

# Good-Bye I/B/E/S (or Not?)

Kelvin K. F. Law

*Nanyang Technological University*

**ABSTRACT:** In October 2018, I/B/E/S claimed to have anonymized the identifications (IDs) of 88 brokers and their analysts in its Detail History Files. This paper examines to what extent the change affects the post-2018 Detail History File. The main findings are as follows. First, there is no evidence of a large-scale ID anonymization for the annual and quarterly EPS forecasts for U.S. firms. Second, the anonymization predominantly affects the forecasts of non-U.S. firms. A significant percentage of forecasts for non-U.S. firms have been anonymized since January 2017 (at least one year earlier than what was claimed). Third, there is no evidence of a large-scale ID reshuffling of forecasts between the 2015 and 2021 Detail History Files. Last, because the anonymization has very little effect on the Recommendations Detail File, I am able to offer detailed steps to reverse engineer the anonymized analyst IDs in the Unadjusted Detail History File.

**JEL Classifications:** G00; M41.

**Keywords:** analysts; I/B/E/S; anonymization; reshuffling; forecasts; recommendations.

## I. INTRODUCTION

In 2018, I/B/E/S announced a change that may have an unprecedented impact on analyst research. Effective October 18, 2018, I/B/E/S was supposed to have anonymized the identifications (IDs) of 88 brokers and their analysts in its Detail History File (see [Appendix A](#) for the ID anonymization notification). After analyzing the data, [Wharton Research Data Services \(WRDS\) \(2023a\)](#) claimed that “It is likely that 13.8% of all broker IDs (ESTIMATOR) and 30.7% of all analyst IDs (ANALYS) have been reassigned.” See [Appendix B](#) for the ID reshuffling notification.<sup>1</sup> Together, these claims suggest that both ID anonymization and ID reshuffling affect post-2018 Detail History Files. As most analyst research depends on the stability and continuation of broker and analyst IDs over time, these two claims seem like bad news for analyst researchers; indeed, they could spell the end of analyst research.

Although these two claims are potentially catastrophic to analyst researchers, neither WRDS nor research studies offer any guidance to clarify the extent of the impact. WRDS does not distribute the analyses underlying its numbers, and I/B/E/S does not disclose the extent of the impact. Hence, the impact of the two significant changes on the I/B/E/S Detail History Files remains unclear. How many forecasts, for example, have been anonymized in the Unadjusted Detail History Files? Which type of forecast was most affected (e.g., earnings per share or cash flow per share)? Were the IDs anonymized retrospectively? Did I/B/E/S also anonymize the broker IDs? Did I/B/E/S reshuffle 30.7 percent of

---

I sincerely thank Clare Wang (editor), two anonymous referees, Robert Bloomfield, Andrew Call, Byung Hun Chung, Alper Darendeli, Jay Jung, Stephannie Larocque, Matthew Pierson, Mark Piorkowski, Jingdan Tan, Jessica Watkins, Yachang Zeng, and seminar participants at the Nanyang Business School Brown Bag and Financial Accounting and Reporting Section (FARS) and the 2022 *Journal of Financial Reporting*: 2nd Annual Research Mini-Conference for their helpful comments. I would like to specifically thank Byung Hun Chung and Yachang Zeng for their generous sharing of analyst data for robustness checks. I acknowledge financial support from Nanyang Technological University’s Start-Up Grants.

Kelvin K. F. Law, Nanyang Technological University, College of Business, Division of Accounting, Singapore.

Supplemental material is available, as linked in the text.

Editor’s note: Accepted by Clare Wang.

*Submitted: December 2021*  
*Accepted: October 2022*  
*Early Access: May 2023*

---

<sup>1</sup> In its publication “I/B/E/S @ WRDS 101: Introduction and Research Guide” ([Wharton Research Data Services \(WRDS\) 2020](#)), WRDS also claims that, through a conference call, I/B/E/S further informed WRDS that individual broker IDs (and all affected analysts) have been and will continue to be subject to reshuffling without warning ([Appendix C](#)).

analyst IDs after October 2018? Did I/B/E/S reshuffle 13.8 percent of broker IDs too? Which type of forecast was most affected by the ID reshuffling? These claims suggest that the changes affect only the Unadjusted Detail History Files, but was the Recommendations Detail File also affected by the ID anonymization and ID reshuffling?

For obvious reasons, these questions are not trivial. First, ID anonymization and ID reshuffling could affect all analyst research papers after October 2018. Assuming that any working paper containing the word “analyst” in the title and circulated on Social Science Research Network (SSRN) in the past three years uses the post-2018 I/B/E/S Detail History File, these lingering questions potentially affect the research design and interpretation of at least 150 working and forthcoming papers.<sup>2</sup> Hence, the scale of the potential impact is massive. Second, the uncertainty of the ID anonymization and ID reshuffling burdens reviewers and may cloud their judgments. Perhaps a better way to gauge the severity of this problem is to refer to a reviewer’s comments I recently received:

First, and foremost, it is no longer possible to cleanly identify analysts in I/B/E/S based on their analyst codes. According to Wharton Research Data Services (WRDS), I/B/E/S scrambled analyst codes starting in ~October 2018—well before the study’s sample period. I include with my review a PowerPoint presentation made available by WRDS; slide 23 suggests that as many as 30 percent of analyst codes may have been scrambled by I/B/E/S. *This does not appear to be an issue that the authors can avoid, should they choose to work with I/B/E/S data.* (emphasis added)

Given the above, I examine whether and to what extent ID anonymization and ID reshuffling affect the Unadjusted Detail History File. For ID anonymization, I examine how many analyst IDs have been anonymized after October 2018. Anonymization means that analyst ID (*ANALYS*) is 0. For ID reshuffling, I compare two vintages of Unadjusted Detail History Files, one downloaded in 2015 and one downloaded in 2021, to identify how many analyst IDs have been reshuffled since October 2018.<sup>3</sup> For ease of reference, I have tabulated the screenshots of a few examples based on the actual I/B/E/S data illustrating how ID anonymization ([Appendix D](#)) and ID reshuffling ([Appendix E](#)) show up in the Unadjusted Detail History File. Finally, I also examine whether and to what extent ID anonymization and ID reshuffling affect the Recommendations Detail File.

The main findings for ID anonymization are as follows. First, there is no evidence of a large-scale ID anonymization for the annual and quarterly earnings per share (*EPS*) forecasts for U.S. firms after October 2018. The ID anonymization has affected less than 0.1 percent of the annual and quarterly earnings forecasts of U.S. firms since 1990.<sup>4</sup>

Second, the ID anonymization predominantly affects the forecasts of non-U.S. firms. A significant percentage of forecasts have been anonymized in the Unadjusted Detail History File since January 2017, at least one year earlier than what was claimed. The percentage of forecasts with anonymized analyst IDs extends to the forecasts of brokers beyond the 88 brokers in the I/B/E/S Product Change Notification ([Thomson Reuters 2018](#)). The percentage of annual earnings forecasts with anonymized analyst IDs reached its highest point of 13.15 percent in August 2021. Overall, the ID anonymization predominantly affects the annual and quarterly earnings forecasts of non-U.S. firms rather than U.S. firms.<sup>5</sup>

Because the ID anonymization does not affect the Recommendations Detail File, I am able to offer detailed steps to reverse engineer some of the anonymized IDs in the Unadjusted Detail History File based on the nonanonymized analyst IDs in the Recommendations Detail File (see the “[Recovering Anonymized Analyst IDs](#)” section for details). Specifically, I reverse engineer the anonymized analyst IDs in the Unadjusted Detail History File (*ANALYS*) based on the nonanonymized analyst IDs in the Recommendations Detail File (*AMASKCD*). On average, following these steps recovers about 55.7 percent of forecasts with anonymized analyst IDs in the Unadjusted Detail History File.

The main findings for ID reshuffling are as follows. First, there is no evidence of a large-scale ID reshuffling of forecasts between the 2015 and 2021 Detail History Files. On average, only 2.87 percent of forecasts have their IDs reshuffled between the 2015 and the 2021 Unadjusted Detail History Files. The percentage is even lower for annual earnings forecasts of U.S. firms at 1.20 percent and for quarterly earnings forecasts of U.S. firms at 0.57 percent.

Second, about 99 percent of forecasts with reshuffled analyst IDs are the forecasts whose analyst IDs were initially assigned to be anonymized. The results suggest that almost all forecasts with reshuffled analyst IDs were forecasts

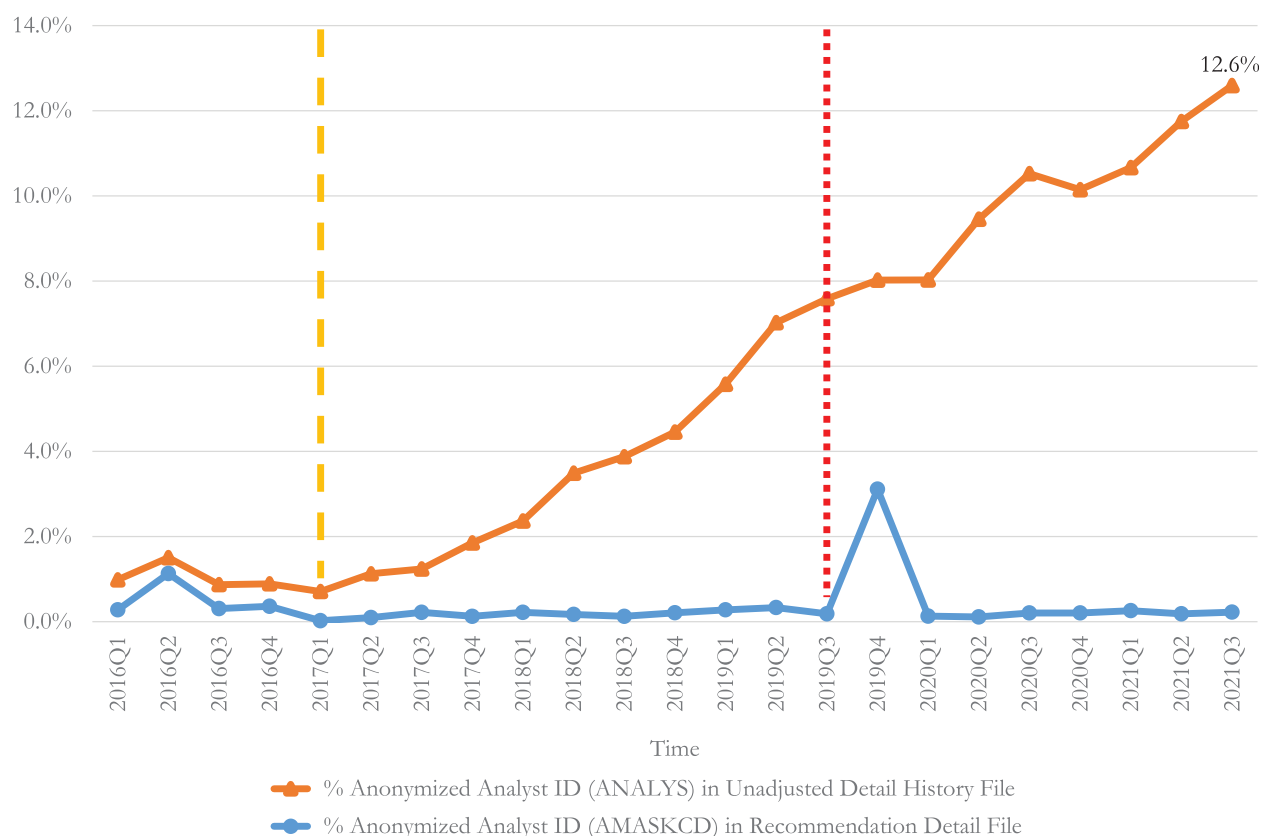
<sup>2</sup> I searched papers in SSRN on August 26, 2021, using the keyword “analyst” and limiting the date of the working papers to the past three years.

