Journal of Accounting and Economics 55 (2013) 23–42

Contents lists available at SciVerse ScienceDirect

Journal of Accounting and Economics

journal homepage: www.elsevier.com/locate/jae

Assessing methods of identifying management forecasts: CIG vs. researcher collected ☆

Elizabeth Chuk [a], Dawn Matsumoto [b], Gregory S. Miller [c,*]

[a] University of Southern California, 3660 Trousdale Parkway HOH 815, Los Angeles, CA 90089-0082, USA
[b] University of Washington Business School, Box 353200, Seattle, WA 98195-3200, USA
[c] University of Michigan, Ross School of Business, 710 East University, Room ER6615, Ann Arbor, MI 48109-1234, USA

ARTICLE INFO

Article history:
Received 8 November 2011
Received in revised form
27 June 2012
Accepted 5 July 2012
Available online 23 July 2012

JEL classifications:
M41
C90

Keywords:
CIG
Forecasts
Management guidance
Accounting disclosure

ABSTRACT

This paper examines the characteristics of management forecasts available on Thomson First Call's Company Issued Guidance (CIG) database relative to a sample of forecasts hand-collected through a search of company press releases. Due to the significantly lower cost of using CIG (relative to hand-collecting data), academics have increasingly relied on this database as a source of management forecasts. However, it is important for researchers to consider the properties of this database (such as coverage, accuracy, and breadth) when evaluating whether it is an appropriate data source for their study. Overall, our results suggest systematic differences between forecasts reported on CIG and forecasts gathered from company press releases. We suggest several sample criteria that will remove or mitigate these biases.

1. Introduction

Management forecasts have long been an area of interest in academic research. Early studies examined, among other things, the accuracy of management forecasts (Copeland and Marioni, 1972; McDonald, 1973; Ruland, 1978); the information content of forecasts (Foster, 1973; Patell, 1976; Waymire, 1984; Pownall and Waymire, 1989); the type of news disclosed (Penman, 1980; Ajinkya and Gift, 1984; Lev and Penman, 1990; Skinner, 1994), and the precision of forecasts (Pownall et al., 1993; Bamber and Cheon, 1998). In most of these early studies, researchers used hand-collected samples of management forecasts gathered from company-issued press releases. However, in recent years, a growing interest in the role of management forecasts in the capital markets – or what has become more commonly referred to as earnings guidance – has fueled the creation of a management forecast database by Thomson First Call (hereafter, First Call) called the Company Issued Guidance (CIG) database.

Not surprisingly, numerous recent studies have relied on this database to address a variety of research questions about earnings guidance as the database provides a relatively large sample of management forecasts in machine-readable format. The appendix presents a list of 23 published studies that rely on the database. In addition, a search of the SSRN

☆ We thank Bok Baik, Bob Bowen, Dave Burgstahler, Shuping Chen, Weili Ge, Frank Hodge, Amy Hutton, Jane Kennedy, Urooj Khan, Rick Mergenthaler, Jonathan Rogers, Lisa Sedor, Terry Shevlin, Doug Skinner, Jerry Zimmerman (editor), Charles Wasley (referee). We also thank workshop participants at the University of Washington and the 2008 AAA conference for their comments. Professor Matsumoto received financial support from the UW Accounting Development Fund and the Emmett S. Harrington Professorship. Professor Miller received financial support from the Michael R. and Mary Kay Hallman Fellowship.

* Corresponding author. Tel.: +1 734 763 6039; fax: +1 734 936 8715.
E-mail addresses: elizabeth.chuk@marshall.usc.edu (E. Chuk), damatsu@u.washington.edu (D. Matsumoto), millerg@umich.edu (G.S. Miller).

0165-4101/$ - see front matter
http://dx.doi.org/10.1016/j.jacceco.2012.07.001

database over the period 2008–2010 identified 49 studies on management forecasts/earnings guidance, of which 37 utilize the CIG database, indicating the reliance on CIG is likely to continue.

To our knowledge, there is no documentation on the process used by CIG to gather forecasts and, while prior studies have provided some evidence of the comparability of hand-collected samples of management forecasts relative to samples generated from the CIG database (Miller, 2002; Anilowski et al., 2007), none have systematically examined this issue. The limited evidence in these studies, however, calls into question the completeness of CIG's coverage of forecasts. For example, Anilowski et al. (2007) find a large spike in the number of forecasts in 1998, which they attribute to more concerted collection efforts on the part of the database provider. Houston et al. (2010) and Chen et al. (2011) both find evidence that firms that were initially thought to have stopped providing earnings guidance based on the CIG database did, in fact, issue forecasts during the purported stoppage period. Our goal is to provide a systematic examination of the differences between the firms and forecasts included on the CIG database and those obtained via hand-collection, providing future researchers with insight into the circumstances under which use of the database must be modified.

We begin by using Lexis Nexis to hand-collect management forecasts from a random sample of Compustat firms and compare these forecasts to data in the CIG database. This comparison provides evidence on the completeness of the CIG database as well as on the factors that influence inclusion in the database. Specifically, we randomly select firms from each size quartile in Compustat in each of the years 1997, 1999, 2001, 2003, 2005, and 2007 and search for press releases until we identify 25 firms in each size quartile with at least one forecast for the year (for a total of 600 firm-years). We identify 1756 press releases containing forecasts of earnings per share (EPS), cash flow from operations (CFO), net income, operating income, EBITDA, revenue, free cash flow, and/or funds from operations (FFO).

We document that the CIG database is far from complete. Of the 600 forecasting firm-years in our full sample, 59% are "covered" by CIG – that is, CIG reports at least one forecast for the firm during that year. Thus, 41% of firm-years in our sample have no coverage on CIG at all during that year. In addition, of the 1756 forecast press releases in our sample, only 51% are represented on the CIG database within a window of ± 5 days of the press release date. Thus, CIG neither covers all firms nor does it capture all the forecast-related press releases issued by firms.

We use logistic regressions to examine whether there are systematic biases in which firms/forecasts are covered by CIG. Some researchers use the CIG database to identify which firms have issued forecasts over a long window, but are not concerned about the exact date the forecast is made. Our firm-level analysis helps researchers to assess whether the CIG database provides an unbiased source for identifying firms' forecasting policies. Other researchers are concerned about whether a forecast is issued on or around a particular date. Our second analysis addresses potential biases in the specific forecast dates represented on CIG.

Using the fiscal year as our long window, we find that the probability of a firm being covered by CIG is greater for firms with high analyst following and high institutional ownership, and lower for firms reporting losses in the recent past. However, after controlling for analyst following and institutional ownership, we do not find any incremental long window bias related to size or growth.

We then examine the factors related to whether a particular forecast-related press release is represented on the database (i.e., whether CIG reports a forecast within five days of the forecast date). We find that forecast press releases are more likely to be represented on the database if the press release contains an EPS forecast and if at least one forecast contains a specific dollar amount. We also find that press releases made by larger firms with high analyst following, high institutional ownership, and fewer prior losses are more likely to be represented on CIG.

We use subsamples to further investigate these biases. Given the importance placed on EPS forecasts, we limit our analysis to press releases containing an EPS forecasts (EPS subsample). We find similar results with this subsample. In addition, we examine the effect of forecast news on CIG coverage by limiting the analyses to forecast-related press releases that do not include an earnings announcement. We find evidence that CIG is biased towards press releases that convey bad news (based on the 3-day return) indicating researchers should consider the impact of news bias, especially if the research question involves comparing the frequencies of good news and bad news forecasts.

For all our analyses, press releases issued after 1997 have higher probabilities of being represented on CIG than press releases issued in 1997, consistent with anecdotal reports that the collection process became more systematic and complete in 1998. The increased coverage represents increased coverage of firms as well as increased coverage of press releases. We do not observe an increase in the completeness of the database around the time of Reg FD.

Combined, these results suggest multiple biases regarding firm and forecast characteristics as well as changes in coverage over time. As we discuss further in section four of the paper, we encourage researchers to evaluate the potential impact of these coverage biases on their studies and, if necessary, supplement CIG data with hand-collected data. In particular, studies of the following four types should not rely solely on CIG data: (1) studies investigating the impact of analyst following, institutional ownership or firm performance, (2) studies where variables of interest are correlated with known coverage biases, (3) studies examining the impact of macroeconomic events on forecasts, and (4) studies focusing on non-EPS or qualitative guidance. We further encourage researchers to test the robustness of their results to the following two tests (even if their study does not appear to fall into one of the previous four scenarios): (1) Limiting the early years in the sample and (2) performing their analysis on a subsample where coverage is known to be higher (e.g., high analyst following observations).

We also investigate whether CIG includes forecasts that would not be identified via hand-collection by randomly generating a sample of management forecasts from the CIG database and searching for the company press release that contains the forecast. In general we find that the forecasts on CIG are valid, particularly post-Reg FD where we find press releases for 96% of the sample. However, our evidence suggests that prior to Reg FD the CIG database likely contain forecasts that are not readily collectible via public sources.

Our study contributes to the literature by providing a systematic analysis of the factors impacting the probability of forecasts being included on the CIG database, an increasingly important database used in accounting research. This analysis is useful in that it allows us to provide guidance to future researchers regarding potential biases that may arise when relying on this database. In particular, we demonstrate that researchers cannot assume that no forecast on the database on a given date means that no forecast was issued – particularly when examining firms with low analyst coverage, low institutional ownership, or poor prior performance; or when the firm does not normally issue EPS forecasts or forecasts with a specific dollar amount. Further, researchers should be aware that CIG is biased toward forecasts containing bad news.

Although the primary purpose of our study is to inform future research, our evidence is potentially useful in evaluating prior research. For example, prior studies that have examined association between certain firm characteristics and the probability of disclosure (e.g., Ajinkya et al., 2005; Karamanou and Vafeas, 2005; Wang, 2007) are potentially biased if the factors examined are related to factors associated with CIG coverage.

In the next section, we discuss our procedure for hand-collecting forecasts and matching them to CIG and provide the results of our analysis. In Section 3, we discuss our procedure for comparing forecasts reported on CIG to public sources. Section 4 presents a discussion of the implication of our findings for researchers. Concluding remarks are presented in Section 5.

2. How complete is the CIG database?

2.1. Hand-collected sample selection procedure and descriptive statistics

To assess the completeness of CIG's coverage, we generate a random sample of firms from Compustat. We perform yearly sorts into four size quartiles based on market capitalization, eliminating foreign firms and firms with a market value of equity of less than $10 million. We then randomly select 25 firms from each size quartile. We search PR Newswire and Business Wire using Lexis Nexis for company-issued press releases that contain a management forecast.[1] If no press release with a management forecast is found for a selected firm-year, we randomly choose another firm from the same size quartile for that year. We repeat this process until we obtain 25 firms per size quartile per year with at least one press release containing a management forecast. This procedure allows us to maintain a sample size of 600 firm-years.

Table 1 Panel A provides descriptive statistics on our sample selection procedure. The first set of columns details the number of firms searched in order to obtain 25 firms with at least one forecast during the year. Dividing the 25 sample firms per size-quartile by the number of firms searched provides an approximation of the number of forecasting firms in each quartile-year. For instance, in size quartile 2 in 1997, we searched 134 firms (resulting in 19% of firms classified as forecasters), whereas we only need to search 33 firms in quartile 4 in 2001 (resulting in 76% of firms classified as forecasters). Not surprisingly, the forecasting percentage is higher for larger firms in all years. We also note that the percentage of forecasting firms increases between 1997 and 2001 across all quartiles. The percentage then levels off and/or declines from 2001 through 2007.

The second set of columns in Panel A presents the average number of press releases found containing forecasts per firm in each quartile-year. Forecasts per firm follow a similar pattern to the number of forecasting firms: generally increasing across size quartiles and increasing from 1997 to 2001 and then declining and/or leveling off from 2001 to 2007. Our sample consists of a total of 1756 press releases.

For comparison, we also estimate the percent of forecasting firms and the average forecast dates per firm based on data from CIG, reported in Panel B. Specifically, we classify all firms on CIG with at least one forecast in a given year into size quartiles based on their market capitalization at the end of the year relative to all firms on Compustat. We divide each number by the total number of firms on Compustat in each size quartile-year (excluding foreign firms and firms with a market value of equity of less than $10 million, similar to the selection criteria used to collect our sample). As shown in Table 1 Panel B, these percentages exhibit a similar trend as those for our hand-collected sample. The percent of forecasting firms increases from 17% in 1997 to 37% in 2001, then falls back to 24% in 2007. However, the percentage of forecasting firms in our hand-collected sample tends to be greater than the percentage of forecasting firms per CIG, particularly from 2001 onward and for the lower size quartiles, suggesting CIG is incomplete with respect to their coverage of smaller firms.

