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

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.