Identification of analyst coverage initiations

JAY HEON JUNG
City, University of London
jay.jung@city.ac.uk

YONG YU
University of Texas at Austin
yong.yu@mccombs.utexas.edu

WUYANG ZHAO
University of Texas at Austin
wuyang.zhao@mccombs.utexas.edu

June 2023

Abstract

Researchers have traditionally inferred analysts' coverage initiations using the first recommendation issued by an analyst, by a broker, or by both an analyst and a broker in IBES. Using a large hand-collected sample of analysts' reports announcing coverage initiations, we examine measurement errors in these traditional methods of inferring initiations. We find that all of these methods generate significant Type I (misclassification) and Type II (omission) errors, and the nature and extent of these errors vary systematically across the methods. We show that the measurement errors introduce substantial sample biases and correcting for the errors can significantly impact research findings. We evaluate the effectiveness of the approaches that prior studies have used to mitigate the omission error in the traditional methods, and propose a new approach that can more effectively reduce the measurement errors.

Keywords: analyst; coverage initiation; measurement error; research implication

JEL Classification Codes: G14, M41

We appreciate helpful comments and suggestions from Mark Bradshaw, Shuping Chen, Peter Easton, Jonas Heese, Victoria Ivashina, Ross Jennings, Minjae Kim, Yongtae Kim, Kevin Li, Siqi Li, Yuan Li, Stanimir Markov, Maria Ogneva, Lakshmanan Shivakumar, Jennifer Wu Tucker, Christopher Yust, and seminar participants at City, University of London, Fudan University, Hong Kong Polytechnic University, Santa Clara University, University of Manchester, University of Texas at Austin, Yonsei University, the 2021 Lone Star Accounting Conference, and 4th Analyst Research Conference. We thank Yiying Chen, Mingqi Li, Wuyou Xie, and Lu Zhao for their excellent research assistance. All errors remain our own.

1. Introduction

Financial analysts play a crucial role as information intermediaries in capital markets, influencing managerial and investors' decisions, firm value, and market efficiency. Numerous studies in accounting and finance have examined analysts' coverage initiations, i.e., analysts adding a stock into their coverage portfolio (see Appendix A). However, due to the lack of readily available machine-readable data on analyst initiations, researchers have traditionally inferred initiations using the first recommendation issued by an analyst (referred to as the "first-by-analyst" method), by a broker (referred to as the "first-by-broker" method), or by both an analyst and a broker (referred to as the "first-by-both" method) in IBES.

However, these methods can generate significant measurement errors. On the one hand, when IBES is incomplete and fails to capture the actual first recommendations, these methods will misidentify subsequent non-initiating recommendations as initiations, resulting in Type I or misidentification errors. On the other hand, analysts and brokers frequently re-initiate coverage on firms that they previously covered and dropped. By focusing on the first recommendation only, these methods will miss some or all of re-initiations, resulting in Type II or omission errors. Specifically, the first-by-analyst method misses re-initiations by the same analyst (i.e., analyst re-initiations), the first-by-broker method misses re-initiations by the same broker (i.e., broker re-initiations), and the first-by-both method misses all re-initiations. To mitigate this omission error, prior studies have modified the traditional methods to include more re-initiations (e.g., Malmendier and Shanthikumar 2007; Ljungqvist et al. 2009; Bernhardt et al. 2016).

Despite the widespread use of these traditional methods, surprisingly no prior research has examined the measurement errors in these methods or the resulting biases in samples and findings. In this paper, we provide the first evidence on these issues. We manually collect initiating reports

– the reports in which analysts announce a coverage initiation – from the Thomson ONE database, and construct a large sample of actual initiations that we directly observe from these initiating reports (observed initiations, hereafter). To ensure that our results do not reflect the data coverage difference between Thomson ONE and IBES, we focus on their intersection, i.e., the common broker-firm-years in both databases, over the period of 2003 – 2017. Within this overlapping universe, we identify 28,139 observed initiations from analysts' initiating reports, and three samples of inferred initiations based on the traditional methods: 29,391 for the first-by-analyst method, 22,823 for the first-by-broker method, and 17,229 for the first-by-both method.

We begin by using our observed initiation sample to provide insights into first-time initiations versus re-initiations. They both account for about half of all observed initiations. Among re-initiations, about 56% are analyst re-initiations, and 67% are broker re-initiations, with the overlapping 23% being re-initiations by the same broker-analyst. While first-time initiations and re-initiations share similarities in reasons for coverage initiation and reporting format, they differ systematically in firm and analyst characteristics. For example, re-initiations are skewed toward older and larger firms with higher institutional ownership and analyst following, and analysts from larger brokers, with all-star status, and providing more accurate and frequent forecasts.

We then turn to our main analyses, focusing on examining the measurement errors in the three traditional methods of inferring initiations. First, we find that each method generates a large omission (Type II) error by missing some or all of re-initiations. The first-by-analyst method includes broker re-initiations, but misses analyst re-initiations; the first-by-broker method includes analyst re-initiations, but misses broker re-initiations; and the first-by-both method misses all re-initiations. Specifically, the first-by-analyst, first-by-broker, and first-by-both methods miss 28% (7,757), 33% (9,308) and 49% (13,841) of all observed initiations, respectively. Second, due to

the incompleteness of IBES (i.e., failing to capture the actual first recommendations), all three methods also generate a significant misidentification (Type I) error. The first-by-analyst, first-by-broker, and first-by-both methods misidentify 31% (9,009), 18% (3,992), and 17% (2,931) of inferred initiations, respectively. Overall, based on the total number of missed and misidentified initiations, the first-by-broker method outperforms the other two traditional methods. In addition, a trend analysis shows that both types of measurement errors are persistent and ongoing.

We next examine sample biases caused by the measurement errors in the traditional methods. We find that each method introduces distinct sample biases, mainly due to omission errors (i.e., missing some or all re-initiations). For firm characteristics, all three methods tilt the sample towards smaller/younger firms, which is expected as re-initiations are associated with larger/older firms. However, for analyst characteristics, the first-by-analyst method skews the sample toward analysts from larger brokers with less experience, while the first-by-broker method tilts the sample toward analysts from smaller brokers with more experience, and the first-by-both method tilts the sample toward analysts from smaller brokers with less experience. Compared with observed initiations, the first-by-analyst method introduces a downward bias in initiating ratings mainly due to misidentification errors, while the first-by-broker and first-by-both methods introduce an upward bias in initiating ratings mainly due to omission errors. These patterns generally hold even after excluding re-initiations with a non-coverage period (i.e., the period between the previous coverage stoppage and the current re-initiation) of less than six months.

To further illustrate the research implications of the measurement errors in the traditional methods, we examine their effects on the findings related to analysts' self-selection (McNichols and O'Brien 1997) and the common determinants of initiations (e.g., Ertimur et al. 2011; Kirk 2011). Compared with observed initiations, the first-by-broker and first-by-both methods overstate

analysts' self-selection by 34% and 31%, respectively, due to omission errors, while the first-by-analyst method understates it by 22% due to misidentification errors. Furthermore, all three methods introduce substantial biases of varying directions and magnitudes to some commonly used determinants of initiations, including analyst following and institutional ownership. These biases mainly stem from omission errors. These results highlight that correcting for the measurement errors in the traditional methods can have a significant impact on research findings.

We then evaluate the effectiveness of three approaches that prior studies have used to mitigate the omission error in the traditional methods: (1) including a broker's new recommendation for a firm since its previous coverage stoppage (Malmendier and Shanthikumar 2007), (2) including a broker's new recommendation for a firm since its previous coverage stoppage or with no preceding recommendation in the past 12 months (Ljungqvist et al. 2009), and (3) including an analyst's recommendation if the analyst did not issue any preceding recommendation in the past 12 months (Bernhardt et al. 2016). We find that while these approaches do significantly reduce the omission error by including more re-initiations, they simultaneously exacerbate the misidentification error to a similar or greater degree. For example, including a broker's new recommendation for a firm since its previous coverage stoppage picks up 8,114 additional re-initiations, but at the same time it also misidentifies 13,936 more initiations. Overall, these prior approaches fail to significantly reduce the overall measurement error.

Finally, we propose a new approach to better mitigate errors in inferring initiations. This approach builds upon the first-by-broker method, which has the lowest overall measurement error among the three traditional methods. It then utilizes earnings and target price forecasts in IBES to mitigate the misidentification error. To reduce the omission error, it further includes the first recommendation issued by a broker since its previous coverage stoppage, conditional on the broker

having issued no earnings or target price forecasts during its non-coverage period. We test varying lengths of non-coverage periods, and find that requiring a minimum non-coverage period of three months is most effective in reducing the overall measurement error. Specifically, it reduces (increases) the number of missed (misidentified) initiations by 6,099 (1,400), resulting in a net decrease of 4,699 in the total number of missed and misidentified initiations.

Our study contributes to the literature by providing the first systematic evaluation and comparison of the traditional methods of inferring analyst initiations from IBES. We provide comprehensive evidence on the nature and extent of measurement errors inherent in these methods, the sample biases they introduce, and their implications for research findings. We also examine the effectiveness of the approaches that researchers have used to mitigate the omission error in the traditional methods. Our results highlight that while these approaches succeed in reducing the omission error, they increase the misidentification error to a similar or greater degree, resulting in no significant decrease in the overall measurement error. We propose a new approach that can mitigate the overall measurement error more effectively. In this way, our study also adds to the growing literature on the use of IBES database (e.g., Ljungqvist et al. 2009; Call et al. 2021; Amiram et al. 2021; Kaplan et al. 2021).

Our findings are relevant to researchers interested in analyzing analysts' initiations. Through a comprehensive evaluation of the traditional methods of inferring initiations, our study enables researchers to assess the potential biases introduced by each method to their samples and results, and consequently make informed decisions when selecting the method that best suits their research questions and settings. Our study shows that the first-by-broker method outperforms the other two traditional methods when equal weight is given to misidentification and omission errors. In cases where different weights are required for the two types of errors, our results on the separate

impact of each error type also allow researchers to select a method that strikes a suitable balance. Furthermore, for researchers seeking to further mitigate the measurement errors in the traditional methods, our proposed new approach provides a solution to achieve this objective.

The next section discusses the traditional methods and their measurement errors. Section 3 describes our sample selection. Sections 4 presents the main results. Section 5 evaluates prior approaches to mitigate the omission error and proposes a new approach. Section 6 concludes.

2. Traditional methods of inferring initiations and their measurement errors

We conduct a comprehensive search of research published between 1990 and 2020 in ten prestigious accounting and finance journals.[1] Specifically, we begin by searching on Google Scholar for all papers in these journals that contain the following two words, "analyst" and "cover," and at least one initiation-related word (i.e., "initiation," "initiate," "initiating" or "initial"). We then read every paper identified from this search to confirm that the paper includes analysts' initiations of coverage in its analyses. This procedure yields 50 papers.[2] We categorize these papers into two groups based on how they identify initiations. The first group infers initiations using an analyst's and/or a broker's first research outputs in databases such as IBES. The second group identifies initiations by directly observing analyst reports or press releases. Appendix A lists the papers along with their respective methods of identifying initiations.

[1]The journals include, in alphabetical order, *Contemporary Accounting Research* (CAR), *Journal of Accounting & Economics* (JAE), *Journal of Accounting Research* (JAR), *Journal of Finance* (JF), *Journal of Financial Economics* (JFE), *Journal of Financial and Quantitative Analysis* (JFQA), *Management Science* (MS), *Review of Accounting Studies* (RAS), *Review of Financial Studies* (RFS), and *The Accounting Review* (TAR).

[2] We acknowledge that our list is not exhaustive. We may have inadvertently overlooked certain papers in these ten journals. Also, we do not consider papers published in other academic journals or unpublished working papers.

Out of the 50 papers, only eight directly identify initiations from analysts' reports, press releases, and business websites, partly due to the concern about the incompleteness of IBES data.[3] However, it is worth noting that due to the cost of data collection, the sample sizes in these studies are relatively small. All these eight studies have sample sizes below 2,000, except for Bradley et al. (2008) with a sample size of 7,487 and Liu and Ritter (2011) with a sample size of 7,319.

To generate a large sample of initiations, the remaining 42 papers rely on the traditional method of inferring initiations based on the first recommendation (or forecast) issued by an analyst, a broker, or both in an analyst database. Except for two early studies (McNichols and O'Brien 1997; Barber et al. 2001), all of these 42 studies use IBES. More specifically, 14 papers use the first-by-analyst method, five papers use the first-by-broker method, and five papers use the first-by-both method. The remaining papers do not provide sufficient information for us to determine which specific traditional method is used.

As discussed in the introduction, the traditional methods of inferring initiations from IBES can generate significant measurement errors. On the one hand, it can misidentify initiations (misidentification or Type I error) due to the incompleteness of the IBES database. Prior research finds that recommendations and forecasts are not always included in IBES for various reasons, such as systematic process errors within IBES (Ljungqvist et al. 2009) and removal requests from brokers (Ljungqvist et al. 2009; Call et al. 2021). Thus, when the actual initiating recommendation is missing from IBES, the first recommendation on IBES will be misidentified as an initiation.[4]

[3] For example, concerned that "IBES is far from comprehensive in its analyst recommendations file," Liu and Ritter (2011) augment IBES recommendations with hand-collected information from Bloomberg, Briefing.com, First Call, Investext, and other online searches to identify analyst initiations.

[4] As we will discuss in Section 3, to ensure that our results are not affected by the coverage differences between Thomson ONE and IBES, our analyses focus on the common coverage overlap between the two databases. Thus, for misidentified initiations, their corresponding true initiations occur prior to the common coverage overlap and thereby are not included in our observed initiation sample.

Exhibit A presents an example of this misidentification error. Exhibit A-1 shows that according to IBES, broker ThinkEquity issued its first recommendation for Cree, Inc. on January 17, 2003. However, this recommendation is not an actual initiation, as shown by its corresponding analyst report retrieved from Thomson ONE in Exhibit A-2. Note that the report explicitly says that "We are raising our forward estimates" and provides its previous forecast estimates, indicating that this recommendation is not an initiation. Exhibit A-3 presents the actual initiation, which was issued earlier on January 9, 2003, but is missing from IBES (unfortunately, we do not have data to find the reason for its absence from IBES). In this case, the traditional method will misidentify the first recommendation in IBES, issued on January 17, 2003, as an initiation.

On the other hand, the traditional methods can miss actual initiations (omission or Type II error). It is common for analysts and brokers to re-initiate coverage on firms that they had previously covered and dropped (e.g., Ljungqvist et al. 2009). However, by focusing on the first recommendation only, the traditional methods miss some or all re-initiations. Specifically, the first-by-analyst method misses re-initiations by the same analyst (i.e., analyst re-initiation), the first-by-broker method misses re-initiations by the same broker (i.e., broker re-initiation), and the first-by-both method misses all re-initiations.

Exhibit B presents an example of this omission error. During our sample period, Jefferies analyst Paul Fremont initiated coverage on Southern Union Company twice. The first initiation occurred on March 12, 2004 (see Exhibit B-2), followed by the stoppage of coverage on April 27, 2005 (based on the IBES recommendation stop file in Exhibit B-3). He provided no coverage over the following six years ending on May 23, 2011 (dated May 24, 2011 in IBES), when he initiated coverage on this firm for the second time (see Exhibit B-4). However, all the three traditional

methods, which focus on the first recommendation in the IBES database, will only capture the first initiation in 2004 and miss the second initiation in 2011.

The measurement error problem in the traditional methods has not gone unnoticed in the literature. To address the omission error and capture more re-initiations, prior studies have employed several approaches to modify the traditional methods. For example, Malmendier and Shanthikumar (2007) additionally include a broker's first recommendation for a firm since its previous stoppage of coverage. Bernhardt et al. (2016) include an analyst's recommendation as an initiation if the analyst has issued no preceding recommendation for the firm in the past year. Ljungqvist et al. (2009) include a recommendation as an initiation if it is issued by a broker that had previously stopped coverage on the firm or if it is not preceded by another recommendation from the same broker in the past 12 months. However, the effectiveness of these approaches in mitigating the measurement errors in the traditional methods remains unknown.

Our study provides the first comprehensive analysis of the measurement errors in the traditional methods of inferring initiations and the resulting sample biases. We also investigate the effectiveness of the prior approaches to mitigate the omission error in the traditional methods, and propose a new approach that can reduce the errors more effectively.

3. Sample

To assemble our sample of observed initiations, we manually download all analyst reports announcing coverage initiations from the Thomson ONE database over the period of 2003 – 2017. This process yields a total of 59,074 initiating reports. To verify that these reports indeed capture analysts' coverage initiations, we conduct a textual analysis of the first two pages of each report,

searching for keywords related to initiation or re-initiation.[5] See Exhibits A and B for the examples of initiating reports. Only 2,177 reports, or 4% of the total, do not contain any keyword. Further investigation reveals that the vast majority of these 2,177 reports are from two small brokers, CFRA Equity Research and BWS Financial, which we exclude from our sample. To further validate our observed initiation sample, we randomly select 100 initiating reports and 100 non-initiating reports and review each one. We find that all the reports are correctly classified. Next, for our samples of inferred initiations, we follow prior research (e.g., McNichols and O'Brien 1997; Kelly and Ljungqvist 2012; Li and You 2015) to use the first recommendation by an analyst (i.e., the first-by-analyst method), by a broker (i.e., the first-by-broker method), or by both (i.e., the first-by-both method) on a firm from IBES. We use first recommendations rather than first earnings forecasts to be consistent with most of the prior literature. For example, out of the 42 studies listed in Appendix A that infer initiations, only seven explicitly use first forecasts. We check a random sample of 100 initiating reports and find that all of these reports include recommendations as well as earnings forecasts.[6]

To ensure that our results do not reflect any database coverage difference between IBES and Thomson ONE, we focus our analyses on the coverage intersection of the two databases, i.e.,

[5] These keywords are initiat, reinitiat, re-initiat, resum, assum, reassum, re-assum, reinstat, re-instat, reinstall, re-install, launch, and initial report.

[6] In untabulated analysis, we also examine how often the first recommendations have same-day first earnings forecasts in IBES. We observe that the populations of firm-broker-analysts in IBES recommendation detail file and EPS detail file differ, indicating that some analysts or brokers with recommendation records in IBES recommendation detail file may not appear in IBES EPS detail file, and vice versa. We focus on the first recommendations whose firm-analysts (for the first-by-analyst method), firm-brokers (for the first-by-broker method), or firm-analyst-brokers (for the first-by-both method) are also covered in the IBES EPS detail file. We find that (1) 74.9%, 88.3% and 89.5% of the first recommendations based on the first-by-analyst, first-by-broker, and first-by-both methods, respectively, have same-day first earnings forecasts, (2) 18.0%, 8.1% and 3.6% of the first recommendations are preceded by the first forecasts, and (3) 7.1%, 3.6%, and 6.9% of the first recommendations are followed by the first forecasts. The fact that first recommendations do not always coincide with or precede the first forecast suggests the possibility that IBES may miss the actual initiating recommendations while having the actual initiating earnings forecasts. We use this insight to introduce an adjustment in our proposed approach to reduce misidentification errors in the traditional methods (see Section 5.2). We are grateful to the referee for raising this important point.

the common broker-firm-years in both databases over 2003-2017. Following Chan et al. (2018), we first consider all brokers that have a minimum of 300 observations in the IBES recommendation file over 2003 – 2017. We then verify whether each broker also appears in Thomson ONE, and identify the overlapping periods of data coverage for 91 brokers with coverage in both databases.

