Rewriting History

ALEXANDER LJUNGQVIST, CHRISTOPHER MALLOY, AND FELICIA MARSTON [*]

ABSTRACT

We document widespread changes to the historical I/B/E/S analyst stock recommendations database. Across seven I/B/E/S downloads, obtained between 2000 and 2007, we find that between 6,580 (1.6%) and 97,582 (21.7%) of matched observations are different from one download to the next. The changes include alterations of recommendation, additions and deletions of records, and removal of analyst names. They are non-random, clustering by analyst reputation, broker size and status, and recommendation boldness, and affect trading signal classifications and back-tests of three stylized facts: The profitability of trading signals, the profitability of consensus recommendation changes, and persistence in individual analyst stock-picking ability.

[*] Ljungqvist is from the New York University Stern School of Business and the Centre for Economic Policy Research, London; Malloy is from Harvard Business School; and Marston is from the University of Virginia McIntire School of Commerce. Thanks for helpful comments go to Campbell Harvey (the Editor), two anonymous reviewers, Viral Acharya, Nick Barberis, Brad Barber, Lauren Cohen, Jennifer Juergens, Jay Ritter, Kent Womack, and to seminar participants at the U.S. Securities and Exchange Commission, the 2006 UNC-Duke Corporate Finance Conference, the 2007 AFA Conference in Chicago, AQR, Oppenheimer Funds, Barclays Global Investors, University of Sydney, University of New South Wales, University of Auckland, Simon Fraser University, University of Virginia, University of Illinois, Dartmouth College, USC, UCLA, Yale University, Tilburg University, London Business School, Wharton, Harvard Business School, and University of Michigan. We are grateful to Ruiming Lin, Pedro Saffi, and Yili Zhang for excellent research assistance. We gratefully acknowledge the contribution of Thomson Financial for providing broker recommendations data, available through the Institutional Brokers Estimate System. Malloy thanks the Economic and Social Research Council for financial support. We are also grateful to many industry insiders for patiently answering our questions. All errors are our own.

Data are the bedrock of empirical research in finance. When there are questions about the accuracy or completeness of a data source, researchers routinely go to great lengths to investigate measurement error, selection bias, or reliability.[1] But what if the very contents of a historical database were to change, in error, over time? Such changes to the historical record would have important implications for empirical research. They could undermine the principle of replicability, which in the absence of controlled experiments is the foundation of empirical research in finance. They could result in over- or underestimates of the magnitudes of empirical effects, leading researchers down blind alleys. And to the extent that financial-market participants use academic research for trading purposes, they could lead to resource misallocation.

Data vendors have little obvious incentive to deliberately change the historical record. However, maintaining large databases of historical records is both costly and technologically demanding, not least in the wake of mergers among data vendors. Given that demand for long time-series of accurate historical financial data (as opposed to real-time information) has traditionally come mainly from academics, who typically pay discounted usage fees,[2] one should not take the integrity of historical data for granted.

In this paper, we demonstrate that the integrity of historical financial data is an important issue for empiricists to consider. On May 22, 2007, and in reaction to an earlier version of this paper, Thomson Financial ("Thomson") began issuing confidential guidance to select clients regarding the integrity of its I/B/E/S historical detail recommendations database.[3] This database contains investment ratings for U.S. listed companies issued by sell-side analysts at most of the brokerage firms active in the U.S. The substance of the guidance, summarized in the Appendix, is that tens of thousands of historical recommendations have inadvertently been added, dropped, or altered, and that the data handling errors that apparently led to these changes have occurred throughout the existence of the database (beginning before 2000 and continuing through the end of 2006). As a result, the actual contents of the recommendations database depend on the precise date when a client downloaded the data. In other words, two clients interested in the same historical time period who obtained the data on different dates would likely have analyzed two quite different sets of data.

We explore the implications of these problems for academic research. The academic literature on analyst stock recommendations, much of which uses I/B/E/S data, is voluminous.[4] Michaely and Womack (2005), in their review of the literature, note that several key topics are each the subject of numerous academic papers. These topics include the compensation, incentives, and biases of analysts; the characteristics of recommended stocks; the investment value of recommendations; and biases and conflicts of interest in the production of recommendations. Given this keen academic interest, as well as the intense scrutiny that research analysts face in the marketplace and from regulators, and the growing popularity of trading strategies based on analyst output, changes to the historical I/B/E/S database are of obvious interest to academics and practitioners alike.

We document that the historical contents of the I/B/E/S recommendations database have been quite unstable over time. Across a sequence of seven nearly annual downloads of the entire I/B/E/S historical recommendations database, obtained between 2000 and 2007, we find that between 1.6% and 21.7% of matched observations are different from one download to the next. For instance, of the 332,145 observations on the 2003 tape, 57,770 (17.4%) are changed in some manner on the 2004 tape. We identify four types of changes which we term alterations, deletions, additions, and anonymizations. For instance, comparing the 2003 tape to the 2004 tape over the period 1993 to 2003, we find 2,411 instances of alterations to a recommendation level (say, turning a “buy” into a “hold”), 3,965 deletions (i.e., records on the 2003 tape that have been deleted from the 2004 tape), 33,335 additions (i.e., records dated 1993 to 2003 that appear on the 2004 tape but not on the 2003 tape), and 18,059 instances where the analyst’s name subsequently went missing from a recommendation. Across all tapes, we find 15,828 alterations, 131,413 deletions, 74,214 additions, and 23,838 anonymizations.

Thomson regard the 2007 tape as purged of the data errors we have identified, except that it continues to include alterations made as a result of broker requests for retrospective changes to their buy/hold/sell recommendation scales. When we undo these retrospective changes to create a true “as-was” 2007 tape, we find that between 10% (on the 2005 tape) and 30% (on the 2004 tape) of all observations are now recorded differently on the 2007 tape. For instance, of the 332,145 records on the 2003 tape, 10,850 appear on the 2007 tape with a corrected recommendation level, 13,892 have been permanently erased from the I/B/E/S historical database, 5,489 records missing from the 2003 tape have been added, and analysts’ names have been reinstated in 6,259 records.

We demonstrate that these changes have a significant and economically important effect on several features of the data that are routinely used by academics and practitioners.

- *Effect on the distribution of recommendations:* Relative to the 2007 tape, recommendations affected by the changes on the 2000, 2001, and 2002 tapes are too optimistic, while those on the 2003, 2004, and 2005 tapes are too pessimistic.
- *Patterns in affected recommendations:* The changes cluster according to three widely used conditioning variables: The analyst’s reputation, the brokerage firm’s size and status, and the boldness of the recommendation. “All-star” analysts and brokerage firms sanctioned under the Global Settlement are overrepresented among affected recommendations on the 2000 and 2001 tapes and underrepresented on later tapes. “Bold” recommendations (those far from consensus) are overrepresented among affected recommendations on all tapes.
- *Effect on trading signals:* Trading signals such as “upgrades” and “downgrades” are the key inputs for a large literature on the economic impact and profitability of analyst research (see Ramnath, Rock, and Shane (2005) for a survey). Depending on the tape, we find that between 2.7% and 23.6% of historic trading signals are reclassified on the 2007 tape.

We illustrate the potential effects these changes can have on research by examining three central tests from the empirical analyst literature: The profitability of trading signals; the profitability of consensus recommendation changes; and the persistence in individual analyst performance. We find that the changes to the I/B/E/S historical record have an economically and statistically significant impact on both calendar-time portfolio returns and three-day event returns to trading signals computed from the different downloads. For example, three-day event returns to upgrades average 3.02% on the 2007 tape but only 2.30% on the 2004 tape (a difference of 72 basis points over three days, and a 31% increase in percentage terms), while three-day event returns to downgrades average -4.72% on the 2007 tape but only -3.79% on the 2004 tape (a difference of 93 basis points, and a 24% decrease). The performance of portfolio strategies based on changes in consensus recommendations (as in Jegadeesh et al. (2004)) shows similar variation across tapes. For instance, we document a temporary boost to the pre-2001 back-testing performance of such strategies on the 2003, 2004, and 2005 tapes relative to the 2002 tape, a boost which then vanishes on the 2007 tape.

The track records of individual analysts are also affected. Analysts' track records are the key variable of interest in several strands of the literature, notably the debate over conflicts of interest[5] in the analyst industry, as well as studies of individual analysts' stock-picking skill. We perform a standard test of persistence in analysts' stock-picking ability on each of our tapes. This test reveals that the 2001 through 2005 I/B/E/S downloads produce inflated estimates of persistence compared to the adjusted 2007 tape.

Taken together, our findings suggest that the pervasive data changes we document in this paper do not simply increase noise; because they have systematic and persistent components, they can and do affect the size of estimated effects. Although we take comfort in the fact that the three tests we examine are generally not overturned directionally across the tapes we examine, the magnitude and significance of the across-tape variation is still disconcerting. Since we did not search over all possible tests using analyst recommendation data, we cannot say to what extent different stylized facts in the literature may or may not be affected by these changes to the historical record. What we can say with certainty is that as a result of our investigation, the quality of post-2006 data downloads will exceed that of any older downloads. Thus, an important lesson for empirical researchers is not to recycle older downloads, even if a fresh download requires substantial investment in routine data cleaning.[6] With regard to "undoing" the broker-requested *retrospective* changes to recommendation scales, we can also report that Thomson is now planning to produce a true "as-was" historical recommendations database in response to our investigation. This should allow future researchers to consistently and accurately replicate any analysis that employs historical analyst recommendations data.

I. Overview of Changes to the I/B/E/S Historical Recommendations Database

A. The Scope of the Problem

Our analysis is based on comparisons of seven snapshots of the entire I/B/E/S U.S. historical detail recommendations database, downloaded at roughly annual intervals between 2000 and 2007. Each snapshot covers the period from the inception of the database (Oct. 29, 1993) to about two months prior to the respective download date. The cutoff dates of our snapshots are 7/20/00 ("2000 tape"), 1/24/02 ("2001 tape"), 7/18/02 ("2002 tape"), 3/20/03 ("2003 tape"), 3/18/04 ("2004 tape"), 12/15/05 ("2005 tape"), and 9/20/07 ("2007 tape"). According to Thomson, the 2007 tape contains data purged of all data errors we have identified, except that it continues to include alterations made as a result of broker requests for retrospective changes to their recommendation scales.

A typical I/B/E/S record includes the analyst's name and her six-digit *amaskcd* identifier as assigned by I/B/E/S; the name of the analyst's employer at the time of the recommendation; the I/B/E/S ticker and historical CUSIP of the company concerned; the date the recommendation was issued; the last date it was considered still in force; and the recommendation itself. Different brokerage firms use different wordings for their recommendations, which I/B/E/S translates into a numerical score on the following scale: Strong buy=1, buy=2, hold=3, sell=4, strong sell=5.

Table I, Panel A examines year-to-year changes to the database by comparing data from adjacent annual downloads, which are merged by standardized brokerage firm code,[7] I/B/E/S ticker, and recommendation date. We focus on the period for which each pair of downloads has overlapping coverage (that is, we ignore recommendations from the later tape dated after the cut-off date of the earlier tape.) Thus, we ask if two researchers, looking at the same time period but working with data obtained on slightly different dates, would face materially different data.

INSERT TABLE I ABOUT HERE

Panel A reveals a disturbingly high incidence of ex post changes to the I/B/E/S recommendations data. Across our sequence of tapes, 10.8%, 8.4%, 13.1%, 17.4%, 21.7%, and 1.6% of observations are changed by our next download date. For instance, of the 450,225 observations on the 2004 tape, 97,582 (21.7%) look different on the 2005 tape. This indicates that the historical contents of the I/B/E/S recommendations database have been quite unstable over time. Only since about Dec. 2005 has the database been relatively stable, with only 6,580 historic observations (1.6%) being changed by Sept. 2007.

