Finance Research Letters 20 (2017) 170–176

Contents lists available at ScienceDirect

Finance Research Letters

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

Reporting errors in the I/B/E/S earnings forecast database: J. Doe *vs.* J. Doe

Tristan Roger

DRM Finance, Université Paris-Dauphine, PSL Research University – Mailing, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE

ARTICLE INFO

Article history:
Received 29 July 2016
Revised 6 September 2016
Accepted 19 September 2016
Available online 28 September 2016

JEL Classification:
C8
E17
G00
G24

Keywords:
Financial analysts
Earnings forecasts
I/B/E/S database
Reporting errors
Namesakes

ABSTRACT

This paper provides evidence of systematic errors in the way I/B/E/S reports analyst earnings forecasts. Analysis of the I/B/E/S earnings forecast database over the 1982–2014 period pinpointed a lack of consistency in the identification of financial analysts, a number of whom are consequently (1) identified by several different codes, and (2) erroneously attributed forecasts that were issued by namesakes. The present empirical investigation reveals that over 10% of the analyst codes in the database are subject to such reporting errors. These reporting errors impact the evaluation of analysts' characteristics, and may bias empirical studies that rely on tracking analysts.

1. Introduction

To date, most empirical studies about financial analysts use the I/B/E/S database.[1] One important feature of the I/B/E/S database is the possibility to track analysts over time. Each analyst is identified by a unique numerical code, enabling researchers to determine the individual characteristics of each analyst (*i.e.*, their experience, the industries in which they are specialized, etc.). However, the efficient tracking of analysts requires a bijection between the set of I/B/E/S identification codes and the set of analysts. A given analyst should be identified by a single code, and a given code should correspond to a single analyst.[2] This article highlights numerous examples where: (1) several codes are used to identify the same analyst, and (2) several analysts are identified with the same code. In other words, the mapping that links the set of codes and the set of analysts is neither surjective nor injective.

E-mail address: tristan.roger@dauphine.fr

[1] For instance, Lin et al. (2013); Bosquet et al. (2014); Medovikov (2014).

[2] The I/B/E/S Detail History User Guide states (on page 7, Chapter 1 - Overview) that "*Among the many possible applications of historical detail data, notables include: [Isolating] a particular estimator or analyst [...]. The accuracy of any individual estimators forecast can be tracked over time. Each estimator, analyst or industry team is assigned a unique and independent identification number.*"

http://dx.doi.org/10.1016/j.frl.2016.09.026
1544-6123/

Imperfect mapping between codes and analysts is likely to result from the I/B/E/S data collection process, during which the attribution of identification codes by I/B/E/S appears to occur after data collection. When collecting the data, I/B/E/S identifies analysts using their last name, their first name initial and their employer. I/B/E/S then assigns codes to analysts. However, this approach leads to noisy matching between codes and analysts and can frequently result in situations such as an analyst being attributed a new code: (1) when moving to a new broker, (2) after changing their name (through marriage, for instance), (3) through spelling mistakes (for instance, an analyst is identified by one given code when his or her name is spelled *McDonald* and by another code when spelled *MacDonald*), and (4) for unidentified reasons. These matching errors are referred to hereafter as *dissociation errors. Namesake errors* constitute the second broad category of matching errors, in which two namesakes (same last name and same first name initial) are identified by the same analyst code.

The I/B/E/S database has already been the subject of criticism (Ljungqvist et al., 2009; Galanti, 2016). Ljungqvist et al. (2009) document that I/B/E/S frequently altered, added and deleted records from one version of the recommendation database to the next between 2000 and 2007. Although I/B/E/S subsequently provided a correction of the data related to recommendations, the earnings forecasts database still contains substantial reporting errors.

This study uses a simple approach to detect dissociation and namesake errors. In a first step, an algorithm is built to detect the two types of errors. The resulting detection process identifies a subset of codes that are potentially flawed. The quality of the algorithm is then verified by manually investigating the activity linked to a subsample of codes flagged by the algorithm in the first step.