The second set of columns in Panel B reports the average number of unique forecast dates per firm on the CIG database (we count multiple forecasts on a given date as one forecast date). As with our hand-collected sample, CIG shows a pattern of increasing forecast dates per firm across size quartiles. However, unlike our hand-collected sample, the forecast dates per firm appear to increase over time from 1997 through 2007, without the same leveling off after 2001. For example,

[1] The search string is: [(forecast, guidance, outlook, expectation, expect, guide, anticipate, expected, anticipated) within 25 words of (earnings, profit, loss, income, sales, EBITDA, revenue, cash flow)].

Table 1
Sample selection and descriptive statistics.

Panel A: Hand-collected sample

	% Firms forecasting						Forecast dates per firm				
	QTL 1 (%)	QTL 2 (%)	QTL 3 (%)	QTL 4 (%)	Total (%)		QTL 1	QTL 2	QTL 3	QTL 4	Total
1997	15	19	20	28	19	1997	1.36	1.48	1.88	1.72	1.61
1999	16	22	37	45	26	1999	1.76	1.64	1.68	2.56	1.91
2001	34	36	53	76	45	2001	2.28	4.04	3.96	4.40	3.67
2003	25	37	58	68	40	2003	2.08	3.04	4.32	4.20	3.41
2005	25	39	57	60	40	2005	3.12	2.96	4.00	4.08	3.54
2007	27	25	45	61	34	2007	3.12	3.08	3.36	4.12	3.42
All years	22	27	39	50	31	All Years	2.29	2.71	3.20	3.51	2.93

Panel B: CIG

	% Firms forecasting						Forecast dates per firm				
	QTL 1 (%)	QTL 2 (%)	QTL 3 (%)	QTL 4 (%)	Total (%)		QTL 1	QTL 2	QTL 3	QTL 4	Total
1997	6	14	19	29	17	1997	1.30	1.37	1.40	1.64	1.49
1999	13	21	31	43	27	1999	1.47	1.66	1.69	2.22	1.87
2001	12	23	45	68	37	2001	1.88	2.10	2.94	3.63	3.04
2003	8	19	41	64	33	2003	1.80	2.81	3.56	4.60	3.84
2005	7	16	35	55	28	2005	2.45	3.17	4.12	4.95	4.30
2007	5	12	29	51	24	2007	2.37	3.70	4.13	5.14	4.51
All years	9	18	33	51	27	All years	1.79	2.34	3.05	3.86	3.21

Panel C: Distribution of the types of forecasts in hand-collected press releases

	1997		1999		2001		2003		2005		2007		All years	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%
EPS only	69	42.9	87	45.5	93	25.3	89	26.1	92	26.0	51	14.9	481	27.4
Revenue only	22	13.7	29	15.2	32	8.7	51	15.0	56	15.8	42	12.3	232	13.2
Net income only	1	0.6	1	0.5	4	1.1	1	0.3	2	0.6	1	0.3	10	0.6
CFO only	8	5.0	3	1.6	7	1.9	1	0.3	1	0.3	0	0.0	20	1.1
EPS+revenue	45	28.0	47	24.6	118	32.2	126	37.0	117	33.1	150	43.9	603	34.3
EPS+NI	2	1.2	4	2.1	1	0.3	10	2.9	5	1.4	0	0.0	22	1.3
EPS+CFO	0	0.0	3	1.6	5	1.4	2	0.6	0	0.0	4	1.2	14	0.8
EPS+revenue+NI	1	0.6	4	2.1	16	4.4	18	5.3	17	4.8	31	9.1	87	5.0
EPS+revenue+CFO	0	0.0	0	0.0	3	0.8	6	1.8	5	1.4	5	1.5	19	1.1
Other	13	8.1	13	6.8	88	24.0	37	10.9	59	16.7	58	17.0	268	15.3
Total	161	100	191	100	367	100	341	100	354	100	342	100	1756	100

Panel D: Distribution of the types of forecasts in CIG

	1997		1999		2001		2003		2005		2007		All years	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%
EPS only	1693	97.4	3279	96.9	6442	96.9	6169	96.9	6304	94.7	5951	94.5	29,838	96.0
EBITDA only	1	0.1	3	0.1	2	0.0	0	0.0	0	0.0	0	0.0	6	0.0
FFO only	35	2.0	55	1.6	169	2.5	132	2.1	193	2.9	271	4.3	855	2.7
Cash EPS only	9	0.5	34	1.0	22	0.3	0	0.0	0	0.0	0	0.0	65	0.2
EPS +EBITDA	0	0.0	2	0.1	0	0.0	0	0.0	0	0.0	14	0.2	16	0.1
EPS +FFO	0	0.0	0	0.0	0	0.0	64	1.0	159	2.4	63	1.0	286	0.9
EPS +cash EPS	0	0.0	0	0.0	6	0.1	0	0.0	0	0.0	0	0.0	6	0.0
Other	1	0.1	10	0.3	10	0.2	0	0.0	0	0.0	0	0.0	21	0.1
Total	1739	100	3383	100	6651	100	6365	100	6656	100	6299	100	31,093	100

Notes: Panel A tabulates the percentage of forecasting firms (left set of columns) and the average number of forecast dates per firm (right set of columns) for the hand-collected sample. Panel B presents the same tabulation for CIG. In Panel A, the percentage of forecasting firms is defined as 25 divided by the number of firms searched in order to obtain 25 firms with at least one forecast during the year, and the number of forecast dates per firm is defined as the number of press releases found containing forecasts for each firm. In Panel B, the percentage of forecasting firms is defined as the number of firms on CIG in each year and size quartile divided by the number of firms on Compustat in the same year and size quartile, and the number of forecast dates per firm is defined as the number of unique forecast dates per firm on CIG. Panel C tabulates the various combinations of types of forecasts in the hand-collected sample. Panel D tabulates the various combinations of types of forecasts in the CIG database.

for firms in quartile 3, the average forecasts per year was approximately four in 2001 based on our hand-collected sample, but the comparable rate on CIG is around three forecasts per year in 2001 and reaches four per year in 2005.

Panel C of Table 1 reports the distribution of the types of forecasts made in each press release in our sample. Of the 1756 press releases collected, 481 (27.4%) contain only EPS forecasts and 232 (13.2%) contain only revenue forecasts. The remaining press releases contain a combination of different types of forecasts with EPS and revenue being the most common combination, comprising 34.3% of our sample (603 press releases).[2] A large portion of our press releases contain non-EPS forecasts, either individually or in combination with an EPS forecast. In contrast, the vast majority of forecasts reported on CIG are EPS forecasts—for 96.0% of the forecast dates CIG reports forecasts of EPS only (Table 1 Panel D). The remaining forecast dates are primarily composed of forecasts of FFO only (2.7%) and a combination of EPS and FFO (0.9%). The CIG database is clearly incomplete with respect to non-EPS forecasts.

2.2. Matching procedure and descriptive statistics

The previous descriptive statistics suggests that not all hand-collected forecasts issued by all firms are included on the CIG database. To provide evidence on which firms and which types of forecasts are represented on the CIG database, we compare our hand-collected sample to the forecasts reported on CIG. We first compare whether the firms in our hand-collected sample have *any* forecast listed on CIG in the same calendar year as the management forecast date in our hand-collected press releases ($\text{MATCH}^{\text{FIRM}}$). We consider this an indication of whether CIG "follows" or "covers" a firm. We then compare whether there is any forecast reported on CIG within ± 5 days of the date of our hand-collected press release date ($\text{MATCH}^{\text{DATE}}$).[3] This match indicates whether CIG data collectors identified a firm's press release as containing a forecast.[4]

The two levels of matching we conduct allow us to systematically address potential issues surrounding the completeness of the CIG database. First, an analysis of $\text{MATCH}^{\text{FIRM}}$ allows us to address the completeness of the database at the firm level—that is, are some firms excluded from coverage by CIG? Analyzing $\text{MATCH}^{\text{DATE}}$ provides evidence on whether certain forecast-related press releases issued by a given a firm are excluded from coverage by CIG. Because researchers use CIG for different purposes, structuring our analyses in this way provides guidance to researchers on potential problems that might arise from relying on the database in different circumstances.

Table 2, Panel A, provides descriptive statistics by year on $\text{MATCH}^{\text{FIRM}}$ (as well as other firm-level variables used in subsequent analyses) and Panel B reports descriptive statistics on $\text{MATCH}^{\text{DATE}}$ (as well as other press-release level variables used in subsequent analyses). The overall means for $\text{MATCH}^{\text{FIRM}}$ and $\text{MATCH}^{\text{DATE}}$ are 59% and 49%, respectively. If we limit the sample to only those press releases that include an EPS forecast (EPS subsample), the match rates for $\text{MATCH}^{\text{FIRM}}$ and $\text{MATCH}^{\text{DATE}}$ are 79% and 65%, respectively (untabulated). In general, these rates suggest that while many firms are "followed" to some degree by CIG during a given calendar year, a significant portion of firms never appear on the database. Moreover, not all forecast-related press releases issued by firms are identified by CIG, even for firms that are presumably "covered" by CIG.

Across both match variables, there is an increase in match rates between 1997 and 1999, consistent with the anecdotal reports of more complete coverage beginning in 1998. For example, $\text{MATCH}^{\text{DATE}}$ increases from 28% to 56.5% between 1997 and 1999. The increase for $\text{MATCH}^{\text{FIRM}}$ is also pronounced—from 45% to 67%. Thus, it appears that the increase in coverage in 1998 relates to both more complete coverage of press releases made by firms that are "followed" by CIG as well as to an increase in the number of firms "followed".

In the next section we discuss potential determinants of a firm being covered by CIG ($\text{MATCH}^{\text{FIRM}}$) as well as determinants of whether a particular press release is represented on the database ($\text{MATCH}^{\text{DATE}}$). In Section 2.4, we discuss our empirical design and results.

2.3. Predictions and variable measurement

2.3.1. Which firms are covered by CIG?

We expect that the CIG decision to "follow" a firm is a function of the demand for information about the firm by purchasers of the CIG database. We identify four factors we expect to be related to this demand from purchasers: (1) analyst following, (2) institutional ownership, (3) firm performance, and (4) growth.

[2] We only code references to net income or earnings as being distinct from EPS if the company provides specific dollar amounts in their guidance and it is clear these amounts are not EPS amounts. For example, "We expect net income to be in the range of $4.5 to 5.0 million." In contrast if the guidance is ambiguous as to whether it refers to net income or EPS – such as, "we expect net income to grow by 12%" – we code these as EPS forecasts.

[3] Using a relatively wide window around the press release date biases in favor of finding a match on CIG. However, an examination of the difference between the press release date and the forecast date on CIG indicates that a vast majority fall on the same date (95.4%). Of the remaining 4.6%, 3.4% have dates within ± 1 day and the remaining 1.2% have dates that differ by 2 or more dates. Interestingly, we find 3.4% of the observations have CIG dates that fall *before* the press release date. We check the specifics of the forecast(s) in the press release against the forecast(s) in CIG and at least one forecast in the press release matches CIG (i.e., the match is appropriate despite the fact that the date on CIG falls before the press release date). Thus, we believe our relatively large window does not result in the classification of a press release as a match on CIG that is not appropriate.

[4] It is important to note that we do not require CIG to report all forecasts included in the press release, but rather only require CIG to report a forecast on that day. For example, if a firm issues multiple forecasts on the same day we would code CIG as covering the press release if at least one forecast were included. Thus, these tests can be interpreted as getting the date correct for a forecast bundle.

Table 2
Match variables and determinants—descriptive statistics.