<sup>3</sup> As [Payne and Thomas \(2003\)](#) point out, the I/B/E/S Adjusted Detail History File accounts for stock splits. In other words, for a stock with a stock split, the same forecast would mechanically have different values in different vintages. Using Adjusted Detail History Files to match forecasts in different vintages, therefore, will overstate the number of unmatched forecasts.

<sup>4</sup> The number of annual and quarterly earnings forecasts of U.S. firms with anonymized analyst IDs (8,694 [Table 2](#), column (1) + 2,182 [Table 2](#), column (3)) ÷ the number of annual and quarterly earnings forecasts of U.S. firms (4,645,545 [Table 2](#), column (1) + 3,654,363 [Table 2](#), column (3)) = 0.1 percent.

<sup>5</sup> Number of annual and quarterly earnings forecasts of non-U.S. firms with anonymized analyst IDs (331,979 [Table 2](#), column (2) + 25,412 [Table 2](#), column (4)) ÷ number of annual and quarterly earnings forecasts of non-U.S. firms (7,035,115 [Table 2](#), column (2) + 857,081 [Table 2](#), column (4)) = 4.5 percent.

**FIGURE 1**  
**Anonymized Analyst IDs**



The following graph plots the percentage of forecasts with anonymized analyst IDs in the Unadjusted Detail History File (the orange line with triangle markers) and the Recommendations Detail File (the blue line with circle markers) in a given quarter. The Unadjusted Detail History File and Recommendations Detail Files were both downloaded in November 2021. The x-axis axis is based on the time stamps of announcement dates. January 2017 falls into 2017 quarter 1 and is marked with a yellow dashed line. October 2018 falls into 2018 quarter 3 and is marked with a red square dotted line. (The full-color version is available online.)

initially anonymized (i.e., the analyst ID was 0) but subsequently backfilled (i.e., the analyst ID was not 0) in later vintages. These findings are important for analyst researchers because, even though ID reshuffling seems problematic, it affects the forecasts that are often already excluded from academic studies. Third, there is no evidence of a large-scale ID reshuffling of recommendations between the 2017 and the 2021 Recommendations Detail Files.

This paper makes the following contributions. First, although there is no shortage of prior studies on I/B/E/S data issues, this paper is the first to examine the impact of ID anonymization and ID reshuffling on analyst research after October 2018.<sup>6</sup> My analyses clarify the extent of the impact, which has important implications for at least 150 working or forthcoming papers that use I/B/E/S data downloaded after October 2018. In [Section IV](#), I offer several specific recommendations for how such studies can best use the insights from this paper. The second contribution is that, because the anonymization has very little effect on the Recommendations Detail File, I am able to offer detailed steps to reverse engineer the anonymized analyst IDs in the Unadjusted Detail History File based on the nonanonymized analyst IDs in the Recommendations Detail File.

<sup>6</sup> For example, [Ljungqvist, Malloy, and Marston \(2009\)](#) find widespread changes to the historical stock recommendations between 2000 and 2007. A recent study by [Call, Hewitt, Watkins, and Yohn \(2021\)](#) compares two vintages of I/B/E/S Detail History Files (2009 versus 2015). They find substantial differences in the attributes of the earnings forecasts in each version. The sample periods of both studies stop well before October 2018, when the I/B/E/S product change took place.

## II. ID ANONYMIZATION

### Overview of Anonymized Analyst IDs

In this section, I examine whether and to what extent ID anonymization affects the Unadjusted Detail History File. I use an I/B/E/S Unadjusted Detail History File downloaded from WRDS in November 2021 to answer this. ID anonymization means that the analyst ID variable (*ANALYS*) of a forecast is 0. What does a “0” code for an analyst ID mean? Wharton Research Data Services (WRDS) (2023b) states that “If the Analyst field is populated by Zeros it means that at that point in time there was no analyst name available.”<sup>7</sup> As mentioned earlier, Appendix D tabulates examples of forecasts whose analyst IDs have been anonymized.

The following provides an overview of the Unadjusted Detail History File. The file contains 119,952,807 forecasts issued between January 1, 1970 and August 31, 2021, and they represent ten types of forecasts: book value per share (*BPS*), cash flow per share (*CPS*), cash earnings per share (*CSH*), dividend per share (*DPS*), earnings per share before goodwill (*EBG*), earnings before interest, taxes, depreciation, and amortization (EBITDA) per share (*EBS*), *EPS*, earnings per share: alternate (*EPX*), funds from operations per share (*FFO*), and GAAP/earnings per share: fully reported (*GPS*). *EPS* forecasts are the most common type; 44.7 percent of the forecasts are *EPS* forecasts. There are 32 different forecast horizons (*FPI*). Annual forecasts (*FPI* = 1) are the most common type; 24.3 percent of the forecasts are annual forecasts, and 6.6 percent of the forecasts are quarterly forecasts (*FPI* = 6). The announcement dates of the forecasts (*ANNDATS*) are between January 1, 1970 and August 31, 2021; 44.3 percent of the forecasts are for U.S. firms (*USFIRM* = 1).

Table 1, column (1) tabulates the number of forecasts with anonymized analyst IDs.<sup>8</sup> To conserve space, I only retain earnings forecasts issued on or after January 1990.<sup>9</sup> The year of a forecast is based on its announcement date (*ANNDATS*). Column (1) shows that some forecasts were anonymized as early as the 1990s. The number of forecasts with anonymized analyst IDs, however, is stable over the sample period, but it significantly increases from 83,128 forecasts in 2017 to 236,929 forecasts in 2018. In terms of year-on-year percentage change, the number of forecasts with anonymized analyst IDs substantially increases by 185 percent  $[= (236,929 - 83,128) \div 83,128]$  from 2017 to 2018. The number of forecasts with anonymized analyst IDs further increases to 450,755 in 2019; 662,297 in 2020; and 531,223 in 2021. There is no clear trend indicating that the IDs were anonymized retrospectively before October 2018.

Call et al. (2021) find that I/B/E/S retroactively removes about 12 percent of forecasts from the historical record. Column (2) examines whether the significant increase is due to a denominator effect in which the number of forecasts in the file has significantly changed since October 2018. Specifically, column (2) divides the number in column (1) by the number of forecasts in a given year. The pattern in column (2) suggests that the increase is not due to a denominator effect because the percentage of forecasts with anonymized analyst IDs also exhibits a significant increase from 1.21 percent in 2017 to 3.51 percent in 2018. The percentage of forecasts with anonymized analyst IDs increases to 7 percent in 2019 and reaches its highest point of 11.57 percent in August 2021.

Table 1, columns (3) and (4) analyze to what extent the ID anonymization affects annual earnings forecasts (*FPI* = 1 and *MEASURE* = *EPS*) and quarterly earnings forecasts (*FPI* = 6 and *MEASURE* = *EPS*), the two most commonly used earnings forecasts in accounting and finance research. Column (3) divides the number of annual earnings forecasts with anonymized analyst IDs by the number of annual earnings forecasts. Column (4) divides the number of quarterly earnings forecasts with anonymized analyst IDs by the number of quarterly earnings forecasts. Column (4) shows that the percentage of annual earnings forecasts with anonymized analyst IDs significantly increases from 1.92 percent in 2017 to 4.81 percent in 2018, reaching its highest point of 13.15 percent in August 2021. The percentage of earnings forecasts with anonymized analyst IDs is lower in quarterly earnings forecasts than in annual earnings forecasts (0.61 percent for quarterly forecasts versus 2.92 percent for annual forecasts). The percentage of quarterly earnings forecasts with anonymized analyst IDs reaches its highest point of 4.41 percent in August 2021 and is lower than the percentage of annual earnings forecasts.

Is the increase driven by the ID anonymization of the 88 brokers that were in the I/B/E/S Product Change Notification? I conduct two tests to examine this. First, I tabulate in column (5) the percentage of forecasts with anonymized analyst IDs issued by one of the 88 brokers. Column (5) shows that only 11 percent of anonymized forecasts were

<sup>7</sup> See <https://wrds-www.wharton.upenn.edu/pages/support/support-articles/ibes/what-does-0-code-analyst-represent> for details.

<sup>8</sup> Number of forecasts with anonymized analyst IDs (3,230,514 in Table 1, column (1)) = number of forecasts (119,952,807) – number of pre-1990 forecasts (1,883,005) – number of forecasts with nonmissing analyst IDs (114,839,288).

<sup>9</sup> The primary reason to exclude the pre-1990 forecasts is to conserve space. My findings are not sensitive to whether the pre-1990 forecasts are included or not because the number of forecasts with anonymized IDs is very small before 1990. For example, there are only 973 annual earnings forecasts (13 quarterly earnings forecasts) with anonymized IDs issued before 1990.