The algorithm used to detect dissociation errors flags analyst codes if the three following conditions are met: (1) the analyst (code) has one or several namesakes in the I/B/E/S database (*i.e.*, one or more code(s) for analysts with the same last name and the same first name initial), (2) the code and the namesakes share a common employer, and (3) the code and the namesakes covered a common sector when employed by the same broker. When these three conditions are fulfilled, there is a very high probability that the code and the namesakes identify the same analyst. The algorithm flags 2169 codes for dissociation errors. Before further analysis, the validity of the detection process is checked by investigating a random sample of 100 codes taken from the set of flagged codes. For this subset of flagged codes, each analysts' employment history is checked using information collected from several websites such as *LinkedIn.com, Brokercheck.finra.org, Bloomberg.com* and *Zoominfo.com*, thus confirming or invalidating links between the flagged code and namesakes for each individual. This manual verification confirms dissociation errors for 98% of the codes. No information was found for the remaining 2% of codes.

When several namesakes are identified by the same code, the analyst (code) appears to be working for several brokers at the same time. Therefore, one possible means to detect namesake errors is to track inconsistent patterns in broker affiliations. For instance, namesake errors can be assessed by flagging analyst codes for which more than one broker affiliation is provided for a given day. Similarly, a code that presents multiple changes in broker affiliation during a short period of time is likely to reveal a namesake error. To ensure that the algorithm detects namesake errors and not simply isolated broker affiliation reporting errors, constraints are added to the algorithm for the sectors covered and the frequency of broker changes. Despite this highly conservative approach, 200 codes are still found to exhibit namesake errors.

Overall, 2288 codes are corrupted by matching errors (dissociation and namesake errors). On average, the yearly proportion of flagged codes is 16.12%. The yearly proportion of forecasts associated with flagged codes is 18.14%. These reporting errors are more frequent at the beginning of the sample period, with proportions of flagged codes reaching values as high as 29.45% in 1986. These findings reinforce previous concerns regarding the poor quality of the reporting in the I/B/E/S database previous to 1990: Hong et al. (2000) and Diether et al. (2002) warn against sparse analyst coverage during this period. The reporting errors that are pinpointed in this article are not limited to this time period, however, and impact analyst codes throughout the entire 1982–2014 period.

Reporting errors such as dissociation and namesake errors have little impact on the results of empirical studies when working at the firm level (*i.e.*, when using consensus forecasts). However, the implications for empirical research at the analyst level can be substantial in studies that rely on tracking analysts. A great number of studies aim to identify factors that determine the accuracy of earnings forecasts. For instance, several studies (Mikhail et al., 1997; Clement, 1999; Jacob et al., 1999) investigate how an analyst's abilities (proxied by the experience) and resources (proxied by the employer size) can influence forecast accuracy. Other studies look at the star status of analysts (Clarke et al., 2007; Emery and Li, 2009). A second stream of research investigates career concerns. Mikhail et al. (1999) and Hong and Kubik (2003) study the link between forecast accuracy and job turnover. Hilary and Hsu (2013) examine how forecast consistency influences the probability of being demoted or gaining star status.

Dissociation and namesake errors lead to erroneous estimates of analyst characteristics such as their experience, the number of firms (and industries) covered, forecast boldness or revision frequency. These reporting errors are also an obstacle to tracking broker changes and prevent the correct identification of star analysts. The key issue for future research is to determine whether these reporting errors simply add noise or whether they have systematic and persistent components that influence the results of empirical studies.

2. Dissociation and namesake errors

This section describes one example of dissociation error and one example of namesake error.