Panel A: All observations—firm level variables

	1997	1999	2001	2003	2005	2007	All years	
	$n=100$	$n=100$	$n=100$	$n=100$	$n=100$	$n=100$	$n=600$	
	Mean (stdev)	Mean (stdev)	Mean (stdev)	Mean (stdev)	Mean (stdev)	Mean (stdev)	Mean (stdev)	Median
MATCH^{FIRM}	0.450	0.670	0.640	0.570	0.560	0.660	0.592	1.000
	(0.500)	(0.473)	(0.482)	(0.498)	(0.499)	(0.476)	(0.492)	
AF	2.578	2.333	3.230	2.539	3.116	3.977	2.962	2.134
	(3.295)	(2.957)	(3.578)	(3.248)	(3.870)	(3.450)	(3.442)	
R_AF	2.294	2.304	2.403	2.195	2.329	2.273	2.299	2.000
	(1.194)	(1.197)	(1.145)	(1.104)	(1.136)	(1.1067)	(1.138)	
INSTIT_OWN	0.285	0.342	0.351	0.362	0.448	0.507	0.382	0.382
	(0.281)	(0.309)	(0.317)	(0.322)	(0.347)	(0.350)	(0.329)	
R_INSTIT_OWN	2.396	2.324	2.242	2.334	2.354	2.248	2.316	2.000
	(1.255)	(1.147)	(1.112)	(1.180)	(1.143)	(1.113)	(1.156)	
N_LOSS	1.746	1.705	2.735	3.125	3.041	2.307	2.449	1.000
	(2.332)	(2.124)	(2.965)	(2.708)	(3.017)	(2.840)	(2.739)	
GROWTH	4.152	2.782	3.151	2.634	5.756	3.879	3.774	2.307
	(3.919)	(3.632)	(4.006)	(4.028)	(9.675)	(5.625)	(5.744)	
R_GROWTH	2.391	2.515	2.569	2.403	2.301	2.412	2.428	2.333
	(1.068)	(1.093)	(1.007)	(1.045)	(1.126)	(1.061)	(1.067)	
MKCAP	5,979	2,183	1.104	960	2,011	2,222	2,410	221
	(29119)	(9561)	(2701)	(2133)	(7140)	(7047)	(13291)	
R_MKCAP	2.500	2.500	2.500	2.500	2.500	2.500	2.500	2.500
	(1.124)	(1.124)	(1.124)	(1.124)	(1.124)	(1.124)	(1.119)	
R&D	0.105	0.052	0.206	0.128	0.946	0.079	0.248	0.007
	(0.507)	(0.119)	(0.871)	(0.611)	(7.729)	(0.239)	(3.145)	
R_R&D	2.258	2.337	2.819	2.326	2.226	2.273	2.369	2.000
	(1.293)	(1.298)	(0.899)	(1.260)	(1.293)	(1.229)	(1.233)	
LITIGATION	0.355	0.318	0.407	0.465	0.410	0.318	0.378	0.000
	(0.481)	(0.468)	(0.494)	(0.502)	(0.495)	(0.468)	(0.485)	

Panel B: All observations—press release level variables

	1997	1999	2001	2003	2005	2007	All years	
	$n=161$	$n=191$	$n=367$	$n=341$	$n=354$	$n=342$	$n=1756$	
	mean (stdev)	mean (stdev)	mean (stdev)	mean (stdev)	mean (stdev)	mean (stdev)	mean (stdev)	median
MATCH^{DATE}	0.280	0.565	0.482	0.525	0.523	0.608	0.514	1.000
	(0.450)	(0.497)	(0.500)	(0.500)	(0.500)	(0.489)	(0.500)	
EPS	0.758	0.780	0.744	0.801	0.718	0.751	0.756	1.000
	(0.430)	(0.415)	(0.437)	(0.400)	(0.451)	(0.433)	(0.429)	
SHORT_HORIZON	0.604	0.773	0.732	0.693	0.629	0.581	0.664	1.000
	(0.491)	(0.420)	(0.444)	(0.462)	(0.484)	(0.494)	(0.472)	
SPEC$	0.547	0.644	0.812	0.880	0.881	0.918	0.817	1.000
	(0.499)	(0.480)	(0.391)	(0.326)	(0.324)	(0.275)	(0.387)	
WITH_EA	0.199	0.319	0.480	0.607	0.672	0.713	0.546	1.000
	(0.400)	(0.467)	(0.500)	(0.489)	(0.470)	(0.453)	(0.498)	
SUPP	0.329	0.325	0.490	0.540	0.458	0.602	0.482	0.000
	(0.471)	(0.469)	(0.501)	(0.499)	(0.499)	(0.490)	(0.500)	
SIGNED_RET	−0.035	−0.064	−0.010	−0.004	−0.012	−0.012	−0.018	−0.009
	(0.113)	(0.145)	(0.136)	(0.104)	(0.108)	(0.099)	(0.118)	

Notes: MATCH^{FIRM} is a binary variable set equal to one if the firm making the forecast in our hand-collected sample has a forecast listed on the CIG database during the same calendar year as our hand-collected forecast. MATCH^{DATE} is a binary variable set equal to one if there is a forecast reported in the CIG database within ± 5 days of the press release date for our hand-collected forecast and zero otherwise. AF is the average number of analysts associated with every mean consensus forecast compiled by First Call during a 90-day period ending with the date of the press release. R_AF is the quartile rank by year obtained by ranking AF into four approximately equal-sized groups, with quartile one (four) representing the lowest (highest) AF. INSTIT_OWN is the average percent of shares held by institutions during a 365-day period ending with the date of the press release. R_INSTIT_OWN is the quartile rank by year obtained by ranking INSTIT_OWN into four approximately equal-sized groups, with quartile one (four) representing the lowest (highest) INSTIT_OWN. N_LOSS is the number of quarters with losses in the past eight quarters, ending in the quarter before the press release date. GROWTH is the market-to-book ratio measured at the end of the most recent quarter prior to the forecast date. R_GROWTH is the quartile rank by year obtained by ranking GROWTH into four approximately equal-sized groups, with quartile one (four) representing the lowest (highest) GROWTH. EPS is a binary variable set equal to one if the press release contains an EPS forecast, and zero otherwise. SHORT_HORIZON is a binary variable set equal to one if the press release contains a forecast pertaining to the upcoming quarter, and zero otherwise. SPEC$ is a binary variable set equal to one if the press release contains a forecast with a specific dollar amount, and zero otherwise. WITH_EA is a binary variable set equal to one if the management forecast is issued on the same day as an earnings announcement, and zero otherwise. SUPP is a binary variable set equal to one if the press release contains an EPS forecast and another type of (non-EPS) forecast, and zero otherwise. SIGNED_RET is the 3-day value-weighted market-adjusted return centered on the date of the press release. MKCAP is market capitalization at the end of the year measured as price per share times the number of shares (Compustat items data25 * datat199). R_MKCAP is the quartile rank by year obtained by ranking MKCAP into four approximately equal-sized groups, with quartile one (four) representing the smallest (largest) firms. R&D is research and development expense (Compustat item 46) scaled by prior year revenue (Compustat item 12). R_R&D is the quartile rank by year obtained by ranking R&D into four approximately equal-sized groups, with quartile one (four) representing the smallest (largest) values of R&D. LITIGATION is a binary variable set equal to one if the firm is in one of the following industries with high litigation risk: 2833–2836 (biotech), 3570–3577 (computers), 7370–7374 (computers), 3600–3674 (electronics), 5200–5961 (retailing), and 8371–8734 (R&D service), and zero otherwise

The CIG database was originated by First Call (prior to the company's merger with Thomson Financial) whose principal business was the collection and dissemination of sell-side analyst forecasts. As such, we predict that the collection of management forecasts focused on companies that were heavily followed by analysts. We measure analyst following (AF) as the average number of analysts associated with every mean consensus forecast issued by First Call during the 90-day period ending with the press release date. We use the quartile rank by year (R_AF) as our measure of analyst following to reduce the impact of outliers. We predict a positive relation between analyst following and the probability of inclusion on the CIG database.

Another major customer group of Thomson Financial's products are institutional investors. Institutions are likely interested in the guidance given by firms in their portfolios and may purchase the CIG database in order to track such information. Thus, we predict that CIG is more likely to follow firms that have greater institutional ownership. We measure institutional ownership (INSTIT_OWN) as the average percent of shares held by institutions during the 365 days prior to the guidance date. We again use the quartile rank by year (R_INSTIT_OWN) to reduce the impact of outliers.

We also expect the demand for information about a firm to be related to the firm's performance. Prior research finds analysts stop covering firms which have unfavorable prospects (McNichols and O'Brien, 1997). If analysts lose interest in firms that are performing poorly, there is also likely a decline in the demand for information on these firms. As a result, First Call may be less likely to collect information on these firms. We measure the number of quarters in the most recent eight quarters prior to the press release date in which the firm reported a net loss (N_LOSS) as our proxy for firm performance.

Finally, we expect the demand for information to be greater for high growth firms. Prior research suggests that the penalty to missing analysts' expectations is greater for high growth firms (Skinner and Sloan, 2002) and providing guidance is one mechanism to avoid such misses. As a result, high growth firms may be more likely to provide guidance (Matsumoto, 2002) and the demand for information about guidance from these firms is likely greater. We measure growth as the market-to-book ratio at the end of the most recent quarter prior to the forecast date and use the quartile rank by year (R_GROWTH) to reduce the impact of outliers.

2.3.2. Which press releases are identified by CIG?

We expect that the likelihood a particular press release issued by a firm is identified by CIG as containing a forecast is a function of the type of information included in the press release. We suggest four factors that likely determine whether a press release is identified as containing a forecast: (1) whether the press release contains an EPS forecast; (2) whether the press release contains a forecast for the upcoming quarter, (3) whether a forecast in the press release contains reference to a specific dollar amount; and (4) whether the press release contains other information, specifically, an earnings announcement.

During our sample period, meeting/beating analysts' EPS forecasts became increasingly important to managers (Brown and Caylor, 2005); thus, it is likely that EPS forecasts are deemed more important by the market and likewise by First Call when collecting forecasts. Thus, we expect First Call to be more likely to identify press releases containing EPS forecasts. We define a indicator variable (EPS) equal to one if a press release contains an EPS forecast and zero otherwise.

We also hypothesize a greater demand for short horizon forecasts. Long-horizon forecasts have greater inherent uncertainty and are potentially discounted by investors relative to short-horizon forecasts. Thus, the incentives to capture short-horizon forecasts is likely greater. We identify press releases that contain a forecast for the upcoming quarter and define a dummy variable (SHORT_HORIZON) equal to one for these press releases (zero otherwise). We expect a positive relation between this variable and the probability of inclusion on the CIG database.

Forecasts can vary significantly in terms of the specificity of the amounts being forecasted. For example, saying, "we expect earnings per share to meet analysts' current expectations" is less specific than saying, "we expect earnings per share to be \$2.58." Both forecasts are technically point estimates but it is likely easier to identify the latter as a management forecast.[5] Thus, we expect press releases that contain specific dollar forecasts are more likely to be identified by First Call's data collectors. We define a variable (SPEC\$) as equal to 1 if the press release contains a forecast with a specific dollar amount. We expect SPEC\$ to be positively related to the probability CIG identifies a press release as containing a forecast(s).

Finally, forecasts are often issued in conjunction with earnings announcements (Anilowski et al., 2007). That is, firms announce earnings for quarter q and, at the same time, provide guidance on earnings for quarter $q+1$. If the primary purpose of the press release is something other than announcing the forecast then we expect that First Call data collectors are less likely to identify the press release as containing a forecast. We therefore predict that forecasts issued in conjunction with an earnings release are less likely to be covered by CIG. We define a variable (WITH_EA) equal to one if the forecast is issued on the earnings report date per Compustat and zero otherwise. We also explore two other potential determinants that are based on prior research but (1) do not have clear cut directional predictions and (2) are specific to a subsample of our press releases. First, prior research suggests that many earnings forecasts are accompanied by other

[5] Prior studies have generally focused on the "precision" of forecasts—i.e., whether forecasts are point, range, minimum, maximum, or qualitative (Bamber and Cheon, 1998). Our interest in this study, however, is less about the "preciseness" of the forecast (i.e., a narrow range vs. a wider range) and more about whether the description of the forecast includes a specific dollar amount.

types of forecasts such as forecasts of sales or cash flows (Hutton et al., 2003). It is unclear whether these supplemental forecasts increase or decrease the likelihood that a press release is represented on the CIG database. On the one hand, these forecasts may increase the credibility of the earnings forecast (Hutton et al., 2003) and it is possible that more credible forecasts are more likely to be identified by CIG. On the other hand, including additional information that may not be the primary focus of the CIG data collectors (who may be primarily looking for EPS forecasts) could lower the probability of identifying a forecast-related press release. We define an indicator variable (SUPP) equal to one for press releases that contain an EPS forecast AND another type of forecast and zero otherwise. We include this variable only when conducting tests on the subsample of press releases that contain an EPS forecast.

Second, prior research often considers the type of news – good or bad – conveyed in a forecast. However, there is no obvious reason why news type would influence inclusion of forecasts on the CIG database. In addition, we face some difficulty in determining the type of news conveyed in a forecast because forecasts of non-EPS amounts cannot be compared to existing analyst forecasts (as most analysts do not forecast non-EPS items). Further, using market reactions to classify news is confounded if other news is released simultaneously (e.g., an earnings announcement). Despite these limitations, studies frequently include type of news and thus we include it in our analyses. We measure the 3-day market-adjusted returns centered on the day of the press release (SIGNED_RET) for all press releases, excluding those issued in conjunction with an earnings announcement.

2.4. Empirical design and results

2.4.1. Descriptive statistics

Table 2, Panel A, provides descriptive statistics by year on the firm-level variables discussed above. Across the six years, our firms have an average of 2.9 analysts following the firm, 38% institutional ownership, 2.5 (out of eight) quarters of prior losses and a market-to-book ratio of 3.8.