Following prior studies (e.g., Crawford et al. 2012), we impose two additional restrictions. First, we require all initiations, both inferred and observed, to be issued after the first three months of a firm's appearance in the CRSP database to exclude coverage initiations for IPO firms. Second, we require all initiations to be issued after the first six months of a broker's appearance in IBES. This requirement prevents the mechanical counting of recommendations issued by newly added brokers in IBES as initiations.

The above procedures yield a sample of 28,139 observed initiations and three samples of 29,391, 22,823, and 17,229 inferred initiations based on the first-by-analyst, first-by-broker, and first-by-both methods, respectively. These samples are derived from the common set of broker-firm-years present in both IBES and Thomson ONE over the period of 2003 – 2017.

4. Results

4.1. Comparing first-time initiations vs. re-initiations using the observed initiation sample

We begin by using our observed initiation sample to provide information on first-time initiations versus re-initiations. Figure 1 shows that out of 28,139 observed initiations, 51% (14,298) are first-time initiations, and 49% (13,841) are re-initiations. Among the 13,841 re-initiations, 4,533 are re-initiations by analysts only, 6,084 are re-initiations by brokers only, and 3,224 are re-initiations by both analysts and brokers. In other words, among re-initiations, 56% (i.e., (4,533+3,224)/13,841) are analyst re-initiations, and 67% (i.e., (6,084+3,224)/13,841) are

broker re-initiations, with the overlapping 23% (i.e., 3224/13,841) being re-initiations by both analysts and brokers.

We manually review a random number of 200 initiating reports for re-initiations and 200 initiating reports for first-time initiations. We find no systematic differences in the format of the reports. Further, we find that analysts' underlying motives for initiating coverage are similar between first-time initiations and re-initiations. In 192 initiating reports for re-initiations, analysts explicitly mention factors such as firms' growth, performance, and management team as the primary reasons for re-initiations, which are very similar to the motives for the first-time initiations that we review and are also consistent with prior findings (e.g., McNichols and O'Brien 1997; Jung et al. 2015).[7] Exhibit B provides such an example.

Table 1 compares initiation keywords and report lengths between first-time initiations and re-initiations. We present the percentage of reports that include the keywords, "initiation," "initiating," "initiate," or other derivatives of "initiation." We find that 99.8% (=14,165/14,197) of first-time initiations include such keywords, and 92.7% (=7,041/7,593) of analyst re-initiations and 86.4% (=7,874/9,110) of broker re-initiations also use the keyword "initiation" or its derivatives in their reports.[8] In terms of report length, we find that the average lengths of reports are 20.5, 16.8, and 19.6 pages for first-time initiations, analyst re-initiations, and broker re-initiations, respectively, which are substantially longer than the average length of 7.6 pages for non-initiation reports in Thomson ONE.[9] In addition, Table 1 also presents the average length of

[7] For the remaining eight reports, seven are sell recommendations and one is a coverage transfer that does not provide motives. The proportion of the re-initiations that contain sell recommendations, 3.5% (=7/200), is similar to what prior studies have documented for first-time initiations (e.g., 3% in Ertimur et al. (2011)).

[8] We note that 463 initiation reports are excluded from the test as we detect no initiation-related keyword from these reports. It is possible that these reports contain keywords less commonly used or unconventional in practice or keywords presented in picture format, making their detection challenging using our textual analysis program.

[9] We scape the "Table of Contents" for the entire Thomson ONE database and collect each analyst report's identifying information (e.g., Firm ticker, broker name, lead analyst name, and the length of reports in number of pages) in Thomson ONE without downloading them.

non-coverage periods – the mean number of days between analysts' or brokers' re-initiations and their preceding recommendations issued on the same firms – for re-initiations.[10] It is 364 days for analyst re-initiations and 846 days for broker re-initiations.

Table 2 compares firm, analyst, and rating characteristics between first-time initiations and re-initiations. Variable definitions are provided in Appendix B. The sample sizes are reduced due to requiring non-missing data for these characteristics. All continuous variables are winsorized at the 1% and 99% levels. We find that most of the firm, analyst, and rating characteristics differ significantly between first-time initiations and re-initiations. For example, for firm characteristics, compared to first-time initiations, analyst and broker re-initiations are associated with firms that are significantly older (*FIRMAGE*) and larger (*Ln* (*MV*)), and have more intangible assets (*INTANGIBLE*), higher institutional ownership (*IO*), higher leverage (*LEVERAGE*), more analysts following (*NANALYST*), more segments (*NSEGMENT*), and better firm performance (*ROA*). We also see significant differences in almost all analyst characteristics. Compared with first-time initiations, re-initiations are associated with analysts who are more likely to have all-star status (*ALLSTAR*), work for larger brokers (*BSIZE*), have more experience in the profession (*GEXP*),[11] cover more firms (*NFIRM*) but fewer industries (*NIND*), issue more accurate earnings forecasts (*ACCURACY*), forecast more items (*NFITEM*), and forecast more frequently (*FREQ*). Additionally, we find significant differences in all analyst characteristics between analyst re-initiations and broker re-initiations. Finally, in terms of stock ratings, compared to first-time

[10] We calculate the non-coverage period using the IBES's review date, i.e., the last confirmed date of an outstanding recommendation. For this test, we exclude 146 analyst re-initiations, for which the non-coverage period is missing due to erroneous review dates.

[11] We do not find a significant difference in analyst experience when comparing analysts associated with first-time initiations and those associated with broker re-initiations.

initiations, re-initiations have a significantly lower percentage of buy ratings (Mean (*BUY RATING*)).[12]

4.2. Evaluating the three traditional methods

4.2.1. Measurement errors of the traditional methods

Figure 2 presents three Venn diagrams that compare the observed initiation sample with each of the three inferred initiation samples. The left circle represents the observed initiation sample, and the right circle represents one of the three inferred initiation samples.

The diagrams show that all three traditional methods of inferring initiations generate a large omission (Type II) error by excluding a portion of or all re-initiations. Specifically, Figure 2-A shows that the first-by-analyst method includes broker re-initiations, but misses 7,757 analyst re-initiations, resulting in an omission (type II) error rate – the number of missed initiations divided by the number of observed initiations (i.e., the sum of correctly identified and missed initiations) – of 28% (7,757/28,139). In contrast, Figure 2-B shows that the first-by-broker method includes analyst re-initiations, but misses 9,308 broker re-initiations, resulting in an omission error rate of 33% (9,308/28,139). Figure 2-C shows that the first-by-both method misses all 13,841 re-initiations, resulting in an omission error rate of 49% (13,841/28,139).

The diagrams also show that all three traditional methods generate significant misidentification (Type I) error. Misidentified initiations refer to the inferred initiations that are not present in the observed initiation sample. To ensure that these inferred initiations are indeed misidentified, we retrieve from Thomson ONE the record of all reports on the same firms by the same analysts or brokers over the period around each inferred initiation. We find that all of these

[12] Following the analyst literature (e.g., Clement and Tse 2003), we also compare relative rank variables of firm and analyst characteristics. The results, presented in the Online Appendix, remain qualitatively similar to those in Table 2, with the exception of *BM* and *BSIZE*, which exhibit significant differences in the opposite direction.

inferred initiations have either a same-day non-initiating report or at least one preceding report by the same analyst or broker within the past six months. This evidence suggests that these inferred initiations are indeed misidentified due to the incompleteness of IBES (i.e., missing the reports prior to these inferred initiations). Figure 2-A shows that the first-by-analyst method generates 9,009 misidentified initiations, resulting in a misidentification (Type I) error rate – the number of misidentified initiations divided by the number of inferred initiations (i.e., the sum of correctly identified and misidentified initiations) – of 31% (9,009/29,391). Figure 2-B shows that the first-by-broker method generates 3,992 misidentified initiations, resulting in a misidentification error rate of 17% (3,992/22,823). Figure 2-C shows that the first-by-both method generates 2,931 misidentified initiations, resulting in a misidentification error rate of 17% (2,931/17,229).[13]

Figure 3 presents the time trends of the misidentification and omission error rates by year. We find that the omission error rates are relatively stable across the years around 28%, 33%, and 49% for the first-by-analyst (Figure 3-A), first-by-broker (Figure 3-B), and first-by-both (Figure 3-C) methods, respectively. In contrast, the misidentification error rates are relatively more volatile over time. For example, in Figure 3-C, it remains between 10% and 20% in most years but shows a spike during the period of 2007 - 2009 with the highest of 32% in 2009.

[13] As we focus on the overlapping coverage universe between Thomson ONE and IBES, the true initiations corresponding to misidentified initiations, which occurred prior to the common coverage period, are not included in our observed initiation sample. However, we are able to locate a small portion of these true initiation reports. For example, for 9,009 misidentified initiations using the first-by-analyst method, we find their respective true initiation reports in Thomson ONE for 5.5% or 496 of them. Using the first-by-broker (first-by-both) method, we find their respective true initiations for 7.5% (8.1%) or 300 (237) of the misidentified initiations. The limited availability of true initiation reports is primarily because Thomson ONE's coverage does not extend far enough. Often, Thomson ONE coverage only precedes IBES coverage for a short period, and consequently, does not include the true initiation report. One possible reason is that both databases started coverage around the same time. In an untabulated analysis, we compare this small sample of misidentified initiations with their corresponding true initiations. We find the mean (median) number of days between a misidentified initiation and its respective preceding true initiation is 57 (31) days, 72 (11) days, and 65 (11) days for the first-by-analyst, first-by-broker, and first-by-both methods, respectively. In terms of firm, analyst, and rating characteristics, we do not find any significant difference, which is not surprising given the small sample size and the short time span. We thank our anonymous referee for suggesting these comparisons.

Overall, these results indicate that all three traditional methods of inferring initiations generate significant misidentification (Type I) and omission (Type II) errors. The first-by-analyst method generates the highest misidentification error rate (31%), while the first-by-both method generates the highest omission error rate (49%). In terms of the total number of missed and misidentified initiations, the first-by-broker method outperforms the other two methods. Moreover, these measurement error rates are persistent and significant over time.

4.2.2. Sample bias generated by the traditional methods

Table 3 compares firm, analyst, and rating characteristics between the observed initiation sample and the three inferred initiations samples. Panels A, B and C present the results for the first-by-analyst, first-by-broker, and first-by-both methods, respectively. The sample sizes in Table 3 are smaller than those in Figure 2 due to requiring non-missing data for measuring these characteristics. To mitigate the influence of outliers, all continuous raw variables are winsorized at the 1% and 99% levels. Detailed variable definitions are provided in Appendix B.

We find that most of the firm, analyst, and rating characteristics differ significantly between the observed initiation sample and the inferred initiation samples. Specifically, Panel A of Table 3 compares observed initiations (Column (1)) with inferred initiations using the first-by-analyst method (Column (3)), with the differences between the two reported in Column (4). Compared with observed initiations, inferred initiations are associated with firms that are younger (*FIRMAGE*) and smaller (*Ln* (*MV*)), have lower institutional ownership (*IO*), lower past 12-month returns (*MOMENTUM*), fewer analysts following (*NANALYST*), and poorer firm performance (*ROA*), and are associated with analysts working for larger brokers (*BSIZE*), having less experience in the profession (*GEXP*), covering fewer firms (*NFIRM*) but more industries (*NIND*), forecasting fewer items (*NFITEM*), and issuing forecasts less frequently (*FREQ*). Furthermore, compared with

observed initiations, inferred initiations exhibit a downward bias in stock ratings, as reflected by the lower percentage of buy ratings (Mean (*BUY RATING*)).

To assess the impacts of misidentification and omission errors, we further compare firm, analyst, and rating characteristics of observed initiations (Column (1)) or inferred initiations (Column (3)) with those of correctly identified initiations (Column (2)). Correctly identified initiations refer to observed initiations that are correctly identified and included in the inferred initiation sample. As observed initiations include both correctly identified and missed initiations, the differences between Columns (1) and (2) capture the impact of omission errors and are reported in Column (5). Similarly, as inferred initiations include both correctly identified and misidentified initiations, the differences between Columns (2) and (3) captures the impact of misidentification errors and are reported in Column (6). We find that for analyst experience (*GEXP*), the number of forecasted items (*NFITEM*), and forecast frequency (*FREQ*), both the omission error (Column (5)) and the misidentification error (Column (6)) introduce downward biases to the inferred initiation sample, resulting in the negative overall impacts (Column (4)) of using the first-by-analyst method. However, when the biases of the two errors are significant but in opposite directions, in all such cases, except for *MOMENTUM*, the omission error dominates as its magnitude is two to three times larger than that of the misidentification error.[14]

Panels B and C repeat the analyses for inferred initiations using the first-by-broker and first-by-both methods, respectively. Regarding firm characteristics, we find similar results to those in Panel A. For example, the first-by-broker and first-by-both methods also generate inferred initiation samples that are biased toward firms that are smaller, younger, and less profitable, and

[14] For example, firm size, *Ln (MV)*, is biased downward by the omission error but upward by the misidentification error. The overall impact on firm size is a downward bias because the omission error (-0.211) is three times larger in magnitude than the misidentification error (0.071).

have lower institutional ownership, and fewer analysts following. These consistent patterns in firm characteristics across all three methods is due to these methods all missing part or all of re-initiations, which are typically issued for firms in the later stage of their life cycle (See Table 2). Consequently, missing re-initiations biases the inferred initiation samples toward firms in the earlier stage of life cycle, which tend to be smaller, younger, underperforming, and less visible.

Interestingly, we find different patterns of biases in analyst and rating characteristics across the three methods. For example, as shown in Panel A, the first-by-analyst method generates biases towards less experienced analysts from larger brokers, whereas the first-by-broker method (Panel B) generates biases towards more experienced analysts from smaller brokers, and the first-by-both method (Panel C) generates biases towards less experienced analysts from smaller brokers. These patterns are also driven by missing re-initiations. The first-by-analyst method excludes analyst re-initiations, which are more likely to be issued by analysts in the later stage of their career, resulting in a lower average analyst experience in its inferred initiation sample. Similarly, the first-by-broker method excludes broker re-initiations, which tend to be issued in the later stage of the broker's life cycle, resulting in a lowering average broker size in its inferred initiation sample. Lastly, the first-by-both method excludes all analysts and broker re-initiations, resulting in a lower average in both analyst experience and broker size in its inferred initiation sample. Furthermore, in terms of stock ratings, the first-by-analyst method generates a downward bias in ratings due to its inclusion of a large number of misidentified initiations, which are less favorable than correctly identified initiations. In contrast, the first-by-broker and first-by-both methods (Panels B and C, respectively) generate an upward bias in ratings due to their omission of broker re-initiations, which are less favorable than first-time initiations and analyst re-initiations (see Table 2). Overall, Table 3 shows

that three different methods generate inferred initiation samples with a similar pattern of biases in firm characteristics but distinct patterns of biases in analyst and rating characteristics.

Next, we estimate the following probit model to examine the relations between the two measurement errors and firm and analyst characteristics:

$$\begin{aligned} &MISIDENTIFIED_{ijtk} \text{ (or } MISSED_{ijtk}) = \\ &\quad BM_{it-1} + FIRMAGE_{it} + INTANGIBLE_{it-1} + IO_{it-1} + LEVERAGE_{it-1} + MOMENTUM_{it-1} \\ &\quad + Ln\ (MV)_{it-1} + NANALYST_{it-1} + NSEGMENT_{it-1} + ROA_{it-1} + ALLSTAR_{jt-1} + BSIZE_{jt-1} \\ &\quad + GEXP_{jt} + NFIRM_{jt-1} + NFITEM_{ijt} + NIND_{jt-1} + ACCURACY_{jt-1} + FREQ_{jt-1} + \varepsilon_{ijtk} \end{aligned} \quad (1)$$

The equation is estimated at the firm-analyst-year-initiation level. The subscripts *i*, *j*, *t*, and *k* denote a firm, analyst, year, and initiation, respectively. The dependent variable is *MISIDENTIFIED* (*MISSED*), an indicator variable that equals one if a recommendation is a misidentified (missed) initiation and zero otherwise. Correctly identified initiations are used as the benchmark group. The independent variables are the commonly examined firm and analyst characteristics as in Table 3.[15] We cluster standard errors by firm (Petersen 2009).

Table 4 reports the regression results. We find that the two measurement errors are systematically associated with various firm and analyst characteristics. For example, Columns (1), (2), and (3) show that *MISIDENTIFIED* is negatively associated with firms' past 12-month return and forecast frequency, and positively associated with analyst following and analysts' all-star status. Columns (4), (5), and (6) indicate that *MISSED* is positively associated with firm age, institutional ownership, leverage, analyst following, the number of firms an analyst covers, and forecast accuracy.[16] Overall, those results are largely consistent with the patterns in Table 3,

[15] We do not include fixed effects in Equation (1) because the goal of the test is to evaluate the direct associations between measurement errors and various firm and analyst characteristics.

[16] Positive associations between *MISSED* and firm and analyst characteristics suggest negative impacts of the omission error on the characteristics (e.g., firm age) because excluding missed initiations will cause a downward bias in the characteristics (e.g., younger firms) for the sample.

indicating that both misidentification and omission errors in the traditional methods of inferring initiations introduce systematic biases in firm- and analyst-level characteristics.

4.2.3. Robustness tests for re-initiations

Despite our findings indicating similarities in initiating reports and motives between first-time initiations and re-initiations, it is still possible that some re-initiations may reflect temporary coverage suspensions rather than coverage stoppages. For example, brokers may temporarily suspend and subsequently resume their coverage on a firm due to reasons such as underwriting business relationships or the implementation of a new rating scale (e.g., Agrawal and Chen 2008; Philippot 2018). To mitigate the potential influence of re-initiations representing temporary suspensions, we conduct a robust test requiring that re-initiations be preceded by a minimum non-coverage period of six months.[17] While the six-month cutoff is rather conservative, it serves the purpose of excluding re-initiations that are likely to represent temporary suspensions.

Table 5 reports the results of repeating the analyses in Table 3 after excluding re-initiations with a non-coverage period shorter than six months. For the first-by-broker and first-by-both methods, the direction of biases remains the same for all the firm- and analyst-level characteristics as well as stock ratings. For the first-by-analyst method, the variable capturing the number of firms an analyst covers (*NFIRM*) becomes significant in the opposite direction. Overall, these results suggest that the sample biases we find stem from the measurement errors inherent in the traditional methods rather than from certain re-initiations that may represent temporary suspensions.