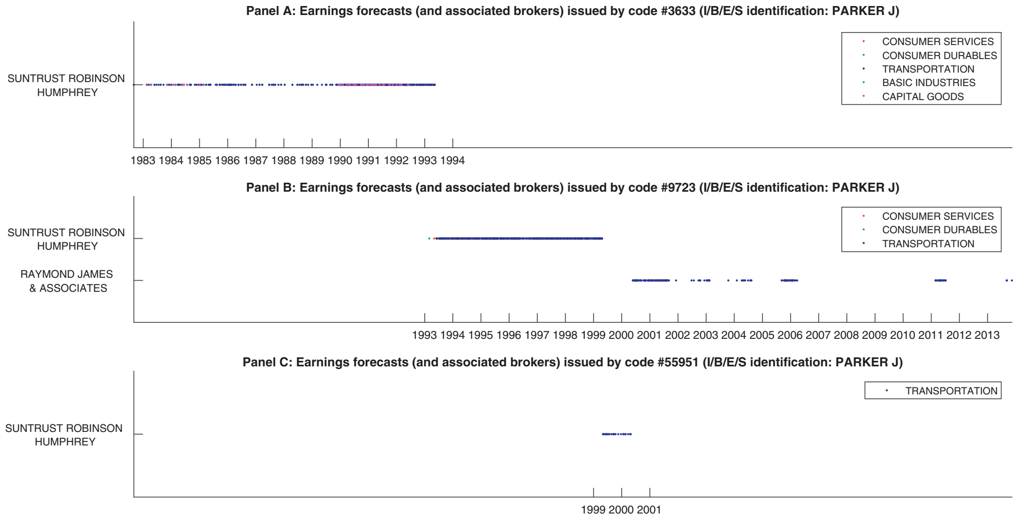

Fig. 1. Dissociation errors: earnings forecasts as reported by I/B/E/S for codes #3633, #9723 and #55951 This figure shows the earnings forecasts associated with the analyst codes #3633, #9723 and #55951. Each observation on the graphs represents a forecast. The *x*-axis represents the forecast issue date and the *y*-axis indicates the broker associated with the forecast. The color associated with each observation represents the sector of the firm for which the forecast was issued. The name provided by I/B/E/S for these three codes is *PARKER J*.

2.1. Dissociation errors

Fig. 1 shows an example of dissociation errors. Earnings forecasts are attributed to three different analyst codes (#3633, #9723 and #55951). Each dot on the graph represents a forecast. The *x*-axis represents the forecast issue date and the *y*-axis indicates the broker associated with the forecast.[3] The color of each dot identifies the sector of the firm for which the forecast was issued. The name provided by I/B/E/S for these three codes is *PARKER J*. Code #3633 and code #9723 share one common broker (*Suntrust Robinson Humphrey*), and code #55951 is employed by *Suntrust Equitable Securities*. Additionally, these three codes appear to specialize in the transportation sector. Research online reveals that these three codes correspond to the same analyst, James D. Parker. His biography describes him as follows: *James D. Parker is an Analyst at Raymond James & Associates, Inc., Research Division. James Parker served as Managing Director of Equity Research at SunTrust Robinson Humphrey, Inc. He was an Equity Analyst of SunTrust Robinson Humphrey, Inc., Research Division and SunTrust Equitable Securities, Research Division*.

2.2. Namesake errors

Fig. 2 shows an example of namesake errors and concerns the forecasts issued by codes #31688 and #72523. I/B/E/S identifies these two codes as *ADER J*. The top and bottom panels provide the forecasts for codes #31688 and #72523, respectively. Hereafter, code #31688 is referred to as J. Ader #1 and code #72523 as J. Ader #2.

In the top panel, the graph suggests that J. Ader #1 was employed simultaneously by two different brokers (*H.C. Wainwright & Co* and *Bear Stearns & Co*) from 1998 to 2000. At the end of 2002 and at the beginning of 2003, J. Ader #1 was issuing forecasts both for *Bear Stearns & Co* and *Thomas Weisel Partners*. At the same time (in 2002 and 2003), J. Ader #2 (bottom panel) was also employed by *Thomas Weisel Partners*.

The sectors covered by the analysts were examined to understand this anomaly. When employed by *Citigroup* (1993 to 1996) and *Bear Stearns & Co* (1999 to 2003), J. Ader #1 covered the finance, consumer services and consumer durables sectors. When employed by *H.C. Wainwright & Co* (1998 to 2000), *Thomas Weisel Partners* (2002 to 2008) and *William Blair & Co* (2008 to 2012), J. Ader #1 covered solely firms from the Technology sector. In the bottom panel, J. Ader #2 only covered firms from the technology sector.