**TABLE 1**  
**Overview of Anonymized Analyst IDs**

Year	Number of Forecasts with Anonymized Analyst IDs (1)	Percentage of Forecasts with Anonymized Analyst IDs (%) (2)	Percentage of Annual Earnings Forecasts with Anonymized Analyst IDs (%) (3)	Percentage of Quarterly Earnings Forecasts with Anonymized Analyst IDs (%) (4)	Percentage of Anonymized Forecasts by 88 Brokers (%) (5)	Number of Brokers that Issued Forecasts with Anonymized Analyst IDs (6)
1990	470	0.11	0.18	0.06	7.87	27
1991	1,762	0.39	0.31	0.00	1.93	23
1992	3,808	0.71	0.75	0.00	1.86	41
1993	5,381	0.82	0.85	0.00	4.55	63
1994	42,943	4.60	1.05	0.00	6.39	301
1995	75,477	6.51	1.71	0.00	7.36	390
1996	81,177	6.40	1.43	0.00	6.88	401
1997	79,832	5.58	1.45	0.00	7.06	433
1998	85,250	5.14	2.09	0.00	7.60	435
1999	88,049	4.65	1.70	0.00	9.54	422
2000	82,791	4.62	2.52	0.08	7.86	415
2001	75,987	3.47	2.42	0.00	7.09	401
2002	53,899	2.24	3.14	0.00	6.75	424
2003	31,333	1.25	2.32	0.00	5.52	362
2004	33,881	1.16	2.04	0.00	3.97	285
2005	32,301	1.05	2.06	0.01	4.01	309
2006	33,995	1.02	2.62	0.02	2.25	331
2007	37,382	1.00	2.10	0.02	2.61	310
2008	30,343	0.67	2.40	0.02	2.42	244
2009	25,968	0.52	2.49	0.01	0.30	76
2010	45,553	0.82	2.29	0.06	4.02	430
2011	57,212	0.93	1.89	0.10	5.74	496
2012	35,098	0.56	1.64	0.09	2.43	413
2013	66,281	1.09	1.35	0.09	10.33	552
2014	39,947	0.61	1.33	0.10	6.06	338
2015	45,853	0.65	1.15	0.12	7.01	380
2016	74,209	1.06	1.64	0.22	11.38	486
2017	83,128	1.21	1.92	0.43	9.80	468
2018	236,929	3.51	4.81	1.42	10.20	579
2019	450,755	7.00	8.26	3.01	13.95	683
2020	662,297	9.48	10.84	3.78	15.92	635
August 2021	531,223	11.57	13.15	4.41	15.45	625
Total	3,230,514	2.74	2.92	0.61	11.36	11,778

Table 1 includes all forecasts issued between January 1, 1990 and August 31, 2021. Column (1) tabulates the number of forecasts with anonymized analyst IDs in the Unadjusted Detail History File. Column (2) divides the number in column (1) by the number of forecasts in a given year. Column (3) divides the number of annual earnings forecasts ( $FPI = 1$  and  $MEASURE = EPS$ ) with anonymized analyst IDs by the number of annual earnings forecasts. Column (4) divides the number of quarterly earnings forecasts ( $FPI = 6$  and  $MEASURE = EPS$ ) with anonymized analyst IDs by the number of quarterly earnings forecasts. Column (5) reports the percentage of forecasts with anonymized analyst IDs issued by one of the 88 brokers in [Online Appendix](#), Table 1. Column (6) reports the number of brokers that issued forecasts with anonymized analyst IDs. Observations before 1990 are not tabulated to conserve space. The shaded rows represent the observations after I/B/E/S allegedly anonymized or removed its observations.

issued by one of the 88 brokers. The percentage of forecasts with anonymized analyst IDs by the 88 brokers is lower than that of the non-88 brokers. In other words, most anonymized forecasts were issued by brokers other than those 88 brokers in the I/B/E/S Product Change Notification.



Second, I examine the change in the number of brokers that issued anonymized forecasts. Column (6) tabulates the number of brokers that issued anonymized forecasts. The number of brokers that issued anonymized forecasts has increased by more than 100, from 468 to 579. The number of brokers that issued anonymized forecasts reaches the highest of 683 in 2019 and hovers around 630 in 2020–2021. The pattern again suggests that the ID anonymization is unlikely to be confined to the forecasts of the 88 brokers. Overall, the results of these two tests show that the ID anonymization extends to the forecasts of brokers beyond the 88 brokers in the I/B/E/S Product Change Notification.

Does the ID anonymization affect broker IDs in the Unadjusted Detail History File (*ESTIMATOR*)? As before, ID anonymization means that a broker ID is 0. I use the same Unadjusted Detail History File to examine this. Untabulated tests show that none of the broker IDs for the 119 million forecasts in the Unadjusted Detail History File are anonymized.

Overall, the evidence shows the following. First, a significant percentage of forecasts have been anonymized since January 2017, at least one year earlier than what was claimed. Second, the ID anonymization is not confined to the forecasts of the 88 brokers; rather, it extends to the forecasts of the non-88 brokers. Third, the percentage of forecasts with anonymized analyst IDs is lower in quarterly earnings forecasts than in annual earnings forecasts. The percentage of annual earnings forecasts with anonymized analyst IDs reached its highest point of 13.15 percent in August 2021.<sup>10</sup>

### U.S. Firms versus Non-U.S. Firms

In this section, I examine whether the ID anonymization equally affects the forecasts of U.S. firms (*USFIRM* = 1) and non-U.S. firms (*USFIRM* = 0). Although I focused on all forecasts in the previous analyses, I now focus on annual and quarterly earnings forecasts, as they are the most frequently used forecast data in analyst research.

Table 2 tabulates the percentage of annual and quarterly earnings forecasts with anonymized analyst IDs by whether a firm is a U.S. firm or not. Columns (1) and (2) (columns (3) and (4)) report the percentage of annual (quarterly) earnings forecasts. Column (1) divides the number of annual earnings forecasts of U.S. firms (*USFIRM* = 1) with anonymized analyst IDs by the number of annual earnings forecasts of U.S. firms in a given year.<sup>11</sup> Column (2) divides the number of annual earnings forecasts of non-U.S. firms (*USFIRM* = 0) with anonymized analyst IDs by the number of annual earnings forecasts of non-U.S. firms in a given year. Column (3) divides the number of quarterly earnings forecasts of U.S. firms with anonymized analyst IDs by the number of quarterly earnings forecasts of U.S. firms in a given year.<sup>12</sup> Finally, column (4) divides the number of quarterly earnings forecasts of non-U.S. firms with anonymized analyst IDs by the number of quarterly earnings forecasts of non-U.S. firms in a given year.

The results in columns (1) and (2) show that the ID anonymization predominantly affects the annual earnings forecasts of non-U.S. firms rather than U.S. firms. In column (1), there is no evidence of a large-scale ID anonymization for the annual *EPS* forecasts for U.S. firms after October 2018. Specifically, only 0.19 percent of the annual earnings forecasts of U.S. firms have been anonymized since 1990.<sup>13</sup> This number is much lower than the 4.72 percent of the annual earnings forecasts of non-U.S. firms anonymized since 1990 in column (2). Indeed, the extent of anonymized annual earnings forecasts for non-U.S. firms increases from 7.81 percent in 2018 to 13.18 percent in 2019, 17.44 percent in 2020, and its highest point of 21.06 percent in 2021. The percentage of anonymized annual earnings forecasts for U.S. firms has been stable at around 1 percent since October 2018. Hence, the ID anonymization affects non-U.S. firms significantly more than U.S. firms.

The quarterly forecasts in columns (3) and (4) show the same pattern. In column (3), there is no evidence of a large-scale ID anonymization for the quarterly *EPS* forecasts for U.S. firms after October 2018. As above, the ID anonymization predominantly affects the quarterly earnings forecasts of non-U.S. firms rather than U.S. firms. For U.S. firms, only 0.06 percent of quarterly earnings forecasts have been anonymized during the sample period, one-third of the magnitude of the ID anonymization of annual earnings forecasts for U.S. firms in column (1). In Table 3 in the [Online Appendix](#), I estimate ordinary least squares (OLS) regressions and document the same findings.

Overall, there is no evidence of a large-scale ID anonymization for the annual and quarterly *EPS* forecasts for U.S. firms after October 2018. The evidence here suggests that the ID anonymization predominantly affects the annual and

<sup>10</sup> In Table 2 in the [Online Appendix](#), I use time series techniques to identify the unknown structural break in data. The test indicates that the break-date occurs at the beginning of 2017, at least one year earlier than October 2018. The results echo the pattern in [Figure 1](#), which shows that the percentage of forecasts starts climbing as early as 2017 quarter 1.

<sup>11</sup> The total number of annual *EPS* forecasts is 11,680,660 (4,645,545 U.S. firms + 7,035,115 non-U.S. firms); 11,680,660 forecasts = total number of forecasts 119,952,807 – non-*EPS* forecasts 66,376,094 – nonannual *EPS* forecasts 41,139,611 – pre-1990 annual *EPS* forecasts 756,442.

<sup>12</sup> The total number of quarterly *EPS* forecasts is 4,511,444 (3,654,363 U.S. firms + 857,081 non-U.S. firms); 4,511,444 forecasts = total number of forecasts 119,952,807 – non-*EPS* forecasts 66,376,094 – nonquarterly *EPS* forecasts 48,858,603 – pre-1990 quarterly *EPS* forecasts 206,666.

<sup>13</sup> As mentioned earlier, my findings are not sensitive to whether the pre-1990 forecasts are included or not.

**TABLE 2**  
**U.S. Firms versus Non-U.S. Firms: Tabulations**

Year	Percentage of Annual Earnings Forecasts with Anonymized Analyst IDs (%)		Percentage of Quarterly Earnings Forecasts with Anonymized Analyst IDs (%)	
	U.S. Firms (1)	Non-U.S. Firms (2)	U.S. Firms (3)	Non-U.S. Firms (4)
1990	0.12	0.24	0.06	0.09
1991	0.01	0.58	0.00	0.00
1992	0.00	1.28	0.00	0.00
1993	0.00	1.44	0.00	0.00
1994	0.00	1.69	0.00	0.00
1995	0.00	2.69	0.00	0.00
1996	0.00	2.24	0.00	0.00
1997	0.00	2.24	0.00	0.00
1998	0.00	3.34	0.00	0.00
1999	0.00	2.70	0.00	0.00
2000	0.00	4.09	0.08	0.11
2001	0.00	3.90	0.00	0.00
2002	0.00	5.16	0.00	0.00
2003	0.00	3.88	0.00	0.05
2004	0.00	3.53	0.00	0.05
2005	0.00	3.67	0.00	0.08
2006	0.00	4.69	0.00	0.14
2007	0.00	3.69	0.00	0.12
2008	0.00	4.06	0.00	0.13
2009	0.00	4.04	0.00	0.07
2010	0.00	3.72	0.00	0.25
2011	0.05	2.98	0.04	0.28
2012	0.04	2.60	0.03	0.25
2013	0.12	2.12	0.04	0.24
2014	0.32	2.01	0.06	0.22
2015	0.23	1.79	0.07	0.27
2016	0.32	2.53	0.11	0.51
2017	0.50	2.84	0.15	1.01
2018	0.29	7.81	0.08	4.40
2019	0.76	13.18	0.25	9.35
2020	1.08	17.44	0.27	12.68
August 2021	0.96	21.06	0.36	14.57
Total	0.19	4.72	0.06	2.96
Number of Forecasts with Anonymized IDs	8,694	331,979	2,182	25,412
Number of Forecasts (Total)	4,645,545	7,035,115	3,654,363	857,081

Table 2 includes all forecasts issued between January 1, 1990 and August 31, 2021. Column (1) divides the number of annual earnings forecasts of U.S. firms (*USFIRM* = 1 and *MEASURE* = *EPS*) with anonymized analyst IDs by the number of annual earnings forecasts of U.S. firms in a given year. Column (2) divides the number of annual earnings forecasts of non-U.S. firms (*USFIRM* = 0 and *MEASURE* = *EPS*) with anonymized analyst IDs by the number of annual earnings forecasts of non-U.S. firms in a given year. Column (3) divides the number of quarterly earnings forecasts of U.S. firms with anonymized analyst IDs by the number of quarterly earnings forecasts of U.S. firms in a given year. Column (4) divides the number of quarterly earnings forecasts of non-U.S. firms with anonymized analyst IDs by the number of quarterly earnings forecasts of non-U.S. firms in a given year. The shaded rows represent the observations after I/B/E/S allegedly anonymized or removed its observations.

