scrutiny and increased internal control requirements stemming from Sarbanes-Oxley (which respondents may or may not view as standard setting). Also, if indeed these questions are linked, one could argue that the combination of "approximately correct amount of discretion currently" in Panel A and decreased discretion in panel B suggests that CFOs used to have too much discretion but that has been reduced to a current level that is roughly appropriate.

3.3.3.3. Absolute estimate question. DGHR also elicit estimates of the prevalence and amount of earnings management. A few points about this elicitation approach are important.

First, DGHR ask participants to assess their impressions of companies in general rather than their own company's practices. Specifically, DGHR obtain a point estimate of the prevalence of earnings management from responses to the following question: "From your impressions of companies in general, in any given year, what percentage of companies use discretion within GAAP to report earnings that misrepresent the economic performance of the business? (emphasis in original)." Thus, CFOs report their impressions of what other CFOs do. This approach sometimes is used with respect to questions about unethical or unsavory behavior, but it necessarily introduces noise—as we discuss above, it is not clear how much CFOs would know about other firms, especially if they have been with one firm for a long time. Second, regarding the prevalence question relative to Table 10, the natural tendency to avoid scale endpoints could add distortion, as could any sample unrepresentativeness stemming from the low response rate. In combination, these concerns suggest that the mean of 18% is at best a very noisy estimate of the prevalence of earnings management.

Regarding the "amount of manipulation" question relevant to Table 11, both the question and the scale used to provide that rating are very odd. The question asks participants to assume that typical EPS is \$1 and then to indicate the number of cents per share that is misrepresented, which seems to be a round-about way of eliciting a percentage of distortion. Then a drop-down menu appears with a scale that features increments of 1 cent, 5 cents, and 10 cents at different parts of the scale, such that spatial distance on the scale is not linearly related to cents. It is unclear how participants reacted to these aspects of the scale.

DGHR make a very strong interpretation of these data. For example, "CFOs believe that ... (iii) about 20% of firms manage earnings to misrepresent economic performance, and for such firms 10% of EPS is typically managed (abstract)" and "... our point estimates of earnings management can be used for the calibration of existing and future models (p. 3)." "One implication of these results is that economy-wide the magnitude of opportunistic earnings management is relatively modest, ... e.g., using our estimate of 20% of firms managing at 10% of earnings implies an economy-wide rate of 2%. Of course, comparisons with existing models of earnings management have to be made with caution because there are many different models, and there are differing constructs of interest. Still, it is probably fair to say that existing models greatly overstate the magnitude of likely earnings malfeasance (p. 24)." When DGHR suggest caution in interpreting these estimates, it appears to be because of differences in earnings management models and constructs to which they are compared. We believe there is likely to be such high error in DGHR's point estimates that they are simply unreliable estimates of the prevalence and amount of earnings management.

4. Conclusion

DGHR provide some useful insights into CFOs' views regarding earnings quality and trends in financial reporting. However, we believe these views should be interpreted in light of CFOs' role and likely incentives in the financial reporting process. We also believe that DGHR's results must be interpreted in light of the strengths and weaknesses of survey approaches in general and in terms of the particular sampling methods and questions that DGHR use. When considered from that perspective, we are skeptical of some of DGHR's claims, particularly their claim to have valid point estimates of the prevalence and magnitude of earnings management.

References

- Ball, R., 2009. Market and political/regulatory perspectives on the recent accounting scandals. *J. Account. Res.* 47, 277–323.
 Bell, T.B., Landsman, W.R., Shackelford, D.A., 2001. Auditors' perceived business risk and audit fees: analysis and evidence. *J. Account. Res.* 39, 35–43.

- Ben-David, I., Graham, J.R., Harvey, C.R., 2013. Managerial miscalibration. Working paper. Ohio State University and Duke University. (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1640552).
- Brav, A., Graham, J.R., Harvey, C.R., Michaely, R., 2005. Payout policy in the 21st century. *J. Financ. Econ.* 77, 483–527.
- Dechow, P., 1994. Accounting earnings and cash flows as measures of firm performance: the role of accounting accruals. *J. Account. Econ.* 18, 3–42.
- Dechow, P., Skinner, D.J., 2000. Earnings management: reconciling the views of accounting academics, practitioners, and regulators. *Account. Horiz.* 14, 235–250.
- Dechow, P., Ge, W., Schrand, C., 2010. Understanding earnings quality: a review of the proxies, their determinants and their consequences. *J. Account. Econ.* 50, 344–401.
- Dichev, I.D., Graham, J., Harvey, C.R., Rajgopal, S., 2013. Earnings quality: evidence from the field. *J. Account. Econ.*, in this issue.
- Durnev, A., Mangen, C., 2009. Corporate investments: learning from restatements. *J. Account. Res.* 43, 679–720.
- Fowler Jr., F.J., 2009. *Survey Research Methods*, 4th ed. SAGE Publications, Inc., Thousand Oaks, CA.
- Gibbins, M., McCracken, S.A., Salterio, S.E., 2005. Negotiations over accounting issues: the congruency of audit partner and chief financial officer recalls. *Auditing: J. Pract. Theory* 24, 171–193.
- Gibbins, M., Newton, J.D., 1994. An empirical exploration of complex accountability in public accounting. *J. Account. Res.* 32, 165–186.
- Gibbins, M., Salterio, S., Webb, A., 2001. Evidence about auditor–client management negotiating concerning clients' financial reporting. *J. Account. Res.* 39, 535–563.
- Graham, J.R., Harvey, C.R., 2001. The theory and practice of corporate finance: evidence from the field. *J. Financ. Econ.* 60, 187–243.
- Graham, J.R., Harvey, C.R., 2002. How do CFOs make capital budgeting and capital structure decisions? *J. Appl. Corp. Financ.* 15, 8–23.
- Graham, J.R., Harvey, C.R., Rajgopal, S., 2005. The economic implications of corporate financial reporting. *J. Account. Econ.* 40, 3–73.
- Graham, J.R., Harvey, C.R., Rajgopal, S., 2006. Value destruction and reporting decisions. *Financ. Anal. J.* 62, 27–39.
- Libby, R., Bloomfield, R., Nelson, M.W., 2002. Experimental research in financial accounting. *Account. Organ. Soc.* 27, 775–810.
- Nelson, M.W., Elliott, J.A., Tarpley, R.L., 2002. Evidence from auditors about managers' and auditors' earnings management decisions. *Account. Rev.* 77, 175–202.
- Ng, T.B.-P., Tan, H.-T., 2003. Effects of authoritative guidance availability and audit committee effectiveness on auditors' judgments in an auditor–client negotiation context. *Account. Rev.* 78, 801–818.
- Penman, S.H., 2013. *Financial Statement Analysis and Security Valuation*, fifth ed. McGraw-Hill, Irwin, New York.
- Watts, R.L., 2006. What has the invisible hand achieved? *Account. Bus. Res.* 36 (Supplement), 51–61.
- Wild, C.J., Seber, G.A.F., 1993. Comparing two proportions from the same survey. *Am. Stat.* 47, 178–181.