4.3. Impact on research findings

[17] If a re-initiation is by both an analyst and a broker, we take the shorter of the two non-coverage periods calculated based on the analyst's and the broker's preceding recommendations.

To further illustrate the research implications of the measurement errors in the traditional methods, we examine the effect of the errors on two sets of prior findings related to analyst initiations. First, we examine how the errors affect the degree of analysts' self-selection. McNichols and O'Brien (1997) hypothesize that due to the significant start-up costs of adding a new firm, analysts require more favorable information to add a new firm than to maintain coverage of previously covered firms. Consistent with this self-selection hypothesis, they find that analysts' initial ratings for newly added firms are significantly more favorable than their ratings for firms they have previously covered. Following their approach, we measure the degree of analyst self-selection using the difference in the proportions of buy ratings between initiations and non-initiations. As in the main analyses, we focus on the common coverage overlap between IBES and Thomson ONE. For the tests using inferred initiations, non-initiations refer to recommendations in IBES that are not inferred as initiations based on the traditional method used in the test. We identify 129,731, 142,941, and 152,557 non-initiations using the first-by-analyst, first-by-broker, and first-by-both methods, respectively. For the tests using observed initiations, non-initiations are 66,139 non-initiating recommendations in Thomson ONE that have valid ratings in IBES.

Panel A of Table 6 presents the results. While the self-selection measure is positive (i.e., initiations have higher proportions of buys than non-initiations do) in all four samples, the degree of self-selection varies significantly across the samples. Compared with the observed initiation sample (self-selection = 14.3%), the inferred initiation samples based on the first-by-broker and first-by-both methods exhibit higher self-selection of 19.1% and 18.8%, overstating the degree of self-selection by 34% and 31%, respectively. The primary reason for this overstatement is that both the first-by-broker and first-by-both methods miss broker re-initiations, whose ratings are significantly less favorable than first-time initiations and analyst re-initiations (see Table 2). In

contrast, the first-by-analyst method generates a lower self-selection of 11.1%, understating the degree of self-selection by 22% due to its inclusion of a considerable number of misidentified initiations with less favorable ratings (see Panel A of Table 3).

Second, we examine how the measurement errors in the traditional methods affect the results concerning the economic determinants of initiations (Ertimur et al. 2011; Kirk 2011) using firm-quarters over 2003 – 2017, as in the main analyses. Initiating coverage on a firm is an important decision for analysts and brokers and has a significant impact on covered firms and investors. We include the most commonly used determinants of initiations: Firm size ($FSIZE_{iq}$), book-to-market ratio (BM_{iq}), sales growth ($SALES\_GR_{iq}$), return on assets (ROA_{iq}), past 12-month return ($Momentum_{iq}$), the number of analysts following ($NANALYST_{iq}$), institutional ownership (IO_{iq}), the number of years since a firm was listed on the stock exchange ($FIRMAGE_{iq}$), and whether a firm undertook an initial public offering within the past one year (IPO_{iq}). The subscripts *i* and *q* denote a firm and a quarter, respectively. Variable definitions are provided in Appendix B.

Panel B of Table 6 presents the logistic regression results on the determinants of initiations. The dependent variable, *INITIATION*, is an indicator that equals one if at least one analyst initiates coverage on a firm in the quarter and zero otherwise. The independent variables are measured at the end of the prior quarter. We require each firm-quarter to have at least one analyst following, as firms not covered by any analyst are likely not under the radar of analysts who initiate coverage. Column (1) reports results based on the observed initiation sample, and Columns (2)-(4) report results based on the inferred initiation samples using the first-by-analyst, first-by-broker, and first-by-both methods, respectively. We find some striking differences across the four columns regarding multiple key variables, including firm size (*FSIZE*), return on assets (*ROA*), the number of analysts following (*NANALYST*), institutional ownership (*IO*), and the number of years since a

firm was listed on the stock exchange (*FIRMAGE*). Most importantly, the coefficients on *NANALYST*, *IO,* and *FIRMAGE* are positive for the observed initiation sample (Column (1)), but mostly negative for the inferred initiation samples (Columns (2)-(4)). These results suggest that the traditional methods introduce a significant downward bias for the relations between these important firm characteristics and analyst initiation. This bias is likely due to the traditional methods missing some or all re-initiations, which are associated with firms with higher institutional ownership, more analyst following, and a longer history (see Table 2).

Overall, the results in Table 6 indicate that the measurement errors and the resulting sample biases in the traditional methods of inferring initiations can have a significant impact on research findings. They highlight the importance of mitigating the errors in inferring initiations.

5. Approaches to mitigate the measurement errors in the traditional methods

5.1. Prior approaches to mitigate omission error

Prior studies have used several different approaches to mitigate the omission error in the traditional methods by including additional re-initiations: Malmendier and Shanthikumar (2007) modify the first-by-broker method by including a broker's new recommendation since its previous coverage stoppage; Ljungqvist et al. (2009) adjust the first-by-broker method by including a broker's new recommendations since its previous coverage stoppage or recommendations with no preceding recommendation issued by the same broker in the past 12 months. Bernhardt et al. (2016) modify the first-by-analyst method by including an analyst's recommendation with no preceding recommendation by the same analyst on the firm in the past 12 months.

Table 7 reports the effectiveness of these prior approaches. We find that all three prior approaches significantly reduce the omission error by picking up additional re-initiations. Specifically, the approaches by Malmendier and Shanthikumar (2007), Ljungqvist et al. (2009),

and Bernhardt et al. (2016) reduce the number of missed initiations by 8,114, 8,464, and 2,096, respectively (see Column (3)). Consequently the number of correctly identified initiations increases from 18,831 to 26,945, from 18,831 to 27,295, and from 20,382 to 22,478 (see Column (1)), resulting in a decrease of the omission error rate – the number of missed initiations divided by the number of observed initiations (i.e., the sum of correctly identified and missed initiations) – from 33% to 4%, from 33% to 3%, and from 28% to 20%, respectively (see Column (5)).

However, at the same time, all three prior approaches also significantly exacerbate the misidentification error. Specifically, the approaches by Malmendier and Shanthikumar (2007), Ljungqvist et al. (2009), and Bernhardt et al. (2016) increase the number of misidentified initiations by 13,936, 14,316, and 1,516, respectively (see Column (2)). Consequently, the misidentification error rate – the number of misidentified initiations divided by the number of inferred initiations (i.e., the sum of correctly identified and misidentified initiations) – increases from 18% to 40% for the approaches by Malmendier and Shanthikumar (2007) and Ljungqvist et al. (2009), and from 31% to 32% for the approach by Bernhardt et al. (2016) (see Column (4)).

Considering the overall measurement error, both of the approaches by Malmendier and Shanthikumar (2007) and Ljungqvist et al. (2009) show a greater increase in the number of misidentified initiations than the decrease in the number of missed initiations, resulting in a significant increase in the total number of misidentified and missed initiations (i.e., an increase of 5,822 = (17,928 – 3,992) – (26,945 – 18,831) and 5,852 = (18,308 – 3,992) – (27,295 – 18,831), respectively). On the other hand, the approach by Bernhardt et al. (2016) performs relatively better, as it reduces the number of missed initiations (2,096 = 10,525 – 9,009) slightly more than it increases the number of misidentified initiations (1,516 = 22,478 – 20,382), resulting in a small decrease in the total number of misidentified and missed initiations (i.e., a decrease of 580 = 1,516

– 2,096). It is worth noting that Bernhardt et al.'s approach builds upon the first-by-analyst method, which generates a higher overall error than the first-by-broker method (see Figure 2).

Taken together, the results in Table 7 suggest that the prior approaches to improve the traditional methods are ineffective in reducing the overall measurement error. While these prior approaches do reduce the omission error, they increase the misidentification error to a similar or greater degree, with the net result of no significant decrease in the overall measurement error.

5.2. A new approach to reduce measurement errors

To help researchers more accurately infer initiations from IBES, we propose a new approach that can more effectively reduce measurement errors in the traditional methods. As shown in Table 8, our new approach begins with the inferred initiation sample based on the first-by-broker method (22,823 inferred initiations). We build upon the first-by-broker method rather than the other two methods, because the first-by-broker method generates the lowest total number of misidentified and missed initiations (see Figure 2).

To mitigate the misidentification error, our new approach excludes inferred initiations accompanied by any earnings per share (EPS) or target price (TP) forecast issued by the same broker for the same firm in the past six months. This step reduces the number of misidentified initiations by 711, while increasing the number of missed initiations by 301.

Next, to mitigate the omission error, our new approach additionally includes broker re-initiations. Specifically, we consider a recommendation as an initiation if it meets all the following criteria: (1) It is the first recommendation by a broker since its previous coverage stoppage on the firm (based on the IBES recommendation stop file);[18] (2) the broker did not issue any EPS or TP

[18] We acknowledge that using the IBES recommendation stop file to identify coverage stoppages may introduce errors, primarily due to the criteria used by IBES to define a stoppage. Specifically, IBES applies a stoppage designation to an outstanding recommendation in specific scenarios, such as when a broker modifies its ratings system, temporarily

forecasts throughout its non-coverage period; and (3) the non-coverage period is longer than a cutoff value. Because requiring a shorter non-coverage period decreases missed initiations but increases misidentified initiations, the "optimal" non-coverage period is an empirical matter. We thus examine various cutoff values ranging from one month to two years.

Similar to the prior approaches in Table 7, this step of our new approach increases misidentified initiations and decreases missed initiations. However, unlike the prior approaches, the decrease in missed initiations dominates the increase in misidentified initiations. Table 8 shows that by requiring a minimum non-coverage period of one month, three months, six months, and 12 months, this step decreases missed initiations by 7,099, 6,464, 5,336, and 3,972, while increasing misidentified initiations by 3,911, 2,111, 1,135, and 732, respectively. Combining the overall impact of both steps, our new approach reduces missed initiations by 6,734, 6,099, 4,971, and 3,607, and increases misidentified initiations by 3,200, 1,400, 424, and 21, respectively. Notably, requiring a minimum 24-month non-coverage period not only reduces missed initiations by 2,384, but also reduces misidentified initiations by 296.

If we give equal weight to omission and misidentification errors and aim to minimize the total number of misidentified and missed initiations, our findings suggest that a non-coverage period of three months yields the best result: It decreases (increases) missed (misidentified) initiations by 6,099 (1,400), resulting in a net decrease of 4,699 in the total number of misidentified and missed initiations. However, we caution that the relative importance of omission vs. misidentification errors may vary depending on specific research questions and settings, and as a

suspends it due to its underwriting relationship with a firm, or experiences the departure of an analyst. Furthermore, if a recommendation remains not updated or confirmed by a broker for more than 180 days, IBES considers it as stopped. Our review of IBES manuals over time reveals that IBES's policy for recommendation stoppages has remained consistent since at least June 17, 2010. However, as the stoppages are defined algorithmically, it is possible that IBES's policies or definitions may have changed in the past or be subject to future modifications.

result, the non-coverage period cutoff needs to be determined based on the relative weight assigned to each type of measurement error.

While our new approach aims to enhance the accuracy of inferring all initiations, including both first-time initiations and re-initiations, from IBES, there may be specific research questions and settings where researchers may want to examine first-time initiations only. In this case, our results indicate that among the three traditional methods, the first-by-both method performs the best in inferring first-time initiations (see Figure 2). Furthermore, for researchers seeking to further mitigate misidentification errors when using the first-by-both method, they can apply the first step outlined in Table 8, i.e., excluding inferred initiations preceded by any EPS or TP forecast from the same broker for the same firm during the past six months. In untabulated tests, we find that this step reduces misidentified initiations in the first-by-both method from 2,931 to 2,436, representing a 17% decrease.

6. Conclusion

Using a large, hand-collected sample of observed initiations directly from analysts' initiating reports, we provide the first systematic examination of the measurement errors in the traditional methods of inferring initiations from IBES (i.e., the first recommendation by an analyst, by a broker, or by both). We find that these methods all generate significant measurement errors. On the one hand, by focusing on the first recommendation only, these methods miss a portion of or all re-initiations, which represent 28 – 49% of all observed initiations. On the other hand, due to the incompleteness of IBES, these methods misidentify a large number of initiations, which represent 17 – 31% of all inferred initiations.

We find that these measurement errors introduce systematic sample biases. While all three methods tilt the sample towards smaller/younger firms, they bias the sample toward analysts with

different characteristics (e.g., broker size and analyst experience) and the distribution of initiating ratings in different directions. We further show that these sample biases can have a significant impact on research findings. Compared with the observed initiations, these methods can overstate the degree of analyst self-selection by up to 34% and introduce biases of varying directions and magnitudes to commonly used determinants of initiations.

We evaluate the effectiveness of three approaches that prior studies have used to mitigate the omission error in the traditional methods. We find that while these approaches successfully reduce the omission error by including more re-initiations, they exacerbate the misidentification error to a similar or greater extent, resulting in no significant decrease in the overall measurement error. We propose a new approach that can more effectively reduce errors in inferring initiation. We show that this new approach can reduce the omission error by two thirds but increase the misidentification error to a much lesser degree.

Our study adds to the literature by providing the first systematic evaluation of the traditional methods of inferring analyst initiations and prior approaches used to mitigate the omission error in these methods. Our findings can help researchers to assess the potential biases introduced by each method and choose one that best suits their research questions and settings. Our new approach offers researchers a means to further mitigate the measurement errors in the traditional methods.

References

Agrawal, A., & Chen, M. A. (2008). Do analyst conflicts matter? Evidence from stock recommendations. *Journal of Law & Economics, 51*(3), 503-537.

Amiram, D., Bozanic, Z., Bradshaw, M., & Rozenbaum, O. (2021). Distribution channels of analyst research: New evidence. Unpublished Working Paper, Tel Aviv University, Florida State University, Boston College, and George Washington University

Anantharaman, D., & Zhang, Y. (2011). Cover me: Manager's responses to changes in analyst coverage in the post-Regulation FD period. *The Accounting Review*, *86*(6), 1851-1885.

Balakrishnan, K., Billings, M. B., Kelly, B., & Ljungqvist, A. (2014). Shaping liquidity: On the causal effects of voluntary disclosure. *The Journal of Finance*, *69*(5), 2237-2278.

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance*, *56*(2), 531-563.

Barber, B. M., Lehavy, R., McNichols, M., & Trueman, B. (2006). Buys, holds, and sells: The distribution of investment banks' stock ratings and the implications for the profitability of analysts' recommendations. *Journal of Accounting and Economics*, *41*(1-2), 87-117.

Becher, D. A., Cohn, J. B., & Juergens, J. L. (2015). Do stock analysts influence merger completion? An examination of postmerger announcement recommendations. *Management Science*, *61*(10), 2430-2448.

Bernhardt, D., Wan, C., & Xiao, Z. (2016). The reluctant analyst. *Journal of Accounting Research*, *54*(4), 987-1040.

Billings, B. K., Buslepp, W. L., & Huston, G. R. (2014). Worth the hype? The relevance of paid-for analyst research for the buy-and-hold investor. *The Accounting Review*, *89*(3), 903-931.

Bradley, D. J., Jordan, B. D., & Ritter, J. R. (2003). The quiet period goes out with a bang. *The Journal of Finance*, *58*(1), 1-36.

Bradley, D. J., Jordan, B. D., & Ritter, J. R. (2008). Analyst behavior following IPOs: the "bubble period" evidence. *The Review of Financial Studies*, *21*(1), 101-133.

Branson, B. C., Guffey, D. M., & Pagach, D. P. (1998). Information conveyed in announcements of analyst coverage. *Contemporary Accounting Research*, *15*(2), 119-143.

Brav, A., & Lehavy, R. (2003). An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. *The Journal of Finance*, *58*(5), 1933-1967.

Brochet, F., Miller, G. S., & Srinivasan, S. (2014). Do analysts follow managers who switch companies? An analysis of relationships in the capital markets. *The Accounting Review*, *89*(2), 451-482.

Call, A. C., Hewitt, M., Watkins, J., & Yohn, T. L. (2021). Analysts' annual earnings forecasts and changes to the I/B/E/S database. *Review of Accounting Studies*, *26*(1), 1-36.

Cen, L., Chen, J., Dasgupta, S., & Ragunathan, V. (2021). Do analysts and their employers value access to management? Evidence from earnings conference call participation. *Journal of Financial and Quantitative Analysis*, *56*(3), 745-787.

Chan, J., Lin, S., Yu, Y., & Zhao, W. (2018). Analysts' stock ownership and stock recommendations. *Journal of Accounting and Economics*, *66*(2-3), 476-498.

Chemmanur, T. J., Hu, G., & Huang, J. (2015). Institutional investors and the information production theory of stock splits. *Journal of Financial and Quantitative Analysis*, 413-445.

Clement, M. B., & Tse, S. Y. (2003). Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? *The Accounting Review, 78*(1), 227-249.

Crawford, S. S., Roulstone, D. T., & So, E. C. (2012). Analyst initiations of coverage and stock return synchronicity. *The Accounting Review*, *87*(5), 1527-1553.
Dambra, M., Field, L. C., Gustafson, M. T., & Pisciotta, K. (2018). The consequences to analyst involvement in the IPO process: Evidence surrounding the JOBS Act. *Journal of Accounting and Economics*, *65*(2-3), 302-330.
Das, S., Guo, R. J., & Zhang, H. (2006). Analysts' selective coverage and subsequent performance of newly public firms. *The Journal of Finance*, *61*(3), 1159-1185.
Driskill, M., Kirk, M. P., & Tucker, J. W. (2020). Concurrent Earnings Announcements and Analysts' Information Production. *The Accounting Review*, *95*(1), 165-189.
Ertimur, Y., Muslu, V., & Zhang, F. (2011). Why are recommendations optimistic? Evidence from analysts' coverage initiations. *Review of Accounting Studies*, *16*(4), 679-718.
Ertimur, Y., Sunder, J., & Sunder, S. V. (2007). Measure for measure: The relation between forecast accuracy and recommendation profitability of analysts. *Journal of Accounting Research*, *45*(3), 567-606.
Gibbons, B., Iliev, P., & Kalodimos, J. (2021). Analyst information acquisition via EDGAR. *Management Science*, *67*(2), 769-793.
Groysberg, B., Healy, P. M., & Maber, D. A. (2011). What drives sell-side analyst compensation at high-status investment banks? *Journal of Accounting Research*, *49*(4), 969-1000.
Guan, Y., Wong, M. F., & Zhang, Y. (2015). Analyst following along the supply chain. *Review of Accounting Studies*, *20*(1), 210-241.
Houston, J., James, C., & Karceski, J. (2006). What a difference a month makes: Stock analyst valuations following initial public offerings. *Journal of Financial and Quantitative Analysis*, 111-137.
Huang, R., & Zhang, D. (2011). Managing underwriters and the marketing of seasoned equity offerings. *Journal of Financial and Quantitative Analysis*, *46*(1), 141-170.
Irvine, P., Lipson, M., & Puckett, A. (2007). Tipping. *The Review of Financial Studies*, *20*(3), 741-768.
Israelsen, R. D. (2016). Does common analyst coverage explain excess comovement? *Journal of Financial and Quantitative Analysis*, *51*(4), 1193-1229.
Jennings, J. (2019). The role of sell-side analysts after accusations of managerial misconduct. *The Accounting Review*, *94*(1), 183-203.
Jennings, J., Lee, J., & Matsumoto, D. A. (2017). The effect of industry co-location on analysts' information acquisition costs. *The Accounting Review*, *92*(6), 103-127.
Juergens, J. L., & Lindsey, L. (2009). Getting out early: An analysis of market making activity at the recommending analyst's firm. *The Journal of Finance*, *64*(5), 2327-2359.
Jung, M. J., Wong, M. F., & Zhang, X. F. (2015). Analyst interest as an early indicator of firm fundamental changes and stock returns. *The Accounting Review*, *90*(3), 1049-1078.
Kaplan, Z., Martin, X., & Xie, Y. (2021). Truncating optimism. *Journal of Accounting Research*, *59*(5), 1827-1884.
Kecskés, A., Michaely, R., & Womack, K. L. (2017). Do earnings estimates add value to sell-side analysts' investment recommendations? *Management Science*, *63*(6), 1855-1871.
Kelly, B., & Ljungqvist, A. (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies*, *25*(5), 1366-1413.
Kim, S. T., Lin, J. C., & Slovin, M. B. (1997). Market structure, informed trading, and analysts' recommendations. *Journal of Financial and Quantitative Analysis*, *32*(4), 507-524.

