The Influence of Financial Data Subscriptions on Analyst Research

Braiden Coleman
University of Georgia
braiden.coleman@uga.edu

Travis Dyer
Brigham Young University
travis.dyer@byu.edu

Mark Lang
University of North Carolina at Chapel Hill
Mark_Lang@unc.edu

May 2025

We use disclosed "source" data from analyst reports to infer brokerage financial data subscriptions (FDS) and investigate their effects on analyst research. When brokerages add FDS, their analysts' forecast accuracy increases. Effect sizes are at least as large as those for analyst experience, busyness, and brokerage size. Benefits are largest for less experienced and busier analysts with less private access to management forecasting over longer horizons. Although adding new FDS benefits individual analysts, there is substantial overlap in FDS across brokerages, leading to homogenized market views. Specifically, when brokerages have overlapping FDS, their analysts' forecasts, timing, boldness, recommendations, report content, and errors all converge. Consistent with subscription overlap negatively affecting the diversity of analyst opinions, consensus estimates become less accurate when there is more overlap in FDS among analysts. Overall, our findings suggest that FDS are an important and overlooked input into analyst research.

Keywords: Data Subscriptions; Sell-side Analysts; Diversity of Opinion; Wisdom of Crowds
JEL Classifications: C81, D83, G17, G23

We thank various anonymous sell-side equity analysts for their insights and helpful comments. We also thank Brett Campbell, John Campbell, Mike Drake, Jeremiah Green, Xu Jiang, Jared Jennings, Michelle Lowry (Discussant), Stan Markov, Bill Mayew, Jeff McMullin, Mark Piorkowski, Bob Resutek, Georg Rickmann (Discussant), Richard Saouma, Roy Schmardebeck, Carol Seregni, Brady Twedt, and workshop participants at Duke University, North Carolina State University, the University of Miami Accounting Conference, the BYU Accounting Symposium, the Washington University in St. Louis Accounting Research Conference, and the Virginia Tech Accounting Research Conference. We thank the University of Georgia Terry College of Business, the BYU Marriott School of Business, and the UNC Kenan-Flagler Business School for financial support. We received excellent research support from Vanessa Johnson and Ham Christensen. All errors are our own.

1. Introduction

A large literature in accounting and finance examines analyst research output as a function of individual analyst characteristics (e.g., experience, education, access to management, gender, and busyness).[1] However, in practice, analysts are likely influenced by factors outside of their control such as the extent to which their brokerages provide access to important datasets through financial data subscriptions (FDS).[2] Despite the ubiquity and potential importance of FDS, we are unaware of any research systematically documenting variation in access to FDS across brokerages and the effects on analyst research, likely due to difficulty in directly observing such subscriptions. In this study, we use disclosed "source" data from a large sample of analyst reports to infer brokerage subscriptions, and we investigate the implications of FDS for analyst research.[3] Specifically, we provide evidence on the primary FDS available to analysts and assess their influence on (a) characteristics of individual analyst forecasts, (b) forecast commonalities across brokerages, and (c) consensus forecast estimates.

To evaluate the effects of FDS, we construct a novel dataset containing the subscriptions analysts identify in their research. More specifically, we rely on the common convention of analysts referencing "source:" followed by a list of data providers used in the report. Exploiting this feature, we extract source references from 595,642 reports written by 3,596 analysts working at 285 brokerages during 2008–2017.[4] Source referencing is very common in analyst reports; approximately 96% of the reports reference data sources and 90% of reports follow the "source:" labeling convention. If an analyst affiliated with a brokerage references a given FDS, we consider that source

[1] See, for example, Bradshaw (2011) and Kothari et al. (2016) for surveys of analyst forecast research.
[2] We use the acronym FDS to differentiate from research on FDPs (forecast data providers), which aggregate individual analyst forecasts into consensus forecasts (Bochkay et al., 2022; Larocque et al., 2023). Conceptually, FDS provide inputs into analyst forecasts, which are in turn aggregated by FDPs into consensus forecasts.
[3] Analyst reports cite sources for several reasons. First, data providers often have licensing agreements that require source attribution (e.g., Thomson Reuters' General Terms and Conditions). Second, ethical guidelines from the CFA Institute and brokerages encourage transparency and credibility in reports. Finally, anecdotal evidence from discussions with a prior UBS equity analyst suggests that analysts also reference sources to increase clients' confidence in the report content, consistent with a credibility motive.
[4] Appendix A provides an example of an analyst report in our sample that includes source references.

"subscribed" to by the brokerage for the quarter before and after the reference. Our conversations with analysts suggest that, due to the high cost, FDS decisions are typically made at the brokerage level, implying that the choice is relatively exogenous from the standpoint of a given analyst. Consistent with FDS decisions being made at the brokerage level, we find a much higher median overlap in sources used among analysts employed by the same brokerage (89.4%) versus analysts randomly assigned at different brokerages (1.9%). Source mentions also exhibit notable persistence, consistent with the notion that these are ongoing data subscriptions.

We focus on the top 100 sources referenced by analysts in our sample. The most common FDS cited are Bloomberg (66% of brokerages), FactSet (47%), S&P Capital IQ (45%), and Thomson (43%). However, many of our FDS are also from smaller and more focused subscriptions (e.g., consumer spending data, niche industry analytics, and alternative financial indicators). The median brokerage subscribes to about seven FDS and brokerages differ systematically in their subscriptions. For example, JPMorgan most often references data provided by Bloomberg, while Morgan Stanley most commonly cites Thomson Reuters. Not surprisingly, variation across brokerages is even higher for FDS from smaller data providers.

We then turn to understanding the relation between FDS and analyst research. We focus initially on forecast accuracy because it provides a consistent and objective measure. To the extent that FDS are an important source of information for analysts, we expect a positive relation between FDS and forecast accuracy (although the significance, magnitude, and cross-sectional variation are open empirical questions).[5] In addition, this analysis helps to validate the underlying source data.

[5] Based on discussions with equity analysts and academics, there are mixed views on how much FDS access is likely to affect forecast accuracy. On its face, increased information access should increase accuracy. An alternative view is that brokerages invest in FDS to ensure a level playing field with competitors, reflecting a shared cost/benefit tradeoff. In the extreme, ready access to standardized datasets could reduce incentives for private information acquisition.

We include *Firm x Year* fixed effects, so the comparison is among analysts following the same firm in the same year with variation coming from their FDS. Overall, there is a strong positive relation between the number of FDS available to an analyst and forecast accuracy. Results are consistent including both *Firm x Year* and *Analyst* fixed effects (i.e., variation is from within-analyst changes in access to FDS, controlling for firm and year); *Firm x Year* and *Brokerage* fixed effects (variation is from within-brokerage changes in FDS); and *Firm x Year* and *Analyst x Brokerage* fixed effects (variation is from changes in FDS for a given analyst employed at a given brokerage). While a positive relation between the number of FDS and forecast accuracy is not necessarily surprising, it adds veracity to our brokerage-level subscription data and confirms the importance of FDS to analyst output. In terms of economic magnitude, after controlling for firm and year fixed effects, the effect of FDS on forecast accuracy is generally larger than analyst experience, brokerage size, or busyness, reinforcing the importance of FDS to analyst research. Results are robust to a range of analyses designed to address potential omitted variables.

In cross-sectional tests, the largest accuracy gains are associated with private FDS (relative to public sources). Additionally, while access to both major FDS (S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, Morningstar) and alternate "minor" FDS increase forecast accuracy, effects are largest for minor FDS, consistent with more novel data sources providing the greatest benefit. The value of FDS also varies predictably based on individual analyst characteristics; additional FDS are most valuable for (1) less experienced analysts with (2) less personal access to management who (3) follow a larger portfolio of stocks, and (4) are forecasting over longer horizons. While the results are largely intuitive, documenting the benefits of FDS is important given the variety of available FDS and their cost.[6] Further, our results highlight the particular types of analysts (busier and less experienced with limited access to private information forecasting over longer horizons) who benefit most from FDS.