Panel B reports similar statistics for our press-release level variables. On average, approximately 75.6% of press releases include an EPS forecast. In addition, we note a trend in the propensity for firms to provide quantitative forecasts over time. In 1997 only 55% of press releases contain a forecast with an explicit dollar amount (SPEC$), vs. 92% in 2007. In addition, consistent with findings in Anilowski et al. (2007), we find an increase in the propensity to issue forecasts in conjunction with an earnings release—in 1997, 20% of our press releases include an earnings announcement vs. 71% in 2007.

2.4.2. Multivariate tests and results

To formally test our predictions, we use the following logistic regressions:

$$\begin{aligned}\mathrm{MATCH}^{\mathrm{FIRM}} = {} & \beta_0+\beta_1 R\_\mathrm{AF}+\beta_2 R\_\mathrm{INSTIT\_OWN}+\beta_3 N\_\mathrm{LOSS}+\beta_4 R\_\mathrm{GROWTH}\\ & +\beta_5 R\_\mathrm{MKTCAP}+\beta_6 \mathrm{YR1999}+\beta_7 \mathrm{YR2001}+\beta_8 \mathrm{YR2003}+\beta_9 \mathrm{YR2005}+\beta_{10} \mathrm{YR2007}+\varepsilon\end{aligned} \tag{1}$$

$$\begin{aligned}\mathrm{MATCH}^{\mathrm{DATE}} = {} & \beta_0+\beta_1 \mathrm{EPS}+\beta_2 \mathrm{SHORT}_{\mathrm{HORIZON}}+\beta_3 \mathrm{SPEC\$}+\beta_4 \mathrm{WITH}_{\mathrm{EA}}++\beta_4 R\_\mathrm{AF}\\ & +\beta_6 R\_\mathrm{INSTIT\_OWN}+\beta_7 N\_\mathrm{LOSS}+\beta_8 R\_\mathrm{GROWTH}+\beta_9 R\_\mathrm{MKTCAP}+\beta_{10} \mathrm{YR1999}\\ & +\beta_{11} \mathrm{YR2001}+\beta_{12} \mathrm{YR2003}+\beta_{13} \mathrm{YR2005}+\beta_{14} \mathrm{YR2007}+\varepsilon\end{aligned} \tag{2}$$

We include R_MKCAP, the market value of equity quartile to which the firm was assigned in year *t*, to control for size. We also include year dummy variables (YR1999–YR2007) to control for differences across the years in the completeness of the database. As discussed previously, anecdotal reports suggest that there was a fundamental shift in the data collection process at First Call in 1998. We also include the variables from Eq. (1) in Eq. (2) since the likelihood that a press release is identified as containing a forecast is likely related to the factors that determine whether the firm is followed by CIG.

Eq. (1) is estimated at the firm-year level. Our sample size for this analysis (538 firm-years) is less than the original 600 firm-years collected due to data requirements to measure our independent variables. In Eq. (2), we have multiple observations for the same firm and many of the independent variables are fairly stable across time for a given firm, suggesting a potential problem with correlated error terms. Thus, we estimate clustered standard errors (clustered by firm) as suggested by Petersen (2009).

Results of Eq. (1) are presented in Table 3. Consistent with our predictions, firms with higher analyst following and higher institutional ownership are more likely to be "followed" by CIG (p-value <0.001 and p-value$=0.02$, respectively). On average, a one quartile increase in analyst following (institutional ownership) results in a 17.9% (5.8%) higher probability of being followed by CIG.[6] We also find that firms reporting more prior quarterly losses are less likely to be included on the CIG database (p-value<0.0001). A firm with four quarters of prior losses would have a 16% smaller probability of being followed than a firm with no prior losses. These percentages are economically significant given the overall rate of not being followed by CIG is 40.8%. We do not, however, find that higher growth leads to an increased

[6] We compute the marginal effects for those independent variables that are continuous as $e^{\beta' X}/(1+e^{\beta' X})^2$ where $\beta' X$ is computed at the mean values of the independent variables (Greene, 1993). For those independent variables that are dummy variables, we calculate the marginal effect as the difference in probability when the variable is equal to 1 versus zero, evaluated at the mean of the other variables.

Table 3
Determinants of Being Followed by CIG.
Model:

$$\begin{aligned}\mathrm{MATCH}^{\mathrm{FIRM}} = {} & \beta_0+\beta_1 R\_\mathrm{AF}+\beta_2 R\_\mathrm{INSTIT\_OWN}+\beta_3 N\_\mathrm{LOSS}+\beta_4 R\_\mathrm{GROWTH}+\beta_5 R\_\mathrm{MKTCAP} \\ & +\beta_6 \mathrm{YR1999}+\beta_7 \mathrm{YR2001}+\beta_8 \mathrm{YR2003}+\beta_9 \mathrm{YR2005}+\beta_{10} \mathrm{YR2007}+\varepsilon \end{aligned} \tag{1}$$

Variable	Pred. sign	All observations (n=538)			Marginal effect
		Coeff.	χ^2	P-value	
Intercept		−2.813	35.34	< 0.0001	
R_AF	+	0.814	35.76	< 0.0001	0.1786
R_INSTIT_OWN	+	0.264	4.08	0.0218	0.0579
N_LOSS	–	−0.189	18.46	< 0.0001	−0.0416
R_GROWTH	+	−0.102	0.82	0.8175	−0.0223
R_MKTCAP	?	0.240	2.27	0.1320	0.0526
YR_1999	+	1.729	20.32	< 0.0001	0.2889
YR_2001	+	1.688	16.94	< 0.0001	0.2841
YR_2003	+	1.159	7.64	0.0029	0.2166
YR_2005	+	0.970	6.17	0.0065	0.1848
YR_2007	+	1.479	14.98	0.0001	0.2582
Adjusted R^2		0.4378			

Notes: Results of estimating the logistic regression in Eq. (1) above using clustered standard errors (clustered by firm) as suggested by Petersen (2009). MATCH$^{\mathrm{FIRM}}$ is a binary variable set equal to one if the firm making the forecast in our hand-collected sample has a forecast listed on the CIG database during the same calendar year as our hand-collected forecast. YR_1999, YR_2001, YR_2003, YR_2005, and YR_2007 are dummy variables equal to 1 if the forecast is issued in 1999, 2001, 2003, 2005 and 2007 respectively, and zero otherwise. See Table 2 for other variable definitions.

probability of being followed by CIG, inconsistent with our predictions. We find significant coefficients on all the year dummies, indicating that the number of firms covered by CIG increased after 1997.[7]

The above analysis indicates that certain firms are more likely to appear on the CIG database. However, this analysis does not necessarily indicate that CIG's coverage of certain firms is more complete—that is, whether CIG's propensity to capture more of the forecasts issued by a given firm varies based on firm characteristics. To provide some evidence on this issue, we calculate a variable, COVERAGE%, equal to the number of forecast-related press releases issued by a firm in a given year that appear on the CIG database divided by the total number of forecast-related press releases issued by a that firm in that year (this equates to the firm-year level mean of MATCH$^{\mathrm{DATE}}$). We then regress this variable on our firm characteristics in Eq. (1). We estimate this regression for the subset of firm-years where MATCH$^{\mathrm{FIRM}}$ equals one because we are interested in whether the same firm characteristics that determine coverage on CIG also determine the *consistency* of coverage on CIG.

Results of this untabulated analysis indicate that firms with greater analyst following and greater institutional ownership have greater coverage consistency (p-values < 0.05), similar to our results on overall coverage (using MATCH$^{\mathrm{FIRM}}$). However, the coefficient on N_LOSS is not statistically significant; suggesting that firm performance does not impact the consistency of coverage, although it does determine overall coverage on CIG. Growth also does not determine the consistency of coverage, similar to overall coverage. Finally, the consistency of coverage increases post-1997, similar to overall coverage.

Overall, these results suggest CIG's coverage of firms differs based on certain firm characteristics. Firms with greater analyst following and higher institutional ownership are more likely to appear on the database and more of these firms' press releases are captured by CIG. Firm performance also influences whether CIG covers a particular firm, although it does not influence whether the consistency with which CIG covers the firm.

We next turn to our analysis of the characteristics that determine whether a particular press release is captured by CIG. The results of estimating Eq. (2) are reported in the first set of columns in Table 4. The probability that First Call identifies a press release as containing a forecast is 53% greater if the press release contains an EPS forecast (coefficient on EPS significant at p-value < 0.0001). In addition, if the press release contains a forecast with a specific dollar amount, there is a 39.5% higher probability of being included on the database (significant at p-value < 0.0001). We do not, however, find that press releases with forecasts for the upcoming quarter (SHORT_HORIZON) are more likely to be included on CIG nor do we find that forecasts bundled with an earnings announcement (WITH_EA) are less likely to be included on the database.

We continue to find significantly positive coefficients on R_AF and R_INSTIT_OWN, indicating that CIG is more likely to identify the forecast-related press releases of firms with higher analyst following and greater institutional ownership.

[7] To address the possibility that our results are driven by CIG coverage in prior years (instead of our hypothesized determinants), we conduct two sets of tests. First, we reperform both our regressions using MATCH$^{\mathrm{FIRM}}$ and MATCH$^{\mathrm{DATE}}$ (Tables 3 and 4, respectively) using a subsample of firms that were not included in CIG in any of the years prior to the year in which the firm is included in our hand-collected sample. Second, we reperform the analyses in Tables 3 and 4 after adding a dummy variable set equal to one if the firm was included in CIG in any year prior to the year that it is included in our hand-collected sample. Results for both sets of tests remain qualitatively similar to the reported results, suggesting that our main results are not driven by inertia in CIG coverage of firms over time.

Table 4
Determinants of a Press Release Date Being Represented on CIG.
Model:

$$\mathrm{MATCH}^{\mathrm{DATE}} = \beta_0 + \beta_1 \mathrm{EPS} + \beta_2 \mathrm{SHORT}_{\mathrm{HORIZON}} + \beta_3 \mathrm{SPEC\$} + \beta_4 \mathrm{WITH}_{\mathrm{EA}} (+\beta_5 \mathrm{SUPP} + \beta_6 \mathrm{SIGNED}_{\mathrm{RET}}) + \beta_7 R\_\mathrm{AF} + \beta_8 R\_\mathrm{INSTIT\_OWN} + \beta_9 N\_\mathrm{LOSS} + \beta_{10} R\_\mathrm{GROWTH} + \beta_{11} R\_\mathrm{MKTCAP} + \beta_{+12} \mathrm{YR1999} + \beta_{13} \mathrm{YR2001} + \beta_{14} \mathrm{YR2003} + \beta_{15} \mathrm{YR2005} + \beta_{16} \mathrm{YR2007} + \varepsilon \quad (2)$$

		Full sample ($n=1577$)				EPS subsample ($n=1240$)				Non-EA subsample ($n=637$)			
Variable	Predicted Sign	Coeff.	χ^2	*P*-value	Marginal effect	Coeff.	χ^2	*P*-value	Marginal effect	Coeff.	χ^2	*P*-Value	Marginal effect
Intercept		−6.923	236.45	<0.001		−4.304	109.21	<0.001		−6.245	100.41	<0.001	
EPS	+	2.576	157.79	<0.001	0.528					2.464	67.00	<0.001	0.518
SHORT_HORIZON	+	0.143	0.93	0.167	0.036	0.162	1.02	0.157	0.034	0.295	1.21	0.135	0.073
SPEC$	+	1.793	73.44	<0.001	0.395	1.790	59.08	<0.001	0.414	1.360	20.82	<0.001	0.313
WITH_EA	−	−0.169	1.33	0.124	−0.042	−0.173	1.10	0.147	−0.036				
SUPP	?					−0.073	0.19	0.665	−0.015				
SIGNED_RET	−									−2.282	6.56	0.005	−0.571
R_AF	+	0.343	20.12	<0.001	0.086	0.426	27.39	<0.001	0.088	0.271	4.93	0.013	0.068
R_INSTIT_OWN	+	0.388	26.78	<0.001	0.097	0.349	17.71	<0.001	0.072	0.426	11.86	<0.001	0.107
N_LOSS	−	−0.097	11.53	<0.001	−0.024	−0.095	9.06	0.001	−0.020	−0.148	8.83	0.002	−0.037
R_GROWTH	+	−0.007	0.01	0.547	−0.002	0.000	0.00	0.499	0.000	0.008	0.01	0.468	0.002
R_MKTCAP	?	0.301	10.13	0.002	0.075	0.256	6.31	0.012	0.053	0.166	1.09	0.296	0.042
YR_1999	+	1.610	26.21	<0.001	0.350	1.464	19.99	<0.001	0.226	1.824	22.80	<0.001	0.395
YR_2001	+	1.101	15.34	<0.001	0.262	0.916	9.88	<0.001	0.166	1.487	17.92	<0.001	0.344
YR_2003	+	1.090	15.10	<0.001	0.259	1.110	13.62	<0.001	0.196	1.547	17.40	<0.001	0.348
YR_2005	+	1.132	15.59	<0.001	0.268	1.120	13.09	<0.001	0.195	1.081	7.83	0.003	0.255
YR_2007	+	1.654	32.26	<0.001	0.370	1.855	32.23	<0.001	0.287	1.262	10.02	0.001	0.290
Adjusted R^2		0.505				0.352				0.507			

Notes: Results of estimating variations on the logistic regression in Eq. (2) on: (1) the full sample, (2) the subsample of press releases that include an EPS forecast, and (3) the subsample of press releases made on non-earnings announcement (EA) dates. The regression on the full sample excludes the variables SUPP and SIGNED_RET. The regression on the EPS subsample excludes the EPS and SIGNED_RET variable but includes the variables SUPP. The non-EA subsample excludes the WITH_EA and SUPP variables but includes the SIGNED_RET variables. All regressions are estimated using clustered standard errors (clustered by firm) as suggested by Petersen (2009). $\mathrm{MATCH}^{\mathrm{DATE}}$ is a binary variable set equal to one if there is a forecast reported in the CIG database within ± 5 days of the press release date for our hand-collected forecast and zero otherwise. YR_1999, YR_2001, YR_2003, YR_2005, and YR_2007 are dummy variables equal to 1 if the forecast is issued in 1999, 2001, 2003, 2005, and 2007, respectively, and zero otherwise. See Table 2 for other variable definitions.