Kirk, M. (2011). Research for sale: Determinants and consequences of paid-for analyst research. *Journal of Financial Economics*, *100*(1), 182-200.
Lang, M. H., Lins, K. V., & Miller, D. P. (2004). Concentrated control, analyst following, and valuation: Do analysts matter most when investors are protected least? *Journal of Accounting Research*, *42*(3), 589-623.
Lawrence, A., Ryans, J. P., & Sun, E. Y. (2017). Investor demand for sell-side research. *The Accounting Review*, *92*(2), 123-149.
Li, K. K., & You, H. (2015). What is the value of sell-side analysts? Evidence from coverage initiations and terminations. *Journal of Accounting and Economics*, *60*(2-3), 141-160.
Liu, X., & Ritter, J. R. (2011). Local underwriter oligopolies and IPO underpricing. *Journal of Financial Economics*, *102*(3), 579-601.
Ljungqvist, A., Malloy, C., & Marston, F. (2009). Rewriting history. *The Journal of Finance*, *64*(4), 1935-1960.
Ljungqvist, A., Marston, F., Starks, L. T., Wei, K. D., & Yan, H. (2007). Conflicts of interest in sell-side research and the moderating role of institutional investors. *Journal of Financial Economics*, *85*(2), 420-456.
Madureira, L., & Underwood, S. (2008). Information, sell-side research, and market making. *Journal of Financial Economics*, *90*(2), 105-126.
Malmendier, U., & Shanthikumar, D. (2007). Are small investors naive about incentives? *Journal of Financial Economics*, *85*(2), 457-489.
McNichols, M., & O'Brien, P. C. (1997). Self-selection and analyst coverage. *Journal of Accounting Research*, *35*, 167-199.
Michaely, R., & Womack, K. L. (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *The Review of Financial Studies*, *12*(4), 653-686.
Mikhail, M. B., Walther, B. R., & Willis, R. H. (1999). Does forecast accuracy matter to security analysts? *The Accounting Review*, *74*(2), 185-200.
Mohanram, P. S., & Sunder, S. V. (2006). How has Regulation FD affected the operations of financial analysts? *Contemporary Accounting Research*, *23*(2), 491-525.
Mola, S., & Guidolin, M. (2009). Affiliated mutual funds and analyst optimism. *Journal of Financial Economics*, *93*(1), 108-137.
O'Brien, P. C., & Tan, H. (2015). Geographic proximity and analyst coverage decisions: Evidence from IPOs. *Journal of Accounting and Economics*, *59*(1), 41-59.
O'Brien, P. C., McNichols, M. F., & Lin, H.-W. (2005). Analyst impartiality and investment banking relationships. *Journal of Accounting Research*, *43*(4), 623-650.
Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*(1), 435-480.
Philippot, A. (2018). Analysts' reinitiations of coverage and market underreaction. *Journal of Banking and Finance, 94*, 208-220.
Tehranian, H., Zhao, M., & Zhu, J. L. (2014). Can analysts analyze mergers? *Management Science*, *60*(4), 959-979.

Appendix A
List of published studies in ten top accounting and finance journals using analysts' initiation of coverage

Panel A: Papers that infer analysts' initiations from databases

		Initiation coverage	
Author(s)	Running title	Importance	Identification method
McNichols and O'Brien (JAR 1997)	Self-Selection and Analyst Coverage	Main analysis	First recommendation by an analyst
Barber et al. (JF 2001)	Can Investors Profit From the Prophets?	Main analysis	First recommendation by an analyst
Brav and Lehavy (JF 2003)	An Empirical Analysis of Analysts' Target Prices	Main analysis	First recommendation (unsure whose)
Lang et al. (JAR 2004)	Concentrated Control, Analyst Following, and Valuation	Additional analysis	First appearance of a firm on IBES
O'Brien et al. (JAR 2005)	Analyst Impartiality and Investment Banking Relationships	Main analysis	First recommendation by an analyst
Barber et al. (JAE 2006)	Buys, Holds, and Sells	Main analysis	First recommendation (unsure whose)
Das et al. (JF 2006)	Analysts' Selective Coverage and Subsequent Performance	Main analysis	First forecast by an analyst
Mohanram and Sunder (CAR 2006)	How Has Regulation FD Affected the Operations of Financial Analysts?	Main analysis	First appearance of an analyst on IBES
Irvine et al. (RFS 2007)	Tipping	Main analysis	First recommendation by an analyst and a broker
Ljungqvist et al. (JFE 2007)	Conflicts of Interest in Sell-Side Research	Control variable	First recommendation by an analyst
Malmendier and Shanthikumar (JFE 2007)*	Are Small Investors Naive About Incentives?	Main analysis	First recommendation by a broker
Madureira and Underwood (JFE 2008)	Information, Sell-Side Research, and Market Making	Main analysis	First forecast or recommendation by a broker
Juergens and Lindsey (JF 2009)	Getting Out Early	Main analysis	First recommendation (unsure whose)
Ljungqvist et al. (JF 2009)*	Rewriting History	Main analysis	First recommendation by a broker
Mola and Guidolin (JFE 2009)	Affiliated Mutual Funds and Analyst Optimism	Additional analysis	First recommendation (unsure whose)
Anantharaman and Zhang (TAR 2011)	Cover Me	Sample selection	No explicit definition
Ertimur et al. (RAS 2011)	Why Are Recommendations Optimistic?	Main analysis	First recommendation by an analyst
Groysberg et al. (JAR 2011)	What Drives Sell-Side Analyst Compensation at High-Status IBs	Main analysis	First recommendation by an analyst
Huang and Zhang (JFQA 2011)	Managing Underwriters and the Marketing of Seasoned Equity Offerings	Main analysis	First recommendation by a broker
Crawford et al. (TAR 2012)	Analyst Initiations of Coverage and Stock Return Synchronicity	Main analysis	First recommendation by an analyst and a broker
Kelly and Ljungqvist (RFS 2012)*	Testing Asymmetric-Information Asset Pricing Models	Sample selection	First recommendation by a broker
Balakrishnan et al. (JF 2014)	Shaping Liquidity	Additional analysis	No explicit definition
Billings et al. (TAR 2014)	Worth the Hype	Sample selection	First recommendation (unsure whose)
Brochet et al. (TAR 2014)	Do Analysts Follow Managers Who Switch Companies?	Main analysis	First forecast by an analyst
Tehranian (MS 2014)	Can Analysts Analyze Mergers?	Main analysis	No explicit definition
Becher et al. (MS 2015)	Do Stock Analysts Influence Merger Completion?	Main analysis	First recommendation (unsure whose)
Chemmanur et al. (JFQA 2015)	Institutional Investors and the Information Production Theory of Stock Splits	Main analysis	First recommendation by an analyst and a broker

Guan et al. (RAS 2015)	Analyst Following Along the Supply Chain	Main analysis	No explicit definition
Jung et al. (TAR 2015)*	Analyst Interest as an Early Indicator of Firm Fundamental Changes	Sample selection	First forecast by an analyst
Li and You (JAE 2015)	What is the Value of Sell-Side Analysts?	Main analysis	First recommendation by an analyst and a broker
O'Brien and Tan (JAE 2015)	Geographic Proximity and Analyst Coverage Decisions	Main analysis	First forecast by an analyst
Bernhardt et al. (JAR 2016)*	The Reluctant Analysts	Main analysis	First recommendation by an analyst
Israelsen (JFQA 2016)	Does Common Analyst Coverage Explain Excess Comovement?	Sample selection	No explicit definition
Jennings et al. (TAR 2017)	The Effect of Industry Co-Location on Analysts' Information Acquisition	Main analysis	First forecast by an analyst
Kecskés et al. (MS 2017)	Do Earnings Estimates Add Value to Investment Recommendations?	Sample selection	No explicit definition
Lawrence et al. (TAR 2017)	Investor Demand for Sell-Side Research	Main analysis	Change in analyst coverage
Chan et al. (JAE 2018)	Analysts' Stock Ownership and Stock Recommendations	Main analysis	No explicit definition
Dambra et al. (JAE 2018)	The Consequences to Analyst Involvement in the IPO Process	Main analysis	First forecast by an analyst
Jennings (TAR 2019)	The Role of Sell Side Analysts After Accusations of Managerial Misconduct	Main analysis	First recommendation by an analyst and a broker
Cen et al. (JFQA 2021)	Do Analysts and Their Employers Value Access to Management?	Main analysis	No explicit definition
Driskill et al. (TAR 2020)	Concurrent Earnings Announcements and Analysts' Information Production	Sample selection	No explicit definition
Gibbons et al. (MS 2021)	Analyst Information Acquisition via EDGAR	Sample selection	No explicit definition

Panel B: Papers that identify analysts' initiations using other sources

		Initiation coverage	
Author(s)	Running title	Importance	Identification method
Kim et al. (JFQA 1997)	Market Structure, Informed Trading, and Analysts' Recommendations	Main analysis	News articles and press releases
Branson et al. (CAR 1998)	Information Conveyed in Announcements of Analyst Coverage	Main analysis	News articles and press releases
Michaely and Womack (RFS 1999)	Conflict of Interest and the Credibility of Underwriter Analyst	Main analysis	Daily commentary provided by a broker
Bradley et al. (JF 2003)	The Quiet Period Goes Out With a Bang	Main analysis	Multiple data sources
Houston et al. (JFQA 2006)	What a Difference a Month Makes	Main analysis	Analyst reports
Bradley et al. (RFS 2008)	Analyst Behavior Following IPOs	Main analysis	Briefing.com
Kirk (JFE 2011)	Research for Sale	Main analysis	Analyst reports
Liu and Ritter (JFE 2011)	Local Underwriter Oligopolies and IPO Underpricing	Main analysis	Multiple data sources

This table presents the list of published studies in ten top accounting and finance journals that use analysts' coverage initiations. * indicates that the paper uses an approach that modifies a traditional method of inferring initiations to mitigate the omission error by including additional re-initiations (e.g., a new recommendation that a broker issues since its previous termination of coverage on the firm as in Malmendier and Shanthikumar (2007) or a recommendation by an analyst who did not issue a recommendation on the firm in the past 12 months as in Bernhardt et al. (2016)).

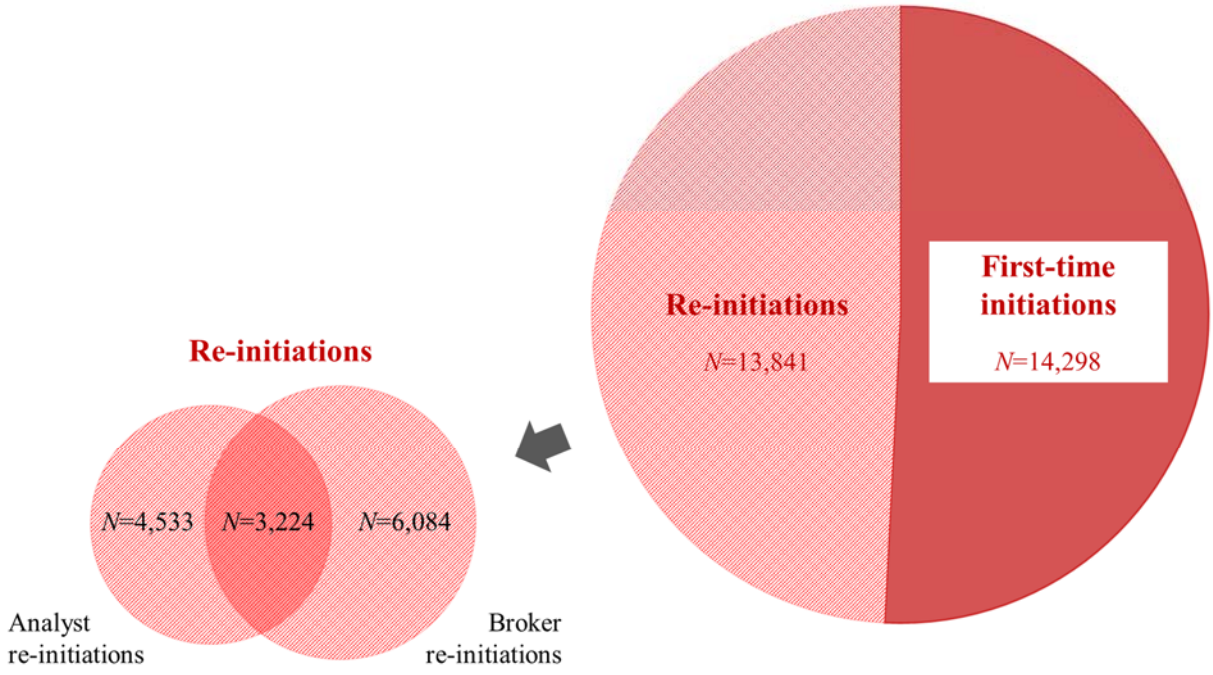
Appendix B
Variable definitions

Variable	Definition
[*Classification of analysts' initiations of coverage*]	
$OBSERVED_{ijtk}$	Indicator variable for an observed initiation that equals one if analyst *j*'s stock recommendation *k* for firm *i* in year *t* is indicated explicitly as an initiating report and zero otherwise. (Data source: Thomson ONE)
$INFERRED_{ijtk}$	Indicator variable for an inferred initiation that equals one if analyst *j*'s stock recommendation *k* for firm *i* in year *t* is the very first recommendation for firm *i* by analyst *j* (for the first-by-analyst method), her broker (for the first-by-broker method), or both (for the first-by-both method) on IBES and zero otherwise. (Data source: IBES)
CORRECTLY $IDENTIFIED_{ijtk}$	Indicator variable for a correctly identified initiation that equals one if analyst *j*'s stock recommendation *k* for firm *i* in year *t* is classified as both an observed initiation (*OBSERVED*=1) and an inferred initiation (*INFERRED*=1) and zero otherwise. (Data source: IBES, Thomson ONE)
$MISSED_{ijtk}$	Indicator variable for a missed initiation that equals one if analyst *j*'s stock recommendation *k* for firm *i* in year *t* is classified as an observed initiation (*OBSERVED*=1) but not as an inferred initiation (*INFERRED*=0) and zero otherwise. (Data source: IBES, Thomson ONE)
$MISIDENTIFIED_{ijtk}$	Indicator variable for a misidentified initiation that equals one if analyst *j*'s stock recommendation *k* for firm *i* in year *t* is classified as an inferred initiation (*INFERRED*=1) but not as an observed initiation (*OBSERVED*=0) and zero otherwise. (Data source: IBES, Thomson ONE)
[*Firm, analyst, and rating characteristics*]	
BM_{it-1}	Ratio of firm *i*'s book value of equity to its market value of equity in year *t*-1. For firm characteristics, year *t* refers to the firm's fiscal year during which an analyst's recommendation is issued. (Data source: Compustat, CRSP)
$FIRMAGE_{it}$	Number of years since firm *i*'s first appearance on CRSP as of year *t.* (Data source: CRSP)
$INTANGIBLE_{it-1}$	Ratio of firm *i*'s intangible assets to its total assets in year *t*-1. (Data source: Compustat)
IO_{it-1}	Mean proportion of firm *i*'s shares held by institutional investors over four quarters in year *t*-1. (Data source: Thomson Reuters 13F Institutional Holdings)
$LEVERAGE_{it-1}$	Ratio of firm *i*'s short- and long-term debts to its total assets in year *t*-1. (Data source: Compustat)
$MOMENTUM_{it-1}$	Cumulative 12-month return of firm *i* in year *t*-1. (Data source: CRSP)
MV_{it-1}	Market capitalization, measured as firm *i*'s stock price times the number of its common shares outstanding, in year *t*-1. (Data source: CRSP)
$NANALYST_{it-1}$	Number of analysts who issue at least one one-year-ahead earnings forecast for firm *i* in year *t*-1. (Data source: IBES)
$NSEGMENT_{it-1}$	Number of firm *i*'s business segments in year *t*-1. (Data source: Compustat)
ROA_{it-1}	Return on assets, measured as firm *i*'s net income divided by total assets, in year *t*-1. (Data source: Compustat)
$ALLSTAR_{jt-1}$	Indicator variable that equals one if analyst *j* is elected as an all-star by *Institutional Investor* magazine in year *t*-1 and zero otherwise. For analyst characteristics, year *t* refers to the calendar year in which the analyst's recommendation is issued. (Data source: *Institutional Investor* magazine)

$BSIZE_{jt-1}$	Number of analysts employed by analyst *j*'s broker in year *t*-1. (Data source: IBES)
$GEXP_{jt}$	Number of years for which analyst *j* has issued at least one one-year-ahead earnings forecast as of year *t*. (Data source: IBES)
$NFIRM_{jt-1}$	Number of firms for which analyst *j* issues at least one one-year-ahead earnings forecast in year *t*-1. (Data source: IBES)
$NFITEM_{ijt}$	Number of forecast items analyst *j* issues for firm *i* in year *t*. We consider the following four forecast items: Earnings per share (EPS), cash flow per share (CPS), long-term growth of earnings (LTG), and target price (TP). (Data source: IBES)
$NIND_{jt-1}$	Number of industries for which analyst *j* issues at least one one-year-ahead earnings forecast in year *t*-1. Industry classifications are based on two-digit SIC codes. (Data source: CRSP, IBES)
$ACCURACY_{jt-1}$	Mean of analyst *j*'s price-scaled forecast accuracy for firms in year *t*-1. Price-scaled forecast accuracy is calculated using analyst *j*'s most recent one-year-ahead earnings forecast for firm *i* in year *t*-1 and measured as negative one times the absolute forecast error (i.e., |actual EPS - forecasted EPS|), scaled by the stock price two trading days prior to the current forecast date. (Data source: CRSP, IBES)
$FREQ_{jt-1}$	Natural logarithm of one plus analyst *j*'s mean forecast frequency for firms in year *t*-1. Forecast frequency is measured as the number of one-year-ahead earnings forecasts an analyst issues for a firm in a year. (Data source: IBES)
$BUY\ RATING_{ijtk}$	Indicator variable that equals one if analyst *j*'s stock recommendation *k* for firm *i* in year *t* is Buy or Strong Buy and zero otherwise. (Data source: IBES)
[P*rior findings on the determinants of analyst coverage initiations (Tests in Table 6)*]	
$INITIATION_{iq}$	Indicator variable that equals one if at least one stock recommendation issued by an analyst for firm *i* in quarter *q* is defined as an observed initiation or an inferred initiation using one of the three traditional methods and zero otherwise. (Data source: IBES)
$FSIZE_{iq}$	Natural logarithm of firm *i*'s total assets in quarter *q*. (Data source: Compustat)
BM_{iq}	Firm *i*'s book value of equity, divided by its market value of equity in quarter *q*. (Data source: Compustat, CRSP)
$SALES\_GR_{iq}$	Firm *i*'s sales growth in quarter *q*, measured as the ratio of sales in the current quarter to the sales in the same quarter in the prior year and minus one. (Data source: Compustat)
ROA_{iq}	Firm *i*'s net income divided by total assets in quarter *q*. (Data source: Compustat)
$MOMENTUM_{iq}$	Firm *i*'s market-adjusted return over the 12 months prior to quarter *q*. (Data source: CRSP)
$NANALYST_{iq}$	Number of analysts who issue at least one one-year-ahead earnings forecast for firm *i* in quarter *q*. (Data source: IBES)
IO_{iq}	Percentage of firm *i*'s shares held by institutional investors in quarter *q*. (Data source: Thomson Reuters 13F Institutional Holdings)
$FIRMAGE_{iq}$	Number of years since firm *i*'s first appearance on CRSP as of quarter *q*. (Data source: CRSP)
IPO_{iq}	Indicator variable that equals one if firm *i* was listed for the first time within the past 365 days and zero otherwise. (Data source: CRSP)

This table shows variable definitions and data sources.