A potential concern with the preceding analyses is that other changes may occur at brokerages that lead to both new FDS and increased accuracy (e.g., the brokerage might be expanding). Because the treatment is at the brokerage level, we cannot entirely dismiss that possibility, but several factors mitigate it: (1) we control for brokerage size in the regressions; (2) consistent with recent prior literature, the relation between brokerage size and accuracy is negative (e.g., Drake et al., 2020; Fang and Hope, 2021; Huang et al., 2022); (3) the results include analyst and brokerage fixed effects so changes in the type of analysts or brokerages covering a firm are controlled; (4) results are robust to alternate controls for changes in brokerage attributes (e.g., growth, percentage of All-Star analysts, and dropping independent analysts); and (5) cross-sectional tests are consistent with predictions (Glaeser and Guay, 2017).

Having established that FDS are associated with the accuracy of analyst forecasts, we next examine the effect of *overlap* in FDS across brokerages on analysts' research. To our knowledge, prior research has not investigated the effects of shared data sources on analyst output. Because there is significant concentration in the financial data industry, with 45.9% of market share concentrated among five financial data providers (Al Bari, 2023), many brokerages overlap in FDS. As a result, overlapping FDS could affect the independence of analyst opinions. Understanding these effects is important for brokerages and analysts, as well as for investors who rely on analyst research for unique insights, and regulators who aim to better understand the operation of financial markets.

[6] Costs include both subscription fees and indirect costs (e.g., time associated with training and familiarization). Annual direct costs range from about $30,000 per user for Bloomberg terminals to more than $100,000 per user for more specialized datasets. See https://www.investopedia.com/terms/b/bloomberg_terminal.asp and https://www.ft.com/content/4897fc5c-374f-4423-b824-21b317e58c83.

The relation between shared FDS across brokerages and analyst output is unclear ex-ante. In our conversations with practitioners and academics, a prominent intuition is that sharing FDS should not have a strong directional effect on analyst output because FDS often rely on the same underlying sources (e.g., EDGAR filings) in developing their products. Further, investors rely on analysts for novel insights not already impounded in price, so analysts have incentives to adjust for the effects of shared information and focus on output differentiation to demonstrate their unique “edge” to clients and superiors (Brown et al., 2015). In the extreme, shared FDS could increase the diversity of research if analysts know peers rely on similar data and seek to differentiate their output and “anti-herd” by biasing forecasts away from shared information (Bernhardt et al., 2006).[7]

Alternatively, overlap in FDS across brokerages could lead to convergence among analysts. Recent evidence suggests that the measurement and disaggregation of some firm and market metrics vary considerably across financial data aggregators even if they use the same underlying sources (Bochkay et al., 2022; Larocque et al., 2023). In addition, financial data aggregators may enhance traditional datasets with “alternative” proprietary data (Bloomberg, 2017; Refinitiv, 2019; Bloomberg, 2023). Similarly, training offered by data providers may lead to standardization in how analysts interpret and incorporate information, encouraging convergence. As a result, we view the overall effect and cross-sectional variation in the effect of shared FDS as open empirical questions.

To evaluate whether overlap in FDS affects the similarity of analyst output, we construct a panel of analyst pairs. We focus on pairs of analysts forecasting annual earnings for the same firm in the same year, resulting in 1,337,709 analyst pair observations. We include *Firm x Year* fixed effects to control for firm and time-related attributes, as well as controls for similarities in brokerage resources, analyst experience, and busyness.

[7] Relatedly, information theory highlights that shared public information can increase disagreement among market participants (Kondor, 2012 and Armstrong et al., 2023).

We examine diversity in analyst opinions by comparing similarity in analyst-pair research outputs for a given firm-year as a function of the extent of overlap in FDS. We find that the similarity of forecast point estimates and boldness both increase significantly once brokerages share common FDS, consistent with shared FDS reducing the diversity of opinions. If the similarity results from shared reliance on FDS, we expect the timing of analyst reports to become more similar since analysts are more likely to issue reports following receipt of new information. We find that the timing of reports also becomes more similar when analysts share FDS, suggesting that overlap in FDS leads to shared information availability and processing. Further, the effect is as large or larger than the effects of similar analyst experience or busyness, suggesting the results are economically significant.

The analyst pairwise design mitigates a variety of endogeneity concerns. When a brokerage subscribes to new FDS, it affects data provider similarity with both subscribing *and* non-subscribing peer brokerages. Hence, selection concerns are at the pairwise, rather than at the individual, level. Results are robust to a specification that includes *Brokerage Pair* fixed effects to hold constant fixed similarities between brokerage pairs, exploiting inter-temporal variation that results from changes in FDS over time. As an alternate source of variation in FDS similarity, we exploit *across* brokerage changes in analyst employment. A benefit of studying employment changes is that FDS are at the brokerage level and are unlikely to change with new analyst hires.[8] We find similar results focusing on variation in FDS similarity caused by employment changes.

We conduct cross-sectional tests to better understand the mechanisms. First, we investigate whether these effects are present across both major data providers (S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar) and smaller FDS. We find that overlap in both major and alternate data providers influences analyst forecast similarity, suggesting that both play important

[8] Our conversations with analysts suggest that new brokerage hires have little sway over data subscriptions, especially for large brokerages. We find similar results when limiting our analysis to large brokerages.

roles in explaining our results. Second, we assess the impact of information access, and find that All-Star analysts, who are more likely to have access to private information, exhibit weaker effects from similarity in FDS. Additionally, the effects are attenuated for analysts with many subscriptions, consistent with greater data availability facilitating more independent opinions.

While analyst forecasts are an important output, they constitute a relatively small component of analyst reports. Further, analysts may rely on FDS primarily for summarizing quantitative data, with the substance of reports being relatively unaffected. Consequently, we expand our analyses to include the broader content of analyst reports (overall stock recommendations and narrative content). We find that buy/sell recommendations become more similar when analysts share FDS and narrative content (as measured by text cosine similarity) converges.

While our evidence suggests that forecasts become more accurate with more FDS and converge when FDS are shared, it could be that forecasts converge *because* they become more accurate (i.e., they converge to the truth). However, it is also possible that forecast errors become more correlated with shared FDS. Inconsistent with forecasts converging toward the truth, we find a significant increase in the correlation of forecast errors for analysts who share FDS, both in terms of the direction and magnitude of forecast errors.

The preceding analyses are conducted at the analyst or analyst-pair level. A related question is how the effects of FDS aggregate to the consensus level. The fact that shared FDS result in correlated forecast errors could result in less accurate consensus forecasts. Determining how FDS affect consensus estimates is important given the central role of consensus estimates in capital markets and academic research. We find that consensus accuracy increases when the underlying analysts have access to more FDS but decreases when there is more overlap in FDS.[9]

[9] If shared FDS results in more shared information sources among analysts, we expect variation in FDS overlap to affect stock return comovement between firms. In the Internet Appendix, we find increased FDS overlap to be associated with greater return comovement.

Our results contribute to several literatures. Most directly, we introduce a novel approach to understanding variation in FDS among analysts using a measure that is available for a large population of analyst reports.[10] We provide descriptive evidence on the array of FDS across brokerages and the effects of FDS on analyst outcomes. Despite the potential importance of FDS to analyst output, to our knowledge ours is the first paper to investigate specific FDS available to analysts and the effects on analyst research. Further, our results should help inform brokerages, investors, and regulators on the potential benefits of FDS and contexts in which FDS are likely to be most beneficial. We believe there is potential for important follow-on research examining determinants and effects of FDS, as well as examining FDS as a potential omitted correlated variable in studies of analyst-specific determinants of forecast properties.

Second, we explore the important effects of overlap in FDS on analyst and market outcomes. In recent years, there has been a substantial increase in the availability of FDS and in concentration among major FDS providers. While the growth and market concentration among FDS will likely continue, the effects are unclear. We provide novel evidence on the effects of FDS on individual analyst output, cross-brokerage correlations, and market consensus forecasts.

Finally, we contribute to an emerging literature examining capital market participants' access to data sources. For example, papers consider the role of conference calls, site visits, and cell phone tracking data. FDS are likely one of the most important sources of information for individual analysts but are outside of their direct control. We add to a growing literature opening the "black box" of analyst research (Bradshaw, 2011; Brown et al., 2015) by developing a new approach to identifying FDS and investigating the effects on analyst research.

[10] We plan to make our data on brokerage subscriptions and analyst sources publicly available once the publication process is complete.