**TABLE 3**  
**Recommendations Detail File**

Year	Number of Recommendations with Anonymized Analyst IDs (1)	Percentage of Recommendations with Anonymized Analyst IDs (%)		
		All Firms (2)	U.S. Firms Only (3)	Non-U.S. Firms Only (4)
1993	4,781	12.8	6.8	16.2
1994	12,810	15.2	7.7	19.0
1995	12,339	14.9	6.2	19.7
1996	14,799	15.6	4.3	20.3
1997	16,640	16.6	3.7	21.5
1998	17,670	15.9	3.5	21.0
1999	15,627	13.9	6.1	17.4
2000	11,655	12.2	5.3	15.4
2001	11,824	11.8	8.7	13.1
2002	6,202	5.0	3.7	5.8
2003	4,258	4.0	3.3	4.3
2004	3,013	2.9	0.5	3.9
2005	2,259	2.2	0.3	2.9
2006	2,625	2.4	0.2	3.2
2007	2,444	2.0	0.1	2.7
2008	2,391	1.8	0.1	2.5
2009	2,656	2.0	0.1	2.6
2010	2,904	2.4	0.3	3.1
2011	2,471	2.0	0.5	2.4
2012	75	0.1	0.1	0.0
2013	101	0.1	0.1	0.1
2014	127	0.1	0.1	0.1
2015	185	0.2	0.2	0.2
2016	520	0.5	0.2	0.6
2017	102	0.1	0.1	0.1
2018	154	0.2	0.1	0.2
2019	838	0.9	1.4	0.8
2020	149	0.2	0.2	0.1
August 2021	125	0.2	0.4	0.2
Total	151,744	5.1	2.3	6.2

Table 3 includes all recommendations issued between October 29, 1993 and August 31, 2021. Column (1) tabulates the number of recommendations with anonymized analyst IDs in the Recommendations Detail File. Column (2) divides the number in column (1) by the number of recommendations in a given year. Column (3) divides the number of recommendations of U.S. firms (*USFIRM* = 1) with anonymized analyst IDs by the number of recommendations of U.S. firms in a given year. Column (4) divides the number of recommendations of non-U.S. firms (*USFIRM* = 0) with anonymized analyst IDs by the number of recommendations of non-U.S. firms in a given year. The shaded rows represent the observations after I/B/E/S allegedly anonymized or removed its observations.

quarterly earnings forecasts of non-U.S. firms rather than U.S. firms. The ID anonymization has affected less than 0.1 percent of the annual and quarterly earnings forecasts of U.S. firms since 1990. Therefore, research studies that use the annual or quarterly earnings forecasts of U.S. firms are unlikely to be affected by the ID anonymization. However, researchers who use the annual or quarterly earnings forecasts of non-U.S. firms need to seriously consider the possible impact of the ID anonymization in their research design.

### Recommendations Detail File

Does the ID anonymization affect the Recommendations Detail File? As above, anonymized analyst IDs mean that the analyst ID variable (*AMASKCD*) of a recommendation is 0. I use an I/B/E/S Recommendations Detail File



**TABLE 4**  
**Recovering Anonymized Analyst IDs**

Descriptions	Number of Forecasts with Anonymized Analyst IDs (1)	Number of Forecasts with Recovered Analyst IDs (2)	Percentage of Recovery Rate (%) (3) = (2) ÷ (1)
(a) All Forecasts	3,230,514	1,798,805	55.7
(b) Annual <i>EPS</i> Forecasts	340,673	128,972	37.9
(c) U.S. Firms	8,694	304	3.5
(d) Non-U.S. Firms	331,979	128,668	38.8
(e) Quarterly <i>EPS</i> Forecasts	27,594	20,616	74.7
(f) U.S. Firms	2,182	260	11.9
(g) Non-U.S. Firms	25,412	20,356	80.1
(h) Annual <i>CPS</i> Forecasts	153,400	73,485	47.9
(i) Annual <i>DPS</i> Forecasts	264,027	93,431	35.4

The data in Table 4 are based on the forecasts issued between January 1, 1990 and August 31, 2021. Column (1) tabulates the number of forecasts with anonymized analyst IDs. Column (2) tabulates the number of forecasts with recovered analyst IDs following the steps detailed in the “Recovering Anonymized Analyst IDs” section. Column (3) divides the number in column (2) by the corresponding number in column (1).

Variable Definitions:

*EPS* = earnings per share;

*CPS* = cash flow per share; and

*DPS* = dividend per share.

downloaded from WRDS in November 2021 to answer this question. The file contains 2,959,702 recommendations. The announcement dates of the recommendations (*ANNDATS*) are between December 14, 1992 and August 31, 2021; 26.9 percent of the recommendations are for U.S. firms (*USFIRM* = 1).

Table 3, column (1) tabulates the number of recommendations with anonymized analyst IDs. The column shows no significant increase in the number of recommendations with anonymized analyst IDs after October 2018. Indeed, the number of recommendations with anonymized analyst IDs has noticeably decreased from 2,471 in 2011 to 75 in 2012, and it is holding in the range of about 125–149 in 2020–2021. Column (2) divides the number in column (1) by the number of recommendations in a given year. The percentage echoes the results in column (1), as the percentage of recommendations with anonymized analyst IDs significantly drops from 2.0 percent in 2011 to 0.1 percent in 2012 and it stays at a relatively low level of about 0.1–0.9 percent between 2012 and 2021.

Column (3) divides the number of recommendations of U.S. firms (*USFIRM* = 1) with anonymized analyst IDs by the number of recommendations of U.S. firms in a given year. Finally, column (4) divides the number of recommendations of non-U.S. firms (*USFIRM* = 0) with anonymized analyst IDs by the number of recommendations of non-U.S. firms in a given year. The results in these two columns suggest that the number of recommendations with anonymized analyst IDs is higher for non-U.S. firms than for U.S. firms, a pattern similar to that for the forecasts in the Unadjusted Detail History File. From January 1993 to August 2021, only 2.3 percent of the recommendations for U.S. firms and 6.2 percent of the recommendations for non-U.S. firms were anonymized.

Does the ID anonymization affect broker IDs in the Recommendations Detail File (*ESTIMID*)? As before, anonymization means that a broker ID is 0. I use the same Recommendations Detail File to examine this. Untabulated tests show that none of the broker IDs (*ESTIMID*) for the three million recommendations in the Recommendations Detail File are anonymized.

Overall, the evidence suggests that, although the ID anonymization is also observed in the Recommendations Detail File, the extent of ID anonymization appears to be negligible in the Recommendations Detail File.

### Unadjusted Price Target File

I examine whether ID anonymization affects the Unadjusted Price Target File and use the Unadjusted Price Target File downloaded on December 28, 2021 to examine this. The announcement dates (*ANNDATS*) of the price targets are between June 1, 1970 and December 16, 2021. The results are tabulated in Table 4 in the Online Appendix to conserve space. There is no evidence of a large-scale ID anonymization in the Unadjusted Price Target File since October 2018.

## Recovering Anonymized Analyst IDs

In this section, I offer detailed steps to reverse engineer the anonymized IDs in the Unadjusted Detail History File. The idea is straightforward. Because (1) the ID anonymization affects the Unadjusted Detail History File, but not the Recommendations Detail File and (2) an analyst of a given broker who issues a recommendation is the same analyst who concurrently issues a forecast, I am able to reverse engineer some of the anonymized analyst IDs in the Unadjusted Detail History File based on the nonanonymized analyst IDs from the Recommendations Detail File. Why can't these steps reverse-engineer all anonymized IDs? It is because, if the analyst IDs in the Unadjusted Detail History File and the Recommendations Detail File are both anonymized, these steps cannot reverse engineer the anonymized IDs in the Unadjusted Detail History File.

The following example illustrates how the algorithm works. Column (1) is an earnings forecast in the Unadjusted Detail History File where the analyst ID is anonymized. Column (2) is the corresponding recommendation issued by the same broker, Goldman Sachs, on the same firm and same date and with the same U.S. firm status. Given that (1) the *ESTIMATOR* of "GOLDMAN" is 118 and (2) *AMASKCD* is the same as *ANALYS*,<sup>14</sup> the anonymized *ANALYS* is 188313, and the analyst who issues the forecast is Mr. Christopher Prykull, who worked as an analyst at Goldman Sachs between December 2011 and February 2020.<sup>15</sup>

I/B/E/S Variables	Recommendations Detail File	Unadjusted Detail History File
<i>ANALYS</i>		0 → 188313
<i>AMASKCD</i>	188313	
<i>ANALYST</i>	PRYKULL, CFA C	
<i>TICKER</i>	SIIX	SIIX
<i>FIRM NAME</i>	Six Flags Entertainment Corporation	Six Flags Entertainment Corporation
<i>USFIRM</i>	1	1
<i>ESTIMATOR/ESTIMID</i>	GOLDMAN	118
<i>ANNDATS</i>	April 6, 2017	April 6, 2017
<i>REVDATS</i>	March 11, 2019	April 6, 2017

There are two main steps to reverse engineer the anonymized analyst IDs. The first step creates a linking table mapping the broker ID in the Unadjusted Detail History File (*ESTIMATOR*) to the broker ID in the Recommendations Detail File (*ESTIMID*). Such a linking table seems trivial but in fact is extremely important because, as illustrated above, the numeric *ESTIMATOR* is different from the textual broker ID in the Recommendations Detail File (*ESTIMID*). I/B/E/S, unfortunately, also does not offer any guidance or linking table to help researchers map these broker IDs. Without such a linking table, it is not possible to reverse engineer the anonymized analyst IDs.

I collect the following variables from the Unadjusted Detail History File to construct the linking table: ticker (*TICKER*), broker ID (*ESTIMATOR*), U.S. firm status (*USFIRM*), announcement date (*ANNDATS*), and review date (*REVDATS*) of all forecasts with nonanonymized IDs (*ANALYS*). I then collect the following variables from the Recommendations Detail File: ticker (*TICKER*), broker ID (*ESTIMID*), U.S. firm status (*USFIRM*), announcement date (*ANNDATS*), and review date (*REVDATS*) of all recommendations with nonanonymized analyst IDs (*AMASKCD*). Next, I merge the forecasts with the recommendations if (1) the ticker (*TICKER*) and U.S. firm status (*USFIRM*) of a forecast are identical to those of a recommendation, (2) the analyst ID in the Unadjusted Detail History File (*ANALYS*) of a forecast is identical to the analyst ID in the Recommendations Detail File (*AMASKCD*), and (3) the first date of the recommendation (*ANNDATS*) falls on or before the last date of the forecast (*REVDATS*) and the last date of the recommendation (*REVDATS*) falls on or after the first date of the forecast (*ANNDATS*).