Figure 1
Composition of observed initiations



This figure depicts the composition of our observed initiation sample, which consists of 28,139 observed initiations. Observed initiations are directly from analysts' initiating reports announcing coverage initiations. 14,298 observed initiations are first-time initiations. 13,841 observed initiations are re-initiations, of which 7,757 are analyst re-initiations and 9,308 are broker re-initiations (the overlap between the two includes 3,224 re-initiations by both analysts and brokers).

Figure 2
Measurement errors in the traditional methods of inferring initiations

Figure 2-A. First-by-analyst method

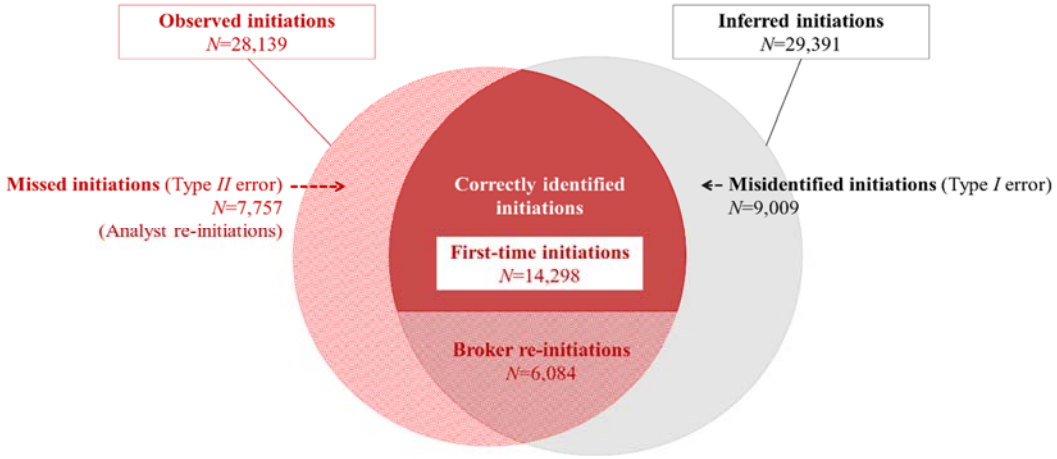


Figure 2-B. First-by-broker method

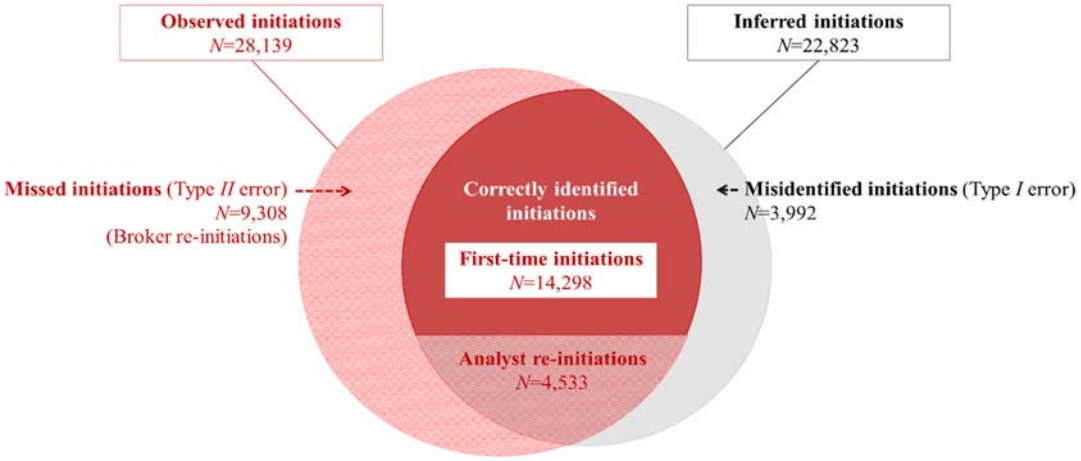


Figure 2-C. First-by-both method

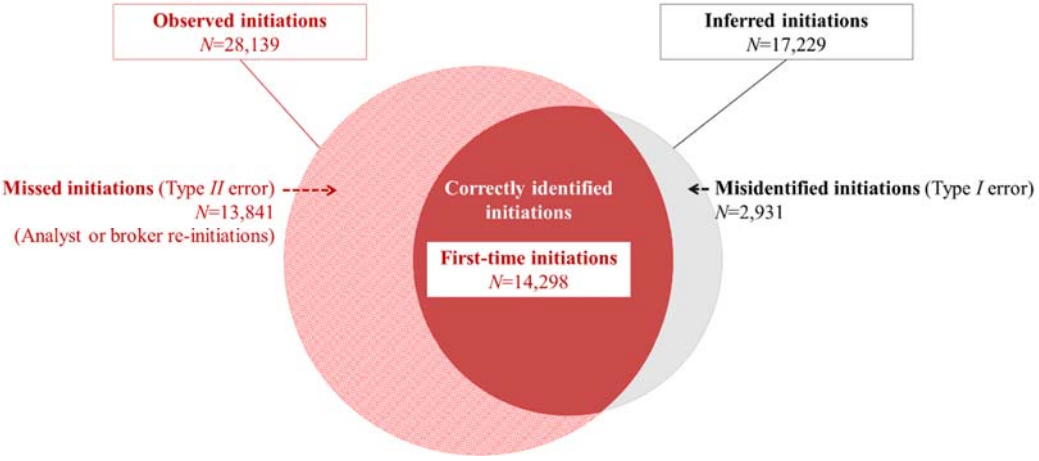


Figure 2 (*Continued*)

This figure shows three Venn diagrams that compare the sample of observed initiations (the red circle on the left) and the sample of inferred initiations (the gray circle on the right). Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. Inferred initiations are the first recommendations on a firm in the IBES database by an analyst (2-A), by a broker (2-B), or by both (2-C). Both the observed and inferred initiation samples are drawn from the common coverage overlap between IBES and Thomson ONE.

Figure 3
Measurement error rates over time

Figure 3-A. First-by-analyst method

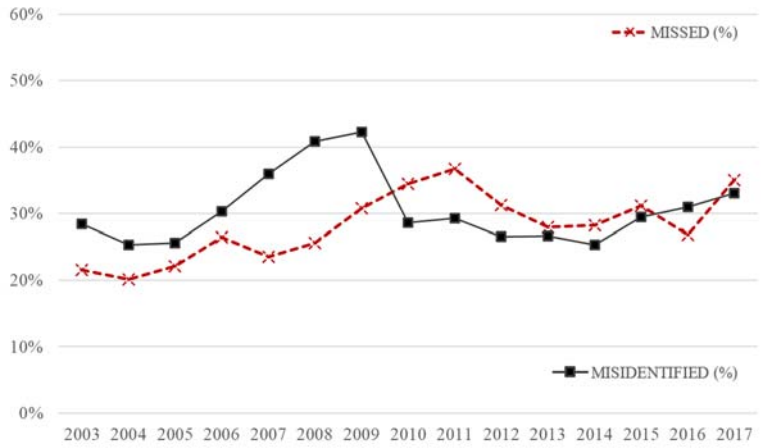


Figure 3-B. First-by-broker method

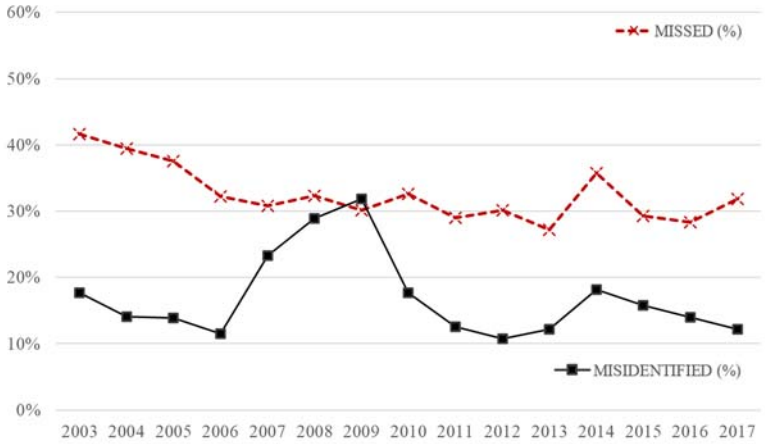


Figure 3-C. First-by-both method

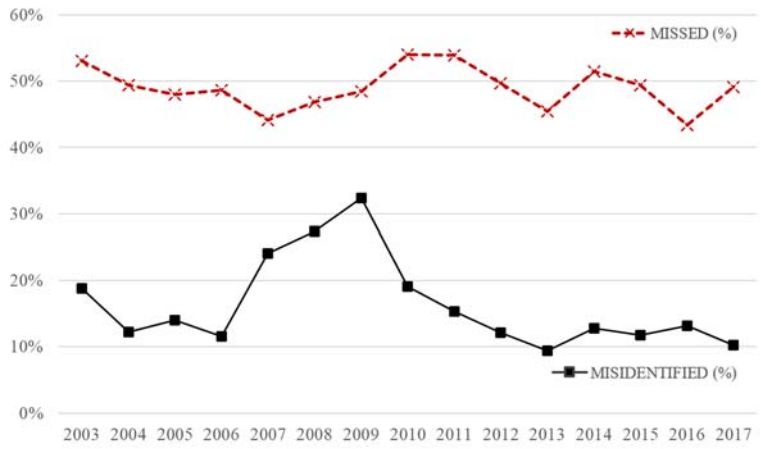


Figure 3 (*Continued*)

This figure plots the rates of measurement errors in the traditional method of inferring initiations by year. The lower black solid line represents the misidentification (Type I) error rate, calculated as the ratio of the number of misidentified initiations to the total number of inferred initiations. The upper red dashed line represents the omission (Type II) error rate, calculated as the ratio of the number of missed initiations to the total number of observed initiations. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. Inferred initiations are the first stock recommendations on a firm in the IBES database by an analyst (3-A), by a broker (3-B), or by both (3-C). Both the observed and inferred initiation samples are drawn from the common coverage overlap between IBES and Thomson ONE. Misidentified initiations refer to inferred initiations that are not actual initiations (i.e., not included in the sample of observed initiations). Missed initiations are observed initiations that are missed by the traditional method of inferring initiations (i.e., included in the sample of observed initiations but not included in the sample of inferred initiations). All frequencies are measured in percentages.

Table 1
Initiation keywords, report lengths, and non-coverage periods of observed re-initiations

Type of observed initiations:	First-time initiations	Re-initiations	
		Analyst	Broker
	(1)	(2)	(3)
[1] Percentage of reports containing the keywords of "initiation", "initiating", "initiate", or other derivatives of "initiation":	99.8% (*N*=14,197)	92.7% (*N*=7,593)	86.4% (*N*=9,110)
[2] Mean number of pages in a report:	20.5 pages (*N*=14,298)	16.8 pages (*N*=7,757)	19.6 pages (*N*=9,308)
[3] Mean number of days between re-initiations and their preceding stoppage:	n/a (*N*=14,298)	364.0 days (*N*=7,611)	846.3 days (*N*=9,308)

This table reports the percentage of reports containing the keywords of "initiation", "initiating", "initiate", or other derivatives of "initiation" and the mean number of pages in a report for first-time initiations and analyst/broker re-initiations. It also reports the mean number of days between re-initiations and their preceding coverage stoppage (i.e., non-coverage period) for analyst/broker re-initiations. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. First-time initiations refer to the initiations of coverage on a firm for the first time by both analysts and brokers. Analyst re-initiations refer to the initiations by analysts on a firm that they had previously covered and dropped. Broker re-initiations refer to the initiations by brokers on a firm that they had previously covered and dropped.

Table 2
Firm, analyst, and rating characteristics of observed initiations

Type of observed initiations:	First-time initiations	Re-initiations		Mean differences		
		Analyst	Broker	[= (2)-(1)]	[= (3)-(1)]	[= (3)-(2)]
Variable (mean)	(1)	(2)	(3)	(4)	(5)	(6)
Firm characteristics:						
BM_{it-1}	0.504	0.484	0.490	-0.020 ***	-0.014 *	0.006
$FIRMAGE_{it}$	16.261	21.884	23.887	5.623 ***	7.626 ***	2.002 ***
$INTANGIBLE_{it-1}$	0.160	0.171	0.177	0.010 ***	0.017 ***	0.007 *
IO_{it-1}	0.633	0.720	0.719	0.087 ***	0.085 ***	-0.002
$LEVERAGE_{it-1}$	0.209	0.221	0.229	0.012 ***	0.019 ***	0.008 **
$MOMENTUM_{it-1}$	0.283	0.217	0.208	-0.066 ***	-0.074 ***	-0.009
$Ln\ (MV)_{it-1}$	13.823	14.781	15.037	0.959 ***	1.215 ***	0.256 ***
$NANALYST_{it-1}$	11.708	18.602	19.684	6.894 ***	7.976 ***	1.082 ***
$NSEGMENT_{it-1}$	2.067	2.219	2.316	0.152 ***	0.249 ***	0.097 **
ROA_{it-1}	-0.006	0.018	0.029	0.025 ***	0.035 ***	0.011 ***
	(*N*=9,546)	(*N*=6,320)	(*N*=6,141)			
Analyst characteristics:						
$ALLSTAR_{jt-1}$	0.015	0.024	0.042	0.009 ***	0.027 ***	0.018 ***
$BSIZE_{jt-1}$	36.577	40.525	60.686	3.948 ***	24.109 ***	20.160 ***
$GEXP_{jt}$	7.728	9.085	7.688	1.357 ***	-0.040	-1.397 ***
$NFIRM_{jt-1}$	13.841	14.968	14.174	1.127 ***	0.333 ***	-0.794 ***
$NFITEM_{ijt}$	2.258	2.347	2.311	0.090 ***	0.053 ***	-0.036 ***
$NIND_{jt-1}$	4.344	4.181	4.081	-0.163 ***	-0.263 ***	-0.099 **
$ACCURACY_{jt-1}$	-0.040	-0.033	-0.031	0.007 ***	0.009 ***	0.003 *
$FREQ_{jt-1}$	1.489	1.552	1.529	0.064 ***	0.040 ***	-0.023 ***
	(*N*=8,743)	(*N*=5,454)	(*N*=4,846)			
Distribution of recommendation ratings:						
Mean (BUY RATING)	0.603	0.541	0.456	-0.062 ***	-0.147 ***	-0.085 ***
	(*N*=14,298)	(*N*=7,757)	(*N*=9,308)			

This table reports the mean firm, analyst, and rating characteristics for first-time initiations, analyst re-initiations, and broker re-initiations in the observed initiation sample. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. First-time initiations refer to the initiations of coverage on a firm for the first time by both analysts and brokers. Analyst re-initiations refer to the initiations by analysts on a firm that they had previously covered and dropped. Broker re-initiations refer to the initiations by brokers on a firm that they had previously covered and dropped. All continuous variables are winsorized at the 1% and 99% levels. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix B.