2. Institutional Setting

2.1 Related Literature and Research Questions

The general topic of how analysts arrive at their forecasts has been a research focus since at least the 1980s (Barry and Brown, 1985). Information used by financial analysts is typically divided between private and public. Prior literature suggests that analysts enhance their research with private access to management (Bowen et al., 2002; Mayew et al., 2013; Green et al., 2014), site visits (Cheng et al., 2016), and government agency FOIA requests (Klein et al., 2020). More recently, researchers have begun to focus on novel “alternative” data sources such as satellite imagery, credit card expenditures, and cellphone data (Cheng et al., 2016; Klein et al., 2020; Chi et al., 2024). In terms of public disclosure, research suggests that analysts who access SEC filings via EDGAR produce more accurate forecasts (Gibbons et al., 2021), that analysts use public, non-financial information in their forecasts (Simpson, 2010), and that analysts’ ability to incorporate newly mandated disclosure depends on the complexity of the disclosure (Plumlee, 2003).

While prior literature highlights the role of a range of inputs on analyst research, perhaps the most fundamental source of information for analysts are brokerage FDS. Because FDS are typically brokerage-wide and costs (both direct and indirect) tend to be high, the subscription choice is likely to be determined at the brokerage level and is largely exogenous from the standpoint of individual analysts. Further, FDS are important to individual analysts because they provide a base level of data that is easily accessible as an input into their models and providers of FDS typically provide training in the use of their products. Even for basic inputs such as financial statement data, analysts are unlikely to refer to the original source but, rather, rely on curated data developed by FDS. Prior research suggests that major financial data providers differ in terms of, for example, the level of aggregation, treatment of non-GAAP earnings, and even their computation of "street earnings" for comparison to consensus forecasts (e.g., Bochkay et al., 2022; Larocque et al., 2023).[11]

Further, brokerage pairs likely vary in the overlap of their FDS. As noted earlier, the industry has become relatively concentrated with the top five data providers comprising nearly 50%. FDS tend to be expensive, so brokerages choose their FDS carefully to gain an advantage over competitors. As a result, for a given brokerage pair, subscriptions likely overlap for only a subset of FDS. Overall, a brokerage's FDS can affect the quality of their analysts' output and the effect of a given data subscription is influenced by a range of factors including the nature of the data (e.g., general or specialized), other FDS to which the brokerage subscribes, analysts' access to private information and attributes of the individual analyst (e.g., experience and busyness). Further, commonalities among pairs of analyst forecasts (e.g., correlations in point estimates, report content, and forecast errors) potentially reflect overlap in their FDS. At an aggregate level, the accuracy of consensus forecasts likely reflects the quantity and overlap in FDS to which the brokerages covering the firm subscribe.

Given the potential importance of FDS to analyst research, correlations across analyst reports, and consensus forecasts, we believe that FDS are an important unexplored input into the analyst production function. Our goal is to take a first step in that direction exploiting FDS disclosed in analyst reports. We begin by providing descriptive evidence on the prevalence and range of FDS as well as the overlap across brokerages. Then we focus on three research questions.[12]

RQ1: *Does FDS availability affect analyst forecast accuracy?*

[11] Bochkay et al. (2022) also provide important evidence on how data providers, such as Thomson Reuters, are capable of influencing analyst research outputs. Specifically, the authors find that changes to the way Thomson Reuters calculates exclusions to GAAP earnings can lead to timelier, more accurate, and less dispersed analyst forecasts.

[12] While we use causal language in stating our research questions, we recognize that causal inference is difficult empirically. We apply a variety of approaches to suggest causal inference but recognize that no approach is perfect.

If incremental FDS provide useful information to analysts, we expect a positive association between the number of FDS available to an analyst and forecast accuracy. While a positive relation between the number of FDS and forecast accuracy is not necessarily surprising, this analysis adds veracity to our brokerage-level subscription data and confirms the importance of FDS to analyst output. Additionally, given the lack of prior research, we view the magnitude and pervasiveness of the effect to be open empirical questions. Perhaps more importantly, we explore the types of FDS for which the relation is strongest and the characteristics of analysts who appear to benefit most.

RQ2: *Does overlap in FDS affect convergence of analyst opinions?*

Brokerages frequently overlap in terms of their FDS and the financial data industry is increasingly concentrated. As argued in the previous section, the effect of overlap in FDS on correlations in opinions across analysts is an empirical question and is, to our knowledge, unexamined in existing research. Our interest is in documenting the extent to which subscription overlap affects correlation in analysts' forecasts, forecast errors, recommendations, and narrative content, and what characteristics of financial data providers, brokerages, and analysts affect the extent of convergence.

RQ3: *Do the quantity and overlap of FDS affect firm-level consensus forecasts?*

Consensus forecasts are a central input in assessing firm performance and implementing valuation models. To the extent that brokerages differ in terms of access to FDS and overlap with other brokerages, we expect effects on consensus forecasts. For example, to the extent that FDS improve individual analyst forecast accuracy, we expect greater consensus accuracy for firms covered by analysts with more FDS. If, however, shared FDS across analysts affects forecast independence among analysts, consensus accuracy may decrease with greater overlap in FDS across brokerages.

2.2 Data and Sample

We begin by extracting the data source references from a sample of 595,642 analyst reports obtained from Thomson ONE, issued during the years 2008-2017.[13] Source referencing is very common in analyst reports; approximately 96% of reports reference the data sources used when preparing the report, and 90% of reports follow the "source:" labeling convention.[14] We extract the 100 characters of text following the reference to "source:" within each report. We then evaluate the most common sources and develop regular expressions to extract the precise source names for the top 100 sources in our sample.[15]

Next, we link the analyst reports and resulting source information to the I/B/E/S detail file. For parsimony, we retain only sources mentioned by at least five brokerages. Using the identified data source references, we construct a panel of brokerage months that includes the active sources for each brokerage for a given month. If any analyst at a brokerage mentions a specific source, we assume this source is available to all analysts at the brokerage for three calendar months before and after the month in which the source is mentioned. Our focus is on brokerage-level FDS for three key reasons: (1) these FDS constitute the data to which analysts have access; (2) our conversations with analysts

[13] Although the financial data vendor market has seen recent consolidation, all major mergers occurred after our sample period.

[14] We randomly sampled 100 analyst research reports and found that (1) analysts almost always cite data sources when preparing reports (96%), and (2) analysts follow similar conventions when citing sources (i.e., 90% followed the "source:" labeling convention). 6% of the random sample reports referenced sources in various ways that are challenging to capture programmatically. For example, one wrote, "The information on which the analysis is based has been obtained from sources believed to be reliable such as, for example, the company's financial statements filed with a regulator, company website, Bloomberg and any other relevant press source." We exclude brokerage reports that do not follow the "source:" labeling convention to mitigate source disclosure selection concerns.

[15] We selected the top 100 sources to make the research process more feasible (i.e., constructing 100 useful regular expressions vs. constructing 3,000+ useful regular expressions). To identify the top 100 sources, we randomly selected 5,000 "Source:" reference examples and had two RAs manually identify the sources referenced. We then identified the most common sources referenced among the random sample. While adding additional sources to our list might reduce measurement error, we are unaware of reasons focusing on common sources would induce bias in our results.

indicated that FDS are a brokerage-level decision; and (3) focusing on FDS has the advantage of mitigating analyst-report-level self-selection concerns.[16]

For brokerages with non-missing information on FDS, we retain the last one-year-ahead annual earnings forecast issued by each analyst ending at least a month before the covered firm's fiscal year-end date from the I/B/E/S detail file (Clement, 1999). We require firms to have positive book-to-market ratios and non-missing forecast values and timestamps. We further require the necessary data to calculate control variables, as described below. Our final sample consists of 213,523 analyst forecasts and 1,337,709 analyst forecast pairs.

2.3 Institutional Descriptive Statistics

Table 1 Panel A provides descriptive evidence on the 15 most common FDS in our sample based on the total number of citing brokerages (e.g., Bloomberg, FactSet, S&P Capital IQ, Thomson Reuters).[17] Table 1 Panel B reports comparisons of source references across analysts within the same brokerage. Consistent with FDS decisions being made at the brokerage level, we find much higher overlap in source use amongst analysts employed by the same brokerage versus randomly assigned analysts at different brokerages. For the median dataset, the probability that analysts at the same brokerage use the same dataset is 89.4% versus 1.9% for randomly-selected analysts who do not share a brokerage, providing strong evidence that FDS are brokerage-level subscriptions. Table 1 Panel C reports a data source transition matrix. Conditional on subscribing (not subscribing) to an FDS, brokerages have an 85.44% (96.38%) likelihood of subscribing (not subscribing) to that FDS the following year, indicating that subscription decisions tend to be relatively persistent.