Overall, these different elements point to a confusion between the data for the earnings forecasts of J. Ader #1 and J. Ader #2. The *LinkedIn* website was used to identify analysts whose first name starts with a J and whose last name is Ader. This careful investigation led to details of a first analyst named Jason Ader, who worked for *Smith Barney* (now *Citigroup*) from 1993 to 1995 and for *Bear Stearns & Co* from 1995 to 2003.[4] The *LinkedIn* profile indicates that this analyst specialized in gaming, leisure and resort industries. A second analyst was also identified, named Jason (Noah) Ader, who worked for

[3] For each forecast, I/B/E/S provides us with details of the analyst's employer (broker) at the time the forecast is issued.

[4] The All-star ranking indicates that J. Ader #1 (code #31688) was an all-star analyst in 1994 when working for *Smith Barney* and from 1995 to 2001 when working for *Bear Stearns & Co*. The different industries in which the analyst was ranked are Gaming and Lodging.

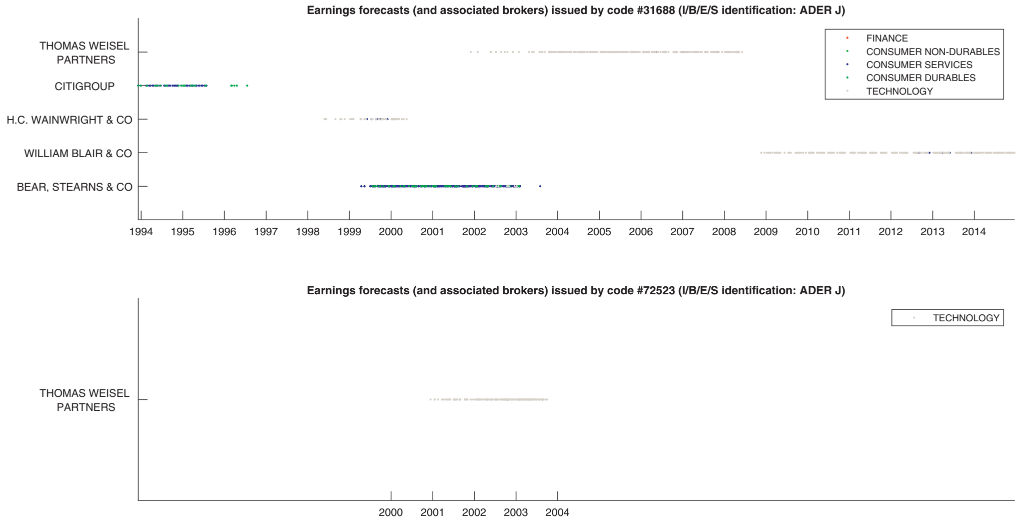

Fig. 2. Namesake errors: earnings forecasts as reported by I/B/E/S for codes #31688 and #72523 This figure shows the earnings forecasts associated with the analyst codes #31688 and #72523. Each observation on the graphs represents a forecast. The *x*-axis represents the forecast issue date and the *y*-axis indicates the broker associated with the forecast. The color associated with each observation represents the sector of the firm for which the forecast was issued. The name provided by I/B/E/S for these two codes is *ADER J*.

H.C. Wainwright & Co from 1998 to February 2000, for *Thomas Weisel Partners* from 2000 to 2008, and has worked for *William Blair & Co* from 2008 to the present day. This second analyst specializes in telecommunications, data networking and computers. The information was cross-checked in paperback volumes of *Nelson's Directory of Investment Research*, and confirms that most of the forecasts issued by analyst J. Ader #2 were erroneously assigned to Jason Ader #1. The two analyst codes, #31688 and #72523, thus exhibit namesake errors.

3. Estimation of the number of analysts with erroneous data

3.1. Data

This analysis uses the earnings forecasts provided by I/B/E/S over the period from 1982 to 2014.[5] Analyst codes for which no name was available were discarded. Teams of analysts were also excluded from the sample by discarding any codes for which the associated name corresponds to a country (*e.g.*, United Kingdom), or an industry (*e.g.*, merchandising), as well as any codes with names containing a slash (/), an ampersand (&), the word "group" or the word "department".[6] The final sample contained over 5,729,587 million earnings forecasts issued by 21,524 analysts.