Prior poor performance (N_LOSS) also decreases the probability of a firm's press release being picked up by CIG.[8] We again find significant coefficients on our year dummy variables, indicating that there was an increase in CIG's coverage of press releases post-1997. Not only did CIG begin covering more *firms* post-1997, they also were more likely to identify more of the press releases of these firms after this period. However, the proportion of matches does not grow over time. Notably, there does not appear to be an *increase* in the proportion of matches between 1999 and 2001—the years before and after the passage of Reg FD. Thus, prior findings of an increase in the frequency of forecasting based on observations on the CIG database are not likely due to higher coverage but rather reflect true increases in the frequency of forecasting.

In the second and third sets of columns in Table 4, we repeat our analysis on two different subsamples. The second set of columns reports the results of conducting our analysis on the EPS subsample, adding the variable SUPP (and excluding the variable EPS). Results are similar to our full sample results—press releases including specific dollar forecasts are more likely to be represented on the database. However, we do not find a significant coefficient on SUPP, indicating that the inclusion of supplemental information to support an EPS forecast has no impact on whether a press release is represented on the database.

The third set of columns reports the results of conducting our analysis on the subsample of press releases that are issued independently of an earnings announcement, including the variable SIGNED_RET (and excluding WITH_EA). Results are again similar to our main results—press releases that include an EPS forecast and a specific dollar amount are more likely to be represented on the database. More importantly, we find a significantly negative coefficient on SIGNED_RET, indicating that forecast-related press releases that convey negative news are more likely to be represented in the database.

The prior results suggest that there are systematic differences between the firms that are covered by CIG and those that are not as well as differences between those forecast-related press releases that are represented on the database and those

---

[8] Note that in our untabulated analysis of coverage consistency, we do not find that poor prior performance is related to coverage consistency, whereas we find a significantly negative coefficient on N_LOSS in our press-release level regression ($\mathrm{MATCH}^{\mathrm{DATE}}$ analysis). This is because our COVERAGE% analysis was conducted on the subset of firms where $\mathrm{MATCH}^{\mathrm{FIRM}}=1$, whereas our $\mathrm{MATCH}^{\mathrm{DATE}}$ analysis is conducted on the full sample of firms.

that are excluded. To provide researchers with further guidance on potential adjustments that could alleviate problems with the completeness of the database, we calculate match rates by year and for different levels of analyst following. Table 5, Panel A reports match rates for all press releases while Panel B reports rates for press releases containing at least one EPS forecast (EPS subsample).

We make two main observations from these tables. First, match rates are generally much lower for 1997 vs. the other years. Even when analyst following is very high (5 or more analysts), the match rates are only 32% (38%) for the full sample (EPS subsample). Second, match rates increase significantly between the subsample with no analyst following and the subsample with at least one analyst following, particularly after 1997. When considering all press releases, match rates range from 57% to 72% for observations with at least one analyst following between the years 1999 and 2007. For the EPS subsample the range is 67–84%. Overall, these results suggest that issues with completeness of the database are particularly severe prior to 1999 and for firms with no analyst following.[9]

Although match rates increase post 1997 and for firms with at least one analyst following, match rates are still far from 100% even for very highly followed firms and even in the most recent year (84% in 2007 for firms with five or more analyst following).[10] This fact may not present problems to researchers as long as there are no systematic biases for these observations. To determine whether this is the case, we re-estimate equation two after eliminating (1) observations in the year 1997 and (2) observations with no analyst following. We then run the analysis for the subsample of firms with one, three, and five analyst following (excluding the variable R_AF).

Results of this analysis are presented in Panel C of Table 5. Both the EPS dummy variable and SPEC$ are significant regardless of the subgroup, indicating these particular biases persist even when analyst following is very high. However, R_INSTIT_OWN is no longer significant when analyst following is five or greater and N_LOSS is not significant in any of the regressions. Thus, if researchers are using samples with relatively high analyst following they do not have to be as concerned that CIG has systematically excluded forecasts issued by firms with lower institutional ownership and/or prior poor performance.[11] Also, there is no difference in coverage over the years once 1997 has been excluded. Thus, it is not the case that CIG coverage has been growing over the years or becoming more complete—rather the beginning years were simply lower in coverage.

Overall, our results indicate that CIG is incomplete with respect to covering certain types of firms as well as certain types of forecasts. We discuss specific suggestions that researchers might consider to address this incompleteness in Section 4.

3. Additional analyses

3.1. Alternative sources of public forecasts

An underlying assumption in our prior analyses is that our hand-collected sample provides a fairly complete set of public forecasts made by our randomly chosen firms. Given that the method we used to gather our forecasts is similar to methods generally employed – i.e., searching press releases available on Lexis-Nexis for certain key words – understanding whether these methods result in complete samples of publicly issued forecasts is likely important to researchers.

To provide some evidence on this issue, we collect a sample of forecasts from an alternative source for forecasts – 8-K filings – and examine the frequency with which forecasts issued in these filings can be found in a press release.[12] Since Reg FD requires managers to issue an 8-K when a forecast is not issued in a public forum, identifying forecasts from 8-K filings would provide a sense of the completeness of samples drawn from press releases (i.e., how frequently firms provide forecasts in 8-K filings without distributing them via press releases). For each of the years 2003, 2005, and 2007, we randomly select 25 firms from Compustat, examine all 8-K filings made by the firm, and identify any forecasts disclosed in the filing. We then determine whether these forecasts are available on PR Newswire and/or Business Wire. Untabulated results indicate that the vast majority of forecasts disclosed in 8-Ks are available via these two newswires. Specifically, we identify 279 instances of guidance in 8-Ks across the three years and are able to find 261 of these forecasts in press releases

[9] We conducted similar analyses for various cut-offs of institutional ownership, market capitalization, and number of prior loss quarters (untabulated). Our basic conclusions are similar. Match rates for 1997 are much lower than other years even for firms with high institutional ownership (>50%), high market capitalization (>$300 M), and good prior performance (0 prior loss quarters). Rates for 1997 were 42%, 30%, and 38% in 1997 for firms in the high institutional ownership, high market capitalization and low prior loss categories versus 69%, 70%, and 72% for the other years. Also, match rates increase significantly between firms with zero institutional ownership (22%) and firms with at least some institutional ownership (61%).

[10] Firms in the highest institutional ownership, highest market capitalization, and lowest number of prior losses groups (as discussed in footnote 9) had match rates of 79%, 75%, and 80% in 2007.

[11] The reduced significance does not appear to be due to smaller sample sizes. We also ran the analysis using dummy variables indicating the different subsamples interacted with the main variables in Eq. (2). We then tested the significance of the coefficients on the main variable and the interacted variable (e.g., the coefficient on EPS+coefficient on EPS^*AF^{group}). Conclusions remain the same: EPS and SPEC$ are significant regardless of the level of analyst following while R_INSTIT_OWN is no longer significant when analyst following is greater than five.

[12] Reg FD restricts the use of outlets that do not result in "broad, non-exclusionary distribution of information to the public" and allows firms to use 8-K filings as an alternative. Thus, post-Reg FD, a management forecast issued in a non-public forum would likely be accompanied by an 8-K filing. However, Reg FD does not *require* managers to make an 8-K filing when issuing a forecast.

Table 5
Analysis of match rates by analyst following.

Panel A: Univariate match rates by year and analyst following—all press releases.

	All observations			No analyst following			Analyst following ≥ 1			Analyst following ≥ 3			Analyst following ≥ 5		
Year	# obs	# matches	% matched	# obs	# matches	% matched	# obs	# matches	% matched	# obs	# matches	% matched	# obs	# matches	% matched
1997	161	45	28	63	6	10	98	39	40	61	27	44	28	9	32
1999	191	108	57	54	9	17	137	99	72	72	55	76	30	22	73
2001	367	171	47	92	13	14	275	158	57	153	101	66	91	61	67
2003	341	176	52	99	23	23	242	153	63	157	104	66	101	69	68
2005	354	185	52	109	25	23	245	160	65	163	122	75	91	73	80
2007	342	208	61	46	14	30	296	194	66	202	156	77	128	108	84

Panel B: Univariate match rates by analyst following—press releases containing at least one EPS forecast

	All observations			No analyst following			Analyst following ≥ 1			Analyst following ≥ 3			Analyst following ≥ 5		
Year	# obs	# matches	% matched	# obs	# matches	% matched	# obs	# matches	% matched	# obs	# matches	% matched	# obs	# matches	% matched
1997	122	44	36	40	6	15	82	38	46	50	26	52	24	9	38
1999	149	98	66	33	6	18	116	92	79	58	49	84	28	22	79
2001	273	156	57	54	10	19	219	146	67	128	91	71	74	53	72
2003	273	169	62	40	20	30	206	149	72	131	101	77	86	68	79
2005	254	175	70	61	19	31	193	159	82	136	121	89	80	73	91
2007	257	204	79	28	11	39	229	193	84	175	155	89	121	108	89

Notes to Panels A and B:
Match rates for all press releases (Panel A) and for press releases including at least one EPS forecast (Panel B) by year and by analyst following. The first set of columns report match rates by year for all observations. The second set of columns report match rates for firms with no analyst following during a 90-day period ending with the date of the press release. The next three sets of columns report match rates for firms with at least one, three, or five analysts with forecasts included in the mean consensus forecast for the firm in the 90-day period ending with the press release date.

Panel C: Determinants of press release date being represented on CIG after eliminating 1997 and conditioning on analyst following

		AF ≥ 1 (n=913)				AF ≥ 3 (n=639)				AF ≥ 5 (n=386)			
Variable	Pred. sign	Coeff.	χ^2	*P*-Value	Marginal effect	Coeff.	χ^2	*P*-value	Marginal effect	Coeff.	χ^2	*P*-value	Marginal effect
Intercept		−2.526	17.35	< 0.001		−3.069	13.05	< 0.001		−2.973	3.67	0.055	
EPS	+	1.777	19.26	< 0.001	0.374	1.717	13.95	< 0.001	0.338	1.999	7.22	0.004	0.381
SHORT_HORIZON	+	0.047	0.04	0.420	0.007	−0.118	0.15	0.653	−0.016	0.008	0.00	0.493	0.001
SPEC$	+	1.483	21.68	< 0.001	0.297	1.299	12.32	< 0.001	0.234	1.367	8.43	0.002	0.23
WITH_EA	−	−0.156	0.47	0.246	−0.023	0.006	0.00	0.508	0.001	−0.330	1.00	0.159	−0.038
R_INSTIT_OWN	+	0.183	3.01	0.041	0.028	0.345	6.62	0.005	0.047	0.169	0.75	0.193	0.020
N_LOSS	−	−0.028	0.25	0.308	−0.004	0.015	0.04	0.582	0.002	0.057	0.24	0.689	0.007
R_GROWTH	+	0.056	0.29	0.295	0.008	0.041	0.11	0.371	0.006	1.180	1.15	0.142	0.021
R_MKTCAP	?	0.238	3.09	0.079	0.036	0.318	3.11	0.078	0.043	0.160	0.31	0.580	0.019
YR_2001	+	−0.705	3.67	0.972	−0.118	−0.763	2.51	0.943	−0.119	−0.151	0.05	0.586	−0.018
YR_2003	+	−0.223	0.27	0.699	−0.035	−0.361	0.43	0.745	−0.053	0.067	0.01	0.463	0.008
YR_2005	+	−0.070	0.02	0.558	−0.011	−0.020	0.00	0.513	−0.003	0.694	0.75	0.193	0.071
YR_2007	+	0.043	0.01	0.459	0.006	0.222	0.16	0.343	0.029	0.594	0.61	0.217	0.065
Adjusted R^2		0.194				0.206				0.229			

Notes to Panel C:
Results of estimating the logistic regression in Eq. (2) for observations with at least one, three, or five analysts with forecasts included in the mean consensus forecast for the firm and after eliminating the year 1997. All regressions are estimated using clustered standard errors (clustered by firm) as suggested by Petersen (2009). YR_2001, YR_2003, YR_2005, and YR_2007 are dummy variables equal to 1 if the forecast is issued in 2001, 2003, 2005, and 2007, respectively, and zero otherwise. YR_1999 is not included in the model because the year 1997 has been eliminated (and, therefore, the year 1999 is captured in the intercept). See Table 2 for other variable definitions.

distributed on PR Newswire and/or Business Wire, representing roughly 94% of the forecasts.[13] This finding suggests that the frequency with which firms release forecasts without issuing a press release is relatively low. Therefore, hand-collecting data by searching company press releases appears to be a robust means of accessing the vast majority of publicly issued forecasts.