Table 3
Firm, analyst, and rating characteristics for observed vs. inferred initiations

Panel A: First-by-analyst method									
Sample:	Observed initiations	Observed initiations, less missed initiations	Inferred initiations						Mean differences
Composition:	Correctly identified + Missed	Correctly identified	Correctly identified + Misidentified	Overall impact [= (3)-(1)]		Impact of type II error [= (2)-(1)]		Impact of type I error [= (3)-(2)]	
Variable (mean)	(1)	(2)	(3)	(4)		(5)		(6)	
Firm characteristics:									
BM_{it-1}	0.495	0.500	0.497	0.003		0.005		-0.002	
$FIRMAGE_{it}$	19.441	18.280	18.841	-0.601	***	-1.161	***	0.561	***
$INTANGIBLE_{it-1}$	0.167	0.165	0.165	-0.002		-0.002		0.000	
IO_{it-1}	0.676	0.655	0.661	-0.015	***	-0.021	***	0.007	**
$LEVERAGE_{it-1}$	0.217	0.215	0.214	-0.002		-0.002		0.000	
$MOMENTUM_{it-1}$	0.248	0.263	0.232	-0.016	**	0.015	**	-0.031	***
$Ln\ (MV)_{it-1}$	14.337	14.126	14.197	-0.140	***	-0.211	***	0.071	***
$NANALYST_{it-1}$	15.213	13.601	14.240	-0.973	***	-1.612	***	0.639	***
$NSEGMENT_{it-1}$	2.171	2.148	2.142	-0.029		-0.023		-0.006	
ROA_{it-1}	0.008	0.003	0.005	-0.003	**	-0.005	***	0.002	
	(*N*=19,611)	(*N*=13,291)	(*N*=19,208)						
Analyst characteristics:									
$ALLSTAR_{jt-1}$	0.022	0.020	0.023	0.001		-0.001		0.002	
$BSIZE_{jt-1}$	42.205	43.000	43.359	1.153	***	0.795	*	0.359	
$GEXP_{jt}$	8.004	7.493	7.262	-0.743	***	-0.511	***	-0.231	***
$NFIRM_{jt-1}$	14.110	13.704	13.976	-0.134	*	-0.406	***	0.272	***
$NFITEM_{ijt}$	2.290	2.263	2.241	-0.049	***	-0.027	***	-0.022	**
$NIND_{jt-1}$	4.239	4.266	4.315	0.076	***	0.027		0.049	
$ACCURACY_{jt-1}$	-0.036	-0.038	-0.038	-0.001		-0.001		0.000	
$FREQ_{jt-1}$	1.509	1.489	1.479	-0.031	***	-0.020	***	-0.010	**
	(*N*=16,982)	(*N*=11,528)	(*N*=16,523)						
Distribution of recommendation ratings:									
Mean (BUY RATING)	0.551	0.555	0.527	-0.024	***	0.004		-0.028	***
	(*N*=28,139)	(*N*=20,382)	(*N*=29,391)						

Table 3 (*Continued*)

Panel B: First-by-broker method									
Sample:	Observed initiations	Observed initiations, less missed initiations	Inferred initiations	Mean differences					
Composition:	Correctly identified + Missed	Correctly identified	Correctly identified + Misidentified	Overall impact [= (3)-(1)]		Impact of type II error [= (2)-(1)]		Impact of type I error [= (3)-(2)]	
Variable (mean)	(1)	(2)	(3)	(4)		(5)		(6)	
Firm characteristics:									
BM_{it-1}	0.495	0.497	0.495	0.001		0.002		-0.001	
$FIRMAGE_{it}$	19.441	17.415	17.953	-1.488	***	-2.027	***	0.539	***
$INTANGIBLE_{it-1}$	0.167	0.162	0.161	-0.006	***	-0.005	**	-0.001	
IO_{it-1}	0.676	0.656	0.659	-0.017	***	-0.020	***	0.002	
$LEVERAGE_{it-1}$	0.217	0.211	0.212	-0.004	**	-0.005	**	0.001	
$MOMENTUM_{it-1}$	0.248	0.266	0.243	-0.005		0.018	***	-0.023	***
$Ln\ (MV)_{it-1}$	14.337	14.018	14.076	-0.261	***	-0.319	***	0.059	***
$NANALYST_{it-1}$	15.213	13.174	13.518	-1.694	***	-2.039	***	0.344	***
$NSEGMENT_{it-1}$	2.171	2.105	2.113	-0.058	***	-0.066	***	0.008	
ROA_{it-1}	0.008	-0.002	-0.001	-0.009	***	-0.010	***	0.001	
	(*N*=19,611)	(*N*=13,470)	(*N*=16,492)						
Analyst characteristics:									
$ALLSTAR_{jt-1}$	0.022	0.013	0.013	-0.009	***	-0.008	***	-0.001	
$BSIZE_{jt-1}$	42.205	34.826	32.785	-9.420	***	-7.379	***	-2.041	***
$GEXP_{jt}$	8.004	8.131	8.118	0.114	**	0.126	**	-0.013	
$NFIRM_{jt-1}$	14.110	14.085	14.060	-0.050		-0.025		-0.025	
$NFITEM_{ijt}$	2.290	2.282	2.278	-0.013		-0.008		-0.004	
$NIND_{jt-1}$	4.239	4.302	4.282	0.043		0.063	**	-0.020	
$ACCURACY_{jt-1}$	-0.036	-0.039	-0.040	-0.003	***	-0.002	**	-0.001	
$FREQ_{jt-1}$	1.509	1.501	1.494	-0.016	***	-0.008	**	-0.008	*
	(*N*=16,982)	(*N*=12,136)	(*N*=14,620)						
Distribution of recommendation ratings:									
Mean (BUY RATING)	0.551	0.598	0.598	0.047	***	0.047	***	0.000	
	(*N*=28,139)	(*N*=18,831)	(*N*=22,823)						

Table 3 (*Continued*)

Panel C: First-by-both method									
Sample:	Observed initiations	Observed initiations, less missed initiations	Inferred initiations	Mean differences					
Composition:	Correctly identified + Missed	Correctly identified	Correctly identified + Misidentified	Overall impact [= (3)-(1)]		Impact of type II error [= (2)-(1)]		Impact of type I error [= (3)-(2)]	
Variable (mean)	(1)	(2)	(3)	(4)		(5)		(6)	
Firm characteristics:									
BM_{it-1}	0.495	0.504	0.503	0.009		0.009		-0.001	
$FIRMAGE_{it}$	19.441	16.261	16.718	-2.724	***	-3.181	***	0.457	**
$INTANGIBLE_{it-1}$	0.167	0.160	0.160	-0.007	***	-0.006	***	0.000	
IO_{it-1}	0.676	0.633	0.636	-0.039	***	-0.043	***	0.003	
$LEVERAGE_{it-1}$	0.217	0.209	0.209	-0.007	***	-0.007	***	0.000	
$MOMENTUM_{it-1}$	0.248	0.283	0.259	0.011		0.034	***	-0.023	**
$Ln\ (MV)_{it-1}$	14.337	13.823	13.870	-0.467	***	-0.514	***	0.048	**
$NANALYST_{it-1}$	15.213	11.708	11.991	-3.222	***	-3.505	***	0.283	**
$NSEGMENT_{it-1}$	2.171	2.067	2.074	-0.096	***	-0.104	***	0.007	
ROA_{it-1}	0.008	-0.006	-0.005	-0.013	***	-0.015	***	0.001	
	(*N*=19,611)	(*N*=9,546)	(*N*=11,596)						
Analyst characteristics:									
$ALLSTAR_{jt-1}$	0.022	0.015	0.014	-0.008	***	-0.007	***	-0.001	
$BSIZE_{jt-1}$	42.205	36.577	34.641	-7.565	***	-5.629	***	-1.936	***
$GEXP_{jt}$	8.004	7.728	7.635	-0.369	***	-0.277	***	-0.093	
$NFIRM_{jt-1}$	14.110	13.841	13.668	-0.442	***	-0.269	***	-0.173	*
$NFITEM_{ijt}$	2.290	2.258	2.252	-0.038	***	-0.032	***	-0.006	
$NIND_{jt-1}$	4.239	4.344	4.313	0.074	**	0.105	***	-0.031	
$ACCURACY_{jt-1}$	-0.036	-0.040	-0.041	-0.005	***	-0.003	***	-0.001	
$FREQ_{it-1}$	1.509	1.489	1.476	-0.034	***	-0.020	***	-0.013	***
	(*N*=16,982)	(*N*=8,743)	(*N*=10,360)						
Distribution of recommendation ratings:									
Mean (BUY RATING)	0.551	0.603	0.605	0.054	***	0.052	***	0.002	
	(*N*=28,139)	(*N*=14,298)	(*N*=17,229)						

Table 3 (*Continued*)

This table reports results of comparing firm, analyst, and rating characteristics between the observed initiation sample and the inferred initiation samples. Panels A, B, and C construct inferred initiation samples using the first-by-analyst, first-by-broker, and first-by-both methods, respectively. Observed initiations (Column (1)) are directly from analysts' initiating reports announcing coverage initiations from the Thomson ONE database. Inferred initiations (Column (3)) are the first stock recommendations on a firm in the IBES database by an analyst, by a broker, or by both. Correctly identified initiations (Column (2)) refer to inferred initiations that are also in the sample of observed initiations. Missed initiations refer to observed initiations that are not in the sample of inferred initiations. Misidentified initiations refer to inferred initiations that are not in the sample of observed initiations. We tabulate the impact of overall errors, type II (omission) errors, and type I (misidentification) errors in Columns (4)-(6), respectively. All continuous variables are winsorized at the 1% and 99% levels. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix B.

Table 4
Associations between measurement errors and firm and analyst characteristics

Dependent variable:	MISIDENTIFIED (Type I error)			MISSED (Type II error)		
Method:	First-by-analyst	First-by-broker	First-by-both	First-by-analyst	First-by-broker	First-by-both
	(1)	(2)	(3)	(4)	(5)	(6)
Firm characteristics						
BM_{it-1}	-0.004	0.048	0.068**	0.037	0.101***	0.061**
	(-0.177)	(1.609)	(1.970)	(1.411)	(3.519)	(2.156)
$FIRMAGE_{it}$	0.003***	0.002**	0.001	0.004***	0.008***	0.008***
	(3.367)	(2.545)	(0.744)	(5.034)	(9.458)	(9.610)
$INTANGIBLE_{it-1}$	-0.088	-0.169**	-0.160*	-0.078	0.054	-0.055
	(-1.453)	(-2.116)	(-1.684)	(-1.320)	(0.867)	(-0.904)
IO_{it-1}	0.016	-0.121**	-0.144**	0.327***	0.151***	0.298***
	(0.368)	(-2.275)	(-2.211)	(7.249)	(3.169)	(6.536)
$LEVERAGE_{it-1}$	0.057	0.039	0.007	0.107**	0.198***	0.176***
	(1.054)	(0.588)	(0.092)	(1.966)	(3.549)	(3.231)
$MOMENTUM_{it-1}$	-0.102***	-0.108***	-0.087***	0.001	-0.019	-0.017
	(-5.188)	(-4.474)	(-3.130)	(0.043)	(-0.910)	(-0.952)
$Ln\ (MV)_{it-1}$	-0.017	0.069***	0.049***	0.004	0.014	0.005
	(-1.465)	(4.589)	(2.604)	(0.305)	(1.121)	(0.379)
$NANALYST_{it-1}$	0.016***	0.005**	0.008***	0.025***	0.030***	0.037***
	(9.596)	(2.384)	(2.882)	(15.278)	(18.096)	(20.017)
$NSEGMENT_{it-1}$	-0.008	0.002	0.004	-0.006	-0.020***	-0.009
	(-1.297)	(0.233)	(0.378)	(-1.036)	(-3.216)	(-1.431)
ROA_{it-1}	0.050	-0.103	0.065	-0.166**	0.030	-0.086
	(0.655)	(-1.137)	(0.591)	(-2.032)	(0.324)	(-1.089)
Analyst characteristics						
$ALLSTAR_{jt-1}$	0.264***	0.475***	0.553***	0.033	0.048	-0.010
	(3.392)	(2.839)	(2.756)	(0.414)	(0.616)	(-0.124)
$BSIZE_{jt-1}$	-0.001	-0.015***	-0.014***	-0.003***	0.012***	0.005***
	(-1.588)	(-13.226)	(-10.306)	(-8.609)	(30.723)	(13.763)
$GEXP_{jt}$	-0.029***	-0.005	-0.008**	0.036***	-0.011***	0.017***
	(-10.707)	(-1.620)	(-2.058)	(15.717)	(-4.362)	(7.099)
$NFIRM_{jt-1}$	0.019***	0.004	-0.003	0.014***	0.005**	0.009***
	(9.945)	(1.590)	(-0.965)	(7.112)	(2.521)	(4.627)
$NFITEM_{ijt}$	-0.087***	-0.005	-0.019	0.105***	-0.046***	0.047***
	(-5.572)	(-0.256)	(-0.802)	(6.756)	(-2.848)	(3.046)
$NIND_{jt-1}$	0.009*	-0.005	0.002	-0.029***	-0.008	-0.025***
	(1.744)	(-0.698)	(0.295)	(-5.297)	(-1.329)	(-4.798)
$ACCURACY_{jt-1}$	0.059	-0.220	-0.269	0.304**	0.240*	0.320**
	(0.431)	(-1.420)	(-1.584)	(2.076)	(1.685)	(2.348)
$FREQ_{jt-1}$	-0.278***	-0.231***	-0.321***	0.084**	-0.063*	-0.038
	(-8.124)	(-5.351)	(-5.881)	(2.486)	(-1.746)	(-1.071)
No. of observations	14,797	12,926	8,799	15,388	15,388	15,388
Pseudo R^2	0.032	0.063	0.063	0.075	0.154	0.112

Table 4 (*Continued*)

This table reports the results of estimating probit models with *MISIDENTIFIED* as the dependent variable in Columns (1), (2), and (3), or *MISSED* as the dependent variable in Columns (4), (5), and (6). Inferred initiations are based on the first-by-analyst method in Columns (1) and (4), the first-by-broker method in Columns (2) and (5), and the first-by-both method in Columns (3) and (6), respectively. The base benchmark group is the correctly identified initiations in all columns. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. Inferred initiations are the first stock recommendations on a firm in the IBES database by an analyst (first-by-analyst), by a broker (first-by-broker), or by both (first-by-both). Correctly identified initiations refer to inferred initiations that are in the sample of observed initiations. Misidentified initiations refer to inferred initiations that are not in the sample of observed initiations. Missed initiations refer to observed initiations that are not in the sample of inferred initiations. The coefficient on the intercept is not reported. In parentheses below coefficient estimates are *t*-statistics based on standard errors clustered by firm. All continuous variables are winsorized at the 1% and 99% levels. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix B.

Table 5
Robustness test: excluding re-initiations with a non-coverage period shorter than six months

Sample:	Observed initiations	Inferred initiations			Mean differences					
Method:		First-by-analyst	First-by-broker	First-by-both	[= (2)-(1)]		[= (3)-(1)]		[= (4)-(1)]	
Variable (mean)	(1)	(2)	(3)	(4)	(5)		(6)		(7)	
Firm characteristics:										
BM_{it-1}	0.506	0.500	0.503	0.503	-0.006		-0.002		-0.002	
$FIRMAGE_{it}$	18.734	18.740	17.611	16.718	0.006		-1.123	***	-2.017	***
$INTANGIBLE_{it-1}$	0.164	0.165	0.160	0.160	0.000		-0.004	*	-0.004	*
IO_{it-1}	0.660	0.659	0.648	0.636	0.000		-0.012	***	-0.023	***
$LEVERAGE_{it-1}$	0.217	0.214	0.213	0.209	-0.003		-0.004	*	-0.007	***
$MOMENTUM_{it-1}$	0.254	0.233	0.241	0.259	-0.022	***	-0.013	*	0.005	
Ln $(MV)_{it-1}$	14.165	14.160	14.008	13.870	-0.005		-0.158	***	-0.295	***
$NANALYST_{it-1}$	13.817	13.949	12.949	11.991	0.132		-0.869	***	-1.827	***
$NSEGMENT_{it-1}$	2.130	2.125	2.091	2.074	-0.005		-0.039		-0.055	**
ROA_{it-1}	0.004	0.004	-0.003	-0.005	0.000		-0.007	***	-0.009	***
	(*N*=14,582)	(*N*=18,286)	(*N*=14,075)	(*N*=11,596)						
Analyst characteristics:										
$ALLSTAR_{jt-1}$	0.020	0.022	0.013	0.014	0.002		-0.007	***	-0.006	***
$BSIZE_{jt-1}$	40.672	42.133	33.061	34.641	1.461	***	-7.611	***	-6.031	***
$GEXP_{jt}$	7.887	7.295	7.991	7.635	-0.591	***	0.105		-0.252	***
$NFIRM_{jt-1}$	13.865	14.009	13.914	13.668	0.143	*	0.049		-0.197	**
$NFITEM_{ijt}$	2.267	2.239	2.262	2.252	-0.028	***	-0.006		-0.015	*
$NIND_{jt-1}$	4.248	4.339	4.267	4.313	0.091	***	0.019		0.065	**
$ACCURACY_{jt-1}$	-0.038	-0.038	-0.041	-0.041	0.000		-0.003	***	-0.003	***
$FREQ_{jt-1}$	1.487	1.478	1.479	1.476	-0.009	**	-0.008	*	-0.012	**
	(*N*=12,656)	(*N*=15,839)	(*N*=12,450)	(*N*=10,360)						
Distribution of recommendation ratings:										
Mean (BUY RATING)	0.563	0.533	0.599	0.605	-0.029	***	0.037	***	0.042	***
	(*N*=21,635)	(*N*=27,905)	(*N*=20,074)	(*N*=17,229)						

Table 5 (*Continued*)

This table reports the results of comparing firm, analyst, and rating characteristics between observed and inferred initiations after excluding re-initiations with a non-coverage period shorter than six months. A non-coverage period is defined as the number of days between a re-initiation and its preceding coverage stoppage. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. Inferred initiations are the first stock recommendations on a firm in the IBES database by an analyst (first-by-analyst; Column (2)), by a broker (first-by-broker, Column (3)), or by both (first-by-both, Column (4)). All continuous variables are winsorized at the 1% and 99% levels. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix B.

Table 6
Results on the degree of self-selection and economic determinants of analyst initiations based on observed vs. inferred initiations

Panel A: Self-selection				
Sample:	Observed initiations	Inferred initiations		
Method:		First-by-analyst	First-by-broker	First-by-both
% of Buys in	(1)	(2)	(3)	(4)
Initiations	55.1	52.7	59.8	60.5
Non-initiations	40.8	41.6	40.7	41.7
Difference:	14.3	11.1	19.1	18.8
(i.e., analysts' self-selection)				
No. of Initiations	28,139	29,391	22,823	17,229
No. of Non-initiations	66,139	129,731	142,941	152,557

Panel B: Determinants of initiations				
Dependent variable: INITIATION				
Definition:	Observed initiations	Inferred initiations		
		First-by-analyst	First-by-broker	First-by-both
	(1)	(2)	(3)	(4)
$FSIZE_{iq}$	-0.052***	0.021***	-0.079***	-0.079***
	(-5.165)	(3.246)	(-11.128)	(-9.887)
BM_{iq}	-0.407***	-0.187***	-0.269***	-0.296***
	(-14.730)	(-9.907)	(-12.845)	(-12.880)
$SALES\_GR_{iq}$	0.130***	0.156***	0.166***	0.172***
	(9.095)	(11.559)	(12.105)	(12.058)
ROA_{iq}	-0.485***	-0.581***	-0.607***	-0.142
	(-2.590)	(-3.601)	(-3.711)	(-0.790)
$MOMENTUM_{iq}$	0.355***	0.282***	0.362***	0.406***
	(20.387)	(17.521)	(21.680)	(23.506)
$NANALYST_{iq}$	0.015***	0.000	0.002*	-0.015***
	(7.402)	(0.038)	(1.767)	(-8.792)
IO_{iq}	0.426***	-0.197***	-0.034	-0.205***
	(9.664)	(-6.469)	(-1.007)	(-5.768)
$FIRMAGE_{iq}$	0.001**	-0.006***	-0.006***	-0.010***
	(1.971)	(-11.081)	(-10.670)	(-12.985)
IPO_{iq}	0.571***	0.544***	0.798***	0.906***
	(13.529)	(14.326)	(20.448)	(22.831)
Intercept	-1.043***	-0.255***	-0.048	-0.142***
	(-16.173)	(-5.753)	(-1.001)	(-2.730)
No. of observations	88,678	88,678	88,678	88,678
Pseudo R^2	0.0180	0.0144	0.0263	0.0445

Table 6 (*Continued*)

This table reports the results on the degree of analyst self-selection (Panel A) and determinants of initiations (Panel B) based on observed vs. inferred initiations. Analyst self-selection is measured by the difference between the percentages of buy ratings for initiations and that for non-initiations. Initiations are observed initiations in Column (1) and inferred initiations based on the first-by-analyst, first-by-broker, and first-by-both methods in Columns (2), (3), and (4), respectively. Non-initiations are the recommendations that are in the common coverage overlap between IBES and Thomson ONE but not defined as initiations in each column. Panel B reports the results of estimating probit models in which the dependent variable is *INITIATION*, an indicator equal to one for initiations and zero otherwise. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. Inferred initiations are the first stock recommendations on a firm in the IBES database by an analyst (first-by-analyst, Column (2)), by a broker (first-by-broker, Column (3)), or by both (first-by-both, Column (4)). The coefficient on the intercept is not reported. In parentheses below coefficient estimates are *t*-statistics based on standard errors clustered by firm. All continuous variables are winsorized at the 1% and 99% levels. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix B.