[16] While we focus on data *available* to analysts, we could have examined the sources *referenced* in each report. In untabulated analyses, we investigated referenced sources and find similar results both in terms of individual analyst forecast accuracy and forecast similarity between analysts.

[17] Some FDS, such Bloomberg, offer a customized subscription model where users select specific datasets that they would like access to. This type of within FDS customization would only bias against finding results (i.e., we would attribute either data access or a connection across brokerages when there is none).

Table 2 Panel A provides basic descriptive statistics on the primary variables used in our tests of forecast accuracy. The median brokerage subscribes to about 7 FDS and roughly 13% of the analysts are classified as All-Star analysts. Table 2 Panel B provides descriptive statistics on the primary variables used in our pairwise tests that examine shared data subscriptions among analysts. Approximately 11% of forecasting pairs are issued by analysts with similar experience, 9% have similar brokerage resources, and 13% have similarly sized analyst portfolios.

3. Brokerage Subscriptions and Analyst Forecasts

3.1 Empirical Design and Results

We first investigate whether access to more FDS is associated with greater accuracy of analysts' forecasts using the following empirical model:

$$Accuracy_{a,f,t} = \alpha_1 NumSubscriptions_{a,t} + \alpha \text{Controls}_{a,f,t} + \beta \text{Fixed Effects}_{f,t} + \varepsilon_{a,f,t} \tag{1}$$

a indexes unique analysts, *f* indexes covered firms, and *t* indexes years. *Accuracy* is the absolute value of the analyst's forecast minus the covered firm's actual earnings, scaled by stock price two trading days prior to the forecast issuance date, multiplied by negative one.[18] Higher values of *Accuracy* indicate more accurate forecasts. We are interested in the coefficient on *NumSubscriptions*, which is the number of FDS to which a brokerage subscribes at time t. A positive coefficient is consistent with greater access to FDS leading to more accurate forecasts.

We include fixed effects and controls to isolate the association between the number of FDS to which an analyst has access and forecast accuracy. First, we control for time-varying characteristics of the covered firm through the inclusion of covered firm-year fixed effects. Additionally, we control for a vector of variables including the forecast's horizon, analyst experience (number of years on

[18] A potential concern is that analysts may differ in terms of the earnings they are forecasting (e.g., EPS forecasts and actuals in I/B/E/S can include non-GAAP adjustments), which could confound inference. Results are very similar if we replace EPS forecasts with sales forecasts, which are less prone to adjustment (see Internet Appendix).

I/B/E/S), brokerage resources (number of analysts employed by the brokerage), and analyst busyness (number of stocks the analyst covers) (Clement, 1999). To mitigate the influence of outliers, we decile rank all continuous variables each year and cluster standard errors by firm-year.[19] Decile ranking also facilitates comparisons of effect sizes across variables in our regressions.

Table 3 Panel A presents our results. In column 1, we include a baseline model with no fixed effects and a vector of control variables as described previously. In column 2, we include firm-year fixed effects and control variables. Across both columns, we find a positive and highly significant coefficient on *NumSubscriptions*. Overall, this result suggests that, as analysts have access to more FDS within their brokerages, forecast accuracy improves. In terms of the control variables, the coefficients are generally consistent with prior research (Drake et al., 2020; Fang and Hope, 2021; Huang et al., 2022). Forecasts issued at longer horizons and by larger brokerages tend to be less accurate, while forecasts issued by more experienced analysts who cover more firms tend to be more accurate (Table 3 column 2).[20] Focusing on column 2, which includes *Firm-Year* fixed effects so the comparison is across analysts following a given firm in a given year, access to FDS is at least as important as analyst experience, busyness, and brokerage size in explaining forecast accuracy.

3.2 Robustness

We consider several robustness analyses to rule out alternative explanations related to analyst and/or brokerage self-selection that could influence our findings. First, if skilled analysts select into brokerages with more FDS, this could manifest in heightened forecast accuracy for such analysts. To mitigate this concern, we re-estimate model (1) and include analyst fixed effects, which control for unobserved analyst heterogeneity, such as skill or talent. Second, brokerages that subscribe to more data providers could have a culture or business model that particularly values forecast accuracy. To

[19] Results are robust to clustering along alternative dimensions, including at the brokerage level.
[20] Clement (1999) documents a positive relation between forecast accuracy and brokerage size, whereas more recent studies (e.g., Drake et al., 2020; Fang and Hope, 2021; Huang et al., 2022) document a negative relation.

alleviate this concern, we re-estimate model (1) with brokerage fixed effects, which control for stable brokerage characteristics, such as resources, business model, and corporate culture. Third, to further reduce concerns related to analyst and/or brokerage selection, we include analyst-brokerage-pair fixed effects, which control for the analyst-brokerage relationship and exploit variation in FDS *within* the pair. Table 3 Panel B reports these results. Across each specification, we continue to observe a positive and significant coefficient on *NumSubscriptions.*

A potential concern is that other changes may occur at brokerages that lead to both new FDS and increased accuracy (e.g., the brokerage might be growing). The previous analyses are robust to *Analyst x Brokerage* fixed effects, so the results do not simply reflect changes in the analysts employed at a brokerage or changes in the mix of brokerages. Further, we control for brokerage size, and the negative coefficient on this variable suggests that increases in brokerage size should work against our findings. Also, the cross-sectional results based on analyst characteristics reported below are most consistent with an information story.

In untabulated robustness, we include controls for brokerage growth, brokerage age, and the proportion of All-Star analysts to capture the possibility that our results might reflect brokerage growth, age, prestige, or clientele. Results are robust. Konfound analysis suggests that any omitted variables would need to have a larger effect than brokerage size, forecast horizon, busyness, and experience (and over 85% of the observations would have to be replaced with an effect of zero to invalidate the inferences).

3.3 Cross-sectional Analysis

We conduct several tests to better understand the mechanisms by which FDS influence forecasting accuracy. The first group of tests evaluates our empirical findings across various categories of data subscriptions, using the following model:

$$Accuracy_{a,\mathrm{f,t}} = \alpha_1 Category1_{a,\mathrm{t}} + \alpha_2 Category2_{a,\mathrm{t}} + \alpha\mathrm{Controls}_{\mathrm{a,f,t}} + \beta\mathrm{Fixed\ Effects}_{\mathrm{f,t}} + \varepsilon_{\mathrm{a,f,t}} \quad (2)$$

Category1 and *Category2* represent placeholders for different categories of data subscriptions, and we include the same control variables and fixed effects as in model (1).

Our first test using this framework investigates whether our results primarily stem from major FDS (S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar) as compared to less common FDS (those not in the top five in terms of market share). The two independent variables of interest are i) *MajorProvider*, which measures the number of FDS to which the brokerage subscribes among the five major data providers and ii) *MinorProvider*, which measures the number of FDS among the minor data providers. Table 4 column 1 presents these results. The point estimate of the coefficient on *MinorProvider* is larger than that of *MajorProvider*, suggesting that nonstandard FDS are particularly important to accuracy, although the difference between the coefficients is insignificant. However, both major and minor FDS have significant effects on forecast accuracy, suggesting that even variation in major FDS is important.

Our second test using the above framework compares the role of FDS to that of public data sources such as EDGAR or other company filings. Here the two independent variables of interest are i) *NumSubscriptions*, as defined previously and ii) *PublicSources*, which represents brokerage references to public sources (e.g., EDGAR filings, conference calls and other company disclosures). Table 4 column 2 reports the results. The coefficient on *NumSubscriptions* remains significantly positive and of similar magnitude relative to our primary analysis, suggesting that the information in FDS is incrementally important relative to public disclosure, while the coefficient on *PublicSources* is insignificant.