Panel A also provides a breakdown of the following four types of ex post changes:

1) Alterations: A broker/ticker/date triad that appears on both tapes but for which the recommendation on one tape is different than on the next tape.

2) Deletions: A broker/ticker/date triad that appears on the earlier tape but not on the later tape.

3) Additions: A broker/ticker/date triad that appears on the later tape but not on the earlier tape.

4) Anonymizations: Cases where the analyst associated with a broker/ticker/date triad is

identified by name on the earlier tape but is anonymous on the later tape.

The number of alterations varies from 121 (between the 2005 and 2007 tapes) to 8,973 (between the 2002 and 2003 tapes). Deletions run in the thousands for every pairwise comparison, peaking in 2005 when 92,244 records – 20.5% of the 450,225 records on the 2004 tape – were deleted. Additions also run in the thousands, peaking at 33,335 between 2003 and 2004. Finally, anonymizations are concentrated between 2002 and 2004: Between 2002 and 2003, 5,000 records were anonymized, followed by a further 18,059 anonymizations between 2003 and 2004.

The evidence in Panel A suggests that two researchers downloading I/B/E/S recommendations a few months apart could face materially different data. However, it does not speak to the question how inaccurate these data might be. Answering that question requires that we compare each download to the "truth". To the extent that the 2007 tape corrects errors arising from accidental deletions and anonymizations, Thomson considers it the most historically accurate record of analyst recommendations. However, the 2007 tape still contains broker-requested *retrospective* changes to recommendation scales, so we reverse these alterations to get back to original, historical data.[8] We refer to this as the "adjusted 2007 tape." In Panel B, we compare each tape to the adjusted 2007 tape to illustrate the extent to which the six earlier tapes were contaminated by data problems.

Panel B points to extensive data problems in each of the earlier tapes. Between 10.0% and 30.0% of observations on the respective tapes have been corrected on the adjusted 2007 tape. For instance, of the 450,225 records on the 2004 tape, 12,682 appear on the adjusted 2007 tape with a different recommendation level (either because Thomson corrected data errors or more often because we undid retrospective rating scale changes), 96,077 are no longer included in the I/B/E/S historical database as of 2007, and 4,381 records that should have been on the 2004 tape (but were not) have been added on the 2007 tape. In addition, 21,902 records that were anonymous on the 2004 tape identify the analyst by name on the 2007 tape.[9]

It is worth noting that the I/B/E/S recommendations database appears to have had the most data problems precisely around the time (namely in 2001 and 2004) when academic interest in analyst recommendations increased in the wake of first Regulation FD and then the Global Settlement.

B. Net Effect of Changes on the Distribution of Recommendations

Table I illustrates that the I/B/E/S recommendations history has changed extensively throughout its existence. We now investigate whether these changes merely add noise to standard empirical tests or whether they are liable to create biases. Under the null that the changes are pure noise, we expect that they leave the recommendation levels of affected records unchanged on average.

Table II suggests that the changes to the I/B/E/S recommendations database have non-random components, both year-to-year (Panel A) and relative to the adjusted 2007 tape (Panel B). In four of the pairwise comparisons shown in Panel A (2000 vs. 2001, 2002 vs. 2003, 2003 vs. 2004, and 2005 vs. 2007), the net effect of the changes is to make the recommendations history look less optimistic. For instance, the average recommendation on the 2002 tape is 2.11 (a little below a “buy” recommendation). The 36,762 records subject to an ex post change have an average recommendation of 1.98 on the 2002 tape. On the 2003 tape, their average is significantly more pessimistic (mean: 2.28), largely because the 2003 deletions are unusually optimistic (mean: 1.63) while the 2003 additions are unusually pessimistic (mean: 2.45). In the two remaining pairwise comparisons (2001 vs. 2002 and 2004 vs. 2005), the net effect of the changes is to make the recommendations history look more optimistic.

INSERT TABLE II ABOUT HERE

Relative to the adjusted 2007 tape, which we regard as more historically accurate, changed recommendations on the first three tapes are too optimistic (i.e., the effect of the corrections on the

2007 tape is to lower the average of these recommendations) while those on the last three tapes are too pessimistic. As we will show in Section II, these apparently systematic patterns in changed recommendations have a direct impact on standard empirical tests.

C. Patterns in Affected Recommendations

In addition to being either systematically optimistic or pessimistic, recommendations affected by the changes to the I/B/E/S recommendations history appear to cluster according to three popular conditioning variables: The analyst's reputation, the brokerage firm's size and status, and the boldness of the recommendation. We measure analyst reputation using all-star status, as designated in the Oct. issue of *Institutional Investor* magazine preceding the recommendation in question. We divide brokerage firms into the 12 (generally large) firms sanctioned under the Global Settlement and all other firms. And we code a recommendation as bold if it was one notch or more above or below consensus (=mean recommendation) computed over the prior three months (requiring at least three outstanding recommendations).

In Table A1 available in an Internet Appendix on www.afajof.org, we compare the frequency of these conditioning variables in the universe of historical recommendations and in the set of changed recommendations. We compare each tape to the next tape as well as to the adjusted 2007 tape.

We find that all-stars are significantly overrepresented among changed recommendations on the 2000 and 2001 tapes, while changed recommendations on the 2002 through 2004 tapes disproportionately come from unrated analysts. Relative to the adjusted 2007 tape, recommendations by unrated analysts are significantly more likely to need correction on every tape except the 2001 tape. Thus, tests comparing all-stars to unrated analysts may yield different results depending on which tape is used. Sanctioned banks are overrepresented among affected recommendations on the 2000 and 2001 tapes, and underrepresented for all later tapes. Relative to the adjusted 2007 tape, sanctioned banks are associated with a significantly lower need for corrections on every tape except the 2001 tape. Finally, bold recommendations are significantly overrepresented among affected records on all tapes. They are also consistently and significantly more likely to be subject to corrections on the adjusted 2007 tape.

II. Impact on Typical Analyses of Stock Recommendations

In this section, we document the potential effects of the I/B/E/S changes for academic research, while bearing in mind that they may also affect the work of regulators, legislators, litigators, and investment professionals, who may also rely on archival databases such as I/B/E/S. We focus on three central findings of the analyst literature: The profitability of trading signals; the profitability of changes in consensus recommendations; and the persistence in individual analyst performance. We stress that we did *not* search over every possible result that might be impacted by the data changes, nor did we necessarily pick the results or the specifications that were most likely to be affected. Our goal was simply to assess if, and by how much, the changes to the historical record that we document might affect key stylized facts in the empirical analyst literature.

A. Effects on Trading Signal Classifications

Besides changing the distribution of recommendation levels, the alterations, deletions, and additions also affect recommendation changes or “trading signals”, the key inputs for a large literature on the profitability of analyst recommendations (see Ramnath, Rock, and Shane (2005) for a review). For each broker/ticker pair, we code trading signals as follows. The first time a broker recommends a stock is an initiation. Subsequent recommendations represent either upgrades, downgrades, or reiterations, as long as no more than 12 months have elapsed since the previous recommendation.[10] Otherwise, they are coded as re-initiations. We also use the I/B/E/S stop file to check for suspensions of broker coverage and broker scale changes, and code resumptions of coverage as re-initiations.[11]

Table III provides a breakdown, for each tape, of the distributions of all trading signals and of those that are affected by the changes to the I/B/E/S database. For instance, of the 222,694 trading signals on the 2000 tape shown in Panel A, 18,737 (31,802 changes less 13,065 additions) are subject to corrections according to the adjusted 2007 tape. When we add the 13,065 additions, we find that 14.3% of the trading signals are different on the 2007 tape than on the 2000 tape, for the exact same time period. The breakdown by type of trading signal shows that 8.9% of the downgrades on the 2000 tape are coded differently on the adjusted 2007 tape, as are 9.4% of upgrades, 23.3% of reiterations, 7% of initiations, and 4.6% of re-iterations.

INSERT TABLE III ABOUT HERE

The right-hand side of Table III provides a transition matrix for the changed trading signals from the earlier tape to the 2007 tape. For instance, 522 recommendations classified as reiterations on the 2000 tape have become downgrades on the 2007 tape, 143 downgrades have become upgrades, and 275 upgrades have become reiterations.

Panels B through F repeat these analyses for the 2001 through 2005 tapes. In each case, a large fraction of trading signals change, ranging from 2.7% on the 2005 tape to 23.6% on the 2004 tape.

B. Effects on Returns to Trading on Upgrades and Downgrades

What is the likely effect of these changes to historic trading signals on backtests of the profitability of strategies that condition on upgrades and downgrades? For brevity, we focus on the 2004 and adjusted 2007 tapes. This is sufficient to illustrate our main point. In unreported tests, we find large and significant differences across a variety of additional pairwise comparisons.

For each tape, we form two portfolios: (1) An upgrade portfolio, consisting of all stocks that at least one analyst upgraded on a given date (e.g., from a buy to a strong buy); and (2) a downgrade portfolio, comprised of all stocks that at least one analyst downgraded on a given date (e.g., from a buy to a hold).[12] Portfolio construction closely follows Barber, Lehavy, and Trueman (2007) and Barber et al. (2006). In the upgrade portfolio, for example, a recommended stock enters the portfolio at the close of trading on the day the recommendation is announced. This explicitly excludes the announcement-day return, on the assumption that many investors likely learn of recommendation changes only with a delay. Each recommended stock remains in the portfolio for the lesser of two weeks or until the stock is downgraded or dropped from coverage by the analyst.[13] If more than one analyst changes a recommendation on a particular stock on a given date, the stock will appear multiple times in the portfolio on that date (once for each recommendation change).

We then compute daily calendar-time buy-and-hold portfolio returns for each tape for the period over which the tapes overlap (that is, Oct. 29, 1993 to Mar. 18, 2004). Assuming an equal dollar investment in each stock, the portfolio return on date *t* is given by $\sum_{i=1}^{n_t} R_{it} x_{it} / \sum_{i=1}^{n_t} x_{it}$, where R_{it} is the date *t* return on stock *i*, n_t is the number of stocks in the portfolio, and x_{it} is the compounded daily return of stock *i* from the close of trading on the day of the recommendation change through day *t*-1. (For a stock recommended on day *t*-1, x_{it} = 1.)

INSERT TABLE IV ABOUT HERE

Panel A of Table IV reports the results for the upgrade portfolio (columns (1)-(3)) and for the downgrade portfolio (columns (4)-(6)). *Ret07* and *Ret04* are the average daily calendar-time portfolio returns (in percent) on the 2007 and 2004 tapes, respectively, and *Diffret* is the average daily return difference between the 2007 and 2004 tapes. We also compute abnormal portfolio returns (*DiffXret)* by estimating “four-factor” alphas (Carhart (1997)), which equal the intercept from a regression of *Diffret* less the risk-free rate on the daily excess return of the market over the risk-free rate (*MKT)* and the return difference between small and large-capitalization stocks (*SMB*), high and low book-to-market stocks (*HML*), and high and low price-momentum stocks (*UMD*).