The reporting errors studied in this article are attributable to the presence of namesakes in the sample of financial analysts. The analysis is therefore based on the comparison of the names associated with the different analyst codes; more specifically, string comparison was used to find namesakes. For this analysis to work correctly, structure had to be identical for all the names. The norm in I/B/E/S is to identify analysts with their last name and their first name initial. For instance, an analyst named John Doe will be identified in I/B/E/S as "*DOE J*". While this structure is the norm, I also find occurrences where the first name initial precedes the last name, or occurrences where I/B/E/S also provides the middle name initial. Other variations include the first name initial being followed by a dot, the last name being followed by a suffix (such as Jr or Sr), or the last name being followed by a title (such as Ph.D. or CFA). To ensure that matching is not impacted by the structure of the name, a common structure was applied for all the names, namely the last name followed by the first name initial. All references to middle names, suffixes and titles were therefore removed. In addition to this common structure rule, it was also important to ensure that the spelling was consistent across names. For instance, a given individual may be referred to as "MCDONALD J", "MACDONALD J", "MC DONALD J" or "MAC DONALD J". The different prefixes (such as Mac/Mc, St/Saint, and so on) were standardized and any spaces between the prefix and the last name were removed. Finally, multiple last names were standardized by placing a hyphen between the different last names. These different corrections were carried out using a semi-automatized approach (the use of an algorithm combined with manual verification).

The descriptive statistics for the frequency of namesakes (same name and same first name initial) in the I/B/E/S/ database are provided in Table 1. There are 6044 analyst codes with at least one namesake in the database. This statistic indicates

[5] This paper uses a download from February 2015.

[6] The different variations of these words are also considered.

Table 1
Frequency of namesakes in the I/B/E/S earnings forecast database.

Number of namesakes	Number of analyst codes	Proportion (in %)
0	15,480	71.92
1	3594	16.7
2	1245	5.78
3	548	2.55
4	300	1.39
5	114	0.53
6	98	0.46
7	32	0.15
8	9	0.04
9	10	0.05
10	44	0.2
11	12	0.06
13	14	0.07
23	24	0.11

This table presents the frequency of namesakes in the I/B/E/S earnings forecast database. A code has a namesake in the database if there is another code with the same last name and the same first name initial. The sample period is 1982–2014.

that the number of dissociation errors is at most equal to 6,044, representing 28% of the total number of codes in the database.

4. Detection of reporting errors

4.1. Dissociation errors

Three conditions are applied to determine whether several codes correspond to the same analyst, namely: (1) the same name is associated with the different codes (both the last name and the first name initial), (2) the different codes exhibit a common broker affiliation at some point in time, and (3) the different analyst codes covered one common sector when employed by the same broker. The likelihood of two namesakes (or more) working for the same broker and covering the same sector is small. These three conditions are therefore considered sufficient to identify dissociation errors. String comparisons are used to check if the first condition is met, retaining any exact matches between the names associated with the different codes. This first condition yields a list of 6044 analysts. The second condition reduces this number to 2317. Finally, when considering the three conditions together, a total of 2169 codes are flagged for dissociation errors.

The validity of the algorithm is checked by the manual collection of information for a sample of 100 analyst codes among the final set of 2169 codes. This information is mainly acquired from newspaper articles and websites such as *LinkedIn.com, Brokercheck.finra.org, Bloomberg.com* and *Zoominfo.com*. Three types of information are used to identify analysts: (1) their names, (2) the different brokers they worked for, and (3) the main sectors they covered. In most cases, these three pieces of information suffice to find qualitative information about their employment history and thus verify their identity. This manual verification confirms the existence of dissociation errors for 98 of the 100 randomly selected codes. No information was found for the two remaining codes.