To better understand the mechanisms underlying the relation between FDS and forecast accuracy and the characteristics of individual analysts who benefit most from FDS, we estimate the following interaction model:

$$Accuracy_{a,\mathrm{f,t}} = \alpha_1 Interaction_{a,\mathrm{t}} \cdot NumSubscriptions_{a,\mathrm{t}} + \alpha_2 Interaction_{a,\mathrm{t}} + \alpha_3 NumSubscriptions_{a,\mathrm{t}} + \alpha \mathrm{Controls}_{\mathrm{a,f,t}} + \beta \mathrm{Fixed\ Effects}_{\mathrm{f,t}} + \varepsilon_{\mathrm{a,f,t}} \quad (3)$$

Interaction is a placeholder for various interaction variables, and we include all controls and fixed effects as in model (1). The interaction variables test the effects of various analyst characteristics. First, we assess whether analysts with greater access to management are less reliant on additional FDS. The interaction variable, *AllStar,* is an indicator equal to one if the analyst receives the All-Star designation during the year (Mayew, 2008; Green et al., 2014). Second, analysts with limited experience likely benefit more from access to FDS. The interaction variable, *LowExperience*, is an indicator variable equal to one if the analyst is in the lowest two deciles of general experience. Third, we expect analysts with heavy workloads to benefit more from additional FDS. The interaction variable, *HighBusyness*, is an indicator equal to one if the number of firms the analyst covers is in the highest two deciles of portfolio size. Fourth, we evaluate whether longer horizon forecasts benefit more from access to additional FDS because analysts have had less time to gather private information. The interaction variable, *LongHorizon,* is an indicator equal to one if the forecast's horizon is in the highest two deciles (i.e., furthest away from the firm's fiscal period end date).

Table 4 columns 3-6 report the results of the interaction tests. As predicted, All-Star analysts appear to benefit less from additional FDS, consistent with these analysts having greater access to soft information and relying less on their brokerages' FDS. In contrast, FDS access appears more important for less experienced analysts, who likely lack the expertise and networks to incorporate private information. Results are also stronger for analysts with heavier workloads, suggesting that analysts with less time to independently source information on individual firms rely more heavily on brokerage FDS. Finally, additional FDS tend to provide more benefit for longer horizon forecasts, consistent with less time to gather private information early in the forecasting cycle.[21] Collectively, these findings highlight the types of settings and analyst characteristics that are associated with greater benefit from FDS. In addition, the consistency of the findings with various cross-sectional predictions helps mitigate concerns over potential endogeneity (Glaeser and Guay, 2017).

4. Subscription Similarity and Forecast Similarity

4.1 Empirical Design and Results

We next investigate whether shared access to FDS affects the attributes of analysts' forecasts across brokerage pairs. We match each analyst forecast for firm *f* with fiscal period end date *t* to all other analyst forecasts issued for the same firm and fiscal period end date. We retain one unique pairing between each analyst forecasting for firm *f* with fiscal period end date *t*. We initially consider three distinct attributes: (1) forecast similarity, (2) forecast timing, and (3) forecast boldness. We use the following model to examine whether FDS similarity is associated with forecast attributes:

$$\text{SimilarAttribute}_{p,f,t} = \alpha_1 \mathit{SubscriptionSimilarity}_{p,t} + \alpha\text{Controls}_{p,f,t} + \beta\text{Fixed Effects}_{f,t} + \varepsilon_{p,f,t} \qquad (4)$$

where *p* indexes unique analyst forecast pairs, *f* indexes the covered firm, and *t* indexes the year. Our primary independent variable of interest is *SubscriptionSimilarity,* which is the number of FDS that both analysts in the pair have access to at their respective brokerages (i.e., the number of overlapping subscriptions), scaled by the number of all possible data subscriptions. We decile rank this variable each year. Thus, higher values of *SubscriptionSimilarity* indicate more overlap in FDS for both analysts in the pair.

We consider three dependent variables that represent attributes of the analysts' forecasts. First, *SimilarForecast* is the absolute value of the difference between the two forecasts in each forecast pair. We scale this difference by the firm's stock price two trading days prior to the first analyst's

[21] We also find that there are diminishing marginal returns to accuracy when adding new data subscriptions. For example, if we include a squared term for the number of subscriptions, it loads negatively (untabulated).

forecast date in the analyst pair and multiply by negative one. We decile rank the resulting value each year. Higher values of *SimilarForecast* indicate more similar forecasts between the two analysts in the pair. A positive coefficient on *SubscriptionSimilarity* (α_1) would be consistent with analysts' earnings point estimates becoming more similar as the analysts have more overlapping FDS.[22]

Second, we examine *SimilarTiming,* which measures how clustered analysts' forecasts are in event time. We decile rank analysts' forecast horizons each year and set *SimilarTiming* equal to one if the forecast horizons in the analyst pair are in the same decile. Forecast horizon is the number of days between the covered firm's fiscal period end and the forecast issuance. A positive coefficient on *SubscriptionSimilarity* (α_1) would be consistent with analysts' forecast horizons becoming more similar as they have more overlapping FDS.

Third, we examine *SimilarBoldness*, which is equal to one if both forecasts in the analyst pair are similar in terms of boldness (i.e., both analysts are bold or both analysts are not bold). We follow Clement and Tse (2005) in calculating forecast boldness, where bold forecasts are those with values that exceed (or are below) both the analyst's prior forecast for the firm and the prevailing consensus forecast at the time; all remaining forecasts are classified as nonbold.[23] A positive coefficient on *SubscriptionSimilarity* (α_1) would be consistent with analysts' forecast boldness becoming more similar as the analysts have more overlapping FDS.

We include fixed effects and controls to better isolate the relationship between shared FDS and analyst forecast attributes. First, we include time-varying covered firm controls including *BTM* (book-to-market ratio), *MVE* (market value of equity), and *ROA* (return on assets). In additional specifications, we introduce firm-year fixed effects to mitigate the impact of stable or time-varying characteristics of covered firms. Because this specification includes a unique fixed effect for each

[22] Results are consistent if we replace earnings forecast similarity with sales forecast similarity (see Internet Appendix).
[23] In cases where we cannot calculate an analyst's forecast boldness because two sequential forecasts are required, we set *SimilarBoldness* equal to zero. Results are robust to dropping these cases.

firm-year, it neutralizes time-varying firm characteristics, rendering firm-year controls redundant. Overall, including firm-year fixed effects is particularly robust, as it ensures that any influence that firm attributes might exert on forecasting behavior within that specific timeframe is held constant.

Next, we control for various characteristics of the analyst and brokerage that vary within the fixed effect structure and which prior studies have shown relate to the attributes of analyst forecasts (Clement, 1999; Cowen et al., 2006). First, we control for whether the analysts have similar forecasting experience. *SimilarExperience* is set equal to one if both analysts in the pair have a similar number of years of experience forecasting on I/B/E/S. Analysts are assumed to have similar forecasting experience if both are in the same experience decile rank, based on the total years forecasting on I/B/E/S as of the prior year, calculated annually. Second, we control for whether the analysts are employed by brokerages with similar resources. *SimilarResources* is set equal to one if both analysts in the pair are employed by brokerages of similar size. Brokerages are assumed to have similar resources if each brokerage is in the same decile rank, based on the number of analysts employed at the brokerage as of the prior year, calculated annually. Third, we control for whether the analysts are similar in terms of busyness. *SimilarBusyness* is set equal to one if both analysts in the pair cover a similar number of firms on I/B/E/S. Analysts are assumed to be similarly busy if both are in the same decile rank, based on the number of covered firms as of the prior year, calculated annually.[24] We cluster standard errors at the firm-year level in each of our estimations.[25]

Table 5 presents the empirical results examining the association between overlap in FDS and forecast similarity. For each dependent variable, we estimate our analysis in two ways: (1) with controls but no fixed effects (Jennings et al., 2023) and (2) with controls and fixed effects.[26] As

[24] Our inferences remain unchanged if we decile rank the absolute differences in analysts' experience, brokerage size, and busyness within each analyst pair.
[25] Results are robust to clustering along alternative dimensions, including at the brokerage-pair level.
[26] We also estimate a specification without controls (Whited et al., 2022) and fixed effects, and our inferences remain unchanged (untabulated).

mentioned previously, we examine three dependent variables that capture unique attributes of forecast similarity based on point estimates, forecast timing, and forecast boldness. Including *Firm x Year* fixed effects makes firm-year controls redundant, so *BTM, MVE,* and *ROA* are dropped when we implement this fixed effects structure.

Table 5 columns 1 and 2 report results for point forecast similarity. Across each column, we find a positive and statistically significant coefficient on *SubscriptionSimilarity*. This suggests that forecast point estimates become more similar as analysts increasingly share the same FDS.[27] In terms of economic magnitude, a one standard deviation increase in FDS similarity equates to about a 15.09% increase in forecast similarity, relative to the mean.[28] FDS similarity has a larger effect size than either analyst experience or busyness, confirming that FDS similarity has an important effect on forecast similarity.