Column (1) indicates that over the full period of overlap (Oct. 29, 1993 to Mar. 18, 2004), upgrades on the adjusted 2007 tape earn 16.1 basis points per day on average, while upgrades on the 2004 tape earn only 14.8 basis points per day. The average daily abnormal return difference (*DiffXret*) between the 2004 and 2007 upgrade samples is 1.3 basis points per day (3.3% annualized). When we split the sample period on Mar. 10, 2000, the day of the Nasdaq peak, we find a substantially larger abnormal return difference of 3.6 basis points per day (9.1% annualized) in the post-"bubble" period (column (2)), and no significant difference in performance prior to Mar. 10, 2000 (column (3)). Thus, the changes to the I/B/E/S 2004 historical record appear to have a disproportionate effect on research that focuses on more recent periods.

Results for downgrades are similar. Downgrades earn -9.5 basis points per day on the adjusted 2007 tape, but only -7.8 basis points on the 2004 tape. The average difference, *DiffXret*, is 1.6 basis points per day (4% annualized) for the whole period and 4 basis points per day (10.1% annualized) for the post-bubble period. As with the upgrade tests, each of these results is highly statistically significant. Prior to Mar. 10, 2000, there is again no significant difference in performance.

Overall, these calendar-time portfolio results indicate that back-tests done using the 2004 data instead of the historically more accurate 2007 data would significantly understate the profitability of trading on both upgrades and downgrades, especially in the period following the bubble.

We next compare the market reaction to upgrades and downgrades across tapes. To do so, we compute three-day raw event return (equal to the geometrically cumulated return for the day before, day of, and day after the recommendation change) and three-day excess returns (equal to the raw stock return less the appropriate size-decile return of the CRSP NYSE/AMEX/NASDAQ index). Panel B of Table IV reports the results for the full sample of upgrades (in the column entitled "All upgrades”) as well as for individual upgrade categories (e.g., “2to1” refers to an upgrade from a buy to a strong buy, while “5to4” refers to an upgrade from a strong sell to a sell). We use the entire period over which the 2004 and adjusted 2007 tapes overlap (i.e., Oct. 29, 1993 to Mar. 18, 2004). For all upgrades, raw three-day event returns average 3.02% on the 2007 tape but only 2.30% on the 2004 tape. *DiffEret*, the average difference in raw event returns between the two tapes, is 72 basis points over the three days (a 31% increase in percentage terms from the 2004 tape to the 2007 tape), while *DiffEXret*, the average difference in excess event returns between the two tapes, is also 72 basis points per day. In addition, we find large and statistically significant differences between the tapes for several of the individual upgrade categories (e.g., “2to1”, “3to2”, “4to2”, and “4to3”).

Panel C shows that the differences across the downgrade samples are equally striking. Three-day event returns on the 2004 tape are -3.79%, versus -4.72% on the adjusted 2007 tape. *DiffEret*, the difference in three-day returns between the two tapes, equals -93 basis points, a 24% decrease in percentage terms from the 2004 tape to the 2007 tape; *DiffEXret* too is large at -89 basis points and statistically different from zero. Several of the individual downgrade categories show large differences between the two tapes (e.g., “2to4”, “3to4”, and “3to5” are each associated with differences in excess of 200 basis points over three days).

C. Effects on Returns to Consensus Recommendations

Another commonly used feature of analyst data is the consensus analyst recommendation for a particular firm. Consensus recommendations are frequently employed in quantitative trading strategies, following evidence that sorting based on consensus recommendations (Barber et al. (2001, 2003)), and particularly on *changes* in consensus recommendations (Jegadeesh et al. (2004)), is a profitable strategy. How do the changes to the I/B/E/S database affect such a strategy?

We employ a standard portfolio classification technique that each day sorts firms into quintiles based on the lagged change in consensus recommendations on the previous day. For this purpose, recommendations are reverse-scored from 5 (strong buy) to 1 (strong sell). The consensus recommendation for a ticker equals the mean outstanding recommendation at the end of a day (based on a minimum of three recommendations).

Table V reports daily portfolio returns for a trading strategy ("spread") that buys stocks in the highest change quintile (Q5) and shorts stocks in the lowest change quintile (Q1). We calculate abnormal portfolio returns by computing daily characteristic-adjusted returns constructed as in Daniel et al. (1997) [henceforth DGTW].[14] DGTW returns are raw returns minus the returns on a value-weighted portfolio of all CRSP firms in the same size, industry-adjusted market-book, and one-year momentum quintiles. The strategy is performed separately (and identically) on the 2002, 2003, 2004, 2005, and adjusted 2007 tapes, and differences across tapes are reported. For ease of comparison with the earlier literature on consensus recommendations, much of which focuses on the period through Dec. 2000, we split the sample in half. Results for the pre-2001 period are in columns (1)-(3) and those for the post-2001 period are in columns (5) to (8).[15]

INSERT TABLE V ABOUT HERE

While the strategy is profitable in the pre-2001 period, according to each data download, it performs significantly *better* on the 2003, 2004, and 2005 tapes than on the 2002 or 2007 tapes, even though we back-test the strategy over the *exact* same time period. The magnitude of these differences is nontrivial, ranging from 1.9 to 2.1 basis points per day (4.8% to 5.3% annualized; see column (4)).[16] This means that the 2003, 2004, and 2005 tapes overstate the profitability of this strategy by 7.1% to 7.8% relative to the performance found on the 2007 tape.

In columns (5) to (8), each tape is compared individually to the adjusted 2007 tape from Jan. 1, 2001 to the cut-off date of the tape in question. Thus, the spread estimates for the 2007 tape shown in column (7) differ depending on the exact period covered by the tape in question. The results suggest that trading on consensus changes continues to produce significant abnormal returns in the post-2001 time period across the various tapes. And while the spread estimates for the 2003, 2004, and 2005 tapes are not significantly different from the 2007 comparison tape, the 2002 spread estimate now is: Trading on consensus changes yielded 6.2 basis points more per day according to the 2002 tape than according to the adjusted 2007 tape (15.6% annualized). This translates into a percentage improvement of 17.3% relative to the performance found on the 2007 tape.

Table V thus reveals a temporary boost to the pre-2001 back-testing performance of the consensus change trading strategy on the 2003, 2004, and 2005 tapes relative to the 2002 tape, a boost that then vanishes on our corrected version of the 2007 tape. By contrast, after 2001, it is the 2002 tape that yields significantly different estimates from the 2007 tape.

D. Effects on Persistence in Analysts' Stock-Picking Ability

Each of the four types of changes to the I/B/E/S database can alter an individual analyst's track record. Several strands of the labor economics, finance, and accounting literatures rely on analyst track records in their empirical tests, and hence are potentially affected by the data changes we document: Studies of analyst career concerns (e.g., Hong, Kubik, and Solomon (2000)), conflicts of interest in the brokerage industry (e.g., Michaely and Womack (1999), Lin and McNichols (1998), Hong and Kubik (2003)), and persistence in individual analysts' stock-picking ability (e.g., Mikhail, Walther, and Willis (2004), Li (2005)).

In this section, we investigate the impact of the data changes on estimates of stock-picking persistence. We perform a standard test (similar to Mikhail, Walther, and Willis (2004)) on each tape. Analysts are grouped into quintiles at the beginning of each half-year period based on the average five-day excess return of their recommendation upgrades and downgrades over the prior half-year period.[17] The excess return is the geometrically cumulated DGTW characteristic-adjusted return for the two days before through the two days after the recommendation change; DGTW returns are constructed as in the previous section. The "persistence spread" equals the difference between the average five-day DGTW-adjusted return of the highest quintile minus the average five-day DGTW-adjusted return of the lowest quintile. The persistence spread measures the extent to which good past performers continue to perform well in the future.

Column (1) of Table VI reports average persistence spreads, where each average is computed over the full available sample period for each tape. Each tape is compared individually to the adjusted 2007 tape; therefore, the estimates for the 2007 tape shown in column (2) differ across the 2000 through 2005 tapes depending on the exact sample period covered by the tape in question. Pairwise differences in persistence spreads compared to the adjusted 2007 tape are reported in column (3).

INSERT TABLE VI ABOUT HERE

Consistent with the findings in Mikhail, Walther, and Willis (2004), column (1) indicates persistence in individual analysts' stock-picking performance in each download, with average five-day persistence spreads of at least 240 basis points across the 2000 through 2005 tapes. However, the magnitude of this spread varies markedly across tapes, and the 2007 tape shows smaller persistence spreads than each of the other tapes (except for the 2000 tape). Column (3) shows that three of the six pairwise comparisons to the 2007 tape yield significant differences in persistence spreads. For example, the difference between the 2001 and 2007 tape is 38.6 basis points, an increase of 15.0% relative to the amount of persistence found on the 2007 tape. Similarly significant differences exist between the 2002 and 2007 tapes (52.3 basis points, a 20.5% increase relative to 2007) and between the 2004 and 2007 tapes (18.4 basis points, a 7.5% increase relative to 2007).

In Table A2 available in an Internet Appendix on www.afajof.org, we show that this result is even more pronounced if we filter on analysts' all-star status (defined as in Section I.C). A common modification to the persistence trading strategy is to buy on recommendations by all-star analysts who are also in quintile 5 and to sell on recommendations by non-all star analysts ranked in quintile 1. This assumes asymmetry in persistence among all-stars: They are likely to repeat good past performance but not poor past performance. Imposing this screen increases the differences in persistence spreads across the tapes. For example, we find a difference between the 2001 and 2007 tapes of 82.0 basis points over five trading days, an increase of 25.3% relative to the amount of persistence found on the 2007 tape. Similarly large differences exist between the 2002 and 2007 tapes (66.3 basis points, a 21.1% increase relative to 2007) and between the 2003 and 2007 tapes (36.6 basis points, a 12.1% increase relative to 2007).

Taken together, our findings suggest that while we continue to find evidence of persistence in analyst performance using the historically more accurate 2007 data, the magnitude of such persistence is substantially lower than if one were to use prior contaminated versions of I/B/E/S.

III. Conclusions

We document widespread ex post changes to the historical contents of the I/B/E/S analyst stock recommendations database. Across a sequence of seven nearly annual downloads of the entire recommendations database, obtained between 2000 and 2007, we find that between 1.6% and 21.7% of matched observations are different from one download to the next. When we use a cleaned-up version of the 2007 tape as a point of comparison, we find that between 10% and 30% of all observations on the earlier tapes are now recorded differently on the 2007 tape.

These changes appear non-random and have a large and significant impact on several features of the data that are routinely used by academics and practitioners. They cluster according to three popular conditioning variables: Analyst reputation, broker status, and boldness. The changes also have systematically optimistic and pessimistic patterns that vary across time and that affect the classification of trading signals. We demonstrate the potential effects these changes have on academic research by examining three central tests from the empirical analyst literature: The profitability of trading signals; the profitability of changes in consensus recommendations; and the persistence in individual analyst performance. In each case, despite examining identical sample periods, we find economically and statistically significant differences in estimated effects across our various downloads.

While most finance empiricists are accustomed to dealing with data issues like selection bias or measurement error, they seldom question the very constancy and veracity of historical data. Given the conflicting incentives of data providers, and the technological demands of handling vast (and increasing) amounts of historical data, however, this tendency may be problematic. Our results demonstrate that the integrity of historical data is an important issue for empiricists to consider.

Appendix: What Happened?

Deletions and Additions

Most additions and deletions are apparently symptoms of a systematic process error that has affected the database throughout its entire existence, until Thomson fixed the process, in response to our enquiries, in the spring of 2007.