Restricting a forecast to fall within a recommendation's window ensures that the time interval for a forecast in the Unadjusted Detail History File and the time interval for a recommendation in the Recommendations Detail File intersect. Restricting a forecast to fall within a recommendation's window is not bulletproof (i.e., it does not guarantee 100 percent accuracy), but it minimizes the misidentification of analyst IDs because—as mentioned earlier—an analyst of a

<sup>14</sup> See <https://wrds-www.wharton.upenn.edu/pages/support/support-articles/ibes/how-do-i-match-analysts-detail-history-estimates-company-associated-recommendations-analyst-company>

<sup>15</sup> See <https://www.linkedin.com/in/christopher-prykull-cfa-cpa-26987816> for Mr. Prykull's employment history. See the Standard & Poor's (S&P) press release covering Mr. Prykull's coverage on Six Flags ([https://www.spglobal.com/marketintelligence/en/news-insights/trending/g9obeld57dl3\\_03ca8lq7q2](https://www.spglobal.com/marketintelligence/en/news-insights/trending/g9obeld57dl3_03ca8lq7q2)).

given broker who issues a recommendation is the same analyst who concurrently issues a forecast. In an untabulated test, I also examine this by relaxing the requirement and find that the accuracy of the linking table generated without such a restriction on announcement dates is much lower.

I then retain both broker IDs (*ESTIMATOR* and *ESTIMID*) and U.S. firm status (*USFIRM*). Keeping *USFIRM* in the linking table is subtle but technically important because the same *ESTIMID* can map to multiple *ESTIMATOR* values (see below for an example) and only an *ESTIMID*–*USFIRM* pair can match a unique *ESTIMATOR*–*USFIRM* pair.

Broker Name	<i>ESTIMID</i>	<i>USFIRM</i>	<i>ESTIMATOR</i>
Goldman Sachs	GOLDMAN	1	118
Goldman Sachs	GOLDMAN	0	119

I then count how many forecasts are issued under a unique *ESTIMATOR*–*ESTIMID*–*USFIRM* trio. After counting, I retain the *ESTIMATOR*–*ESTIMID*–*USFIRM* trio with the highest count of *ESTIMATOR*–*USFIRM* because a unique *ESTIMATOR*–*ESTIMID*–*USFIRM* trio represents the most likely mapping between the broker ID in the Unadjusted Detail History File (*ESTIMATOR*) and the broker ID in the Recommendations Detail File (*ESTIMID*). Retaining the highest count is similar to choosing the trio with the highest probability because other trios with lower counts represent less likely mappings that are more prone to error.

Before proceeding further, I use the I/B/E/S Broker Translation File to verify the accuracy of the linking table. Although I/B/E/S stopped updating its I/B/E/S Broker Translation File since 2007, the file contains the real mapping of *ESTIMATOR*–*ESTIMID*–*USFIRM*.<sup>16</sup> Hence, the information in the I/B/E/S Broker Translation File would help verify the accuracy of my linking table at least through the period before the I/B/E/S Broker Translation File stopped updating the file. I find that, conditional on observing the same broker ID (*ESTIMID*–*USFIRM*) in the I/B/E/S Broker Translation File and the linking table, my linking table correctly identifies 98.5 percent of the real mapping of *ESTIMATOR*–*ESTIMID*–*USFIRM* in the I/B/E/S Broker Translation File.

At the end of these steps, this final linking table covers 2,252 *ESTIMID* values; 3,067 *ESTIMATOR* values; 2,971 unique *ESTIMID*–*USFIRM* pairs; and 3,178 unique *ESTIMATOR*–*USFIRM* pairs.<sup>17</sup>

The second step is to reverse engineer the anonymized IDs in the Unadjusted Detail History File. I go back to the Unadjusted Detail History File to identify those forecasts with anonymized analyst IDs (*ANALYS*). I then merge the forecasts with anonymized IDs with the recommendations in the Recommendations Detail File if (1) the ticker (*TICKER*) and U.S. firm status (*USFIRM*) of a forecast are identical to those of a recommendation, (2) the broker ID in the Unadjusted Detail History File (*ESTIMATOR*–*USFIRM*) in the linking table corresponds to the broker ID in the Recommendations Detail File (*ESTIMID*–*USFIRM*), and (3) the first date of the recommendation (*ANNDATS*) falls on or before the last date of the forecast (*REVDATS*) and the last date of the recommendation (*REVDATS*) falls on or after the first date of the forecast (*ANNDATS*).

Recall that the analyst ID variable in the Unadjusted Detail History File (*ANALYS*) is identical to the analyst ID variable in the Recommendations Detail File (*AMASKCD*). Hence, I reverse engineer the anonymized analyst IDs in the Unadjusted Detail History File (*ANALYS*) based on the nonanonymized analyst IDs in the Recommendations Detail File (*AMASKCD*).

Before reporting the recovery rate of the reverse-engineering exercise on those forecasts with anonymized analyst IDs, I test the accuracy of the reverse engineering. I use the 114,839,288 forecasts with nonanonymized analyst IDs in the Unadjusted Detail History File issued on or after 1990. I then apply the algorithm and find that 31 percent (36,014,963 ÷ 114,839,288) of these forecasts have corresponding recommendations with nonanonymized analyst IDs in

<sup>16</sup> My untabulated investigation into the I/B/E/S Broker Translation File shows that the number of brokers newly added in the file significantly decreased in 2007. The decrease corresponds to the year when I/B/E/S is reported to have ceased to provide an updated file (see <https://wrds-www.wharton.upenn.edu/pages/support/support-articles/ibes/ibes-broker-transaction-files> for details). Hence, the result suggests that the I/B/E/S Broker Translation File stops updating from 2008 onward.

<sup>17</sup> I do not use *EMASKCD* in the Recommendations Detail file for two reasons. First, it is more likely to observe one *EMASKCD* matched to multiple *ESTIMID* values than one *ESTIMID* matched to multiple *EMASKCD* values. Hence, using *ESTIMID* is less likely to have incorrect matches than using *EMASKCD*. Second, the original I/B/E/S Broker Translation File does not use *EMASKCD* for broker IDs. To verify the accuracy of the linking table, I need to use the information in the original I/B/E/S Broker Translation File, so I use *ESTIMID* (but not *EMASKCD*) to construct the linking table.

**TABLE 5**  
**Percentage of Forecasts with Reshuffled Analyst IDs**

**Panel A: Forecasts**

Descriptions	Number of Forecasts in the OLD Detail History File (1)	Number of Forecasts with Reshuffled Analyst IDs between the OLD and the NEW Detail History Files (2)	Percentage of Forecasts with Reshuffled Analyst IDs between the OLD and the NEW Detail History Files (%) (3) = (2) ÷ (1)
(a) All forecasts	84,730,387	2,432,375	2.87
(b) Annual <i>EPS</i>	10,011,136	248,870	2.49
(c) U.S. Firms	4,314,473	51,875	1.20
(d) Non-U.S. Firms	5,696,663	196,995	3.46
(e) Quarterly <i>EPS</i>	3,641,541	48,185	1.32
(f) U.S. Firms	3,151,827	18,075	0.57
(g) Non-U.S. Firms	489,714	30,110	6.15
(h) Annual <i>CPS</i>	2,694,483	63,886	2.37
(i) Annual <i>DPS</i>	2,307,216	89,459	3.88

**Panel B: Transition Matrix**

U.S. and Non-U.S. Firms			
Forecasts with Anonymized Analyst IDs in the OLD Detail History File?	Forecasts with Anonymized Analyst IDs in the NEW Detail History File?		
	No	Yes	Total
Yes	2,401,237 (98.7%)	0 (0.0%)	2,401,237 (98.7%)
No	25,905 (1.1%)	5,233 (0.2%)	31,138 (1.3%)
Total	2,427,142 (99.8%)	5,233 (0.2%)	2,432,375 (100.0%)
Non-U.S. Firms Only			
Yes	2,207,543 (99.3%)	0 (0.0%)	2,207,543 (99.3%)
No	10,602 (0.5%)	5,227 (0.2%)	15,829 (0.7%)
Total	2,218,145 (99.8%)	5,227 (0.2%)	2,223,372 (100.0%)
U.S. Firms Only			
Yes	193,694 (92.7%)	0 (0.0%)	193,694 (92.7%)
No	15,303 (7.3%)	6 (0.0%)	15,309 (7.3%)
Total	208,997 (100.0%)	6 (0.0%)	209,003 (100.0%)

Table 5, Panel A summarizes the percentage of forecasts whose analyst IDs were reshuffled between the OLD and NEW Detail History Files. See the [Online Appendix](#), Table 5, for more details on Panel A. Panel B tabulates the transition matrix. The 2,432,375 forecasts in Panel B are forecasts with reshuffled analyst IDs between the OLD and the NEW Detail History Files. The percentage is relative to the total number of forecasts.

**TABLE 6**  
**WRDS Release Notes on Comparing I/B/E/S Vintages**

Old I/B/E/S	New I/B/E/S	Release Note	Release Date	Number of Brokers Added				Number of Brokers Removed			
				<i>EPS</i>	<i>EPS</i>	Non- <i>EPS</i>	Non- <i>EPS</i>	<i>EPS</i>	<i>EPS</i>	Non- <i>EPS</i>	Non- <i>EPS</i>
				Forecasts of U.S. Firms	Forecast of Non-U.S. Firms	Forecasts of U.S. Firms	Forecasts of Non-U.S. Firms	Forecasts of U.S. Firms	Forecast of Non-U.S. Firms	Forecasts of U.S. Firms	Forecasts of Non-U.S. Firms
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
2016-10	2017-05	Yes	2017-05	18	24	13	23	1	0	1	0
2017-05	2017-08	Yes	2017-08	4	11	5	11	0	0	0	0
2017-08	2018-01	Yes	2018-01	12	15	12	17	1	0	0	1
2018-01	2018-05	Yes	2018-05	5	10	2	13	0	0	0	1
2018-05	2019-06	No	—	—	—	—	—	—	—	—	—
2019-06	2019-09	Yes	2019-09	11	13	9	12	1	2	1	2
2019-09	2020-03	Yes	2020-03	160	214	125	189	98	114	95	105
2020-03	2020-06	Yes	2020-06	3	11	5	7	0	0	0	0
2020-06	2020-09	Yes	2020-09	6	9	5	9	1	1	1	1
2020-09	2020-12	Yes	2020-12	13	27	13	29	9	15	10	14
2020-12	2021-03	Yes	2021-03	17	26	16	24	13	18	12	18
2021-03	2021-06	Yes	2021-06	3	8	3	6	0	0	0	0
2021-06	2021-09	Yes	2021-09	3	6	3	5	0	1	0	0

Table 6 summarizes the findings of the Release Notes prepared by Wharton Research Data Services (WRDS) that compare I/B/E/S updates. The Release Notes are available at <https://wrds-www.wharton.upenn.edu/pages/support/manuals-and-overviews/i-b-e-s/ibes-estimates/release-notes>.

the Recommendations Detail File. When I exclude the forecasts matched to multiple analyst IDs in the Recommendations Detail File, the algorithm correctly predicts the analyst IDs of 83 percent of the 36 million forecasts.