Table 5 columns 3 and 4 report results for forecast timing. We find a positive and statistically significant coefficient on *SubscriptionSimilarity,* indicating that shared FDS not only influences forecast point estimates but also influences the timing of the forecasts. These results suggest that the shared timing of information arrival associated with shared FDS influences the *processing* of information and the timing of forecasts. A one standard deviation increase in FDS similarity equates to a 3.66% increase in the probability of sharing a similar horizon decile.[29] Again, the effect of FDS similarity is larger than that of analyst experience or busyness.

Table 5 columns 5 and 6 report results for forecast boldness. Across each column, we find a positive and statistically significant coefficient on *SubscriptionSimilarity*. A one standard deviation increase in FDS similarity equates to a 3% increase in the probability that both analysts are similar in

---

[27] While we consider forecast horizon to be an outcome of interest (*SimilarTiming*), results in Table 5 columns 1 and 3 are robust to constraining to analyst pairs that release their forecasts on exactly the same day.

[28] $0.053 \cdot 3.33 \cdot (0.0089/0.0104) = 15.09\%$; 0.0089 and 0.0104 are the mean and average decile change in forecast similarity (unranked), respectively.

[29] $0.011 \cdot 3.33 = 3.66\%$.

the boldness of their forecasts.[30] Again, the effect of similarity in FDS is larger than that of analyst experience or busyness. Overall, our collective evidence is consistent with similarity in FDS influencing the similarity of analysts' forecasts in terms of point estimates, timing, and boldness.

4.2 Robustness Tests

While our pairwise design and *Firm x Year* fixed effects alleviate various concerns with endogeneity, in additional analyses we address potential alternative explanations. Specifically, although we control for brokerage similarity, other similarities across brokerages may correlate with both similarity in FDS and forecast similarity. In the next test, we hold brokerage-pairs constant and exploit inter-temporal variation in *SubscriptionSimilarity*. To the extent that correlated omitted variables are fixed between brokerage pairs or are uncorrelated with *changes* in similarity in FDS, such inter-temporal variation in FDS similarity can help rule out these alternative explanations. We estimate a specification including brokerage pairwise fixed effects by creating a distinct fixed effect for each brokerage pair in our sample. Results reported in Table 6 columns 1, 3, and 5 are robust to inclusion of brokerage pair fixed effects, suggesting that fixed pairwise attributes between brokerages do not drive our findings.

To refine these inferences, we note that changes in data subscription similarity measured at the analyst level, holding brokerage pairs constant, can come from two sources: (1) brokerages changing their FDS and (2) analysts changing brokerages. To investigate the effects of intertemporal changes in analyst access to FDS at brokerages, we impose a more robust fixed effects design. Specifically, we interact brokerage-pair fixed effects with analyst-pair fixed effects, thus constraining variation to be for analyst pairs with no employment changes. Another benefit of adding analyst pair fixed effects is that it helps mitigate concerns that fixed similarities between *analysts* are driving our

[30] $0.009 \cdot 3.33 = 3.00\%$.

results. Table 6 columns 2, 4, and 6 report the results. We find that when brokerages change their FDS intertemporally, the results are consistent with similarity in FDS influencing forecast similarity.

Finally, to capture the effects of *employment change* on data subscription similarity, we control for similarity in FDS at analysts' former employers in the current time period (*OldSubscriptionSimilarity*). To the extent that current access to FDS drives our inferences, we expect to see effects from *SubscriptionSimilarity* and not from *OldSubscriptionSimilarity*. A unique benefit from this employment change analysis is that similarity in FDS *between* brokerages is unlikely to change systematically when hiring new analysts.[31] Table 6 Panel B reports results from this analysis. We continue to find a positive and significant coefficient on *SubscriptionSimilarity*, while the coefficient on *OldSubscriptionSimilarity* is insignificant. We also observe that *SubscriptionSimilarity* is statistically different from *OldSubscriptionSimilarity*. Overall, this reinforces our main conclusion that brokerage data subscriptions influence analyst forecasting behavior and helps mitigate a variety of omitted variable bias concerns.[32]

4.3 Cross-sectional Analysis

To better understand the mechanisms by which similarity in FDS influences analyst research, we conduct cross-sectional tests exploiting variation in the underlying nature of the FDS and in analyst characteristics. We first test whether the results primarily stem from major FDS (S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar) compared to "minor" FDS with lower market share. Given increasing concentration in market share among the top data providers, it is important for regulators, investors, and brokerages to understand the effects of industry consolidation

[31] While analysts select which brokerages to work for, and those decisions may be based in part on subscription access, such selection preferences are unlikely to *change* between brokerages within analyst.

[32] Several FDS provide consensus analyst forecasts, which might encourage analyst herding. Our results are robust to dropping such FDS from our analysis (per the listing in Larocque et al., 2023).

on forecast attributes. In addition, it provides insight into the generalizability of our results across types of FDS. We estimate the following model:

$$\text{SimilarAttribute}_{p,f,t} = \alpha_1 \textit{MajorSubscriptionSimilarity}_{p,t} + \alpha_2 \textit{MinorSubscriptionSimilarity}_{p,t} + \alpha\text{Controls}_{p,f,t} + \beta\text{Fixed Effects}_{f,t} + \varepsilon_{p,f,t} \quad (5)$$

For each forecast pair, *MajorSubscriptionSimilarity* measures the brokerages' FDS similarity among the five major data providers, while *MinorSubscriptionSimilarity* measures the brokerages' FDS similarity among the more minor FDS. We include all controls and fixed effects as in model (4).

Column 1 in Table 7 panels A through C present results for each forecast attribute. The effect of FDS similarity on forecast similarity is evident for both major and minor FDS. Interestingly, effect sizes for major FDS are at least as large as those for minor FDS. Collectively, these results highlight two key insights. First, major FDS appear to play an important role in shaping the documented effects on forecasting convergence, which is important given their substantial market share. Further, the fact that overlap in major FDS affects forecast similarity suggests that they do not simply provide homogenous background information (e.g., financial statement data, stock prices, etc.), but rather provide differentiated information. This result is also consistent with research suggesting that FDS differ in their level of aggregation, treatment of non-GAAP earnings, and computation of "street earnings" relative to consensus forecasts (e.g., Bochkay et al., 2022; Larocque et al., 2023). Second, FDS similarity effects generalize across both major and minor data providers, suggesting the results are not simply a byproduct of access to less conventional FDS.

While we focus specifically on similarity in FDS amongst the brokerages in our sample, it is likely that some analysts also have preferential access to soft information (e.g., via a relationship with management). Accordingly, we next examine how soft information might affect analysts' anchoring on hard data from FDS. We estimate the following model:

$$\text{SimilarAttribute}_{p,f,t} = \alpha_1 \textit{SubscriptionSimilarity}_{p,t} \cdot \textit{AllStars}_{p,t} + \alpha_2 \textit{SubscriptionSimilarity}_{p,t} + \alpha_3 \textit{AllStars}_{p,t} + \alpha \text{Controls}_{p,f,t} + \beta \text{Fixed Effects}_{f,t} + \varepsilon_{p,f,t} \quad (6)$$

As in our accuracy analyses, we use analysts' All-Star status as a proxy for soft information access (Mayew, 2008; Green et al., 2014). Specifically, we interact *SubscriptionSimilarity* with an indicator, *AllStars*, that is set equal to one if both analysts in the pair receive the All-Star designation during the year. To the extent that All-Star analysts have greater access to soft information, the effect of FDS should be attenuated as reflected in a negative coefficient on *SubscriptionSimilarity · AllStars* (α_1). We include all controls and fixed effects as in model (5).

Column 2 in Table 7 panels A through C present results for each forecast attribute. Across each of our three main dependent variables (*SimilarForecast*, *SimilarTiming*, and *SimilarBoldness*), we find a negative and significant coefficient on *SubscriptionSimilarity · AllStars*. Overall, this suggests that All-Star analysts are less influenced by data provider similarity, which is consistent with these analysts having greater access to soft information and therefore relying less on FDS.