The error concerns the broker recommendation translation table which maps each broker's recommendation scale onto the familiar five-point I/B/E/S scale. Recommendations enter the database by broker, ticker, and recommendation only (for example, "ABC, MSFT, market perform"). This information is then matched up by broker to a broker translation table, in which ABC's recommendation of "market perform" is translated as I/B/E/S recommendation level 3. Thomson contends that its data entry clerks occasionally overwrote existing entries in the translation table when faced with variations or changes in wording of the broker's recommendation. For example, if ABC changes its "market perform" recommendations to "mkt. performer", a clerk may overwrite broker ABC's "market perform" entry when adding the "mkt. performer" entry to the table. As a result, the next time the historical recommendations database is created for export to clients, the translation table will fail to translate any of ABC's historic "market perform" recommendations. From a client's point of view, these records will appear to have been deleted. Additions occur when another data entry clerk, by chance or because he has noticed the missing recommendations, at some later point adds the "market perform" entry back into the broker translation table.

Thus, an entire level of a broker's historic recommendations (e.g., every "sell") can go missing for some time and then reappear. In this sense, additions are reversals of past deletions. To illustrate, in Sept. 2001, I/B/E/S lost all 1,716 historic "market perform" recommendations of a particular broker. They were restored in a Nov. 2002 cleanup when Thomson noticed that thousands of recommendations were missing. Subscribers were apparently not notified. However, the Nov. 2002 cleanup did not address the cause of the deletions, which only came to light in the spring of 2007, as a result of our investigation. Thus, the database continued to experience deletions and additions until recently.

Besides problems with the broker translation table, most remaining additions and deletions between 2003 and 2005 were caused by the erroneous inclusion of recommendations issued by eight quantitative research groups.[18] According to Thomson, these recommendations were not supposed to be viewable by its clients yet became part of the database some time between 2003 and 2004. They were subsequently permanently removed at some point between 2004 and 2005.[19]

Anonymizations

Thomson's database stores recommendations by broker and not by analyst. To add the analyst's identity, Thomson combines data from the recommendations database with data from the coverage table that records which analyst covers which tickers at which broker between which dates.

During 2003, Thomson undertook a major review of the coverage table in an effort to reconcile the I/B/E/S and First Call databases and to remove invalid coverage assignments. In the process, the start and end dates of various analyst/broker/ticker triads were changed. This apparently resulted in some historic recommendations no longer being associated with an analyst and hence being "anonymized." Separately, Thomson attempted to consolidate instances of multiple analyst codes for a given analyst but in the process removed the entire coverage history for some analysts.

In response to an earlier version of this paper, in December 2006, Thomson changed the file generation process such that anonymizations should not occur in the future.

Alterations

Brokerage firms often tweak their rating scales. To illustrate, in the wake of the Global Settlement, many firms moved from a five- or four-point scale to a simpler three-point scale (say, buy/hold/sell). When brokers adopt new rating scales, they sometimes request that Thomson restate, *retroactively*, their entire history of recommendations in an effort to make past and future recommendations appear on the same scale. According to Thomson, the vast majority of alterations result from such requests. The remainder are the result of errors made by Thomson in effecting these requests.[20] From a research point of view, retrospective ratings changes are problematic, as the recommendation recorded in the database no longer matches the recommendation market participants had access to at the time.

References

Barber, Brad, Reuven Lehavy, and Brett Trueman, 2007, Comparing the stock recommendation performance of investment banks and independent research firms, *Journal of Financial Economics* 85, 490-517.

Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman, 2001, Can investors profit from the prophets? Security analyst recommendations and stock returns, *Journal of Finance* 56, 531-564.

Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman, 2003, Prophets and losses: Reassessing the returns to analysts' stock recommendations, *Financial Analysts Journal* 59, 88-96.

Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman, 2006, Buys, holds, and sells: The distribution of investment banks' stock ratings and the implications for the profitability of analysts' recommendations, *Journal of Accounting and Economics* 41, 87-117.

Bennin, Robert, 1980, Error rates in CRSP and COMPUSTAT: A second look, *Journal of Finance* 35, 1267-1271.

Canina, Linda, Roni Michaely, Richard Thaler, and Kent Womack, 1998, Caveat compounder: A warning about using the daily CRSP equal-weighted index to compute long-run excess returns, *Journal of Finance* 53, 403-416.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.

Cowen, Amanda, Boris Groysberg, and Paul M. Healy, 2006, Which types of analyst firms make more optimistic forecasts? *Journal of Accounting and Economics* 41, 119-146.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristics-based benchmarks, *Journal of Finance* 52, 1035-1058.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2001, A first look at the accuracy of the CRSP Mutual Fund Database and a comparison of the CRSP and Morningstar Mutual Fund Databases, *Journal of Finance* 56, 2415-2430.

Hong, Harrison, and Jeffrey D. Kubik, 2003, Analyzing the analysts: Career concerns and biased forecasts, *Journal of Finance* 58, 313-351.

Hong, Harrison, Jeffrey D. Kubik, and Amit Solomon, 2000, Security analysts' career concerns and herding of earnings forecasts, *RAND Journal of Economics* 31, 121-144.

Jegadeesh, Narasimhan, Joonghyuk Kim, Susan D. Krische, and Charles Lee, 2004, Analyzing the analysts: When do recommendations add value?, *Journal of Finance* 59, 1083-1124.

Li, Xi, 2005, The persistence of relative performance in stock recommendations of sell-side financial analysts, Journal of Accounting and Economics 40, 129-152.

Lin, Hsiou-wei, and Maureen F. McNichols, 1998, Underwriting relationships, analysts' earnings forecasts and investment recommendations, *Journal of Accounting and Economics* 25, 101-127.

Michaely, Roni, and Kent L. Womack, 1999, Conflict of interest and the credibility of underwriter analyst recommendations, *Review of Financial Studies* 12, 653-686.

Michaely, Roni, and Kent L. Womack, 2005, Market efficiency and biases in brokerage recommendations, book chapter published in *Advances in Behavioral Finance II*, edited by Richard Thaler.

Mikhail, Michael B., Beverly R. Walther, and Richard H. Willis, 2004, Do security analysts exhibit persistent differences in stock picking ability? *Journal of Financial Economics* 74, 6-91.

Ramnath, Sundaresh, Steve, Rock, and Philip Shane, 2005, A review of research related to financial analysts’ forecasts and stock recommendations, mimeo, Georgetown University.

Rosenberg, Barr and Michel Houglet, 1974, Error rates in CRSP and Compustat data bases and their implications, *Journal of Finance* 29, 1303-1310.

Shumway, Tyler, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327-340.

Shumway, Tyler and Vincent A. Warther, 1999, The delisting bias in CRSP’s Nasdaq data and its implications for interpretation of the size effect, *Journal of Finance* 54, 2361-2379.

Endnotes

[1] See, for instance, Rosenberg and Houglet (1974), Bennin (1980), Shumway (1997), Canina et al. (1998), Shumway and Warther (1999), and Elton, Gruber, and Blake (2001). See http://www.kellogg.northwestern.edu/rc/crsp-cstat-references.htm for a summary of academic work on problems with financial databases.

[2] The recent rise in popularity of quantitative investment strategies may have increased demand for historical data.

[3] The guidance is available only to clients, only on request, and only upon signing of a non-disclosure agreement. Thomson have shared their findings with us, and we are not bound by any non-disclosure agreement, though we are unable to quote verbatim from Thomson's report. Interested readers who are Thomson clients are advised to obtain the report directly from Thomson.

[4] As of September 4, 2008, Google Scholar identifies 1,110 articles and working papers using the keywords "I/B/E/S", "analysts", and "recommendations."

[5] See, for example, Michaely and Womack (1999), Lin and McNichols (1998), and Hong and Kubik (2003), among others. As of September 4, 2008, Google Scholar lists 285 articles and working papers containing the key words "analysts", "conflicts of interest", and "I/B/E/S".

[6] For example, I/B/E/S periodically changes its historical broker (bmaskcd) and analyst (amaskcd) codes, so programs that adjust for broker mergers or that track analysts across brokers typically need updating after every fresh download.

[7] In some cases, I/B/E/S uses multiple codes to identify the same brokerage firm (e.g., NOMURA and NOMURAUS both decode to Nomura Securities). We standardize such name variations before merging the downloads.

[8] This adjusted version of the 2007 tape corresponds to the “as-was” historical recommendations database which Thomson intends to make available to researchers in response to our investigation.

[9] The 2007 tape reverses not only all the 23,828 anonymizations shown in Panel A, but also adds analyst names for 28,199 broker/ticker/date triads that originally appeared without names on the earlier tapes. While welcome, such “de-anonymizations” may affect the replicability of tests that rely on tracking analysts (e.g., models of career concerns).

[10] We use the I/B/E/S field “revdats” to check whether the previous recommendation continues to be in effect.

[11] When a scale change occurs, Thomson places a stop on the broker’s outstanding recommendations. After a day or so, recommendations are re-started at the new scale level in the detail recommendations file. Thus, in Table III we code the first recommendation after a scale change as a re-initiation.

[12] We have experimented with other portfolio classifications (such as including initiations at buy or strong buy in the upgrade portfolio and including initiations at hold, sell, or strong sell in the downgrade portfolio) with similar results.

[13] The choice of a two-week cutoff point is arbitrary but not selective. We have experimented with a variety of holding periods, from three trading days up to one calendar year, and the differences across tapes vary significantly across holding periods, further highlighting our main insight. These results, excluded for brevity, are available on request.

[14] We obtain similar results when we estimate abnormal returns relative to a four-factor model constructed as in Section II.B (available on request).

[15] We drop the 2000 tape from this analysis as it ends before the end of 2000 and so covers a shorter time period than the other tapes. Similarly, we drop the 2001 tape for lack of sufficient data in the post-2001 time period.

[16] Using a monthly rebalancing rule yields similar results on the differences across tapes (available on request). Note that by using daily rebalancing, our estimates of the consensus spread itself are quite large since they ignore the large transactions costs that such a strategy would entail. Our focus, however, is on the *differences* across tapes, and these differences are significant for a variety of different rebalancing rules.

[17] In unreported tests we find that using quarterly or annual (rather than semi-annual) windows to measure the past performance of individual analysts yields similar results (available on request).

[18] Note that the quantitative research groups produce algorithmic recommendations constrained to be symmetrically distributed. Thus, tests that include these data points will face lower average recommendation levels.

[19] In addition, some records were permanently deleted between 2000 and 2007 at the request of brokerage firms that no longer wished their data to be available through I/B/E/S. In such instances, Thomson issues a notification to its clients. Since the 2007 tape is purged of prior errors, most of the deletions on the 2007 tape relative to earlier tape comparisons represent broker removals. An exception is 2004, a year in which there were erroneous *additions* that are also deleted on the 2007 tape.

[20] Thomson estimates that approximately 20% of the alterations that occurred between 2002 and 2004 are due to errors it made in restating broker recommendations retroactively.

Table I. Overview of Changes to the I/B/E/S Recommendations History.