3.2. Does CIG include forecasts not available through hand-collection?

Our prior analyses focus on the *completeness* of the CIG database. A separate issue is whether the data included on CIG database are representative of what would be identified through hand-collection. In other words, are there forecasts on CIG that are not available via hand-collection? And, are the data reported on CIG accurately reported? To investigate this issue, we randomly select forecasts reported on CIG and attempt to trace them to press releases. Specifically, we partition firms on the CIG database into four size quartiles based on market capitalization from Compustat (eliminating foreign firms and firms with a market value of equity below $10 million) and randomly select 100 observations from each size quartile in each of the years 1997, 1999, 2001, 2003, 2005, and 2007 (for a total of 2,400 observations). We then search Business Wire and PR Newswire on Lexis-Nexis using the same search string that we used to generate our hand-collected sample. We read the press releases found and compare the guidance given in the press release to that reported on the CIG database.

Untabulated results indicate that most forecasts on CIG can be found in press releases using hand-collection methodologies typically employed by researchers, especially after Reg FD. In particular, after adjusting for firm name changes and eliminating the use of a search string, we find press releases in 91% of the cases. If we limit the sample to post-FD years, the rate is 96%. Thus, the majority of forecasts on CIG are found in press releases, particularly post-FD. However, in many cases, the forecasts reported on CIG are for a different line item, time horizon, amount, or specificity. Overall, we find an error rate of approximately 11%.

Of the 9% of forecasts that are not matched to a press release, 5% are available through an alternative source—a news article, 8-K, and/or conference call transcript.[14] Researchers should be aware of these alternative sources but overall, the proportion of forecasts available through these sources (and not through press releases) is not large.

Finally, we note that for 4% of our sample, we are not able to find the source for the forecast. The rate in the pre-FD time period is much higher (8.5%) relative to the post-FD period (1.7%). It is possible that CIG picked up these forecasts through an alternative source that we are unable to identify. Alternatively, these observations may be errors by CIG. The fact that this situation occurs in the post-Reg FD era when private dissemination of information is not allowed (thereby limiting the possible alternative sources available), suggests the possibility of errors in the database.[15] At a minimum, these forecasts on CIG do not represent widely disseminated public forecasts. Depending on the research question being addressed, this fact could represent another problem with relying on the CIG database. We make specific suggestions for future research in the next section of the paper.

4. Research implications

4.1. Demonstrations of potential biases

The analyses in this paper suggest several potential issues with relying on the CIG database as a source of management forecasts. It is difficult, however, to definitively ascertain whether the conclusions in prior studies are affected by their reliance on CIG because direct replication of a prior study is infeasible. Hand-collection of forecasts is time-consuming and the samples in many prior studies are large (which is the reason they rely on the CIG database). However, to provide some

[13] Specifically, we find 96% (86/90), 95% (92/97), and 90% (83/92) of forecasts on PR Newswire and Business Wire in 2003, 2005, and 2007, respectively. With the exception of 2007, we accessed PR Newswire and Business Wire using Lexis Nexis. In 2007, however, there were 12 forecasts that we were not able to find on PR Newswire using Lexis-Nexis but which we were able to find on that newswire using Factiva. All these forecasts were made between July and October leading us to believe that there was some incompleteness in Lexis-Nexis's coverage of PR Newswire during those months. It does not appear that this represents a systematic problem with Lexis-Nexis and we do not believe this introduces any systematic bias in our prior findings regarding the completeness of the CIG database.

[14] Pre-FD, the percent that are found through alternative sources is 10%, while post-FD the rate is 2% (for an average of 5% across the entire sample). Pre-FD, forecasts were found in news articles or 8-Ks only, while post-FD, some forecasts were found in conference call transcripts in addition to news articles or 8-Ks (conference call transcripts are available only from 2001 onward). For those found in news articles only, it is not clear where the authors of these articles obtained information about these forecasts but it does not appear that the forecast was widely disseminated through another mechanism, e.g., over a broad-based newswire. CIG may have used the articles themselves as their source. Alternatively, CIG may have access to the same alternative source(s) used by the authors of these business press articles. The latter possibility suggests CIG includes forecasts that are not readily available through public sources.

[15] To provide some evidence on this possibility, we calculate the 3-day, market-adjusted, absolute returns centered around the purported forecast date per CIG to assess whether there appears to have been an information event on that day. We exclude observations whose forecast date coincides with an earnings announcement date because we would not be able to attribute the market reaction to the forecast. Results (untabulated) indicate that the average absolute returns on these dates are significantly *less* than the average absolute returns of those forecasts for which we are able to easily identify the related press release. However, the returns are significantly greater than zero. Thus, while it appears that an information event did occur on that date for those observations where we are unable to find a source document, it is unlikely to represent a forecast similar to the ones for which we are able to easily identify a source document.

Table 6
Percentage of forecasting firms using hand-collected (HC) sample and CIG partitioned on commonly used determinants of forecasting.

	HC	CIG	Diff: HC vs. CIG
Low AF	24.56%	14.57%	−9.99%***
High AF	52.85%	51.97%	−0.88%
Diff: High vs. Low AF	28.30%***	37.41%***	9.11%***
Low MKTCAP	24.31%	12.40%	−11.91%***
High MKTCAP	43.86%	43.42%	−0.44%
Diff: High vs. Low MKTCAP	19.55%***	31.02%***	11.47%***
Low INSTIT_OWN	20.50%	10.68%	−9.82%***
High INSTIT_OWN	47.82%	43.07%	−4.76%*
Diff: High vs. Low INSTIT_OWN	27.32%***	32.39%***	5.07%*
Low N_LOSS	31.61%	30.08%	−1.54%
High N_LOSS	34.80%	22.02%	−12.78%***
Diff: High vs. Low N_LOSS	3.19%	−8.06%***	−11.25%***
Low GROWTH	30.77%	23.08%	−7.69%***
High GROWTH	44.28%	38.86%	−5.42%**
Diff: High vs. Low GROWTH	13.51%***	15.78%***	2.27%
Pre-Reg FD	22.08%	19.32%	−2.76%
Post-Reg FD	39.53%	27.17%	−12.35%***
Diff: Pre vs. Post Reg FD	17.45%***	7.86%***	−9.59%***

*, **, and *** indicate the differences are significant at p-values less than 10%, 5%, and 1%, respectively.
The variables above are partitioned as follows. For AF, INSTIT_OWN, MKTCAP, GROWTH, firms are categorized in the "high" group for firms in the top two quartiles, and in the "low" group for the bottom two quartiles. For N_LOSS, firms with more than one loss out of the past eight consecutive quarters are categorized in the "high" group, and firms with one or less loss quarters are categorized in the "low" group.

evidence on the *potential* effects on conclusions from using the CIG database, we compute the proportion of firms issuing a forecast in a given year based on our hand-collected sample, for various subsets of firms.[16] We then compare these proportions to the proportions that would be obtained from relying on the CIG database. The subsets examined are based on the firm characteristics we have predicted will be related to bias in coverage (e.g., analyst following and institutional ownership). The goal of these comparisons is to allow readers to see how the implications of studies examining relations between these firm characteristics and management forecasts are influenced by the choice to use the CIG database.

Results of these univariate comparisons are reported in Table 6. A clear pattern emerges from a review of the data. The firms for which you would expect better coverage – larger, better performing, more highly followed firms – have forecasting proportions per CIG that are generally very close to those reported using hand-collected data (generally statistically equivalent). For example, 52.85% of firms with high analyst following are classified as having forecasted using hand-collected data vs. 51.97% using CIG. Similarly, 43.86% of high market cap firms are classified as forecasters using hand-collection vs. 43.42% using CIG. In both cases the forecasting rates are statistically equivalent. However, firms with "low coverage" attributes have forecasting proportions on CIG that are significantly below what would be found using hand-collected data – as much as 50% lower. For example, CIG shows a 14.57% forecasting rate for firms with low analyst following, while hand-collected data finds the forecasting rate to be much higher: 24.56%. Similarly, hand-collection shows a 24.31% forecasting rate for low market cap firms compared to 12.40% reported on CIG. Both differences are statistically significant.

Moreover, while the difference in forecasting proportions between the types of firms (e.g., low/high analyst following) are statistically significant using either CIG or hand-collected data, the differences are much larger using CIG data. For example, the difference in forecasting proportions for firms with high and low analyst following is 37.4% using CIG data but only 28.3% using hand-collected data, and this difference is statistically significant. The larger difference for CIG is because of the under-representation of low analyst following firms in the CIG database. It is clear that researchers at a minimum would draw erroneous conclusions regarding the *magnitude* of differences in forecasting across firm types.[17] Table 6 also shows that CIG systematically under-estimates the impact of Reg FD on the proportion of forecasting firms.

[16] As detailed in Table 1, we searched the press releases of 1918 firm-years in order to identify the 600 firm-years used in our previous analysis. We, therefore, use the 1918 firm-years in our analysis, classifying the 600 firm-years as forecasting years based on the results of our hand-collection efforts. We then take these same 1918 firm-years and use the CIG database to identify which firm-years would be classified as forecasting firm-years based on CIG.

[17] In untabulated analyses we performed a stacked regression comparing the coefficients on each of these variables when handcollection is used rather than CIG. Our results are similar in significance and magnitude, though the colinearity between analysts following, institutions and size causes some variables to be insignificant in certain specifications (though always significant when included individually). We present the univariate results as they make the potential magnitude of the bias more apparent.

Table 7
Changes around Reg FD in the percentage of forecasting firms using hand-collected (HC) sample and CIG partitioned on commonly used determinants of forecasting.

	Panel A: Handcollected sample (HC)			Panel B: CIG sample		
	Pre-Reg FD	**Post-Reg FD**	**Diff: pre vs. post**	**Pre-Reg FD**	**Post-Reg FD**	**Diff: pre vs. post**
Low AF	17.66%	30.74%	13.08%***	11.87%	16.99%	5.12%***
High AF	36.28%	67.64%	31.36%***	43.26%	59.75%	16.5%***
Diff: High vs. Low AF	18.62%***	36.90%***	18.27%***	31.39%***	42.76%***	11.37%***
Low MKTCAP	17.70%	29.90%	12.20%***	11.86%	12.86%	1.00%
High MKTCAP	29.33%	58.31%	28.98%***	31.67%	55.10%	23.43%***
Diff: High vs. Low MKTCAP	11.63%***	28.41%***	16.79%***	19.81%***	42.25%***	22.43%***
Low INSTIT_OWN	15.17%	26.46%	11.29%***	11.42%	9.85%	−1.57%
High INSTIT_OWN	36.52%	54.96%	18.44%***	35.84%	47.63%	11.79%***
Diff: High vs. Low INSTIT_OWN	21.35%***	28.50%***	7.15%*	24.42%***	37.78%***	13.36%***
Low N_LOSS	26.49%	36.68%	10.19%***	26.05%	34.06%	8.01%***
High N_LOSS	19.32%	45.97%	26.64%***	17.29%	25.43%	8.14%***
Diff: High vs. Low N_LOSS	−7.17%**	9.29%***	16.45%***	−8.76%***	−8.63%***	0.13%
Low GROWTH	22.08%	38.31%	16.23%***	18.18%	27.32%	9.14%***
High GROWTH	31.72%	55.21%	23.50%***	34.30%	42.82%	8.51%**
Diff: High vs. Low GROWTH	9.64%***	16.90%***	7.27%	16.12%***	15.50%***	−0.63%

*, **, and *** indicate the differences are significant at p-values less than 10%, 5%, and 1%, respectively.
The partitioning of the "high" and "low" groups is defined in Table 6.

The increase in the proportion of forecasting firms before and after Reg FD is 7.9% per CIG but is 17.5% using hand-collected data (a statistically significant difference).