In addition, brokerages have access to varying levels of FDS, with the importance of specific FDS likely to vary based on the availability of other FDS. For instance, analysts at brokerages with more FDS can rely on a more diverse set of inputs. Following this intuition, we expect the effect of sharing FDS to be weaker for analysts with access to more FDS. To examine this, we rank the brokerages each year based on the number of available FDS. If the analysts in a given pair are both employed by brokerages in the upper 50th percentile of FDS, we set the variable *HighSubscriptionAccess* equal to one and zero otherwise. We interact *HighSubscriptionAccess* with *SubscriptionSimilarity* and include these variables in a modified version of model (5). A negative and significant coefficient on *SubscriptionSimilarity · HighSubscriptionAccess* would be consistent with our results being attenuated when analysts have access to a greater number of FDS.

Column 3 in Table 7 panels A through C present these results. Across each of our three main dependent variables (*SimilarForecast, SimilarTiming,* and *SimilarBoldness*), we find a negative and significant coefficient on *SubscriptionSimilarity · HighSubscriptionAccess.* Overall, this suggests that data subscription similarity is less important for analysts with access to a greater number of FDS.[33]

4.4 Analyst Report Content

While analyst forecasts are important, they form only a small portion of analyst reports. The broader content, including overall stock recommendations and narrative text, provides a more comprehensive view of the analysts' perspectives and the potential influence of FDS. Analysts might use FDS mainly for quantitative summaries or benchmarking purposes, leaving the content of reports relatively unaffected. By examining the similarities in stock recommendations and narrative content, we can better understand the breadth of the effects that overlapping FDS have on the homogenization of viewpoints among analysts.

To examine similarity in analysts' stock recommendations, we obtain the analysts' most recent outstanding recommendation that was active at the time their last annual forecast was issued, at least 30 days prior to the fiscal year end. If the recommendations are the same for both analysts in the pair, we set *SimilarRecommendations* equal to one and zero otherwise. To test whether sharing similar data subscriptions affects the written narrative of analyst reports, we analyze the cosine similarity between the analysts' written reports. Similar to our assessment of analyst recommendations, we focus on those reports that are issued surrounding the analysts' last annual forecasts, issued at least 30 days prior to the fiscal year end for each firm. We perform standard document cleaning procedures, including removing tables, stop words, and words containing

[33] Because larger brokerages generally have access to more FDS than smaller brokerages, in untabulated analyses we evaluate whether our results are robust to examining forecasts issued by analysts *only* at large brokerages. We find similar inferences as our main pairwise results reported in Table 5. Thus, while our results are attenuated for analysts with access to a greater number of sources, the effect is still present when examining only large brokerages.

numbers. Since many analyst reports include boilerplate disclosures at the end of the report (e.g., legalese), we also remove those sections of the report using common regular expressions. After doing so, we calculate a new variable, *SimilarReports*, which is the cosine similarity between the analyst reports in the pair, decile ranked by year. Higher values of *SimilarReports* indicate that the written content between the two reports is more similar.

The results in Table 8 show that *SubscriptionSimilarity* is significantly positively related to both recommendation and report text similarity. These findings reinforce the conclusion that overlapping FDS contribute significantly to the convergence of broader analyst report content, highlighting the influence of FDS on the homogeneity of analyst research.

5. Data Subscription Consequences

5.1 Correlated Forecast Errors

One significant consequence of shared FDS among analysts is the potential for correlated forecast errors. When multiple analysts rely on the same FDS, their forecasts tend to converge (Table 5), which could either reflect convergence toward the "truth" or correlation in forecast errors. Understanding how FDS contribute to this phenomenon is important as correlated forecasts can amplify the impact of errors on market perceptions, leading to less accurate market consensus views and potentially creating correlated risk when analysts rely on the same FDS.

We test this using model (4) with three new dependent variables. *SimilarError* equals one if both analysts in the pair have the same directional forecast error (i.e., both analysts over- or under-forecast actual earnings). A positive coefficient on *SubscriptionSimilarity* is consistent with shared FDS leading to correlated forecast errors. Next, we examine the error magnitude. *SimilarErrorMagnitude* is an indicator that is equal to one if both analysts in the pair have an absolute forecast error, scaled by actual announced earnings, that is in the same yearly decile rank. Finally, we examine the interaction between *SimilarError* and *SimilarErrorMagnitude* and form a new variable, *SimilarError&Magnitude,* which is equal to one if both analysts in the pair have the same directional earnings forecast error and the magnitude of the error is in the same decile rank.

Table 9 reports the results across each of these three outcome variables. Column 1 indicates that sharing FDS is associated with making identical directional forecasting errors (*SimilarError*). Column 2 shows that similar FDS are linked to forecast errors of comparable magnitudes. In column 3, the results reveal that shared FDS are related to both similar directional errors and similar error magnitudes. These findings suggest that reliance on the same FDS leads to correlated errors, implying that analyst forecasts are not independent when they rely on similar underlying FDS.

5.2 Consensus Forecast Accuracy

Consensus analyst forecasts constitute an important market metric in forecasting and benchmarking firm performance, used by both market participants and academics. Results in Table 3 suggest that a larger pool of FDS enriches the information available on a firm, potentially improving the precision of consensus forecasts. However, results in Tables 5 and 9 suggest that, when analysts share FDS, the benefits of this diversity and independence may be compromised. In line with this reasoning, the wisdom of crowds theory suggests that, as opinions are less diverse, the crowd becomes less "wise," leading to deterioration in the accuracy of the crowd forecast (Surowiecki, 2005). In this section, we explore how the number of FDS and the degree of overlap among analysts affect the reliability of consensus forecasts, offering insights into the effects of FDS quantity and information overlap. We estimate the following model:

$$ConsensusAccuracy_{f,t} = \alpha_1 NumConsensusSubscriptions_{f,t} + \alpha_2 AvgSubscriptionSimilarity_{f,t} + \alpha\text{Controls} + \beta\text{Fixed Effects}_{f,t} + \varepsilon_{f,t} \quad (7)$$

where *f* indexes firm and *t* indexes year. *NumConsensusSubscriptions* is the number of unique FDS available to consensus analysts, decile ranked by year. *AvgSubscriptionSimilarity* is defined as the average value of *SubscriptionSimilarity* across consensus analyst pairs, prior to its decile ranking, calculated at the firm-year level. The variable is then decile ranked by year. Increases in *AvgSubscriptionSimilarity* reflect less FDS diversity (i.e., analysts contributing to the consensus share more FDS as *AvgSubscriptionSimilarity* increases). The dependent variable, *ConsensusAccuracy,* is the absolute value of the difference between the covered firm's reported earnings and the median analyst consensus forecast, scaled by the firm's stock price from the most recent quarter, multiplied by negative one, and decile ranked by year. Higher values of *ConsensusAccuracy* indicate a more accurate consensus. If data subscription quantity (similarity) leads to a more (less) accurate consensus forecast, we would observe a positive (negative) coefficient on *NumConsensusSubscriptions*; α_1 (*AvgSubscriptionSimilarity*; α_2).

Given that this model includes only one observation at the firm-year level, we are unable to include firm-year fixed effects. Instead, we include firm and year fixed effects. We also control for time-varying characteristics of the analysts and forecasts that form the consensus. Specifically, we control for the average experience of analysts contributing to the consensus (*AvgExperience*), the average brokerage size for analysts contributing to the consensus (*AvgBrokerageSize*), and the average horizon for each forecast forming the consensus (*AvgHorizon*). We also include several additional control variables related to the covered firms that are associated with consensus forecast accuracy. We include the firms' book-to-market ratio (*BTM*), size (*MVE*), profitability (*ROA*), and an indicator for whether the firm reports a loss (*Loss*).

Table 10 reports the results. In column 1, we include firm fixed effects and the control variables. In column 2, we include firm and year fixed effects, along with the control variables. Across both columns, we find a positive and significant coefficient on *NumConsensusSubscriptions*, suggesting that as the group of consensus analysts has access to more FDS, the consensus accuracy improves. We also find a negative and significant coefficient on *AvgSubscriptionSimilarity*. Overall, this suggests that, as analysts' consensus forecasts utilize more overlapping FDS, the accuracy of the consensus forecast degrades.[34] This finding highlights the importance of FDS independence when forming consensus opinions.[35]

6. Conclusion

Financial data subscriptions (FDS) constitute costly and potentially important inputs into analyst research that have largely been unexplored in prior literature. We use a comprehensive and novel dataset of nearly 600,000 equity research reports to identify brokerage FDS. We evaluate the effect of FDS on the characteristics of individual analyst forecasts, correlation among forecasts across analysts, and effects on consensus forecasts. We provide consistent evidence that availability of FDS significantly enhances the accuracy of analyst forecasts, on par with analyst experience, busyness, and brokerage size. Analysts who benefit most from FDS are less experienced, busier (cover more portfolio firms), have fewer private information sources, and forecast over longer horizons.