The table documents the extent, types, and time profile of changes to the I/B/E/S historical recommendations database. In Panel A, we examine year-to-year changes to the database by comparing data from adjacent annual downloads. We focus on the period for which each pair of downloads has overlapping coverage (that is, we ignore recommendations from the later tape that are dated after the cut-off date of the earlier tape.) The cutoff dates of our tapes are 7/20/00 ("2000 tape"), 1/24/02 ("2001 tape"), 7/18/02 ("2002 tape"), 3/20/03 ("2003 tape"), 3/18/04 ("2004 tape"), 12/15/05 ("2005 tape"), and 9/20/07 ("2007 tape"). According to Thomson, the 2007 tape contains data purged of all data errors we have identified, except that it continues to include broker-requested, retrospective changes to recommendation scales. In Panel B, we compare the 2000 through 2005 tapes to the 2007 tape, after reversing the broker-requested, retrospective changes to recommendation scales. This adjusted version of the 2007 tape corresponds to the "as-was" historical recommendations database which Thomson intends to make available to researchers in response to our investigation. The comparisons in Panel B therefore show the extent to which the earlier tapes were contaminated by data errors compared to the most accurate available historic record. We define an *alteration* as a broker/ticker/date triad that appears on both tapes but for which the recommendation on one tape is different than on the other tape. A *deletion* is a broker/ticker/date triad that appears on the earlier tape but not on the later tape to which it is compared. An *addition* is a broker/ticker/date triad that appears on the later comparison tape but not on the earlier tape. In Panel A, *anonymizations* refer to cases where the analyst associated with a broker/ticker/date triad is identified by name on the earlier tape but is anonymous on the later tape. In Panel B, *de-anonymizations* refer to cases where the analyst associated with a broker/ticker/date triad is identified by name on the 2007 tape but is anonymous on the earlier tape. We make this switch because as of Sept. 2007, Thomson has reversed not only the anonymizations shown in Panel A but has also added analyst names for 28,199 broker/ticker/date triads that originally appeared without names on the earlier tapes.

Panel A: Breakdown of types of change in adjacent annual downloads

Comparison tapes	No. of obs. on earlier tape	All ex post changes		Alterations		Deletions		Additions		Anonymizations	
		No.	%	No.	%	No.	%	No.	%	No.	%
2000 vs. 2001	222,694	24,116	10.8%	2,241	1.0%	13,049	5.9%	8,647	3.9%	179	0.1%
2001 vs. 2002	266,619	22,473	8.4%	493	0.2%	13,302	5.0%	8,661	3.2%	17	0.0%
2002 vs. 2003	280,567	36,762	13.1%	8,973	3.2%	4,318	1.5%	18,471	6.6%	5,000	1.8%
2003 vs. 2004	332,145	57,770	17.4%	2,411	0.7%	3,965	1.2%	33,335	10.0%	18,059	5.4%
2004 vs. 2005	450,225	97,582	21.7%	1,589	0.4%	92,244	20.5%	3,208	0.7%	541	0.1%
2005 vs. 2007	414,881	6,580	1.6%	121	0.0%	4,535	1.1%	1,892	0.5%	32	0.0%

Panel B: Breakdown of types of change relative to adjusted 2007 tape

Comparison tapes	No. of obs. on earlier tape	All ex post changes		Alterations		Deletions		Additions		De-anonymizations	
		No.	%	No.	%	No.	%	No.	%	No.	%
2000 vs. 2007	222,694	29,101	13.1%	1,531	0.7%	14,281	6.4%	13,065	5.9%	224	0.1%
2001 vs. 2007	266,619	46,217	17.3%	2,178	0.8%	19,819	7.4%	23,714	8.9%	506	0.2%
2002 vs. 2007	280,567	33,982	12.1%	2,265	0.8%	11,395	4.1%	19,756	7.0%	566	0.2%
2003 vs. 2007	332,145	36,490	11.0%	10,850	3.3%	13,892	4.2%	5,489	1.7%	6,259	1.9%
2004 vs. 2007	450,225	135,042	30.0%	12,682	2.8%	96,077	21.3%	4,381	1.0%	21,902	4.9%
2005 vs. 2007	414,881	41,516	10.0%	12,522	3.0%	4,535	1.1%	1,889	0.5%	22,570	5.4%

Table II. Mean Recommendation Levels by Type of Change.

The table reports mean recommendation levels among changed recommendations. In Panel A, changes are defined by reference to the next available tape. In Panel B, changes are defined by reference to the adjusted 2007 tape, after reversing the broker-requested, retrospective changes to recommendation scales on the 2007 tape; see Table I. Recommendations are scored by I/B/E/S on a five-point scale, where 1=strong buy and 5=sell. We test for differences in mean recommendations using standard two-sample *F*-tests. The tests compare mean recommendation levels among changed recommendations before and after the changes (column (1) vs. (2) and column (3) vs. (4)). In the last two columns, we compare average recommendation levels among deletions and additions (column (5) vs. (6)). Under the null that the changes affecting the I/B/E/S recommendations history are pure noise, we expect to find no significant changes in recommendation levels. Statistically significant differences in recommendation levels at the 5% level are indicated in bold typeface.

				Average rec. (all changes)		Average rec. (alterations only)		Average rec.	
Comparison tapes	No. of obs on earlier tape	Average rec.	No. of ex post changes	before (1)	after (2)	before (3)	after (4)	deletions (5)	additions (6)
Panel A									
2000 vs. 2001	222,694	2.11	24,116	**2.28**	**2.41**	**2.03**	**2.68**	2.33	2.35
2001 vs. 2002	266,619	2.11	22,473	**2.28**	**2.08**	**1.74**	**2.34**	**2.30**	**2.06**
2002 vs. 2003	280,567	2.11	36,762	**1.98**	**2.28**	**2.07**	**2.01**	**1.63**	**2.45**
2003 vs. 2004	332,145	2.18	57,770	**2.17**	**2.70**	**1.79**	**2.34**	**2.49**	**3.01**
2004 vs. 2005	450,225	2.36	97,582	**2.89**	**1.78**	**1.42**	**2.10**	**2.92**	**1.54**
2005 vs. 2007	414,881	2.24	6,580	**2.15**	**2.36**	**1.98**	**2.89**	**2.15**	**2.33**
Panel B									
2000 vs. 2007	222,694	2.11	29,101	**2.16**	**2.30**	**1.89**	**2.15**	**2.20**	**2.33**
2001 vs. 2007	266,619	2.11	46,217	**2.23**	**2.28**	**2.47**	**2.15**	**2.21**	**2.29**
2002 vs. 2007	280,567	2.11	33,982	**2.24**	**2.38**	**2.64**	**1.98**	**2.18**	**2.44**
2003 vs. 2007	332,145	2.18	36,490	**2.22**	**2.07**	**2.03**	**2.08**	**2.39**	**1.93**
2004 vs. 2007	450,225	2.36	135,042	**2.68**	**2.06**	**2.03**	**1.99**	**2.89**	**1.74**
2005 vs. 2007	414,881	2.24	41,516	**2.13**	**2.10**	**2.09**	**1.97**	**2.15**	**2.33**

Table III. Effect of Alterations, Deletions, and Additions on Trading Signals.

We compare trading signals on the 2000 through 2005 tapes to the adjusted version of the 2007 tape, described in Table I. Tapes are matched up by standardized brokerage firm name, I/B/E/S ticker, and recommendation date. Observations on the 2007 tape dated after the cut-off date of the earlier tape are ignored. Trading signals are constructed on a per-broker and per-I/B/E/S-ticker basis using a 12-month look-back window. For instance, a downgrade is defined as a negative change from a recommendation issued by the same broker for the same I/B/E/S ticker within the previous 12 months. If the previous recommendation was issued more than 12 months ago, or was stopped according to the I/B/E/S stop file, the current recommendation is defined to be a reinitiation. If there is no previous recommendation, the current recommendation is defined to be an initiation. The table also provides a transition matrix for the changed trading signals from the earlier tape to the 2007 tape.

	Orig. tape	All changes		**Trading signal according to adjusted 2007 tape**					
Trading signal as of original tape	No.	No.	%	downgrade	upgrade	reiteration	initiation	reinitiation	deleted
Panel A: Migrations in trading signals (2000 tape vs. 2007 tape)									
downgrade	50,866	4,508	8.9%		143	168	5	14	4,178
upgrade	44,427	4,176	9.4%	124		275	15	18	3,744
reiteration	10,957	2,549	23.3%	522	606		36	22	1,363
initiation	89,065	6,242	7.0%	715	605	298	0	94	4,530
reinitiation	27,379	1,262	4.6%	344	335	115	2		466
added by 2007		13,065		3,473	2,489	1,336	4,409	1,358	
all signals	222,694	31,802	14.3%	5,178	4,178	2,192	4,467	1,506	14,281
Panel B: Migrations in trading signals (2001 tape vs. 2007 tape)									
downgrade	65,403	6,988	10.7%		125	536	13	37	6,277
upgrade	52,831	5,859	11.1%	68		492	8	31	5,260
reiteration	12,901	3,417	26.5%	433	939		25	19	2,001
initiation	100,605	7,671	7.6%	585	911	431	0	114	5,630
reinitiation	34,879	2,073	5.9%	489	544	389	0		651
added by 2007		23,714		7,043	3,725	1,511	7,324	4,111	
all signals	266,619	49,722	18.6%	8,618	6,244	3,359	7,370	4,312	19,819
Panel C: Migrations in trading signals (2002 tape vs. 2007 tape)									
downgrade	67,912	4,110	6.1%		149	522	22	64	3,353
upgrade	54,155	3,254	6.0%	68		517	21	65	2,583
reiteration	14,127	3,042	21.5%	510	1,234		43	74	1,181
initiation	103,462	6,276	6.1%	673	1,188	532	0	136	3,747
reinitiation	40,911	2,280	5.6%	550	738	450	11		531
added by 2007		19,756		6,161	2,668	1,583	6,688	2,656	
all signals	280,567	38,718	13.8%	7,962	5,977	3,604	6,785	2,995	11,395

	Orig. tape	All changes		Trading signal according to adjusted 2007 tape					
Trading signal as of original tape	No.	No.	%	downgrade	upgrade	reiteration	initiation	reinitiation	deleted
Panel D: Migrations in trading signals (2003 tape vs. 2007 tape)									
downgrade	79,772	4,027	5.0%		18	560	12	22	3,415
upgrade	62,108	3,200	5.2%	61		520	10	22	2,587
reiteration	21,632	5,234	24.2%	1,552	1,254		53	57	2,318
initiation	111,577	4,421	4.0%	434	9	36	0	15	3,927
reinitiation	57,056	1,865	3.3%	187	10	23	0		1,645
added by 2007		5,489		768	1,364	365	1,444	1,548	
all signals	332,145	24,236	7.3%	3,002	2,655	1,504	1,519	1,664	13,892
Panel E: Migrations in trading signals (2004 tape vs. 2007 tape)									
downgrade	111,370	26,609	23.9%		14	612	5	2	25,976
upgrade	94,072	27,341	29.1%	48		570	2	7	26,714
reiteration	35,073	16,217	46.2%	1,937	1,587		33	29	12,631
initiation	143,546	28,877	20.1%	450	14	45	0	17	28,351
reinitiation	66,164	2,711	4.1%	209	41	56	0		2,405
added by 2007		4,381		703	1,305	292	1,299	782	
all signals	450,225	106,136	23.6%	3,347	2,961	1,575	1,339	837	96,077
Panel F: Migrations in trading signals (2005 tape vs. 2007 tape)									
downgrade	103,086	2,045	2.0%		14	567	3	2	1,459
upgrade	82,579	1,625	2.0%	16		535	4	1	1,069
reiteration	26,347	3,955	15.0%	1,735	1,626		29	28	537
initiation	130,502	1,295	1.0%	3	6	18	0	1	1,267
reinitiation	72,367	218	0.3%	0	4	11	0		203
added by 2007		1,889		520	458	113	438	360	
all signals	414,881	11,027	2.7%	2,274	2,108	1,244	474	392	4,535

Table IV. Effect of Changes on the Abnormal Returns to Upgrades and Downgrades.