In Table 7, we extend our univariate analysis by interacting the firm characteristics partitions with the Reg FD split. The goal of this analysis is to provide an example of how an issue of interest to researchers – the impact of one of the largest disclosure regulatory changes in U.S. history – can be impacted by the choice of using CIG to study the issue. Consistent with the evidence in Table 6, we find that hand-collection and CIG provide very different insights into the impact of Reg FD. In most cases, CIG understates the impact of Reg FD (the pre- vs. post- differences tend to be larger for our hand-collected sample in Panel A than the CIG sample in Panel B). However, the understatement is not always equivalent across different types of firms. For example, CIG data shows roughly the same increase as hand-collected data for large firms (23.43% for CIG vs. 28.98% for hand-collected); however, for small firms, CIG shows a minor and statistically insignificant increase, while hand-collected data shows a large and statistically significant increase (12.2%). CIG also shows a similar increase using hand-collected data for firms with high institutional ownerships but for firms with low institutional ownership, hand-collected data show a statistically significant increase while CIG shows a decrease (though insignificant). These differences could significantly impact the conclusions researchers would draw from their studies. While we have not taken these analyses to the level of a full study (multivariate analyses, multiple predictions, etc.), they demonstrate the potential for reaching misinformed conclusions when relying on CIG in research.

4.2. Recommendations to researchers

Our prior analyses demonstrate that the CIG database has incomplete coverage of forecasts and that CIG's coverage varies in systematic ways. Thus, if a researcher's sample includes firms and/or years that are known to have lower coverage, researchers should consider supplementing their CIG samples with hand-collected data. This endeavor is labor intensive but is likely the only solution to alleviate concerns about measurement error and bias.[18] In many cases this approach will still allow the researcher to gain substantial time savings by using the CIG data as a starting point relative to hand-collecting the entire sample. In other cases, the researchers may find that the necessary level of "supplementing" would be so great that hand-collecting the entire sample is more efficient. We strongly encourage researchers to supplement CIG data with hand-collected data or to entirely hand-collect their samples in each of the following four circumstances:

1. *Studies where the primary focus is on the impact of analysts, institutional ownership, or earnings performance on forecasting:* Our evidence indicates strong cross-sectional differences in both inclusion of firms on the database during a given period and in the consistency of their coverage even if they are covered on the database during some period. This skewed coverage has the

[18] For example, both Houston et al. (2010) and Chen et al. (2011) examine the determinants of firms' decisions to stop providing earnings guidance. One potential determinant of stopping guidance is poor performance. Therefore, relying solely on the CIG database would be problematic because of CIG's bias against covering poorly performing firms. Both studies search company press releases to ensure that the firms with no forecasts listed on CIG did not actually issue a forecast that was missed by CIG. Without this step, the conclusion that poorly performing firms are more likely to stop providing earnings guidance would be suspect because the relation could be driven by the fact that CIG has lower coverage of forecasts issued by poorly performing firms.

potential to bias studies that attempt to draw an association (or causation) between forecasting and these firm characteristics.[19] For example, using CIG to examine the relation between the level (or change) in analyst following and the propensity to issue a forecast (or the change in the propensity to issue a forecast) will undoubtedly lead to biased results. Researchers should note that earnings performance skews coverage both due to historical performance and the performance being forecasted. In the former case, we show consistent and robust evidence that firms with a history of loss are less likely to be covered on the database. Thus, any attempt to study the relation of forecasting and past earnings will result in bias. However, we also show that on the forecast level CIG is more likely to cover forecasts of bad news. As discussed previously in the paper, it is more difficult to test this bias due to the need to identify the type (sign) of the news being forecasted. However, the evidence we are able to provide suggests that researchers will need to either supplement or entirely hand-collect in these situations as well.

2. *Studies where variables of interest are correlated with known coverage biases*: We identify a number of variables that influence CIG coverage but biases are not limited to situations in which these specific variables are the variables of interest. Biases can manifest indirectly through *correlations* with the researcher's variables of interest. Because we cannot anticipate an exhaustive list of future variables of interest (to directly test coverage biases), we recommend researchers examine whether their variables of interest are correlated with CIG coverage biases (e.g., analyst following, institutional ownership, firm performance). If correlations are significant, relying solely on CIG could lead to biased inferences and we recommend supplementing CIG with hand-collection.

   For example, Choi et al. (2011) examine the effect of management forecasts on future earnings response coefficients. Prior research suggests that earnings response coefficients are lower for poorly performing firms, such as loss firms (Hayn, 1995). If the authors were to find a significant correlation between future earnings response coefficients (variable of interest) and performance, we would recommend they supplement CIG with hand-collection to ensure the validity of their findings.

   The correlations can often be complex and not readily apparent. For example, Brown and Zhou (2011) study the relation between analyst forecast *efficiency* (i.e., how well analysts' forecasts incorporate prior signals) and managers' propensity to issue an earnings forecasts. The authors measure the efficiency of the *first* forecast issued after an earnings release. However, since the timing of a forecast is likely related to analyst coverage (i.e., greater analyst following likely leads to more timely forecasts following an earnings release) and also to forecast efficiency (i.e., forecasts issued earlier in the quarter are likely less efficient), it is possible that analyst following (a known coverage bias) is correlated with forecast efficiency (the variable of interest). If the correlation is significant, supplementing with hand-collected data is necessary.

   We also note that the "variable of interest" may actually be an economic event and, as a result, the research design may examine *changes* in forecasting for a given firm over time. As with a cross-sectional research design, it is important to assess whether the economic event is correlated with any of the known coverage biases – that is, whether the event is correlated with changes in analyst following, institutional ownership, performance, etc. For example, Feng et al. (2011) examine whether accounting misconduct (the variable of interest) affects managers forecasting behavior. However, because accounting misconduct often occurs when firms are performing poorly, coverage biases would likely have impacted their results. The authors recognize this potential and, as a result, hand-collect forecasts.

3. *Studies examining the impact of macroeconomic events on forecasting:* Our prior analyses show that CIG coverage is far from complete. Moreover, the level of completeness varies across years. If the purpose of a study is to examine the impact of an event on market-wide forecasting behavior, issues of bias arise when CIG is relied upon to represent the universe of forecasting firms (because it does not). Therefore, studies cannot draw conclusions about market-wide forecasting behavior (or the impact of such patterns on the macro-economy), particularly conclusions about changes in these patterns over time. The CIG database represents only a subset of forecasts issued by a subset of firms and changes in these patterns over time may also be the result of changes in the CIG selection process.[20]

   In addition, many macroeconomic studies examine cross-sectional differences in the impact of the event. Because our analyses show that overall CIG coverage issues can differ across macro-economic events (such as Reg FD), we again recommend researchers avoid relying solely on CIG data if the cross-sectional differences examined are related to coverage biases (such as analyst following, institutional ownership, etc.). For example, using hand-collected data, we find increases in forecasting around Reg FD for both large and small firms (and firms with high and low institutional ownership), while the evidence using CIG data shows no change in forecasting for small firms (and firms with low institutional ownership). Thus, a study hypothesizing differential effects of Reg FD across firms of different sizes (or institutional following) would need to either hand-collect or supplement CIG data with hand-collected data to avoid biased inferences.

4. *Studies focusing on non-EPS or qualitative guidance*: CIG's coverage of non-EPS forecasts is sparse and the bias toward coverage of EPS forecasts is pervasive (even for firms with very high analyst following). The same is true of qualitative

[19] Our reference to earnings performance includes both the history of performance and the performance being forecasted (i.e., forecast news). We are able to measure the former for our entire sample (using the number of loss quarters over the prior two years) and include this variable in all our tests. We find it consistently creates a bias. The bias caused by the latter is more difficult to assess due to the constraints in identifying news type. Thus, we are limited to testing this bias on a subsample of forecasts. However, the evidence we provide indicates researchers must be concerned about bias in this situation as well.

[20] Even if data from CIG is reported primarily for descriptive purposes it can be misleading to readers if it is reported in such a way as to purport to represent the universe of forecasts. For example, Balakrishnan et al. (2012) discuss the pattern of forecasting around Reg FD based on the CIG database. It is not the primary focus of their paper, but is used to validate their use of CIG and their subsample. Our analyses at several places in the paper show that CIG understated the impact of Reg FD.

forecasts and forecasts that exclude a specific dollar amount. Thus, if a study is focused on non-EPS forecasts or qualitative forecasts, we believe hand-collecting or supplementing CIG with hand-collection is necessary.

If a study falls into one of the four categories above, we suggest not relying solely on CIG. In these situations, it is likely infeasible or impractical to find a solution that will alleviate concerns about the impact of coverage biases on the inferences of a study without some hand-collection efforts. In our list of published papers that have used CIG (Appendix A) we indicate (in bold) those studies that contain a hypothesis that falls into one of the four conditions where we believe supplementing CIG data with hand-collection would have been preferable.

We also caution researchers to avoid assessing the potential biases that might exist in their studies by only focusing on the simple univariate relationship between known coverage biases and the variables of interest in their study. The impact of biases in a multivariate setting will depend on the specific firms being studied, the bias in coverage of those firms/attributes and the functional form of the econometric tests being used. Thus, even if the researchers' variable(s) of interest do not have significant univariate correlations with any known coverage bias, we encourage researchers to conduct robustness tests to address the potential that a more complex correlation structure biases their results. We make two main suggestions[21]:

1. *Limit early years in the sample*: The extent of coverage on the CIG database changed significantly between 1997 and 1999. The coverage rates are low even for firms with high analyst coverage (Table 5, Panel A). Thus, researchers should test the sensitivity of their results to eliminating years prior to 1998. The appendix indicates (with an asterisk) those prior studies that rely on CIG data from years prior to 1998.
   Researchers should also consider the fact that prior to Reg FD, forecasts reported on CIG were potentially not disseminated over a broad-based newswire. This fact can be problematic if studies are assuming forecasts on CIG are *public* forecasts. For example, Wang (2007) relies on CIG data in her study of the effects of Reg FD on firms' disclosure decisions. In particular, she assumes forecasts on CIG in the pre-FD period represent *public* forecasts and uses the number of CIG forecasts to classify firms as private vs. public guiders. The evidence in this paper suggests that the assumption that forecasts reported on CIG are all public forecasts is erroneous, particularly in the pre-FD era.
2. *Conduct analyses on subsets of data*: Given the known bias toward the coverage of certain types of firms and certain types of forecasts, researchers can reduce concerns that their results are attributable to these coverage biases by conducting their analyses on subsets of data where coverage is known to be better. For example, in Table 5, Panel C, we note that certain biases (i.e., poor performance and institutional ownership) are reduced for the subsets of data with greater analyst following. Thus, researchers can reduce concerns that their results are driven by coverage biases by demonstrating that their results hold for subsets of the data where coverage is known to be better (e.g., for firms with high analyst following). Specifically, we recommend that researchers partition their sample into high and low portfolios (for instance, using the median as a cutoff) along the dimensions of known coverage biases (e.g., analyst following, institutional ownership, performance). If results using the high coverage partition are significantly different from results using the full sample, the study's findings are likely to be biased by CIG coverage.

We conclude with a final word of caution. While our prior discussion focuses heavily on the potential for bias caused by coverage issues with CIG, there also exists a prevailing measurement issue with the database. The CIG data is incomplete and, in some cases, significantly so. Even if these missing forecasts are not systematically related to a variable of interest in a study, the measurement error inherent in the data can result in a failure to reject the null when, in fact, an effect exists. If this null effect is used to support a researcher's thesis, erroneous inferences will follow. Alternatively, this failure to reject the null could cause researchers to prematurely abandon research projects. Moreover, this measurement error likely grows as the data is aggregated across forecasts for a given firm. For example, our overall match rate is 51.4% across all press releases. If the probability of coverage on CIG were independent across all press releases, then for a firm issuing two press releases there would be a 73.6% probability ($1-0.514^2$) that at least one of the two press releases would be missed by CIG. If the firm issued three press releases, the probability of missing at least one grows to 86.4%.[22] Similarly, if a researcher is aggregating across time periods—say, calculating the percent of forecasting years—the same issue occurs: the probability of improperly classifying at least one period grows as the number of periods increases.[23] Examples of prior studies that have aggregated CIG data include Bhojraj et al. (2010), Hu and Jiang (2009), Brochet et al. (2011), and Rogers et al. (2009). If researchers are not careful, such measurement error can lead to erroneous inferences.

[21] Both suggestions will limit the sample size and generalizability of the study if such procedures are applied to the primary sample. To balance the goals of internal and external validity, we recommend that researchers conduct their main tests using broad samples and conduct sensitivity tests on subsamples of data as discussed below.

[22] Of course, it is likely that each press release is not an independent draw. Still, if there is any randomness to the probability of coverage across observations for a given firm, aggregating data at the firm level is likely to increase the measurement error in the data. For example, the percent of firms with a 100% coverage rate on CIG (i.e., every press release made by the firm is picked up by CIG) among those firms with at least one forecast picked up by CIG ($MATCH^{FIRM}=1$) is 54%. Thus, it is not the case that firms who are covered by CIG have all their forecasts picked up by CIG, suggesting there is some randomness in coverage within a firm.