4.2. Namesake errors

There are three main types of broker affiliation error in the I/B/E/S database. The first is the isolated reporting error, indicated by the occurrence of an isolated broker affiliation. The second type of erroneous broker affiliation is typically found at times when analysts change brokers. During this period when analyst i moves from broker A to broker B, I/B/E/S wrongly associates some of the forecasts to the former broker (broker A) rather than to the new broker (broker B). Multiple occurrences of these “noisy broker changes” were identified in the database. Finally, the last type of broker affiliation error is related to namesake errors, by which two or more analysts working for different brokers are identified using the same code. The two first types of broker affiliation errors are less problematic because, in these cases, only the broker affiliation is erroneous. In the case of namesake errors, the broker affiliation is correct but the analyst code is corrupted. The following algorithm was used to detect broker affiliation errors. $B_{i,t}$ represents the broker affiliation associated with analyst code i for a forecast issued at time t. A broker affiliation error is considered to have occurred if the following condition is met:

$$B_{i,t_1} \neq B_{i,t_2} \cap B_{i,t_1} = B_{i,t_3}$$
with $t_1 \leq t_2 \leq t_3$ and $t_3 - t_1 < 3$ months.

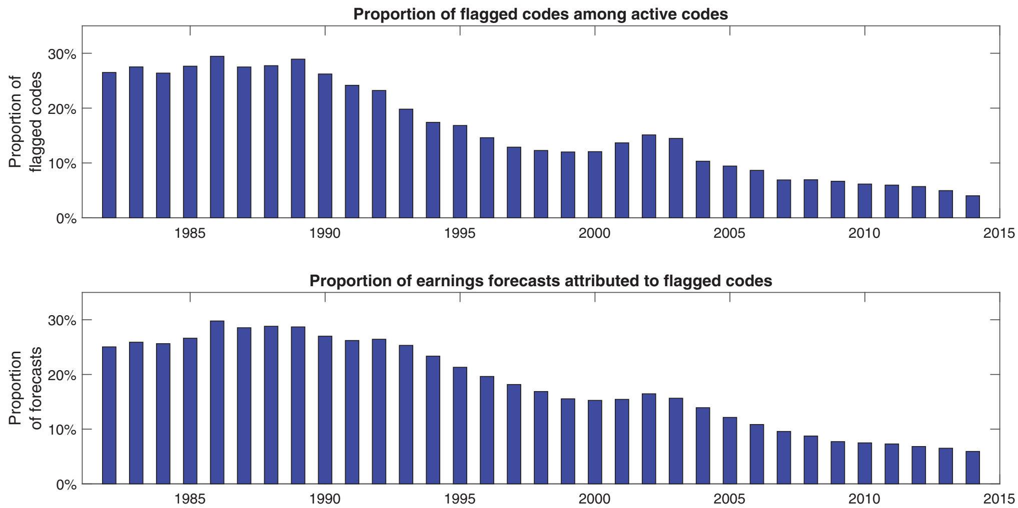

Fig. 3. Statistics on flagged codes This figure presents statistics on codes flagged for either dissociation or namesake errors. The top graph shows, for each year of the sample period, the ratio of the number of active flagged codes to the total number of active codes. A code is considered to be active in a given year if there is at least one associated earnings forecast. The bottom graph gives the yearly proportion of earnings forecast issued by flagged codes. The sample period is 1982–2014.

4.3. Statistics on flagged analysts

The above condition is sufficient to characterize broker affiliation errors. Indeed, it is extremely unlikely that an analyst will change employer and return to her initial employer within a three-month period. However, the aim here is to identify namesake errors. To eliminate isolated broker affiliation errors and noisy broker changes, an additional condition was imposed for the sectors of the covered firms. If an analyst (code) appears to work simultaneously for broker *A* and broker *B*, the sector(s) of the firms covered under broker affiliation *A* must be different from the sector(s) of the firms covered under broker affiliation *B*. This additional condition is conservative but it ensures that namesake errors are retained rather than isolated broker affiliation errors or noisy broker changes. Overall, the algorithm identified 338 flagged codes.

Online resources were used to check the validity of the algorithm. Namesake errors were confirmed for 143 codes. For 57 of the remaining cases, the repetitive nature of the broker affiliation errors (*e.g.*, an analyst code presents simultaneous broker affiliations during several months or years) strongly suggests the occurrence of namesake errors. Finally, for the 138 remaining cases, it is impossible to confirm whether there are several analysts associated to the codes (primarily due to an insufficient number of forecasts to identify the analysts).