Now, how well does such an algorithm reverse engineer the anonymized analyst IDs in the Unadjusted Detail History File? Table 4 tabulates the recovery rates of the forecasts with anonymized analyst IDs. Column (1) tabulates the number of forecasts with anonymized analyst IDs. Column (2) tabulates the number of forecasts with recovered analyst IDs following the above steps. Column (3) divides the number in column (2) by the corresponding numbers in column (1). In column (3), following these steps, the algorithm reverse engineers about 55.7 percent of forecasts with anonymized analyst IDs. For annual *EPS* forecasts, the recovery rate is about 37.9 percent; for quarterly *EPS* forecasts, it is 74.7 percent. The recovery rates are higher for the annual *EPS* forecasts of non-U.S. firms than for those of U.S. firms (i.e., 38.8 percent versus 3.5 percent), and the same is true for quarterly *EPS* forecasts (i.e., 80.1 percent versus 11.9 percent). The lower recovery rates for the forecasts of U.S. firms are not surprising because my earlier analyses show that the ID anonymization predominantly affects the annual and quarterly earnings forecasts of non-U.S. firms. Last, the algorithm reverse engineers a similar percentage of annual *CPS* forecasts and *DPS* forecasts with anonymized analyst IDs (47.9 percent and 35.4 percent, respectively).

Overall, this section shows that it is possible to reverse engineer the anonymized IDs in the Unadjusted Detail History File. Such recovery is especially helpful for research studies that use the forecasts of non-U.S. firms.

### III. ID SHUFFLING

#### Overview of the OLD Detail and Recommendations Detail Files

In this section, I examine whether and to what extent ID reshuffling affects the Unadjusted Detail History File. ID reshuffling means that the analyst ID variable (*ANALYS*) of a forecast has changed since October 2018. As mentioned earlier, Appendix E tabulates examples of forecasts whose analyst IDs have been reshuffled.



In addition to the I/B/E/S Unadjusted Detail History File and the Recommendations Detail File downloaded from WRDS in November 2021, I use two old files to answer this question. The first file is an Unadjusted Detail History File downloaded on September 3, 2015. The file contains 84,730,387 forecasts representing the ten types of forecasts: *BPS*, *CPS*, *CSH*, *DPS*, *EBG*, *EBS*, *EPS*, *EPX*, *FFO*, and *GPS*. *EPS* forecasts are the most numerous, accounting for 47.6 percent of the forecasts. The file contains 32 types of forecasts covering different ranges of forecast horizons (*FPI*). Annual forecasts (*FPI* = 1) are the most numerous, accounting for 25.1 percent of the forecasts; 6.7 percent of the forecasts are quarterly forecasts (*FPI* = 6). The announcement dates of the forecasts (*ANNDATS*) are between January 1, 1970 and April 30, 2015; 44.3 percent of the forecasts are forecasts for U.S. firms (*USFIRM* = 1). Table 5, Panel A, in the [Online Appendix](#) tabulates the forecasts by year and measure, and Panel B tabulates the forecasts by forecast horizon and measure. The second file is a Recommendations Detail File downloaded on November 15, 2017. The file contains 2,753,714 recommendations. The announcement dates of the forecasts (*ANNDATS*) are between December 14, 1992 and July 20, 2017; 28.1 percent of the forecasts are recommendations for U.S. firms (*USFIRM* = 1). [Online Appendix](#) Table 6 tabulates the recommendations by year.

For ease of reference, I refer to this Unadjusted Detail History File as the OLD Detail History File and this Recommendations Detail File as the OLD Recommendations Detail File. I also refer to the Unadjusted Detail History File and Recommendations Detail File that I downloaded in November 2021 as the NEW Detail History File and the NEW Recommendations Detail File, respectively.

### Percentage of Forecasts with Reshuffled Analyst IDs

I now investigate how many forecasts in the OLD Detail History File have been reshuffled with new analyst IDs in the NEW Detail History File. The details of the matching exercise are as follows. First, there are 21 variables in both the OLD and NEW Detail History Files (i.e., *TICKER*, *CUSIP*, *OFTIC*, *CNAME*, *ACTDATS*, *ESTIMATOR*, *ANALYS*, *CURRFL*, *PDF*, *FPI*, *MEASURE*, *VALUE*, *CURR*, *USFIRM*, *FPEDATS*, *ACTTIMS*, *REVDATS*, *REVTIMS*, *ANNDATS*, *ANNTIMS*, and *REPORT\_CURR*).<sup>18</sup> For each forecast in the OLD Detail History File, I track down the same forecast with identical values in all of these variables in the NEW Detail History File and exclude them from subsequent analyses (because they are identical in both files, so there is no ID reshuffling for the forecasts).

Second, for the remaining unmatched forecasts in the first step, I again track down the same forecast with identical values in 20 of these variables (there is one less variable because now *ANALYS* is excluded). The probability of observing two forecasts with identical values in these 20 variables in the OLD Detail History File is 0. Hence, these 20 variables together serve as a unique fingerprint for each forecast in the OLD and NEW Detail History Files. Third, I compare the analyst IDs in the OLD and the NEW Detail History Files of the forecasts from the second step. If the analyst IDs of the same forecast in the OLD and the NEW Detail History Files are different, I classify such a forecast as having a reshuffled analyst ID.

Before proceeding, it is important to acknowledge the limitation of the matching exercise. First, such a matching exercise cannot track down forecasts in the OLD Detail History File that are subsequently deleted from the NEW Detail History File. [Call et al. \(2021\)](#), for example, find that 11.68 percent of annual earnings forecasts from the 2009 Detail History File are deleted from the 2015 Detail History File. If a forecast in the OLD Detail History File is subsequently deleted in the NEW Detail History File, the matching exercise will undercount the number of forecasts with reshuffled analyst IDs between the OLD and the NEW Detail History Files. Second, if the analyst ID of a forecast concurrently changed with the value of any of the 19 variables, the matching exercise will undercount the number of forecasts with reshuffled analyst IDs between the OLD and the NEW Detail History Files.

[Table 5](#), Panel A tabulates the outcome of the matching exercise. On average, only 2.87 percent of forecasts have their analyst IDs reshuffled between the OLD and the NEW Detail History Files. The percentage is slightly lower for annual earnings forecasts (i.e., 2.49 percent) and is even lower for quarterly earnings forecasts (i.e., 1.32 percent). For the annual (quarterly) earnings forecasts for U.S. firms, the percentage of forecasts with reshuffled analyst IDs is very low at 1.20 percent (0.57 percent). These figures are substantially lower than the 30.7 percent of analyst IDs reported by WRDS. The percentage of forecasts with reshuffled analyst IDs is higher for non-U.S. firms than for U.S. firms (rows (c) and (f) versus rows (d) and (g)). For annual *CPS* forecasts, 2.37 percent of forecasts have their IDs reshuffled between the OLD and the NEW Detail History Files. The percentage is higher for annual *DPS* forecasts at 3.88 percent. For more details on the outcome of the matching exercise for different types of forecasts, please see [Online Appendix](#) Table 7.

<sup>18</sup> See <https://wrds-www.wharton.upenn.edu/pages/get-data/ibes-thomson-reuters/ibes-academic/unadjusted-detail/history> for the descriptions of each variable.

Given the earlier analyses on ID anonymization in Section II, a natural question is whether the ID reshuffling occurs because nonanonymized analyst IDs in the OLD Detail History File have been anonymized in the NEW Detail History File (or *vice versa*). To examine this possibility, I use the sample of 2,432,375 forecasts (in Table 5, Panel A, column (2)) whose IDs have been reshuffled between the OLD and the NEW Detail History Files for analyses.

Table 5, Panel B shows whether the analyst ID of a forecast is anonymized in the OLD and the NEW Detail History Files. There are several findings. First, the transition matrix indicates that only 0.2 percent of forecast IDs have been reshuffled from nonanonymized (i.e., the analyst ID is not 0) in the OLD Detail History File to anonymized (the analyst ID is 0) in the NEW Detail History File. Second, about 99 percent of forecasts with reshuffled analyst IDs are the forecasts whose analyst IDs were initially assigned as anonymized and only assigned subsequently in the NEW Detail History File (i.e., the analyst IDs changed from 0 to non-0). Third, the ID reshuffling is observed mainly among the forecasts of non-U.S. rather than U.S. firms. Fourth, it is very rare to observe forecasts with reshuffled new analyst IDs from non-0 analyst IDs to non-0 analyst IDs (e.g., *ANALYS* from 3195 to 43110 in Appendix E).

Overall, the results suggest that almost all forecasts with reshuffled analyst IDs were forecasts that were initially anonymized (i.e., the analyst ID was 0) in the OLD Detail History File but were subsequently backfilled (i.e., the analyst ID was not 0) in later vintages. These findings are important for analyst researchers because, even though ID reshuffling seems problematic, it affects the forecasts that are often already excluded from academic studies.

Although the above analyses focus on analyst IDs, I also examine whether and to what extent ID reshuffling affects the broker IDs (instead of analyst IDs) in the Unadjusted Detail History File. I continue my earlier matching exercise. First, for each forecast in the OLD Detail History File, I track down the forecasts with identical values in all 21 variables in the NEW Detail History File and exclude them (because they are identical in both files and there is no ID reshuffling issue).

Second, for the remaining unmatched forecasts in the first step, I again track down the same forecasts with identical values in 20 of these variables (there is one less variable because *ESTIMATOR* is now excluded). Third, I compare the broker IDs in the OLD and the NEW Detail History Files. If the broker IDs of the same forecast in the OLD and the NEW Detail History Files are different, I classify the forecast as having a reshuffled broker ID.

An untabulated test shows that, out of 84 million forecasts in the OLD Detail History File, 703 of them (i.e., less than 0.001 percent) have been reshuffled with new broker IDs in the NEW Detail History File. I also relax the matching criteria to exclude two variables, *ESTIMATOR* and *ANALYS*, at the same time. In another untabulated test, I find that, out of 84 million forecasts in the OLD Detail History File, only 2.9 percent ( $2,435,508 \div 84,730,387$ ) have been reshuffled and have both new broker IDs and new analyst IDs in the NEW Detail History File.

Overall, the evidence above suggests the following. First, there is no evidence of a large-scale ID reshuffling of forecasts between the OLD and the NEW Detail History Files after October 2018. On average, only 2.87 percent of forecasts have their IDs reshuffled between the OLD and the NEW Detail History Files. For the annual (quarterly) earnings forecasts for U.S. firms, the percentage of forecasts with reshuffled analyst IDs is very low at 1.20 percent (0.57 percent). Second, only 0.2 percent of forecasts have their IDs reshuffled from nonanonymized in the OLD Detail History File to anonymized in the NEW Detail History File. About 99 percent of forecasts with reshuffled analyst IDs are the forecasts whose analyst IDs were initially assigned as anonymized (i.e., analyst IDs changing from 0 to non-0).