Table 7
The effectiveness of prior approaches to mitigate the measurement errors in the traditional methods

	Inferred initiations			Measurement errors	
	Correctly identified	Misidentified	Missed	Misidentification (Type I) error rate = (2)/[(1)+(2)]	Omission (Type II) error rate = (3)/[(1)+(3)]
Prior approach	(1)	(2)	(3)	(4)	(5)
[1] Malmendier and Shanthikumar (2007)					
[Base] *First-by-broker* method:	18,831	3,992	9,308	17.5%	33.1%
[Mitigation] Additionally including a broker's new recommendation since its previous coverage stoppage:					
	26,945	17,928	1,194	40.0%	4.2%
[2] Ljungqvist et al. (2009)					
[Base] *First-by-broker* method:	18,831	3,992	9,308	17.5%	33.1%
[Mitigation] Additionally including a broker's new recommendation since its previous coverage stoppage or that has no preceding recommendation in the past 12 months:					
	27,295	18,308	844	40.1%	3.0%
[3] Bernhardt et al. (2016)					
[Base] *First-by-analyst* method:	20,382	9,009	7,757	30.7%	27.6%
[Mitigation] Additionally including an analyst's recommendation that has no preceding recommendation in the past 12 months:					
	22,478	10,525	5,661	31.9%	20.1%

This table reports results of assessing the effectiveness of prior approaches that modify the traditional methods of inferring initiations to mitigate the omission error. Columns (1)-(5) report changes in the numbers of correctly identified initiations, misidentified initiations, and missed initiations, as well as the misidentification and omission error rates for applying each of these prior approaches, respectively. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. Inferred initiations are the first stock recommendations on a firm in the IBES database by a broker (the first-by-broker method, as in Malmendier and Shanthikumar 2007; Ljungqvist et al. 2009) or by an analyst (the first-by-analyst method, as in Bernhardt et al. 2016). Note that each prior approach further complements the inferred initiation sample by including additional recommendations that meet their own condition as described in the table. Correctly identified initiations refer to inferred initiations that are also in the sample of observed initiations. Missed initiations are observed initiations that are not in the sample of inferred initiations. Misidentified initiations are inferred initiations that are not in the sample of observed initiations.

Table 8
A new approach to mitigate the measurement errors in the traditional methods

	Inferred initiations			Two measurement errors	
	Correctly identified	Misidentified	Missed	Misidentification (Type I) error rate = (2)/[(1)+(2)]	Omission (Type II) error rate = (3)/[(1)+(3)]
	(1)	(2)	(3)	(4)	(5)
[Base] *First-by-broker* method:					
	18,831	3,992	9,308	17.5%	33.1%
[Step 1] Remove inferred initiations with any EPS or TP forecasts issued by the same broker for the same firm within the past six months:					
	18,466	3,281	9,673	15.1%	34.4%
[Step 2] Additionally include a recommendation that meets ALL the following criteria:					
(1) It is a new recommendation by a broker since its previous stoppage.					
(2) The broker did not issue any EPS or TP forecasts throughout its non-coverage period.					
(3) The non-coverage period is longer than:					
30 day (1 month)	25,565	7,192	2,574	22.0%	9.1%
91 days (3 months)	24,930	5,392	3,209	17.8%	11.4%
182 days (6 months)	23,802	4,416	4,337	15.6%	15.4%
365 days (12 months, 1 year)	22,438	4,013	5,701	15.2%	20.3%
730 days (24 months, 2 years)	21,215	3,696	6,924	14.8%	24.6%

This table reports the results of assessing the effectiveness of a new approach that modifies the traditional methods of inferring initiations to mitigate the measurement errors. Columns (1)-(5) report changes in the numbers of correctly identified, misidentified, and missed initiations, as well as the misidentification and omission error rates after applying our new approach, respectively. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. Inferred initiations are the first stock recommendations on firms in the IBES database by brokers (the first-by-broker method). Note that we further modify the inferred initiation sample by including additional recommendations that meet our conditions outlined in the table. Correctly identified initiations refer to inferred initiations that are also in the sample of observed initiations. Missed initiations are observed initiations that are not in the sample of inferred initiations. Misidentified initiations are inferred initiations that are not in the sample of observed initiations.

Exhibit A
An example of the misidentification (Type I) error

Exhibit A-1: The first recommendation for a firm by both a broker and an analyst on IBES is inferred as an initiation.

IBES Ticker Symbol	Company Name	Announce Date, SAS Format	Review Date, SAS Format	Estimator ID	Analyst Name	Estimator Text
CREE	CREE INC	17JAN2003	31JUL2003	HALLUM	SEPENZIS T	EQUALWEIGHT
CREE	CREE INC	25SEP2003	02JUL2004	HALLUM	SEPENZIS T	OVERWEIGHT
CREE	CREE INC	15JUL2004	02JUN2006	HALLUM	SEPENZIS T	BUY
CREE	CREE INC	13JUL2006	20APR2007	HALLUM	BURTON M	ACCUMULATE
CREE	CREE INC	06JUN2007	22APR2009	HALLUM	BURTON M	SELL
CREE	CREE INC	23JUN2009	09SEP2009	HALLUM	RAKESH V	ACCUMULATE
CREE	CREE INC	21OCT2009	13NOV2009	HALLUM	RAKESH V	BUY
CREE	CREE INC	09DEC2009	20JAN2010	HALLUM	RAKESH V	BUY
CREE	CREE INC	08SEP2011	19OCT2011	HALLUM	RUSCH C	HOLD
CREE	CREE INC	18JAN2012	18JAN2012	HALLUM	RUSCH C	BUY
CREE	CREE INC	18APR2012	20APR2012	HALLUM	RUSCH C	HOLD
CREE	CREE INC	28JUN2012	11AUG2012	HALLUM	RUSCH C	BUY

Exhibit A-2: The inferred initiation, issued on January 17, 2003, is misidentified. It is not an actual initiation.

ThinkEquity
partners

Technology:
Reported Results **January 17, 2003**

Cree (CREE)

Equal Weight

Price	$18.39
Price Target	$25.00
52 Week High	$33.00
52 Week Low	$8.98
Shr.O/S-Basic (Mils.)	73.1
Market Cap.(Mils.)	$1,344.0
Avg Dly Vol (3 Mo Mils)	2.829

CREE Tops Estimates - Raising Guidance

Cree beat top and bottom line estimates, both ours and consensus, and has raised guidance for the March quarter on strong LED growth driven by wireless handsets and automotive applications. We are raising our forward estimates and target price.

Fiscal Year Jun	F02A	F03E	F04E	2Q03A	3Q03E	3Q02A
Revenue(Current $)	$155.40	$231.10	$296.90	$56.70	$60.70	$33.40
Revenue(Previous $)	$155.40	$205.50	$262.30	$50.40	$51.40	$33.40
EPS(Current $)	$0.20	$0.47	$0.77	$0.12	$0.14	$0.00
EPS(Previous $)	$0.20	$0.34	$0.60	$0.08	$0.10	$0.00
P/S(Current)	8.6	5.8	4.5	--	--	--
P/E(Current)	92.0	39.1	23.9	--	--	--

Exhibit A-3: The actual initiation was issued earlier on January 9, 2003. It is missed in IBES.

ThinkEquity
partners

Technology:
Initiation of Coverage

January 09, 2003

Cree (CREE)

Equal Weight

Price	$18.00
Price Target	$20.00
52 Week High	$33.00
52 Week Low	$8.98
Shr.O/S-Basic (Mils.)	74.5
Market Cap.(Mils.)	$1.3
Avg Dly Vol (3 Mo Mils)	2.850
Net Cash (Mils)	$96

Cree - Blue Light Special

Cree represents one of the few manufacturing companies that has had to add capacity in the past two years, with market leading LED products and singular process technologies. While the company is fairly valued at current estimates, continued growth in the LED segment could keep the stock moving higher in the long term. We are initiating coverage of CREE with an Equal Weight rating and a $20.00 target price.

Fiscal Year Jun	F02A	F03E	F04E	1Q03A	2Q03E	2Q02A
Revenue(Current $)	$155.40	$205.50	$262.30	$48.80	$50.40	$41.10
Revenue(Previous $)	--	--	--	--	--	--
EPS(Current $)	$0.20	$0.34	$0.60	$0.05	$0.08	$0.08
EPS(Previous $)	--	--	--	--	--	--
P/S(Current)	0.0	0.0	0.0	--	--	--
P/E(Current)	90.0	52.9	30.0	--	--	--

Exhibit B
An example of the omission (Type II) error

Exhibit B-1: The first recommendation for a firm by both a broker and an analyst on IBES is inferred as an initiation.

IBES Ticker Symbol	Company Name	Announce Date, SAS Format	Review Date, SAS Format	Estimator ID	Analyst Name		Estimator Text
SUG	SOUTHN UNION CO	12MAR2004	17MAR2005	JEFFEREG	FREMONT	P	BUY
SUG	SOUTHN UNION CO	24MAY2011	24MAY2011	JEFFEREG	FREMONT	P	BUY
SUG	SOUTHN UNION CO	17JUN2011	06JUL2011	JEFFEREG	FREMONT	P	HOLD

Exhibit B-2: The inferred initiation, issued on March 12, 2004, is correctly identified. It is an actual initiation.

JEFFERIES

Electric Utilities **Initiation Report – March 12, 2004**

Paul B. Fremont (212) 284-2466 pfremont@jefco.com
Brooke K. Glenn (212) 284-2467 bglenn@jefco.com
Debra E. Bromberg (212) 284-2452 dbromberg@jefco.com

Southern Union Company

NYSE: SUG - $18.56

Rating: Buy

		FY June	2003A	2004E	2005E	2006E
52 Week Range	$11.50–19.75	1Q	($0.29)A	($0.10)A	–	–
Shares Out - FD (MM)	73.1	2Q	$0.50A	$0.47A	--	--
Float (MM)	55.5	3Q	$0.79A	$1.00E	--	--
Average Daily Vol (000)	237	4Q	($0.20)A	($0.05)E	--	--
Equity Market Cap (MM)	$1,360	EPS	$0.74	$1.30	$1.45	$1.55
Price Target	$22.00	P/E	25.1x	14.3x	12.8x	12.0x
		($MM), except per share data.				

SUG: Low Risk Operations with Above Average Growth

We are initiating coverage of Southern Union Company with a **Buy** rating and a $22.00 price target. We are initiating our FY2004, FY2005 and FY2006 EPS estimates of $1.30, $1.45 and $1.55. The company currently trades at a Price/Earnings ratio of 12.8x, which equates to a 15.0% discount to its gas distribution peers. We believe that the company should trade at the gas distribution group average multiple. While the company has higher leverage than its peers and does not pay a dividend, we believe that these negative factors are offset by above-average growth. The company's management currently has reinvestment options for its cash and therefore has chosen to redeploy the cash into the business as opposed to paying a dividend. The company's 12% CAGR through 2007 compares favorably to the industry average rate of 3–4%. While the leverage is currently high, the company's management has not only committed to improving the balance sheet but also has the means to do so through the redeployment of its predictable, regulated cash flow streams. We see the company's adjusted leverage improving to 56% by the end of FY 2005. While a portion of the company's operations are located in Missouri, which historically has been a difficult jurisdiction, our estimates continue to reflect low returns on equity and therefore we have not attributed a regulatory discount in the valuation. Using a P/E multiple equal to gas distribution group average results in a 15.0x PE multiple, which when combined with our 2005 EPS estimate of $1.45 results in a Price Target of $22.00. Based on yesterday's closing price of $18.56 results in potential price appreciation of 18.5%.

Exhibit B-3: IBES stop file indicates that the coverage is terminated on April 27, 2005.

IBES Ticker Symbol	CUSIP/SEDOL	Company Name	Official Ticker Symbol	Activation Date, SAS Format	Estimator ID
SUG	84403010	SOUTHN UNION CO	SUG	27APR2005	JEFFEREG
SUG	84403010	SOUTHN UNION CO	SUG	16DEC2011	JEFFEREG

Exhibit B-4: A new initiation is issued six years later (dated May 24, 2011 on IBES).

COMPANY NOTE

Initiating Coverage

USA | Energy | Electric Utilities

May 23, 2011

Jefferies

Southern Union Co. (SUG)

Initiate with Buy: This Is No Pipe Dream

BUY
Price target $33.50
Price $29.57

Key Takeaway

In our opinion, Southern Union is under-valued and levered to changes in oil and natural gas prices. We believe that Southern Union's transmission business ("pipe") segment is trading at a 15% discount to our group average multiple for MLP stocks adjusted for taxes. If the company were able to execute on its plan to build an LNG export facility our valuation would increase by $2.00 per share.

The proposed LNG terminal at Lake Charles, LA, could provide the company with additional growth that is not included in our forecast, which could add $2 to the valuation.

Southern Union generates $200-$250 million of positive free cash flow. Jefferies sees this cash as the primary source of funding for new projects.

Roughly 10% of Southern Union's EBITDA remains exposed to changes in commodity prices for oil and gas. Commodity price exposure relates primarily to Southern Union's gas and processing business.

Valuation/Risks

Valuation: The $33.50 price target is based on our revised EV/EBITDA analysis. For Southern Union's transmission segment, we are using a 2012 average multiple of 10.7x, which is a 5% discount to our MLP group average adjusted for taxes. On the processing segment, we apply a 10.9x multiple and the gas distribution assets are valued based on an 8.5x multiple. Together, we arrive at a combined company price target of $33.50. Based on yesterday's closing price of $29.44, this should result in 12-month price appreciation and total return potential of 13.8% and 15.8%, respectively, including the company's current dividend yield of 2.0%. Our valuation does not include the proposed LNG export facility, which could add $2.00 per share.

Financial Summary	
Book Value (MM):	$109.6
Book Value/Share:	$20.62
Net Debt (MM):	$4,800.0
Dividend Yield:	2.0%

Market Data	
52 Week Range:	$30.38 - $20.00
Total Entprs. Value (MM):	$8,499.2
Market Cap. (MM):	$3,699.2
Insider Ownership:	7.2%
Institutional Ownership:	71.0%
Shares Out. (MM):	125.1
Float (MM):	109.6
Avg. Daily Vol.:	876

Paul B. Fremont *
Equity Analyst
(212) 284-2466 pfremont@jefferies.com

Debra E. Bromberg *
Equity Analyst
(212) 284-2452 dbromberg@jefferies.com

Online Appendix to

"Identification of analyst coverage initiations"

JAY HEON JUNG
City, University of London
jay.jung@city.ac.uk

YONG YU
University of Texas at Austin
yong.yu@mccombs.utexas.edu

WUYANG ZHAO
University of Texas at Austin
wuyang.zhao@mccombs.utexas.edu

June 2023

OA.1 Calculation of relative ranks

We transform each raw variable to a ranked variable ranging from 0 (min) to 1 (max). The transformation is performed either among analysts following the same firm in the same year (for analyst characteristics, such as brokerage size (*BSIZE*)) or across firms followed by the same analyst in the same year (for firm characteristics, such as institutional ownership (*IO*)) (Clement and Tse 2003). The relative rank variables are identified by the prefix "R_" in their variable names (e.g., *R_BSIZE*).

To calculate the rank variables for analyst characteristics, we use the formula of Equation (OA.1). Similarly, for firm characteristics, we employ the formula of Equation (OA.2):

$$R\_VAR_{ijt} = [VAR_{ijt} - \text{Min}\ (VAR_{ijt})_{it}]/\ [\text{Max}\ (VAR_{ijt})_{it} - \text{Min}\ (VAR_{ijt})_{it}] \quad \text{(OA.1)}$$

$$R\_VAR_{ijt} = [VAR_{ijt} - \text{Min}\ (VAR_{ijt})_{jt}]/\ [\text{Max}\ (VAR_{ijt})_{jt} - \text{Min}\ (VAR_{ijt})_{jt}] \quad \text{(OA.2)}$$

The subscripts *i*, *j*, and *t* denote a firm, analyst, and year, respectively. In Equation (OA.1), Max $(\cdot)_{it}$ and Min $(\cdot)_{it}$ refer to the maximum and minimum values of an analyst characteristic variable among all analysts covering firm *i* in year *t*. Likewise, in Equation (OA.2), Max $(\cdot)_{jt}$ and Min $(\cdot)_{jt}$ denote the maximum and minimum values of a firm characteristic variable across all firms followed by analyst *j* in year *t*.

When computing a relative rank variable for forecast accuracy, we use a modified version of Equation (OA.1) to improve the interpretability of the variable. The modified formula is as follows: $R\_ACCU_{ijt} = [\text{Max}\ (AFE_{ijt})_{it} - AFE_{ijt}]/\ [\text{Max}\ (AFE_{ijt})_{it} - \text{Min}\ (AFE_{ijt})_{it}]$, where AFE_{ijt} is the absolute forecast error, measured as the absolute difference between firm *i*'s actual earnings per share (EPS) and analyst *j*'s forecasted EPS in year *t*. We consider analyst *j*'s most recent one-year-ahead EPS forecast for firm *i* in year *t*. By employing this modified formula, we assign a value of 1 (0) to an analyst with the highest (lowest) forecast accuracy within the same firm and year.