This table compares the abnormal returns to upgrades and downgrades for the 2004 and 2007 I/B/E/S tapes using two different approaches. Panel A reports average daily percentage buy-and-hold abnormal returns for simple calendar-time portfolios based on portfolios of upgrades and downgrades. *Diffret* is average daily return difference between the 2004 portfolio (*Ret04*) and the corresponding 2007 portfolio (*Ret07*). *DiffXret* is the average excess return difference between the same 2004 and 2007 portfolios. Excess returns are equal to the intercept from a regression of *Diffret* (less the riskfree rate) on (i) the excess of the market return over the risk-free rate, (ii) the difference between the daily returns of a value-weighted portfolio of small stocks and one of large stocks (SMB), (iii) the difference between the daily returns of a value-weighted portfolio of high book-to-market stocks and one of low book-to-market stocks (HML), and (iv) the difference between the daily returns of a value-weighted portfolio of high price momentum stocks and one of low price momentum stocks (UMD). Column (1) reports the average daily returns for the entire sample period over which the 2004 and 2007 tapes overlap (Oct. 29, 1993 to Mar. 18, 2004); columns (2) and (3) report the average daily returns for the "post-bubble" period (i.e., the period subsequent to Mar. 10, 2000, the date of the NASDAQ market peak) and the "pre-bubble" period (the period prior to Mar. 10, 2000). Columns (4)-(6) are defined similarly for downgrades. Panels B and C report differences in the three-day event-time returns between the 2004 and 2007 tapes for upgrades and downgrades, respectively. The column labeled "2to1" refers to upgrades from I/B/E/S recommendation code 2 (i.e., "buy") to I/B/E/S code 1 (i.e., "strong buy") only; other columns are defined analogously. *ERet04* and *ERet07* are the three-day raw event returns, calculated as the geometrically cumulated return for the day before, day of, and day after the recommendation, using data from the 2004 and 2007 tapes, respectively. *DiffEret* then equals the average difference between *ERet04* and *ERet07*. Analogously, we compute the three-day excess event return as the raw stock return less the appropriate size-decile return of the CRSP NYSE/AMEX/NASDAQ index (not shown for brevity) and report *DiffEXret*, the average difference between the three-day excess return samples. *t*-statistics are in parentheses, and 5% statistical significance is indicated in bold typeface.

Panel A: Daily calendar-time portfolio returns (in %): 2004 versus 2007 tapes

	Upgrades			Downgrades		
	Full period (1)	Post- "bubble" (2)	Pre- "bubble" (3)	Full period (4)	Post- "bubble" (5)	Pre- "bubble" (6)
Ret07	**0.161**	**0.191**	**0.142**	**-0.095**	**-0.141**	**-0.065**
	(6.76)	(3.89)	(6.06)	(-3.68)	(-2.51)	(-2.93)
Ret04	**0.148**	**0.159**	**0.142**	**-0.078**	-0.101	**-0.063**
	(6.37)	(3.36)	(6.02)	(-3.10)	(-1.87)	(-2.79)
Diffret	**0.012**	**0.032**	-0.000	**-0.017**	**-0.040**	-0.002
	(3.65)	(3.99)	(-0.01)	(-4.88)	(-4.85)	(-1.10)
DiffXret	**0.013**	**0.036**	0.000	**-0.016**	**-0.040**	-0.002
	(3.90)	(4.58)	(0.22)	(-4.70)	(-4.89)	(-1.00)

Table IV. Continued.

Panel B: Three-day upgrade event returns (in %): 2004 versus 2007 tapes

	All	Upgrades to strong buy				Upgrades to buy			… to hold		to sell
	upgrades	*2to1*	*3to1*	*4to1*	*5to1*	*3to2*	*4to2*	*5to2*	*4to3*	*5to3*	*5to4*
ERet07	**3.016**	**3.040**	**3.068**	**3.061**	**1.836**	**3.097**	**4.524**	1.272	**2.740**	**2.118**	0.885
	(82.91)	(44.36)	(41.12)	(4.21)	(4.19)	(53.46)	(6.38)	(1.90)	(14.92)	(11.33)	(1.39)
ERet04	**2.304**	**2.853**	**2.997**	**1.971**	**1.475**	**2.366**	**1.961**	0.398	**1.054**	**1.698**	0.130
	(78.47)	(46.10)	(42.04)	(4.56)	(4.21)	(50.31)	(7.22)	(1.12)	(14.11)	(10.87)	(0.89)
DiffEret	**0.712**	**0.187**	0.071	1.090	0.361	**0.731**	**2.563**	0.875	**1.686**	0.420	0.755
	(15.37)	(2.03)	(0.69)	(1.29)	(0.64)	(9.90)	(3.37)	(1.17)	(9.93)	(1.74)	(1.21)
DiffEXret	**0.724**	**0.204**	0.093	1.180	0.518	**0.686**	**2.828**	0.577	**1.850**	0.474	0.657
	(15.63)	(2.25)	(0.90)	(1.40)	(0.88)	(9.28)	(3.36)	(0.79)	(10.69)	(1.90)	(1.05)

Panel C: Three-day downgrade event returns (in %): 2004 versus 2007 tapes

	All										from
	down-	Downgrades from strong buy				Downgrades from buy			… from hold		sell
	grades	*1to2*	*1to3*	*1to4*	*1to5*	*2to3*	*2to4*	*2to5*	*3to4*	*3to5*	*4to5*
ERet07	**-4.720**	**-4.045**	**-5.342**	**-6.079**	**-4.680**	**-4.925**	**-6.531**	**-3.441**	**-4.130**	**-3.851**	-0.584
	(-103.34)	(-53.20)	(-53.01)	(-6.31)	(-6.47)	(-70.20)	(-10.95)	(-3.65)	(-13.67)	(-16.07)	(-0.55)
ERet04	**-3.794**	**-3.756**	**-5.169**	**-5.425**	**-3.352**	**-4.102**	**-3.018**	**-1.278**	**-1.387**	**-2.868**	0.177
	(-99.21)	(-51.39)	(-54.49)	(-9.03)	(-5.43)	(-68.11)	(-10.03)	(-2.52)	(-11.21)	(-14.82)	(0.97)
DiffEret	**-0.926**	**-0.289**	-0.173	-0.654	-1.328	**-0.823**	**-3.513**	**-2.163**	**-2.743**	**-0.983**	-0.761
	(-15.66)	(-2.74)	(-1.25)	(-0.60)	(-1.40)	(-8.95)	(-5.81)	(-2.20)	(-10.00)	(-3.23)	(-0.70)
DiffEXret	**-0.890**	**-0.263**	-0.241	-0.991	-1.191	**-0.754**	**-3.175**	**-2.169**	**-2.776**	**-0.958**	-0.887
	(-14.74)	(-2.50)	(-1.48)	(-0.85)	(-1.16)	(-8.09)	(-4.90)	(-2.14)	(-9.46)	(-3.01)	(-0.76)

Table V. Effect of Alterations, Additions, and Deletions on Consensus Trading Strategies.

This table reports daily portfolio returns (in %) for a trading strategy ("spread") based on changes in consensus analyst recommendations. We use all I/B/E/S recommendations that have been outstanding for less than one year. The consensus recommendation for a ticker equals the mean outstanding recommendation at the end of a calendar day, based on a minimum of three recommendations. Firms are grouped into quintiles at the beginning of the next day based on the change in consensus. We compute daily portfolio returns by buying stocks in the highest consensus change quintile (Q5) and shorting stocks in the lowest consensus change quintile (Q1). Daily Daniel et al. (1997, "DGTW") characteristic-adjusted returns are defined as raw portfolio returns minus the returns on a value-weighted portfolio of all CRSP firms in the same size, (industry-adjusted) market-book, and one-year momentum quintiles. The strategy is performed separately on the 2002, 2003, 2004, 2005, and 2007 tapes, and differences across tapes are reported. We split the sample into two sub-periods, 1993-2000 ("pre-2001") and 2001 to the end of a tape's time window ("2001-onward"). In the latter case, the exact sample period for the 2007 comparison tape extends from Jan. 1, 2001 to the end of the tape in question, so the estimates for the 2007 tape shown in columns (3) and (7) are different for each comparison. *t*-statistics are in parentheses, and 5% statistical significance is indicated in bold typeface.

	Pre-2001				2001-onwards			
	Spread (Q5-Q1) in raw portfolio return	Spread (Q5-Q1) in DGTW adjusted returns	Spread (Q5-Q1) in DGTW returns, 2007 tape	Difference in DGTW spread: 2007 minus 200(X)	Spread (Q5-Q1) in raw portfolio return	Spread (Q5-Q1) in DGTW adjusted returns	Spread (Q5-Q1) in DGTW returns, 2007 tape	Difference in DGTW spread: 2007 minus 200(X)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2002 tape	**0.272**	**0.266**	**0.269**	0.003	**0.477**	**0.427**	**0.364**	**-0.062**
	(9.76)	(9.50)	(10.64)	(0.04)	(8.01)	(7.22)	(6.09)	(2.10)
2003 tape	**0.292**	**0.289**	**0.269**	**-0.020**	**0.406**	**0.383**	**0.386**	0.003
	(12.69)	(11.26)	(10.64)	(-2.26)	(8.01)	(7.80)	(8.610	(0.11)
2004 tape	**0.294**	**0.29**	**0.269**	**-0.021**	**0.428**	**0.365**	**0.409**	0.044
	(12.72)	(11.21)	(10.64)	(-2.23)	(10.91)	(8.830	(10.78)	(1.50)
2005 tape	**0.289**	**0.288**	**0.269**	**-0.019**	**0.476**	**0.429**	**0.426**	-0.003
	(12.42)	(11.22)	(10.64)	(-2.54)	(15.34)	(13.95)	(13.94)	(-0.36)

Table VI. Effect of Changes on Persistence in Individual Analyst Performance.

The table reports tests of persistence in individual analysts' stock-picking skills. These tests measure the extent to which good past performers continue to perform well in the future. Tests are performed separately on the 2000, 2001, 2002, 2003, 2004, 2005, and 2007 tapes. For each analyst, we compute the average five-day DGTW-adjusted return of all upgrades and downgrades issued by that analyst over the previous six months; in doing so, we assume that we buy on upgrades and sell on downgrades. We then rank analysts into quintiles in January and July of each year based on their average five-day DGTW-adjusted returns over the prior six months. Next, we compute a "persistence spread" equal to the difference between the average five-day DGTW-adjusted return of analysts in the highest quintile (Q5) minus the average five-day DGTW-adjusted return of analysts in the lowest quintile (Q1), in each case computed over the following six months. The five-day return is the geometrically cumulated DGTW-adjusted return for the two trading days before through the two trading days after the recommendation. Daily DGTW characteristic-adjusted returns are defined as raw returns minus the returns on a value-weighted portfolio of all CRSP firms in the same size, (industry-adjusted) market-book, and one-year momentum quintile. We report persistence spreads for each I/B/E/S tape from 2000 through 2005 (shown in column (1)) and for the 2007 tape (shown in column (2)). Note that each tape is compared over its full available sample period to the 2007 tape, so the estimates for the 2007 tape shown in column (2) are different for each comparison tape. In column (3), we report differences between each tape and the 2007 tape. *t*-statistics are shown in parentheses, and 5% statistical significance is indicated in bold typeface.