### Percentage of Recommendations with Reshuffled Analyst IDs

I now examine how many recommendations in the OLD Recommendations Detail File have been reshuffled with new analyst IDs in the NEW Recommendations Detail File. Again, I follow the same matching procedures. There are 19 variables in both OLD and NEW Recommendations Detail Files (i.e., *TICKER*, *CUSIP*, *CNAME*, *OFTIC*, *ACTDATS*, *ESTIMID*, *ANALYST*, *ERECCD*, *ETEXT*, *IRECCD*, *ITEXT*, *EMASKCD*, *AMASKCD*, *USFIRM*, *ACTTIMS*, *REVDATS*, *REVTIMS*, *ANNDATS*, and *ANNTIMS*).<sup>19</sup> For each recommendation in the OLD Recommendations Detail File, I track down the same recommendation with identical values in all 17 variables (i.e., two variables less because *AMASKCD* and *ANALYST* in the Recommendations Detail File are now excluded). I then compare the analyst IDs (*AMASKCD*) in the OLD and the NEW Recommendations Detail Files. If the analyst IDs of the same recommendation in the OLD and the NEW Recommendations Detail Files are different, I classify such a recommendation as a recommendation with a reshuffled analyst ID. Again, the probability of observing two recommendations with identical values in these 17 variables in the OLD Detail History File is 0. Hence, these 17 variables together

<sup>19</sup> See <https://wrds-www.wharton.upenn.edu/pages/get-data/ibes-thomson-reuters/ibes-academic/recommendations/detail> for the description of each variable.

serve as a unique fingerprint for each recommendation in the OLD and NEW Recommendations Detail Files. The limitations of the matching exercise for forecasts discussed above continue to apply.

Out of 2,753,714 recommendations, I find that only 146 recommendations have reshuffled analyst IDs. These 146 recommendations account for about 0.01 percent of all recommendations. As in the analyses based on forecasts, there is no evidence suggesting a large-scale ID reshuffling of recommendations between the OLD and the NEW Recommendations Detail Files after October 2018.

### Percentage of Unadjusted Price Targets with Reshuffled Analyst IDs

I conduct the same analyses using the Unadjusted Price Target File discussed earlier. I also obtained the Unadjusted Price Target File downloaded in April 2010. I find that the analyst IDs (*ALYSNAM*) of about 4.2 percent of 1,899,396 price targets have changed from 2010 to 2021. My inspection, however, reveals that most of the changes are corrections of errors (e.g., from “SCHNEIDER J” to “SCHNEIDER, PH.D. J”). Hence, there is no evidence of large-scale ID reshuffling in the Unadjusted Price Target File.

### Reconciliation with the WRDS’s Claims

As mentioned in [Section I](#), [WRDS \(2023a\)](#) claims that “It is likely that 13.8% of all broker IDs (*ESTIMATOR*) and 30.7% of all analyst IDs (*ANALYS*) have been reassigned.” So what drives such a discrepancy between this statement and my findings?

With my colleagues’ help, I obtained additional historical I/B/E/S files to conduct a fresh investigation. The first file is an Unadjusted Detail History File downloaded on September 26, 2019 from the September 2019 vintage (“2019 Unadjusted U.S. Detail History File”). The 2019 Unadjusted U.S. Detail History File includes 27,926,376 *EPS* forecasts of U.S. firms across different forecast horizons issued between January 28, 1980 and May 16, 2019. The smallest and the largest broker IDs (*ESTIMATOR*) are 3 and 91645, respectively. The smallest and the largest analyst IDs (*ANALYS*) are 0 and 702031, respectively.<sup>20</sup>

The second file is a Recommendations Detail File downloaded on the same date from the September 2019 vintage (“2019 Recommendations Detail File”). The 2019 Recommendations Detail File includes 2,812,735 recommendations of U.S. firms ( $n = 796,533$ ) and non-U.S. firms ( $n = 2,016,202$ ) issued between December 14, 1992 and May 16, 2019. Although 2,100 brokers (*EMASKCD*) are in the 2019 Recommendations Detail File, the range of the IDs is from 1 (the smallest) to 91638 (the largest).

Now I return to my analyses and summarize some stylized facts.

1. The range of broker IDs (*ESTIMATOR*) in the 2019 Unadjusted U.S. Detail History File is very large. In the historical I/B/E/S data available to me, no broker IDs in the Unadjusted Detail History Files are larger than 4519. In the 2019 Unadjusted U.S. Detail History File, however, 95 new brokers have broker IDs larger than 4519.
2. The announcement dates of the forecasts issued by analysts of these 95 new brokers (out of 1,519 brokers in total) are between January 28, 1980 and May 16, 2019; 34.9 percent (11,595 analysts from these 95 brokers ÷ 33,196 analysts in total) of analysts come from these 95 brokers. The analysts of these 95 brokers issued 35.7 percent of forecasts.
3. The range of broker IDs (*EMASKCD*) in the 2019 Recommendations Detail File is also very large. In the historical I/B/E/S data available to me, no broker IDs in the Recommendations Detail File are larger than 42400. In the 2019 Recommendations Detail File, however, 101 new brokers have broker IDs larger than 42400. The broker IDs (*ESTIMID*) of these 101 new brokers all start with “PRMDN” (e.g., PRMDN002 and PRMDN003), and the analyst names (*ANALYST*) of these 101 new brokers are all redacted with “PERMDENIED.” Sixteen of these brokers are not among the 88 brokers in the I/B/E/S Product Change Notification (see [Online Appendix](#), Table 1, Panel B for the list). The analyst IDs (*AMASKCD*) of these 101 new brokers range between 500005 and 656427.
4. The announcement dates of the recommendations issued by analysts of these 101 new brokers (out of 2,100 brokers in total) are between December 14, 1992 and May 16, 2019. These 101 brokers issued 33.1 percent of

<sup>20</sup> Before using the 2019 Unadjusted U.S. Detail File, I use another file to verify the data in the 2019 Unadjusted U.S. Detail File. That file is an Unadjusted Detail History File downloaded on October 1, 2019. The file includes 6,404,599 annual ( $FPI = 1$ ) and quarterly ( $FPI = 6$ ) *EPS* forecasts of U.S. firms that were issued between October 20, 1997 and June 1, 2015. The broker ID (*ESTIMATOR*) with the smallest value is 3, and the one with the largest value is 3961. If I focus on the annual and quarterly *EPS* forecasts (i.e., either  $FPI = 1$  or  $FPI = 6$ ) and restrict the announcement dates to be between October 20, 1997 and June 1, 2015 in the first file, the broker ID (*ESTIMATOR*) with the smallest value is 3 and the ID with the largest value is 3961. The smallest and the largest analyst IDs are 0 and 702031. The identical range of broker and analyst IDs provides independent verification of the data completeness and accuracy of the 2019 Unadjusted U.S. Detail File.

recommendations; 31.4 percent (21,908 analysts from these 101 brokers  $\div$  69,664 analysts in total) of analysts come from these 101 brokers.

5. WRDS has issued Release Notes from October 2016 to September 2021 (see [Table 6](#)). The only exception is the June 2019 vintage (which was scheduled to be released in WRDS in June 2019), because WRDS did not issue a Release Note to compare the June 2019 vintage with the May 2018 vintage.
6. In the March 2020 Release Note, WRDS stated that the IDs of 111 brokers were removed from I/B/E/S data. All of the 95 new brokers in the 2019 Unadjusted U.S. Detail History File were removed in the March 2020 vintage. However, 160 brokers were added for the *EPS* forecasts of U.S. firms (see [Table 6](#)).

The most likely explanation consistent with these stylized facts is as follows. In the June 2019 vintage, I/B/E/S revised the broker IDs (*ESTIMATOR*) and the analyst IDs (*ANALYS*) of at least 106 brokers (88 brokers + 16 undisclosed brokers) in the Detail History File and the Recommendations Detail File. In the Detail History File, the old analyst and broker IDs were replaced with new IDs never used before.

However, in the Recommendations Detail File, the old broker IDs were given completely new IDs starting with “PRMDN” (e.g., PRMDN002 and PRMDN003). The old analyst IDs of these affected brokers were suppressed. As a result, these new IDs were not replaced prospectively from October 2018 but were replaced retrospectively all the way back to the 1980s for all affected brokers and their analysts. Due to the significant number of revisions in broker IDs, WRDS did not issue a Release Note to compare the May 2018 and June 2019 vintages.

In the September 2019 vintage, the situation remained unchanged, so the number of brokers added and removed was minimal.

In the March 2020 vintage, I/B/E/S decided to reverse the revisions and restore these brokers’ original IDs and their analyst IDs. The significant number of new brokers added and removed in the March 2020 vintage represents the restoration of the original broker IDs and analyst IDs. As the number of brokers added is significantly more than the number of brokers removed in the March 2020 vintage, the asymmetry suggests that, in some cases, multiple original broker IDs were initially combined and given one new broker ID. For example, a back-of-envelope estimation would indicate that, on average, 1.7 original broker IDs (160 brokers added back  $\div$  95 brokers observed in the 2019 Unadjusted U.S. Detail History File) were merged into one new broker ID in the June 2019 vintage. As WRDS examined a larger sample covering both U.S. and non-U.S. firms, it observed that 13.8 percent of all broker IDs and 30.7 percent of all analyst IDs (*ANALYS*) were revised in June 2019, which is consistent with my findings above. However, I/B/E/S reversed these revisions in the March 2020 vintage. Instead, I/B/E/S has started to anonymize analyst IDs since January 2017. Overall, it would have been best for WRDS to update its guidance (or even remove the claim) back in March 2020 because the claim was no longer valid when WRDS received the March 2020 vintage.<sup>21</sup>

#### IV. CONCLUSION AND RECOMMENDATIONS

This paper examines whether and to what extent the ID anonymization and ID reshuffling affect the post-2018 Detail History File. The main findings are as follows. First, there is no evidence of a large-scale ID anonymization for the annual and quarterly *EPS* forecasts for U.S. firms after October 2018. The ID anonymization has affected less than 0.1 percent of the annual and quarterly earnings forecasts of U.S. firms since 1990. Second, the ID anonymization predominantly affects the forecasts of non-U.S. firms. A significant percentage of forecasts have been anonymized in the Unadjusted Detail History File since January 2017, at least one year earlier than what was claimed. Third, there is no evidence of a large-scale ID reshuffling of forecasts between the 2015 and 2021 Detail History Files. Finally, because ID anonymization has very little effect on the Recommendations Detail File, I am able to offer detailed steps to reverse engineer the anonymized IDs in the Unadjusted Detail History File based on the nonanonymized analyst IDs in the Recommendations Detail File.

Given the above findings, I offer the following recommendations:

1. Because the ID anonymization significantly affects the forecasts for non-U.S. firms, researchers who use the non-U.S. forecasts should consider following the steps in this paper to reverse engineer the anonymized analyst IDs in the Unadjusted Detail History File based on the nonanonymized analyst IDs in the Recommendations Detail File. Researchers should report the number of forecasts with recovered analyst IDs following the steps detailed above.