OA.2 Additional tests using relative rank variables for firm and analyst characteristics

In this section, we extend our main analyses by examining relative rank variables. Table OA.1 presents the results of comparing relative rank variables for firm and analyst characteristics across three types of observed initiations: first-time initiations, analyst re-initiations, and broker re-initiations. We observe similar patterns to those using raw continuous variables for firm characteristics (see Table 2). Specifically, analyst and broker re-initiations are associated with firms that are significantly older (*R_FIRMAGE*), larger (*R_MV*), have more intangible assets (*R_INTANGIBLE*), higher institutional ownership (*R_IO*), higher leverage (*LEVERAGE*), more analysts following (*R_NANALYST*), and better firm performance (*R_ROA*) among all firms followed by the same analyst in the same year. It is important to note that relative rank variables allow for comparisons within the same analyst-year. Regarding analyst characteristics, there are some similarities in patterns compared to Table 2, but also differences in statistical significance or sign when comparing the results.

Table OA.2 presents the results of comparing relative rank variables for firm and analyst characteristics between observed initiations and inferred initiations, as well as between correctly identified initiations and either observed initiations or inferred initiations. Inferred initiations are based on the first-by-analyst, first-by-broker, and first-by-both methods in Panels A, B, and C of Table OA.2, respectively. Again, we find similar results for certain firm and analyst characteristics compared to those reported in Table 3. Regardless of the traditional method used to infer initiations, inferred initiations are associated with significantly younger firms (*R_FIRMAGE*), lower institutional ownership (*R_IO*), and fewer analysts following (*R_NANALYST*) compared to observed initiations. Regarding analyst characteristics, we also find that the main patterns hold similar: inferred initiations using the first-by-analyst method are associated with analysts from

larger brokers (*R_BSIZE*) but shorter experience (*R_GEXP*) (refer to Panel A), whereas inferred initiations using the first-by-broker method are associated with analysts from smaller brokers (*R_BSIZE*) but longer experience (*R_GEXP*) (refer to Panel B). However, we caution that the results using relative rank variables in Table OA.2 require careful interpretation as the variables capture analysts' or firms' relative positions within the same firm- or analyst-year.

Table OA.3 presents the results of a robustness test excluding re-initiations with a non-coverage period shorter than six months. Overall, we find that the results using relative rank variables remain qualitatively similar and robust when requiring a minimum six-month non-coverage period for re-initiations. Compared to the results in Table OA.2, there are a few changes in statistical significance for several rank variables for firm and analyst characteristics. However, we note that the rank variable for the number of analysts following (*R_NANALYST*) is the only one that flips its statistical significance to the opposite direction.

In summary, we find consistent evidence of systematic biases associated with the traditional methods of inferring initiations using relative rank variables. However, we caution readers in interpreting these results due to the complex nature of relative rank variables and the interplay of various factors influencing relative positions within the same firm-year or the same analyst-year.

Table OA.1
Relative rank variables for firm and analyst characteristics (observed initiations)

Type of observed initiations:	First-time initiations	Re-initiations		Mean differences					
		Analyst	Broker	[= (2)-(1)]		[= (3)-(1)]		[= (3)-(2)]	
Variable (mean)	(1)	(2)	(3)	(4)		(5)		(6)	
Firm characteristics (Relative rank variables):									
R_BM_{ijt-1}	0.390	0.395	0.392	0.005		0.002		-0.003	
$R\_FIRMAGE_{ijt}$	0.356	0.428	0.464	0.072	***	0.108	***	0.037	***
$R\_INTANGIBLE_{ijt-1}$	0.334	0.345	0.361	0.011	**	0.028	***	0.016	***
R_IO_{ijt-1}	0.602	0.676	0.672	0.074	***	0.070	***	-0.004	
$R\_LEVERAGE_{ijt-1}$	0.363	0.373	0.392	0.009	*	0.029	***	0.019	***
$R\_MOMENTUM_{ijt-1}$	0.476	0.439	0.447	-0.037	***	-0.029	***	0.009	
R_MV_{ijt-1}	0.207	0.255	0.298	0.048	***	0.091	***	0.042	***
$R\_NANALYST_{ijt-1}$	0.339	0.492	0.525	0.154	***	0.186	***	0.032	***
$R\_NSEGMENT_{ijt-1}$	0.349	0.351	0.371	0.002		0.022	***	0.020	***
R_ROA_{ijt-1}	0.525	0.543	0.548	0.018	***	0.023	***	0.005	
	(*N*=9,546)	(*N*=6,320)	(*N*=6,141)						
Analyst characteristics (Relative rank variables):									
R_BSIZE_{ijt-1}	0.265	0.208	0.290	-0.056	***	0.025	***	0.081	***
R_GEXP_{ijt}	0.436	0.477	0.400	0.041	***	-0.036	***	-0.077	***
R_NFIRM_{ijt-1}	0.413	0.418	0.393	0.005		-0.019	***	-0.025	***
R_NFITEM_{ijt}	0.645	0.654	0.619	0.009		-0.026	***	-0.035	***
R_NIND_{ijt-1}	0.398	0.376	0.365	-0.022	***	-0.033	***	-0.011	*
Mean $(R\_ACCU)_{jt-1}$	0.813	0.813	0.818	0.000		0.005	***	0.005	**
Mean $(R\_FREQ)_{jt-1}$	0.432	0.440	0.427	0.008	**	-0.005		-0.012	***
	(*N*=8,743)	(*N*=5,454)	(*N*=4,846)						

This table reports the mean levels of relative rank variables for firm and analyst characteristics across three types of observed initiations: first-time initiations, analyst re-initiations, and broker re-initiations. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. First-time initiations refer to the initiations of coverage on a firm for the first time by both analysts and brokers. Analyst re-initiations refer to the initiations by analysts on a firm that they had previously covered and dropped. Broker re-initiations refer to the initiations by brokers on a firm that they had previously covered and dropped. The prefix "R_" for a continuous variable indicates that the variable is scaled to range from 0 to 1, either among analysts following the same firm in the same year or across firms followed by the same analyst in the same year. The specific formulas for calculating the relative rank variables are described in Section OA.1. Note that we cannot calculate rank variables for *ALLSTAR* and *Mean (BUY RATING)*. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix B.

Table OA.2
Relative rank variables for firm and analyst characteristics (observed vs. inferred initiations)

Panel A: First-by-analyst method												
Sample:	Observed initiations		Observed initiations, less missed initiations		Inferred initiations		Mean differences					
Composition:	Corrected identified + Missed		Corrected identified		Corrected identified + Misidentified		Overall impact [= (3)-(1)]		Impact of type II error [= (2)-(1)]		Impact of type I error [= (3)-(2)]	
Variable (mean)	(1)		(2)		(3)		(4)		(5)		(6)	
Firm characteristics (Relative rank variables):												
R_BM_{ijt-1}	0.391		0.390		0.384		-0.007	**	-0.002		-0.005	
$R\_FIRMAGE_{ijt}$	0.401		0.388		0.393		-0.007	**	-0.013	***	0.006	
$R\_INTANGIBLE_{ijt-1}$	0.343		0.342		0.340		-0.003		-0.001		-0.002	
R_IO_{ijt-1}	0.638		0.619		0.629		-0.008	**	-0.018	***	0.010	***
$R\_LEVERAGE_{ijt-1}$	0.373		0.373		0.371		-0.002		0.000		-0.002	
$R\_MOMENTUM_{ijt-1}$	0.460		0.470		0.462		0.002		0.010	***	-0.008	**
R_MV_{ijt-1}	0.240		0.233		0.238		-0.002		-0.007	*	0.005	
$R\_NANALYST_{ijt-1}$	0.418		0.383		0.407		-0.011	***	-0.035	***	0.024	***
$R\_NSEGMENT_{ijt-1}$	0.356		0.358		0.357		0.001		0.002		-0.001	
R_ROA_{ijt-1}	0.536		0.533		0.539		0.003		-0.003		0.006	*
	(*N*=19,611)		(*N*=13,291)		(*N*=19,208)							
Analyst characteristics (Relative rank variables):												
R_BSIZE_{ijt-1}	0.254		0.275		0.266		0.012	***	0.022	***	-0.009	***
R_GEXP_{ijt}	0.436		0.416		0.399		-0.037	***	-0.019	***	-0.018	***
R_NFIRM_{ijt-1}	0.408		0.403		0.405		-0.003		-0.005		0.002	
R_NFITEM_{ijt}	0.643		0.637		0.625		-0.018	***	-0.005		-0.012	***
R_NIND_{ijt-1}	0.384		0.388		0.391		0.007	*	0.004		0.003	
Mean $(R\_ACCU)_{jt-1}$	0.813		0.814		0.813		0.000		0.000		0.000	
Mean $(R\_FREQ)_{jt-1}$	0.431		0.427		0.419		-0.012	***	-0.004	*	-0.008	***
	(*N*=16,982)		(*N*=11,528)		(*N*=16,523)							

Table OA.2 (*Continued*)

Panel B: First-by-broker method									
Sample:	Observed initiations	Observed initiations, less missed initiations	Inferred initiations	Mean differences					
Composition:	Corrected identified + Missed	Corrected identified	Corrected identified + Misidentified	Overall impact [= (3)-(1)]		Impact of type II error [= (2)-(1)]		Impact of type I error [= (3)-(2)]	
Variable (mean)	(1)	(2)	(3)	(4)		(5)		(6)	
Firm characteristics (Relative rank variables):									
R_BM_{ijt-1}	0.391	0.391	0.390	-0.002		0.000		-0.001	
$R\_FIRMAGE_{ijt}$	0.401	0.371	0.376	-0.024	***	-0.029	***	0.005	
$R\_INTANGIBLE_{ijt-1}$	0.343	0.335	0.335	-0.008	**	-0.008	**	0.000	
R_IO_{ijt-1}	0.638	0.622	0.624	-0.014	***	-0.016	***	0.002	
$R\_LEVERAGE_{ijt-1}$	0.373	0.364	0.367	-0.006		-0.009	**	0.003	
$R\_MOMENTUM_{ijt-1}$	0.460	0.466	0.463	0.003		0.006		-0.002	
R_MV_{ijt-1}	0.240	0.214	0.223	-0.017	***	-0.026	***	0.009	**
$R\_NANALYST_{ijt-1}$	0.418	0.370	0.382	-0.036	***	-0.048	***	0.013	***
$R\_NSEGMENT_{ijt-1}$	0.356	0.349	0.350	-0.006	*	-0.007	*	0.000	
R_ROA_{ijt-1}	0.536	0.530	0.532	-0.004		-0.006		0.001	
	(*N*=19,611)	(*N*=13,470)	(*N*=16,492)						
Analyst characteristics (Relative rank variables):									
R_BSIZE_{ijt-1}	0.254	0.240	0.223	-0.031	***	-0.014	***	-0.017	***
R_GEXP_{ijt}	0.436	0.450	0.447	0.012	***	0.014	***	-0.003	
R_NFIRM_{ijt-1}	0.408	0.414	0.407	-0.001		0.006	*	-0.007	*
R_NFITEM_{ijt}	0.643	0.652	0.647	0.004		0.009	**	-0.005	
R_NIND_{ijt-1}	0.384	0.392	0.388	0.004		0.008	**	-0.004	
Mean $(R\_ACCU)_{jt-1}$	0.813	0.812	0.810	-0.003	***	-0.002		-0.002	
Mean $(R\_FREQ)_{jt-1}$	0.431	0.432	0.426	-0.005	**	0.001		-0.006	**
	(*N*=16,982)	(*N*=12,136)	(*N*=14,620)						

Table OA.2 (*Continued*)

Panel C: First-by-both method									
Sample:	Observed initiations	Observed initiations, less missed initiations	Inferred initiations	Mean differences					
Composition:	Corrected identified + Missed	Corrected identified	Corrected identified + Misidentified	Overall impact [= (3)-(1)]		Impact of type II error [= (2)-(1)]		Impact of type I error [= (3)-(2)]	
Variable (mean)	(1)	(2)	(3)	(4)		(5)		(6)	
Firm characteristics (Relative rank variables):									
R_BM_{ijt-1}	0.391	0.390	0.389	-0.003		-0.001		-0.001	
$R\_FIRMAGE_{ijt}$	0.401	0.356	0.360	-0.040	***	-0.045	***	0.004	
$R\_INTANGIBLE_{ijt-1}$	0.343	0.334	0.334	-0.009	**	-0.009	**	0.001	
R_IO_{ijt-1}	0.638	0.602	0.605	-0.032	***	-0.035	***	0.003	
$R\_LEVERAGE_{ijt-1}$	0.373	0.363	0.366	-0.006		-0.009	**	0.003	
$R\_MOMENTUM_{ijt-1}$	0.460	0.476	0.475	0.015	***	0.016	***	-0.001	
R_MV_{ijt-1}	0.240	0.207	0.218	-0.023	***	-0.033	***	0.011	**
$R\_NANALYST_{ijt-1}$	0.418	0.339	0.351	-0.067	***	-0.080	***	0.012	***
$R\_NSEGMENT_{ijt-1}$	0.356	0.349	0.351	-0.006		-0.007		0.001	
R_ROA_{ijt-1}	0.536	0.525	0.529	-0.007	*	-0.011	***	0.004	
	(*N*=19,611)	(*N*=9,546)	(*N*=11,596)						
Analyst characteristics (Relative rank variables):									
R_BSIZE_{ijt-1}	0.254	0.265	0.247	-0.006	*	0.011	***	-0.017	***
R_GEXP_{ijt}	0.436	0.436	0.430	-0.006		0.000		-0.006	
R_NFIRM_{ijt-1}	0.408	0.413	0.404	-0.004		0.005		-0.009	**
R_NFITEM_{ijt}	0.643	0.645	0.642	-0.001		0.003		-0.004	
R_NIND_{ijt-1}	0.384	0.398	0.393	0.008	**	0.014	***	-0.005	
Mean $(R\_ACCU)_{jt-1}$	0.813	0.813	0.811	-0.003	**	-0.001		-0.002	
Mean $(R\_FREQ)_{jt-1}$	0.431	0.432	0.423	-0.008	***	0.001		-0.009	***
	(*N*=16,982)	(*N*=8,743)	(*N*=10,360)						

Table OA.2 (*Continued*)

This table reports results of comparing relative rank variables for firm and analyst characteristics between the observed and inferred initiation samples. Panels A, B, and C construct inferred initiation samples using the first-by-analyst, first-by-broker, and first-by-both methods, respectively. Observed initiations (Column (1)) are directly from analysts' initiating reports announcing coverage initiations from the Thomson ONE database. Inferred initiations (Column (3)) are the first stock recommendations on a firm in the IBES database by an analyst, by a broker, or by both in Panels A, B, and C, respectively. Correctly identified initiations (Column (2)) refer to inferred initiations that are also in the sample of observed initiations. Missed initiations refer to observed initiations that are not in the sample of inferred initiations. Misidentified initiations refer to inferred initiations that are not in the sample of observed initiations. We tabulate the impact of overall errors, type II (omission) errors, and type I (misidentification) errors in Columns (4), (5), and (6), respectively. The prefix "R_" for a continuous variable indicates that the variable is scaled to range from 0 to 1, either among analysts following the same firm in the same year or across firms followed by the same analyst in the same year. The specific formulas for calculating the relative rank variables are described in Section OA.1. Note that we cannot calculate rank variables for *ALLSTAR* and *Mean (BUY RATING)*. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix B.

Table OA.3
Robustness test using relative rank variables: excluding re-initiations with a non-coverage period shorter than six months

Sample:	Observed initiations	Inferred initiations			Mean differences		
Method:		First-by-analyst	First-by-broker	First-by-both	[= (2)-(1)]	[= (3)-(1)]	[= (4)-(1)]
Variable (mean)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Firm characteristics (Relative rank variables):							
R_BM_{ijt-1}	0.394	0.385	0.390	0.389	-0.009 **	-0.004	-0.006
$R\_FIRMAGE_{ijt}$	0.392	0.391	0.371	0.360	0.000	-0.020 ***	-0.031 ***
$R\_INTANGIBLE_{ijt-1}$	0.341	0.339	0.335	0.334	-0.002	-0.006	-0.007
R_IO_{ijt-1}	0.623	0.628	0.615	0.605	0.005	-0.009 **	-0.018 ***
$R\_LEVERAGE_{ijt-1}$	0.375	0.370	0.369	0.366	-0.005	-0.006	-0.009 **
$R\_MOMENTUM_{ijt-1}$	0.468	0.463	0.467	0.475	-0.004	-0.001	0.007 *
R_MV_{ijt-1}	0.232	0.235	0.222	0.218	0.003	-0.010 ***	-0.015 ***
$R\_NANALYST_{ijt-1}$	0.387	0.401	0.370	0.351	0.014 ***	-0.016 ***	-0.036 ***
$R\_NSEGMENT_{ijt-1}$	0.358	0.357	0.351	0.351	-0.002	-0.008 *	-0.008 *
R_ROA_{ijt-1}	0.531	0.537	0.528	0.529	0.006 *	-0.003	-0.002
	(N=14,582)	(N=18,286)	(N=14,075)	(N=11,596)			
Analyst characteristics (Relative rank variables):							
R_BSIZE_{ijt-1}	0.262	0.264	0.231	0.247	0.001	-0.031 ***	-0.015 ***
R_GEXP_{ijt}	0.435	0.400	0.444	0.430	-0.035 ***	0.009 **	-0.005
R_NFIRM_{ijt-1}	0.404	0.406	0.403	0.404	0.002	-0.001	-0.001
R_NFITEM_{ijt}	0.641	0.626	0.642	0.642	-0.015 ***	0.001	0.001
R_NIND_{ijt-1}	0.386	0.393	0.388	0.393	0.006	0.002	0.006
Mean $(R\_ACCU)_{jt-1}$	0.812	0.813	0.809	0.811	0.001	-0.003 **	-0.001
Mean $(R\_FREQ)_{jt-1}$	0.422	0.419	0.420	0.423	-0.003	-0.002	0.001
	(N=12,656)	(N=15,839)	(N=12,450)	(N=10,360)			

This table reports the results of comparing relative rank variables for firm and analyst characteristics between observed and inferred initiations after excluding re-initiations with a non-coverage period shorter than six months. A non-coverage period is defined as the number of days between a re-initiation and its preceding coverage stoppage. Observed initiations are directly from analysts' initiating reports announcing coverage initiations in the Thomson ONE database. Inferred initiations are the first stock recommendations on a firm in the IBES database by an analyst, by a broker, or by both in Columns (2), (3), and (4), respectively. The prefix "R_" for a continuous variable indicates that the variable is scaled to range from 0 to 1, either among analysts following the same firm in the same year or across firms followed by the same analyst in the same year. The specific formulas for calculating the relative rank variables are described in Section OA.1. Note that we cannot calculate rank variables for *ALLSTAR* and *Mean (BUY RATING)*. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Variable definitions are provided in Appendix B.