[23] Note that the probability of correctly classifying a given period as containing a forecast increases as the number of forecast increases (because the probability of missing all the forecasts is small). However, the issue here is not about correctly classifying a given period but about aggregating across periods.

In summary, it is important to reiterate that we are not advocating that researchers *never* use the CIG database but rather we are encouraging them to carefully consider potential biases introduced by the coverage issue when designing their studies.

5. Conclusion

The purpose of this paper is to conduct a systematic analysis of the completeness and accuracy of forecasts reported on First Call's CIG database relative to those hand-collected via press releases. The CIG database provides researchers with data on management forecasts in an easy-to-use machine readable format and is, therefore, a convenient source for data that was previously difficult and time-consuming to collect. However, the increasing reliance on the database to address various questions related to managers' disclosure decisions raises questions about the appropriateness of the database in addressing these questions.

Comparing our hand-collected sample of forecasts to the CIG database, we find that a majority of the firms (59%) in our sample are represented on the CIG database within the same calendar year as the hand-collected press release date but only about half (51%) of press release dates are represented. A multivariate logit analysis indicates that the firms that are not represented on the database during a given year are systematically different from those that are: they have less analyst following and institutional ownership, and poorer prior performance. In addition, whether a press release containing a forecast is represented on CIG depends on whether the press release contains a forecast of EPS, whether the forecast is in specific dollar amounts, and the news in the press release. Overall, our results suggest that researchers can inadvertently introduce biases into their studies if they fail to consider these factors when designing their studies. In many cases, the researcher should consider supplementing their data with hand-collected data to alleviate concerns over coverage biases.

Appendix A

See Table A1.

Table A1
List of published studies using CIG.

	Author(s)	Title
Panel A: CIG used in main analysis		
*	**Ajinkya et al. (JAR 2005)**	**The Association between Outside Directors, Institutional Investors and the Properties of Management Earnings Forecasts**
*	**Karamanou and Vafeas (JAR 2005)**	**The Association between Corporate Boards, Audit Committees, and Management Earnings Forecasts: An Empirical Analysis**
	Heflin et al. (TAR 2003)	**Regulation FD and the Financial Information Environment: Early Evidence**
*	**Wang (TAR 2007)**	**Private Earnings Guidance and Its Implications for Disclosure Regulation**
	Brochet et al. (JAR 2011)	Earnings Guidance Following Top Executive Turnovers
*	Rogers et al. (JAE 2009)	Earnings Guidance and Market Uncertainty
*	Xu (JAE 2010)	Do Management Earnings Forecasts Incorporate Information in Accruals?
*	Feng and McVay (TAR 2010)	Analysts' Incentives to Overweight Management Guidance when Revising Their Short-Term Earnings Forecasts
	Lennox and Park (JAE 2006)	The Informativeness of Earnings and Management's Issuance of Earnings Forecasts
*	Bamber et al. (TAR 2010a)	Managers' EPS Forecasts: Nickeling and Diming the Market?
*	Bamber et al. (TAR 2010b)	What's My Style? The Influence of Top Managers on Voluntary Corporate Financial Disclosure
	Feng et al. (JAE 2009)	Internal Control and Management Guidance
	Ali et al. (JAE 2007)	Corporate Disclosures by Family Firms
*	Rogers and Stocken (TAR 2005)	Credibility of Management Forecasts
*	Seybert and Yang (Management Science, 2012))	The Party's Over: The Role of Earnings Guidance in Unraveling Sentiment-Driven Overvaluation
*	Anilowski et al. (JAE 2007)	Does Earnings Guidance Affect Market Returns? The Nature and Information Content of Aggregate Earnings Guidance
	Choi et al. (RAST 2011)	Do Management EPS Forecasts Allow Returns to Reflect Future Earnings? Implications for the Continuation of Management's Quarterly Earnings Guidance
	Runyan and Smith (International Journal of Accounting, Auditing and Performance Evaluation, 2007)	The Effect of Multi-Nationality on the Precision of Management Earnings Forecasts
	Krishan et al. (Journal of Accounting and Public Policy 2011)	How do Auditors View Managers' Voluntary Disclosure Strategy? The Effect of Earnings Guidance on Audit Fees
Panel B: CIG used in supplemental analysis or as starting point for hand-collection		
*	Miller (JAR 2002)	Earnings Performance and Discretionary Disclosure
	Houston et al. (CAR 2010)	To Guide or Not to Guide? Causes and Consequences of Stopping Quarterly Earnings Guidance

Table A1 (*continued*)

Panel B: CIG used in supplemental analysis or as starting point for hand-collection		
	Chen et al. (JAE 2011)	Is Silence Golden? An Empirical Analysis of Firms that Stop Giving Quarterly Earnings Guidance
*	Wasley and Wu (JAR 2006)	Why Do Managers Voluntarily Issue Cash Flow Forecasts?

Notes:
This appendix includes published papers that use CIG.
Bolded text indicates that the paper contains a hypothesis about analyst following, institutional ownership, firm performance, a macroeconomic event, non-EPS forecasts, or qualitative forecasts.
* indicates that the paper uses a sample period prior to 1998.

References

Ajinkya, B., Gift, M., 1984. Corporate managers' earnings forecasts and symmetrical adjustments of market expectations. Journal of Accounting Research 22 (2), 425–444.

Ajinkya, B., Bhojraj, S., Sengupta, P., 2005. The association between outside directors, institutional investors, and the properties of management earnings forecasts. Journal of Accounting Research 43 (3), 343–376.

Ali, A., Chen, T., Radhakrishnan, S., 2007. Corporate disclosures by family firms. Journal of Accounting and Economics 44, 238–286.

Anilowski, C., Feng, M., Skinner, D.J., 2007. Does earnings guidance affect market returns? The nature and information content of aggregate earnings guidance. Journal of Accounting and Economics 44, 36–63.

Balakrishnan, K., Billings, M., Kelly, B., Ljungqvist, A., 2012. Shaping Liquidity: On the Causal Effects of Voluntary Disclsoure, Working Paper New York University.

Bamber, L., Cheon, Y.S., 1998. Discretionary management earnings forecast disclosures: antecedents and outcomes associated with forecast venue and forecast specificity choices. Journal of Accounting Research 36 (2), 167–190.

Bamber, L., Hui, K., Yeung, P., 2010a. Managers' EPS forecasts: nickeling and diming the market? The Accounting Review 85 (1), 63–95.

Bamber, L., Jiang, J., Wang, I., 2010b. What's my style? The influence of top managers on voluntary corporate financial disclosure. The Accounting Review 85 (4), 1131–1162.

Bhojraj S., R. LibbyH. Yang. 2010. Analyzing guidance at the firm level: the association between guidance frequency, guidance properties, and market reactions. Working Paper, Cornell University.

Brochet, F., Faurel, L., McVay, S., 2011. Manager-specific effects on earnings guidance: an analysis of top executive turnovers. Journal of Accounting Research 49 (5), 1123–1162.

Brown, L., Caylor, M., 2005. A temporal analysis of quarterly earnings thresholds: propensities and valuation consequences. The Accounting Review 80 (2), 423–440.

Brown, L., L. Zhou. 2011. Interactions between analyst earnings forecasts and management earnings forecasts. Working Paper, Georgia State University.

Chen, S., Matsumoto, D., Rajgopal, S., 2011. Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance. Journal of Accounting and Economics 51 (1–2), 134–150.

Choi, J., Myers, L., Zang, Y., Ziebart, D., 2011. Do management EPS forecasts allow returns to reflect future earnings? Implication for the continuation of management's quarterly earnings guidance. Review of Accounting Studies 16, 143–182.

Copeland, R., Marioni, R., 1972. Executives' forecasts of earnings per share versus forecasts of naïve models. Journal of Business October, 497–512.

Feng, M., W. Ge, C. Li, N. Nagarajan. 2011. Repeated lies? How earnings manipulators guide investors. Working Paper. University of Pittsburgh.

Feng, M., Li, C., McVay, S., 2009. Internal control and management guidance. Journal of Accounting and Economics 48, 190–209.

Feng, M., McVay, S., 2010. Analysts' incentives to overweight management guidance when revising their short-term earnings forecasts. The Accounting Review 85 (5), 1617–1646.

Foster, G., 1973. Stock market reactions to estimates of earnings per share by company officials. Journal of Accounting Research, 25–37.

Greene, W.H., 1993. Econometric Analysis. Macmillan, New York, NY.

Hayn, C., 1995. The information content of losses. Journal of Accounting and Economics 20 (2), 125–153.

Heflin, F., Subramanyam, K.R., Zhang, Y., 2003. Regulation FD and the financial information environment: early evidence. The Accounting Review 78 (1), 1–37.

Houston, J., Lev, B., Tucker, J., 2010. To guide or not to guide? Causes and consequences of stopping quarterly earnings guidance. Contemporary Accounting Research 27 (1), 143–185.

Hu, B. , C. Jiang. 2009. The impact of earnings guidance cessation on information asymmetry. Working Paper, University of Memphis.

Hutton, A., Miller, G., Skinner, D., 2003. The role of supplementary statements with management earnings forecasts. Journal of Accounting Research 41 (5), 867–890.

Karamanou, I., Vafeas, N., 2005. The association between corporate boards, audit committees, and management earnings forecasts: an empirical analysis. Journal of Accounting Research 43 (3), 453–486.

Krishan, G., Pevzner, M., Sengupta, M., 2011. How do auditors view managers' voluntary disclosure strategy? The effect of earnings guidance on audit fees. Journal of Accounting and Public Policy 30, 1–24.

Lennox, C., Park, C., 2006. The informativeness of earnings and management's issuance of earnings forecasts. Journal of Accounting and Economics 42, 439–458.

Lev, B., Penman, S., 1990. Voluntary forecast disclosure, nondisclosure, and stock prices. Journal of Accounting Research 28 (1), 49–76.

Matsumoto, D., 2002. Management's incentives to avoid negative earnings surprises. The Accounting Review 77 (3), 483–514.

McDonald, C.L., 1973. Examination of the reliability of published predictions of future earnings. The Accounting Review July, 502–510.

McNichols, M., O'Brien, P., 1997. Self-selection and analyst coverage. Journal of Accounting Research 35, 167–199.

Miller, G.S., 2002. Earnings performance and discretionary disclosure. Journal of Accounting Research 40 (1), 173–204.

Patell, J., 1976. Corporate forecasts of earnings per share and stock price behavior: empirical test. Journal of Accounting Research 14 (2), 246–276.

Penman, S., 1980. An empirical investigation of the voluntary disclosure of corporate earnings forecasts. Journal of Accounting Research 18 (1), 132–160.

Petersen, M., 2009. Estimating standard errors in finance panel data sets: comparing approaches. Review of Financial Studies 22 (1), 435–480.

Pownall, G., Wasley, C., Waymire, G., 1993. The stock price effects of alternative types of management earnings forecasts. The Accounting Review 68 (4), 896–912.

Pownall, G., Waymire, G., 1989. Voluntary disclosure credibility and security prices: evidence from management earnings forecasts, 1969–73. Journal of Accounting Research 27 (2), 227–245.

Rogers, J., Skinner, D., Van Buskirk, A., 2009. Earnings guidance and market uncertainty. Journal of Accounting and Economics 48, 90–109.

Rogers, J., Stocken, P., 2005. Credibility of management forecasts. The Accounting Review 80 (4), 1233–1260.

Ruland, W., 1978. The accuracy of forecasts by management and by financial analysts. The Accounting Review April, 439–447.
Runyan, B., Smith, L., 2007. The effect of multi-nationality on the precision of management earnings forecasts. International Journal of Accounting, Auditing and Performance Evaluation 4 (6), 572–588.
Seybert, N., Yang, H., 2012. The party's over: the role of earnings guidance in unraveling sentiment-driven overvaluation. Management Science 58, 308–319.
Skinner, D., 1994. Why firms voluntarily disclose bad news. Journal of Accounting Research 32 (1), 38–60.
Skinner, D., Sloan, R., 2002. Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. Review of Accounting Studies 7 (2–3), 289–312.
Wasley, C., Wu, J., 2006. Why do managers voluntarily issue cash flow forecasts? Journal of Accounting Research 44 (2), 389–429.
Wang, I.Y., 2007. Private earnings guidance and its implications for disclosure regulation. The Accounting Review 82 (5), 1299–1332.
Waymire, G., 1984. Additional evidence on the information content of management earnings forecasts. Journal of Accounting Research 22 (2), 703–718.
Xu, H., 2010. Do management earnings forecasts incorporate information in accruals? Journal of Accounting and Economics 49, 227–246.