<sup>21</sup> My discussion with WRDS personnel also reveals another minor reason why our findings differ. WRDS classified a forecast as a forecast without ID reshuffling if the forecast has the same value in two different vintages of Adjusted Detail History Files. WRDS should use Unadjusted Detail History Files, but not Adjusted Detail History Files, for comparison because Adjusted Detail History Files account for stock splits. In other words, for a stock with a stock split, the same forecast would mechanically have different values in different vintages of Adjusted Detail History Files. Because the WRDS analyses used Adjusted Detail History Files, its analyses wrongly overstated the extent of ID reshuffling.



2. Researchers who use the forecasts for U.S. firms and choose to reverse engineer anonymized analyst IDs should report the number of forecasts with recovered analyst IDs following the steps detailed above. Researchers who use the forecasts for U.S. firms but choose not to reverse engineer anonymized analyst IDs should include a caveat on the ID anonymization.
3. After reverse engineering anonymized analyst IDs, researchers who use the Unadjusted Detail History Files should not include any forecasts with anonymized analyst IDs (i.e., analyst ID is 0). Including forecasts with anonymized analyst IDs would erroneously treat those forecasts as coming from the same analyst. Including these forecasts in one's research design may inadvertently affect the interpretation of research findings.
4. Researchers should closely inspect the broker and analyst IDs in each new I/B/E/S vintage for anomalies before using them. Extremely large broker IDs (e.g., *ESTIMATOR* above 4519) or analyst IDs in an unusual range (e.g., *AMASKCD* or *ANALYS* in the range between 500000 and 656400) should be flagged for further investigation.
5. Researchers should clearly state when they downloaded the I/B/E/S data. Researchers should not use the I/B/E/S June 2019 and September 2019 vintages or the I/B/E/S vintage downloaded between June 2019 and March 2020 (both months inclusive). This will help eliminate any concerns stemming from the now outdated WRDS claims described in the "[Reconciliation with WRDS's Claims](#)" section.
6. Researchers should not use Adjusted Detail History Files and should only use Unadjusted Detail History Files.

## V. LOOKING FORWARD

In the future, I/B/E/S might increase the anonymization of forecasts and/or recommendations. There are two possible scenarios.

*Scenario 1 (the best-case scenario):* I/B/E/S increases the anonymization of forecasts, but not recommendations. In this scenario, analyst researchers could still use the detailed steps to reverse engineer the anonymized IDs in the Unadjusted Detail History File.

*Scenario 2 (the worst-case scenario):* I/B/E/S increases the anonymization of both forecasts and recommendations simultaneously. In this scenario, the number of anonymized forecasts and recommendations would significantly grow over time. It is also less likely that analyst researchers will be able to reverse engineer the IDs of anonymized forecasts in the Unadjusted Detail History File over time. Specifically, in the extreme case where I/B/E/S simultaneously anonymized all analyst IDs in the Recommendations Detail File for those forecasts with anonymized IDs in the Unadjusted Detail History File, my code will not recover any anonymized IDs in the Unadjusted Detail History File. In this scenario, this paper's findings and conclusions on ID anonymization might not hold for future vintages of the I/B/E/S files.

Although it is hard to predict which scenario will occur, I commit to updating the pattern of ID anonymization and the linking table at least once per calendar year (or until the next I/B/E/S structural shift). The updates will be made available at <https://sites.google.com/site/kelvinlaw2012/home/analysts-good-bye-ibes>. I hope they will benefit other researchers who are interested in analyst research.

## REFERENCES

- Call, A. C., M. Hewitt, J. Watkins, and T. L. Yohn. 2021. Analysts' annual earnings forecasts and changes to the I/B/E/S database. *Review of Accounting Studies* 26 (1): 1–36. <https://doi.org/10.1007/s11142-020-09560-x>
- Ljungqvist, A., C. Malloy, and F. Marston. 2009. Rewriting history. *The Journal of Finance* 64 (4): 1935–1960. <https://doi.org/10.1111/j.1540-6261.2009.01484.x>
- Payne, J. L., and W. B. Thomas. 2003. The implications of using stock-split adjusted I/B/E/S data in empirical research. *The Accounting Review* 78 (4): 1049–1067. <https://doi.org/10.2308/accr.2003.78.4.1049>
- Thomson Reuters. 2018. Product change notification. [https://wrds-www.wharton.upenn.edu/documents/1030/Product\\_Change\\_Notification-IBES\\_Detail\\_History-PreApproval\\_Contributor...pdf](https://wrds-www.wharton.upenn.edu/documents/1030/Product_Change_Notification-IBES_Detail_History-PreApproval_Contributor...pdf)
- Wharton Research Data Services (WRDS). 2020. I/B/E/S @ WRDS 101: Introduction and research guide. [https://wrds-www.wharton.upenn.edu/documents/1372/IBES\\_RUI.pptx](https://wrds-www.wharton.upenn.edu/documents/1372/IBES_RUI.pptx)
- Wharton Research Data Services (WRDS). 2023a. Data vendors: I/B/E/S. <https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/vendor-partner-ibes/>
- Wharton Research Data Services (WRDS). 2023b. Knowledgebase: I/B/E/S. <https://wrds-www.wharton.upenn.edu/login/?next=/pages/support/support-articles/ibes/what-does-0-code-analyst-represent/> (last accessed April 24, 2023).



---

## APPENDIX A

### I/B/E/S Product Change Notification (Excerpt from Two Consecutive Pages)

#### Description of Change:

With the onset of the MiFID II and continuing focus on regulatory compliance, Thomson Reuters is modifying the display of detailed estimates data for PreApproval contributors in the I/B/E/S Detail History Product Suite – FTP, I/B/E/S v1 over QA Direct & Market QA and I/B/E/S Point in Time (PIT).

Effective October 18, 2018, the contributor and analyst names of 88 PreApproval contributors will be



anonymized for all clients accessing detailed estimates data over the I/B/E/S Detail History Product Suite, regardless of individual client entitlements.

(The full-color version is available online.)

---

## APPENDIX B

### WRDS Announcement of I/B/E/S Changes (Excerpt)



## I/B/E/S

I/B/E/S International Inc. created their Academic Research Program over 30 years ago to provide both summary and individual analyst forecasts of company earnings, cash flows, and other important financial items, as well as buy-sell-hold recommendations. In 2000, I/B/E/S was integrated with Thomson Reuters / First Call, and in 2012 First Call was discontinued.

As of October 18, 2018, Thomson-Reuters changed the identifiers of a large number of brokers and analysts in I/B/E/S. It is likely that 13.8% of all broker IDs (ESTIMATOR) and 30.7% of all analyst IDs (ANALYS) have been reassigned. While the October 2018 change affected a large portion of the sample, WRDS has also become aware of a different issue wherein individual **broker IDs** (and all affected analysts) have been and **will continue to** be subject to **reshuffle without warning**. The safest course of action may be to treat each data vintage as an entirely separate sample. See [this document](#) for more detail.

Additionally, UBS Equities was removed from I/B/E/S Detail History.

Source: <https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/vendor-partner-ibes/>

(The full-color version is available online.)

---

## APPENDIX C

### WRDS Slides (Excerpt)

#### Encrypted History (Cont.)

- Pierson (WRDS 2020) compares two I/B/E/S detailed files from 2014 and 2019 vintage to calibrate the impact made in Oct 2018.
  - The data from two vintage are matched based on estimated amount, announcement data, security, etc. except analyst and estimator codes.
  - It is likely that 13.8% of all broker IDs (ESTIMATOR) has been modified, consistent to the listed 89 brokers
  - Also I/B/E/S may have resigned up to 30.7% of all analyst IDs (ANALYS), many of whom are not necessarily associated with those 89 brokers
- Fortunately, the changes are only made to detailed estimate datasets, presumably due to regulatory concerns.

*"There will be no change to the I/B/E/S Summary History estimates product (consensus). Detailed estimates from all Pre-Approval brokers, including UBS, will remain within all summary/consensus calculations in accordance with existing methodology."*

#### Reference:

- Thomson Reuters Product Change Notification ref: CN 082718

23



Rui Dai, Ph.D. CFA

(The full-color version is available online.)

## APPENDIX D

### Examples of ID Anonymization

This is a screenshot of forecasts issued by Goldman Sachs (*ESTIMATOR* = 118), where the analyst IDs of forecasts issued after October 2018 have been anonymized.

```
. list anndats acttims ticker estima~r measure fpi analys value
```

	anndats	acttims	ticker	estima~r	measure	fpi	analys	value
1.	14may2013	17:39:47	GLNG	118	EPS	1	115781	.93
2.	04oct2013	08:00:56	GLNG	118	EPS	1	115781	.89
3.	14may2013	17:39:47	GLNG	118	EPS	2	115781	1.19
4.	04oct2013	08:00:56	GLNG	118	EPS	2	115781	1.24
5.	14may2013	17:39:47	GLNG	118	EPS	3	115781	1.8
6.	04oct2013	08:00:56	GLNG	118	EPS	3	115781	1.88
7.	10dec2019	04:22:57	GLNG	118	EPS	1	0	-1.88
8.	10dec2019	04:22:57	GLNG	118	EPS	6	0	-.38
9.	10dec2019	04:22:57	GLNG	118	EPS	2	0	-.21
10.	10dec2019	04:22:57	GLNG	118	EPS	3	0	-.17

---

## APPENDIX E

### Examples of ID Reshuffling

These are the screenshots of example forecasts in which analyst IDs are reshuffled between the 2015 and 2020 vintages.

Example 1: JP Morgan (*ESTIMATOR* = 797)

OLD Detail History File

```
. list anndats acttims ticker estimator measure analys value
```

	anndats	acttims	ticker	estima~r	measure	analys	value
1.	18feb2011	06:40:32	@JOE	797	EPS	3195	.05

NEW Detail History File

```
. list anndats acttims ticker estimator measure analys value
```

	anndats	acttims	ticker	estima~r	measure	analys	value
1.	18feb2011	06:40:32	@JOE	797	EPS	43110	.05

Example 2: Goldman Sachs (*ESTIMATOR* = 118)

OLD Detail History File

```
. list anndats acttims ticker estimator measure analys value
```

	anndats	acttims	ticker	estima~r	measure	analys	value
1.	18feb1994	17:21:20	CFVR	118	EPS	32499	.08
2.	27apr1994	17:03:38	CFVR	118	EPS	32499	.1
3.	15aug1994	15:38:45	CFVR	118	EPS	32499	.09
4.	17mar1995	16:01:36	CFVR	118	EPS	32499	.06

NEW Detail History File

```
. list anndats acttims ticker estimator measure analys value
```

	anndats	acttims	ticker	estima~r	measure	analys	value
1.	18feb1994	17:21:19	CFVR	118	EPS	15000	.08
2.	27apr1994	17:03:38	CFVR	118	EPS	15000	.1
3.	15aug1994	15:38:45	CFVR	118	EPS	15000	.09
4.	17mar1995	16:01:36	CFVR	118	EPS	15000	.06

---

Copyright of Journal of Financial Reporting is the property of American Accounting Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.