Average five-day event returns (in %) from persistence quintiles

	Persistence spread (Q5-Q1)	Persistence spread (Q5-Q1) from 2007 tape	Difference in persistence spreads, 2007-200X
	(1)	(2)	(3)
2000 tape	**2.432**	**2.480**	0.047
	(5.62)	(8.14)	(0.21)
2001 tape	**2.960**	**2.574**	**-0.386**
	(8.13)	(9.21)	(-3.40)
2002 tape	**3.079**	**2.556**	**-0.523**
	(7.75)	(9.68)	(-2.22)
2003 tape	**2.673**	**2.490**	-0.183
	(9.14)	(9.65)	(-1.65)
2004 tape	**2.645**	**2.461**	**-0.184**
	(9.95)	(10.49)	(-2.18)
2005 tape	**2.561**	**2.444**	**-**0.118
	(11.07)	(11.76)	(**-**1.86)

For Supplements and Dataset Section of www.afajof.org

Internet Appendix to

Rewriting History

Alexander Ljungqvist
Stern School of Business
New York University
and *CEPR*

Christopher Malloy
Harvard Business School

Felicia Marston
McIntire School of Commerce
University of Virginia

September 10, 2008

Ljungqvist, Malloy, and Marston (2008) document widespread changes to the historical I/B/E/S analyst stock recommendations database. This document contains supplementary tests and results.

A. Patterns in Popular Conditioning Variables

Recommendations affected by the changes to the I/B/E/S recommendations history appear to cluster according to three popular conditioning variables: The analyst's reputation, the brokerage firm's size and status, and the boldness of the recommendation. We measure analyst reputation using all-star status, as designated in the Oct. issue of *Institutional Investor* magazine preceding the recommendation in question. We divide brokerage firms into the 12 (generally large) firms sanctioned under the Global Settlement and all other firms. And we code a recommendation as bold if it was one notch or more above or below consensus (=mean recommendation) computed over the prior three months (requiring at least three outstanding recommendations).

In Table A1, we compare the frequency of these conditioning variables in the universe of historical recommendations and in the set of changed recommendations. We compare each tape to the next tape as well as to the adjusted 2007 tape.

INSERT TABLE A1 ABOUT HERE

We find that all-stars are significantly overrepresented among changed recommendations on the 2000 and 2001 tapes, while changed recommendations on the 2002 through 2004 tapes disproportionately come from unrated analysts. Relative to the adjusted 2007 tape, recommendations by unrated analysts are significantly more likely to need correction on every tape except the 2001 tape. Thus, tests comparing all-stars to unrated analysts may yield different results depending on which tape is used. Sanctioned banks are overrepresented among affected recommendations on the 2000 and 2001 tapes, and underrepresented for all later tapes. Relative to the adjusted 2007 tape, sanctioned banks are associated with a significantly lower need for corrections on every tape except the 2001 tape. Finally, bold recommendations are significantly overrepresented among affected records on all tapes. They are also consistently and significantly more likely to be subject to corrections on the adjusted 2007 tape.

B. Further Evidence on the Effect of the Changes on Trading Signals

To investigate some of the results in Ljungqvist, Malloy, and Marston’s (2008) Tables IV-VI in greater depth, we employ a series of additional tests. First, we explore subsets of Table IV, Panels B and C by removing movements to and from one category at a time. For example, to determine if the rating category 4 (strong sell) is driving the results, we re-run the analysis for “All upgrades” and “All downgrades” by removing any upgrades/downgrades that involve movements to and from the rating category 4. We repeat this analysis for movements to and from each individual rating category (1-5). Panel A of Table A2 reports summary statistics from this analysis.

INSERT TABLE A2 ABOUT HERE

The results indicate that while removing upgrades to and from category 4 does decrease the difference in event returns between the two tapes for “All upgrades”, the remaining difference is still large economically and statistically significant (0.426%, *t*=8.67). Removing movements to and from categories 3 and 5 has a similar, but somewhat smaller dampening effect, while removing movements to and from categories 1 and 2 has the opposite effect (i.e., it increases the difference in event returns between the two tapes). Overall, the results in Table IV, Panel B do not appear to be driven by any particular rating category.

When we repeat this analysis for downgrades (Panel B of Table A2), we find similar results: Although removing movements to and from category 4 has the biggest impact on the differential between the two tapes, it does not drive the differences between the two tapes. The differences in event returns for downgrades are still large and significant across the two tapes, regardless of which category we choose to remove.

To recap, the differences in event returns for both upgrades and downgrades remain economically large and statistically significant across the two tapes, regardless of which category we choose to remove. However, removing category 4 does have a non-trivial impact on these results. To understand why this is the case, we also examined some observations by hand. In doing so, we discovered that the reason category 4 has a non-trivial impact in this particular case is that a large number of recommendations issued by eight quantitative research groups were erroneously included in the database around this time, and are captured on the 2004 tape. The quantitative research groups produce algorithmic recommendations constrained to be symmetrically distributed; hence they include more category 4 and 5 recommendations than usual. According to Thomson, these recommendations were not supposed to be viewable by its clients yet were added to the database sometime between 2003 and 2004. They were subsequently deleted at some point between 2004 and 2005. This suggests that deletions from the 2007 tape are the main drivers of the changes we document in Table IV, Panels B and C.

To demonstrate this more formally, we take a different approach that seeks to trace back the impact of each of the four types of data changes that we document in Tables 1 and 2 (i.e., the additions, deletions, alterations, and anonymizations) on the results in Tables IV-VI. Note that since upgrades and downgrades are computed at the broker level, anonymizations will have no impact on the Table IV results. To measure the impact of each of the other three types of changes individually, we create three hypothetical 2004 tapes, each of which simulates what the 2004 tape would have looked like if one type of change had *not* occurred. We then use each of these new hypothetical tapes as the 2004 tape, recalculate all the trading signals, and re-compute the 2004-2007 differences reported in Table IV, Panels A and B. Specifically, we compute the 2004 tape: a) with "no deletions", meaning that we omit on the 2004 tape those records that are later deleted on the 2007 file, b) with "no alterations", meaning that we re-instate altered records to original values on the 2004 tape as we have done on the 2007 tape, and c) with "no additions", meaning that we take those records that were added on the 2007 tape and add them back to the 2004 tape. Note that we do not make any changes to the 2007 tape, so the 2007 event returns stay the same across all comparisons. We report these results in Table A3.

INSERT TABLE A3 ABOUT HERE

Consistent with our explanation above, this table demonstrates that the differences reported in Table IV, Panels B and C for the 2004-2007 comparison are primarily due to the impact of deletions from the 2007 tape (of which the quantitative research groups' recommendations are the primary subset; the remainder involve the removal of certain brokerage firm histories by request of the broker as of 2007). Once these deletions are removed from the 2004 snapshot, the event returns are not statistically different between the 2004 and 2007 tapes.

This raises the question of whether all of the results are driven by recommendations that are subsequently removed on the 2007 tape. To test this conjecture, we randomly pick three additional pairwise comparisons for the tests in Tables IV-VI, and perform the same in-depth decompositions as above for these new cases. (Assembling the hypothetical tapes and running the analysis is very time-consuming, which is why we choose three random comparisons.) For Table IV, we choose the 2000 tape, for Table V we choose the 2003 tape, and for Table VI we choose the 2002 tape, each of which is compared to the 2007 tape.

Untabulated statistics indicate that as with the 2004-2007 results reported in Table IV, the differences in 3-day event returns between the 2000 and 2007 tape are significant (although smaller in magnitude). However, unlike the 2004-2007 differences, which were driven mainly by deletions from the 2007 tape, as demonstrated above, the 2000-2007 differences are driven primarily by additions to the 2007 tape (i.e., records that are not on the 2000 tape but that have been added back on the 2007 tape). Only when these 2007 additions are added back to the 2000 tape does the spread in event returns between the two tapes become insignificant. In magnitude, only 1% of the difference in upgrade event returns between the 2000 and 2007 tapes is due to deletions (21% for downgrades), while 7% of the difference is due to alterations (3% for downgrades), and 85% of the difference is due to additions (76% for downgrades).[1] Thus the *addition* of records to the 2007 tape appears to be the primary cause of the significant differences in event returns between the 2000 and 2007 tapes.

In Ljungqvist, Malloy, and Marston’s (2008) Table V tests, when we conduct similar breakdowns for the 2003 vs. 2007 comparison, we find that most of the difference in the pre-2001 DGTW spread (=-0.020 from the table) is due to alterations: Only when we remove the alterations from the 2003 tape does this spread become insignificant. In magnitude,

[1] Note that these numbers do not have to add up to 100% since the removal of one category at a time does not constitute a complete decomposition (since removing two categories at a time can result in additional differences as well).

approximately 12% of the difference in the spread is due to deletions, 53% is due to alterations, and 22% is due to additions.

In Ljungqvist, Malloy, and Marston’s (2008) Table VI tests, we need to compute four hypothetical 2002 tapes rather than three, because anonymizations can impact individual analyst-level persistence (even though they cannot impact the measures of consensus recommendations from Table V, nor the upgrade/downgrade measures from Table IV). We find that most of the difference in the persistence spread is again due to additions: Only when we add the 2007 additions back to the 2002 tape does this spread become insignificant. In magnitude, approximately 21% of the difference in the spread is due to deletions, 9% is due to alterations, 36% is due to additions, and 1% is due to anonymizations.

In summary, across all tests, differences between the various pairwise comparisons appear to be caused by different combinations of the four types of data changes.

C. Further Evidence on the Persistence in Analyst Stock-picking

Ljungqvist, Malloy, and Marston (2008) find evidence of persistence in individual analysts’ stock-picking performance on every I/B/E/S tape, but the extent of persistence varies markedly across tapes. In Table A4, we show that this result is even more pronounced if we filter on analysts’ all-star status (defined as in Section I.C). A common modification to the persistence trading strategy is to buy on recommendations by all-star analysts who are also in quintile 5 and to sell on recommendations by non-all star analysts ranked in quintile 1. This assumes asymmetry in persistence among all-stars: They are likely to repeat good past performance but not poor past performance. Imposing this screen increases the differences in persistence spreads across the tapes. For example, we find a difference between the 2001 and 2007 tapes of 82.0 basis points over five trading days, an increase of 25.3% relative to the amount of persistence found on the 2007 tape. Similarly large differences exist between the 2002 and 2007 tapes (66.3 basis points, a 21.1% increase relative to 2007) and between the 2003 and 2007 tapes (36.6 basis points, a 12.1% increase relative to 2007).

INSERT TABLE A4 ABOUT HERE

Table A1. Patterns in Popular Conditioning Variables.

The table documents patterns in the changes to the I/B/E/S historical recommendations database analyzed in Ljungqvist, Malloy, and Marston (2008). We examine year-to-year changes to the database by comparing data from adjacent annual downloads. We focus on the period for which each pair of downloads has overlapping coverage (that is, we ignore recommendations from the later tape that are dated after the cut-off date of the earlier tape.) The cutoff dates of our tapes are 7/20/00 ("2000 tape"), 1/24/02 ("2001 tape"), 7/18/02 ("2002 tape"), 3/20/03 ("2003 tape"), 3/18/04 ("2004 tape"), 12/15/05 ("2005 tape"), and 9/20/07 ("2007 tape"). According to Thomson, the 2007 tape contains data purged of all data errors we have identified, except that it continues to include broker-requested, retrospective changes to recommendation scales. We also compare the 2000 through 2005 tapes to the 2007 tape, after reversing the broker-requested, retrospective changes to recommendation scales. This adjusted version of the 2007 tape corresponds to the "as-was" historical recommendations database which Thomson intends to make available to researchers in response to our investigation. In Ljungqvist, Malloy, and Marston (2008), we document four types of changes to the I/B/E/S recommendations data. We define an *alteration* as a broker/ticker/date triad that appears on both tapes but for which the recommendation on one tape is different than on the other tape. A *deletion* is a broker/ticker/date triad that appears on the earlier tape but not on the later tape to which it is compared. An *addition* is a broker/ticker/date triad that appears on the later comparison tape but not on the earlier tape. *Anonymizations* refer to cases where the analyst associated with a broker/ticker/date triad is identified by name on the earlier tape but is anonymous on the later tape. In this table, we compare the frequency of three popular conditioning variables in the universe of historical recommendations and in the set of recommendations subject to ex post changes (due to alterations, deletions, additions, or anonymizations). The three variables of interest condition on whether the analyst has all-star status (the top three rated analysts in each sector, as designated in the Oct. issue of *Institutional Investor* magazine preceding the recommendation in question), whether the brokerage firm is among the 12 firms sanctioned under the Global Settlement, and whether the recommendation was "bold", where bold is an indicator equaling one if the recommendation was one notch or more above or below consensus (=mean recommendation) computed over the prior three months (requiring at least three outstanding recommendations). We test for differences in fractions using standard two-sample *F*-tests of equal proportions. The tests compare the universe to the set of changed recommendations. Statistically significant differences at the 5% level are indicated in bold typeface.