Fig. 3 shows the proportion of codes that are flagged each year for either dissociation errors or namesake errors.[7] The top graph represents the yearly percentage of active codes flagged for having erroneous forecasts at some point in time. The bottom graph reports the yearly proportion of forecasts issued by flagged analysts. Overall, Fig. 3 indicates that a large proportion of codes exhibit either dissociation or namesake errors. The proportion of yearly active flagged codes is 16.12% on average, with a minimum of 4% in 2014 and a maximum of 29.45% in 1986. On average, 18.14% of the yearly forecasts are associated with flagged codes. This proportion is as high as 29.78% in 1986. The global trend on these two graphs indicates that a large number of reporting errors are concentrated in the first half of the sample period. However, when evaluating analyst characteristics such as experience, reporting errors made in the first part of the sample period continue to have an impact at a later date. While these statistics cannot be used to analyze how reporting errors may impact the results of empirical analyses that rely on tracking analysts, they do show that reporting errors affect a large amount of the data in the I/B/E/S earnings forecast database.

5. Conclusion

This paper provides evidence of substantial reporting errors in the I/B/E/S earnings forecast database. These errors result from I/B/E/S incorrectly identifying analysts. As a result, some analysts are identified by several different codes throughout their forecasting history. Additionally, I/B/E/S erroneously attributes forecasts to some analysts that were in fact issued by namesakes. Overall, this analysis indicates that at least 2288 analyst codes are corrupted by such errors. These errors lead to erroneous estimates of analysts' characteristics and may have an influence on the results found in empirical studies that rely on tracking analysts. Further analysis is required to evaluate the implications of such reporting errors for academic research.

[7] The total number of analyst codes flagged for either dissociation or namesake errors is equal to 2288.

References

Bosquet, K., de Goeij, P., Smedts, K., 2014. Gender heterogeneity in the sell-side analyst recommendation issuing process. Finance Res. Lett. 11, 104–111.
Clarke, J., Khorana, A., Patel, A., Rau, R., 2007. The impact of all-star analyst job changes on their coverage choices and investment banking deal flow. J. Financ. Econ. 84 (3), 713–737.
Clement, M.B., 1999. Analyst forecast accuracy: do ability, resources, and portfolio complexity matter? J. Account. Econ. 27 (3), 285–303.
Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. J. Finance 57 (5), 2113–2141.
Emery, D.R., Li, X., 2009. Are the wall street analyst rankings popularity contests? J. Financ. Quant. Anal. 44 (2), 411–437.
Galanti, S., 2016. Archival data of financial analysts' earnings forecasts in the euro zone: problems with euro conversions. Res. Int. Bus. Finance 38, 466–473.
Hilary, G., Hsu, C., 2013. Analyst forecast consistency. J. Finance 68 (1), 271–297.
Hong, H., Kubik, J.D., 2003. Analyzing the analysts: career concerns and biased earnings forecasts. J. Finance 58 (1), 313–351.
Hong, H., Lim, T., Stein, J.C., 2000. Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. J. Finance 55, 265–295.
Jacob, J., Lys, T.Z., Neale, M.A., 1999. Expertise in forecasting performance of security analysts. J. Account. Econ. 28 (1), 51–82.
Lin, W.-C., Chang, S.-C., Chen, S.-S., Liao, T.-L., 2013. The over-optimism of financial analysts and the long-run performance of firms following private placements of equity. Finance Res. Lett. 10, 82–92.
Ljungqvist, A., Malloy, C., Marston, F., 2009. Rewriting history. J. Finance 64 (4), 1935–1960.
Medovikov, I., 2014. Can analysts predict rallies better than crashes? Finance Res. Lett. 11, 319–325.
Mikhail, M.B., Walther, B.R., Willis, R.H., 1997. Do security analysts improve their performance with experience? J. Account. Res. 35, 131–157.
Mikhail, M.B., Walther, B.R., Willis, R.H., 1999. Does forecast accuracy matter to security analysts? Account. Rev. 74 (2), 185–200.