	All-star analysts			**Global Settlement banks**			**Bold recommendations**		
	Share of recom-mendations universe	Share of changed recommendations		Share of recom-mendations universe	Share of changed recommendations		Share of recom-mendations universe	Share of changed recommendations	
		relative to next tape	relative to 2007 tape		relative to next tape	relative to 2007 tape		relative to next tape	relative to 2007 tape
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2000 tape	15.3%	**19.6%**	**9.1%**	24.3%	**37.6%**	**10.4%**	30.6%	**31.8%**	**35.9%**
2001 tape	13.3%	**23.1%**	**16.3%**	23.2%	**44.2%**	**28.7%**	31.0%	**33.1%**	**35.6%**
2002 tape	15.0%	**10.5%**	**8.6%**	26.9%	**16.0%**	**12.4%**	30.4%	**33.6%**	**37.3%**
2003 tape	15.0%	**6.4%**	**9.2%**	28.8%	**13.7%**	**19.9%**	31.6%	**39.6%**	**36.2%**
2004 tape	11.5%	**0.4%**	**1.4%**	23.8%	**1.7%**	**6.6%**	32.5%	**36.3%**	**35.6%**
2005 tape	13.9%	14.0%	**4.0%**	30.0%	**12.7%**	**19.1%**	32.3%	**30.5%**	**33.7%**

Table A2. Effect of Changes on the Abnormal Returns to Upgrades and Downgrades, Removing One Rating Category at a Time.

This table compares the event returns to upgrades and downgrades for the 2004 and 2007 I/B/E/S tapes, and complements Table IV in Ljungqvist, Malloy, and Marston (2008). Panels A and B report differences in the three-day event-time returns between the 2004 and 2007 tapes for upgrades and downgrades, respectively. The first column includes all upgrades (downgrades), while the subsequent columns exclude all upgrades (downgrades) to and from a particular rating category (e.g., I/B/E/S recommendation code 2 = "buy"). *ERet04* and *ERet07* are the three-day raw event returns, calculated as the geometrically cumulated return for the day before, day of, and day after the recommendation, using data from the 2004 and 2007 tapes, respectively. *DiffEret* then equals the average difference between *ERet04* and *ERet07*. Analogously, we compute the three-day excess event return as the raw stock return less the appropriate size-decile return of the CRSP NYSE/AMEX/NASDAQ index (not shown for brevity) and report *DiffEXret*, the average difference between the three-day excess return samples. *t*-statistics are in parentheses, and 5% statistical significance is indicated in bold typeface.

Panel A: Three-day upgrade event returns (in %): 2004 versus 2007 tapes

	All Upgrades	All upgrades except to and from category:				
		1	*2*	*3*	*4*	*5*
ERet07	**3.016**	**2.986**	**2.864**	**3.037**	**3.022**	**3.065**
	(82.91)	(55.41)	(44.36)	(46.40)	(82.01)	(82.17)
ERet04	**2.304**	**1.874**	**1.928**	**2.467**	**2.596**	**2.284**
	(78.47)	(50.52)	(41.65)	(43.46)	(79.25)	(78.61)
DiffEret	**0.712**	**1.111**	**0.935**	**0.570**	**0.426**	**0.654**
	(15.37)	(17.48)	(12.02)	(6.62)	(8.67)	(13.65)
DiffXret	**0.724**	**1.122**	**1.003**	**0.595**	**0.421**	**0.662**
	(15.63)	(17.49)	(12.57)	(6.95)	(8.61)	(13.82)

Panel B: Three-day downgrade event returns (in %): 2004 versus 2007 tapes

	All downgrades	All downgrades except to and from category:				
		1	*2*	*3*	*4*	*5*
ERet07	**-4.720**	**-4.783**	**-4.992**	**-4.140**	**-4.733**	**-4.761**
	(-103.34)	(-70.97)	(-55.32)	(-54.39)	(-103.84)	(-102.03)
ERet04	**-3.794**	**-3.296**	**-3.586**	**-3.383**	**-4.197**	**-3.933**
	(-99.21)	(-65.91)	(-53.16)	(-50.67)	(-101.90)	(-98.92)
DiffEret	**-0.926**	**-1.487**	**-1.406**	**-0.756**	**-0.536**	**-0.827**
	(-15.66)	(-18.08)	(-12.73)	(-7.50)	(-9.75)	(-13.57)
DiffXret	**-0.890**	**-1.423**	**-1.445**	**-0.710**	**-0.503**	**-0.787**
	(-14.74)	(-16.85)	(-12.57)	(-7.00)	(-8.09)	(-12.67)

Table A3. Effect of Changes on the Abnormal Returns to Upgrades and Downgrades, Removing One Type of Data Change at a Time.

This table compares the event returns to upgrades and downgrades for the 2004 and 2007 I/B/E/S tapes, and complements Table IV in Ljungqvist, Malloy, and Marston (2008). The table reports differences in the three-day event-time returns between the 2004 and 2007 tapes for upgrades and downgrades. To measure the impact of each of the three types of data changes (deletions, alterations, additions) individually, we create three hypothetical 2004 tapes, each of which simulates what the 2004 tape would have looked like if one type of data change had *not* occurred. We then use each of these new hypothetical tapes as our 2004 tape, recalculate all the trading signals, and re-compute the 2004-2007 differences reported in Table IV, Panels A and B. Specifically, we compute the 2004 tape: a) with "no deletions", meaning that we omit on the 2004 tape those records that are later deleted on the 2007 file, b) with "no alterations", meaning that we re-instate altered records to original values on the 2004 tape as we have done on the 2007 tape, and c) with "no additions", meaning that we take those records that were added on the 2007 tape and add them back to the 2004 tape. Note that we do not make any changes to the 2007 tape, so the 2007 event returns stay the same across all comparisons. *ERet04* and *ERet07* are the three-day raw event returns, calculated as the geometrically cumulated return for the day before, day of, and day after the recommendation, using data from the 2004 and 2007 tapes, respectively. *DiffEret* then equals the average difference between *ERet04* and *ERet07*. Analogously, we compute the three-day excess event return as the raw stock return less the appropriate size-decile return of the CRSP NYSE/AMEX/NASDAQ index (not shown for brevity) and report *DiffEXret*, the average difference between the three-day excess return samples. *t*-statistics are in parentheses, and 5% statistical significance is indicated in bold typeface.

Three-day upgrade event returns (in %): 2004 hypothetical tapes versus 2007 tape

	All upgrades				All downgrades			
	All changes	No deletions	No alterations	No additions	All changes	No deletions	No alterations	No additions
ERet07	**3.016**	**3.016**	**3.016**	**3.016**	**-4.720**	**-4.720**	**-4.720**	**-4.720**
	(82.91)	(82.91)	(82.91)	(82.91)	(-103.34)	(-103.3)	(-103.34)	(-103.34)
ERet04	**2.304**	**3.063**	**2.295**	**2.304**	**-3.794**	**-4.784**	**-3.773**	**-3.797**
	(78.47)	(82.50)	(78.60)	(78.84)	(-99.21)	(-102.7)	(-99.16)	(-99.95)
DiffEret	**0.712**	-0.048	**0.721**	**0.711**	**-0.926**	0.069	**-0.946**	**-0.923**
	(15.37)	(0.93)	(15.61)	(15.39)	(-15.66)	(0.71)	(-16.50)	(-15.65)
DiffXret	**0.724**	-0.028	**0.728**	**0.722**	**-0.890**	0.054	**-0.908**	**-0.885**
	(15.63)	(0.61)	(15.75)	(15.61)	(-14.74)	(0.81)	(-15.21)	(14.70)

Table A4. Effect of Changes on Persistence in Individual Analyst Performance.

The table reports tests of persistence in individual analysts' stock-picking skills. The table complements Table VI in Ljungqvist, Malloy, and Marston (2008). Tests are performed separately on the 2000, 2001, 2002, 2003, 2004, 2005, and 2007 tapes. For each analyst, we compute the average five-day DGTW-adjusted return of all upgrades and downgrades issued by that analyst over the previous six months; in doing so, we assume that we buy on upgrades and sell on downgrades. We then rank analysts into quintiles in January and July of each year based on their average five-day DGTW-adjusted return over the prior six months. Next we compute a "persistence spread" equal to the difference between the average five-day DGTW-adjusted return of analysts in the highest quintile (Q5) minus the average five-day DGTW-adjusted return of analysts in the lowest quintile (Q1), in each case computed over the following six months. The five-day return is the geometrically cumulated DGTW-adjusted return for the two trading days before through the two trading days after the recommendation. Daily DGTW characteristic-adjusted returns are defined as raw returns minus the returns on a value-weighted portfolio of all CRSP firms in the same size, (industry-adjusted) market-book, and one-year momentum quintile. We report persistence spreads for each I/B/E/S tape from 2000 through 2005 (shown in column (1)) and for the 2007 tape (shown in column (2)). Note that each tape is compared over its full available sample period to the 2007 tape, so the estimates for the 2007 tape shown in column (2) are different for each comparison tape. In column (3), we report differences between each tape and the 2007 tape. The results in this table are computed identically to those in Table VI in Ljungqvist, Malloy, and Marston (2008), except that we impose an additional all-star filter: we restrict quintile 5 to be the subset of quintile 5 analysts who are also all-star analysts (as designated in the preceding Oct. issue of *Institutional Investor* magazine), and we restrict quintile 1 to be the subset of quintile 1 analysts who are *not* also all-star analysts. *t*-statistics are shown in parentheses, and 5% statistical significance is indicated in bold typeface.

Average five-day event returns (in %) from persistence quintiles with all-star screens included

	Persistence spread (Q5-Q1) (1)	Persistence spread (Q5-Q1) from 2007 tape (2)	Difference in persistence spreads, 2007-200X (3)
2000 tape	**3.049**	**3.158**	0.109
	(4.61)	(7.60)	(0.22)
2001 tape	**4.059**	**3.239**	**-0.820**
	(9.32)	(8.74)	(-2.80)
2002 tape	**3.811**	**3.149**	**-0.663**
	(9.11)	(8.72)	(-2.57)
2003 tape	**3.404**	**3.038**	-0.366
	(10.01)	(8.46)	(-1.94)
2004 tape	**3.131**	**2.964**	-0.168
	(9.32)	(9.03)	(-1.07)
2005 tape	**2.991**	**2.897**	-0.094
	(10.43)	(9.94)	(-0.82)