However, the extent of overlap in brokerage FDS also appears to affect the similarity of analyst research across the brokerages. Specifically, analysts at brokerages with similar FDS exhibit greater similarity in their forecast values, timing, and boldness. These effects are attenuated when analysts have access to soft information or have access to a greater number of unique FDS. Furthermore, our results suggest that both large and small FDS (in terms of market share) influence analyst forecasting behavior. Evidence of the impact of shared FDS extends beyond analyst forecasts, also manifesting in increased similarity in forecast errors, stock recommendations, and report text. Finally, consensus analyst forecasts tend to be more accurate when the analysts covering the firm have access to more FDS but less accurate the more those FDS overlap across brokerages.

[34] Results are very similar for sales forecasts, with accuracy increasing in the total number of FDS but decreasing in FDS overlap (see Internet Appendix).

[35] A related question is whether the effects of shared FDS find their way into stock pricing as reflected in return co-movement. The Internet Appendix reports results for return correlation for pairs of firms as a function of shared FDS including a variety of fixed effects, indicating that return co-movement increases as analysts following firms share more FDS, consistent with pricing effects.

We believe our findings provide novel insights into an important, and thus far unexplored, input into analyst research—brokerage FDS. Prior research has largely focused on analyst-specific inputs while FDS, which are largely exogenous from the standpoint of the individual analyst, are likely a starting point and central input into analyst research. Our findings have implications for brokerages in understanding the potential benefits of additional FDS and the types of analysts that benefit most, but also the limitations associated with shared FDS across brokerages. Similarly, there are potential implications for investors and regulators interested in the effects of increasing FDS availability and concentration on analyst research and consensus forecasts.

In addition, our novel approach for identifying FDS and analyst sources suggests potential avenues for future research. For example, given that most analyst reports provide "source" information, it may be possible to more directly evaluate the inputs into analysts' research. Similarly, the ability to identify brokerage FDS opens the possibility of further exploring cross-brokerage variation in information resources and research priorities. Further, while we focus on aggregated data, there is room for research investigating determinants and effects of specific types of FDS.

Several caveats are worth noting. First, we focus on sell-side analysts. Although our results suggest that FDS influence various attributes of analyst forecasts, the dynamics we observe may not directly translate to other market participants. Second, we base our inferences on FDS disclosed in analyst reports. While most analyst reports indicate sources and we do not expect omitted brokerages to systematically bias inference given our analyst-pairwise research design, we cannot definitively rule out selection concerns. Lastly, and perhaps most importantly, while we attempt to rule out alternative explanations using analyst and brokerage pair fixed effects, changes in analysts' employment, and cross-sectional analyses, we acknowledge the potential for other explanations. However, alternative explanations would need to align with the totality of our results. At a minimum, we believe our results provide important initial evidence that we hope will encourage future research.

Bibliography

Al Bari, S. (2023). Financial Data Service Providers in the US (IBISWorld Industry Report OD5491). Retrieved from IBISWorld database.

Armstrong, C., Heinle, M. S., & Luneva, I. (2023). Financial Information and Diverging Beliefs. Review of Accounting Studies.

Barry, C., & Brown, S. (1985). Differential information and security market equilibrium. Journal of Financial and Quantitative Analysis, 20(4), 407-422.

Bernhardt, D., Campello, M., & Kutsoati, E. (2006). Who herds? Journal of Financial Economics, 80(3), 657-675.

Bloomberg (2017). Finding Novel Ways to Trade on Sentiment Data. Retrieved March 6, 2024, from https://www.bloomberg.com/company/stories/finding-novel-ways-trade-sentiment-data/.

Bloomberg (2023). Bloomberg Makes Alternative Data Accessible Alongside Traditional Financial Data. Retrieved March 6, 2024, from https://www.bloomberg.com/company/press/bloomberg-makes-alternative-data-accessible-alongside-traditional-financial-data/

Bochkay, K., Markov, S., Subasi, M., & Weisbrod, E. (2022). The roles of data providers and analysts in the production, dissemination, and pricing of street earnings. Journal of Accounting Research, 60(5), 1695-1740.

Bowen, R., Davis, A., & Matsumoto, D. (2002). Do conference calls affect analysts' forecasts? The Accounting Review, 77(2), 285-316.

Bradshaw, M. (2011). Analysts' forecasts: what do we know after decades of work? Available at SSRN 1880339.

Brown, L., Call, A., Clement, M., & Sharp, N. (2015). Inside the "black box" of sell-side financial analysts. Journal of Accounting Research, 53(1), 1-47.

Cheng, Q., Du, F., Wang, X., & Wang, Y. (2016). Seeing is believing: Analysts' corporate site visits. Review of Accounting Studies, 21, 1245-1286.

Chi, F., Hwang, B. H., & Zheng, Y. (2024). The use and usefulness of big data in finance: Evidence from financial analysts. Management Science.

Clement, M. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? Journal of Accounting and Economics, 27(3), 285-303.

Clement, M., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. The Journal of Finance, 60(1), 307-341.

Cowen, A., Groysberg, B., & Healy, P. (2006). Which types of analyst firms are more optimistic? Journal of Accounting and Economics, 41(1-2), 119-146.

Drake, M., Joos, P., Pacelli, J., & Twedt, B. (2020). Analyst forecast bundling. Management Science, 66(9), 4024-4046.

Fang, B., & Hope, O. K. (2021). Analyst teams. Review of Accounting Studies, *26*, 425-467.

Gibbons, B., Iliev, P., & Kalodimos, J. (2021). Analyst information acquisition via EDGAR. Management Science, 67(2), 769-793.

Glaeser, S., & Guay, W. R. (2017). Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017). Journal of Accounting and Economics, 64(2-3), 305-312.

Green, T. C., Jame, R., Markov, S., & Subasi, M. (2014). Access to management and the informativeness of analyst research. Journal of Financial Economics, 114(2), 239-255.

Huang, A. H., Lin, A. P., & Zang, A. Y. (2022). Cross-industry information sharing among colleagues and analyst research. Journal of Accounting and Economics, 74(1), 101496.

Jennings, J., J.M. Kim, J. Lee, & D. Taylor. "Measurement error, fixed effects, and false positives in accounting research." Review of Accounting Studies (2023): 1-37.

Klein, A., Li, T., & Zhang, B. (2020). Seeking out non-public information: Sell-side analysts and the Freedom of Information Act. The Accounting Review, 95(1), 233-257.

Kondor, P. (2012). The more we know about the fundamental, the less we agree on the price. The Review of Economic Studies, 79(3), 1175-1207.

Kothari, S.P., So, E. & Verdi, R. (2016). Analysts’ Forecasts and Asset Pricing: A Survey. Annual Review of Financial Economics, 8, 197-219.

Larocque, A., Watkins, J., & Weisbrod, E. (2023). Consensus? An Examination of Differences in Earnings Information Across Forecast Data Providers. Working Paper.

Markowitz, Harry M., and ANDRÉF PEROLD. "Portfolio analysis with factors and scenarios." The Journal of Finance 36, no. 4 (1981): 871-877.

Mayew, W. (2008). Evidence of management discrimination among analysts during earnings conference calls. Journal of Accounting Research, 46(3), 627-659.

Mayew, W., Sharp, N., & Venkatachalam, M. (2013). Using earnings conference calls to identify analysts with superior private information. Review of Accounting Studies, 18, 386-413.

Plumlee, M. (2003). The effect of information complexity on analysts' use of that information. The Accounting Review, 78(1), 275-296.

Refinitiv (2019). Refinitiv makes strategic investment in BattleFin and partners to incorporate alternative datasets within investor workflow. Retrieved March 6, 2024, from https://www.lseg.com/en/media-centre/press-releases/refinitiv/2019/june/refinitiv-makes-strategic-investment-in-battlefin-and-partners-to-incorporate-alternative-datasets-within-investor-workflow

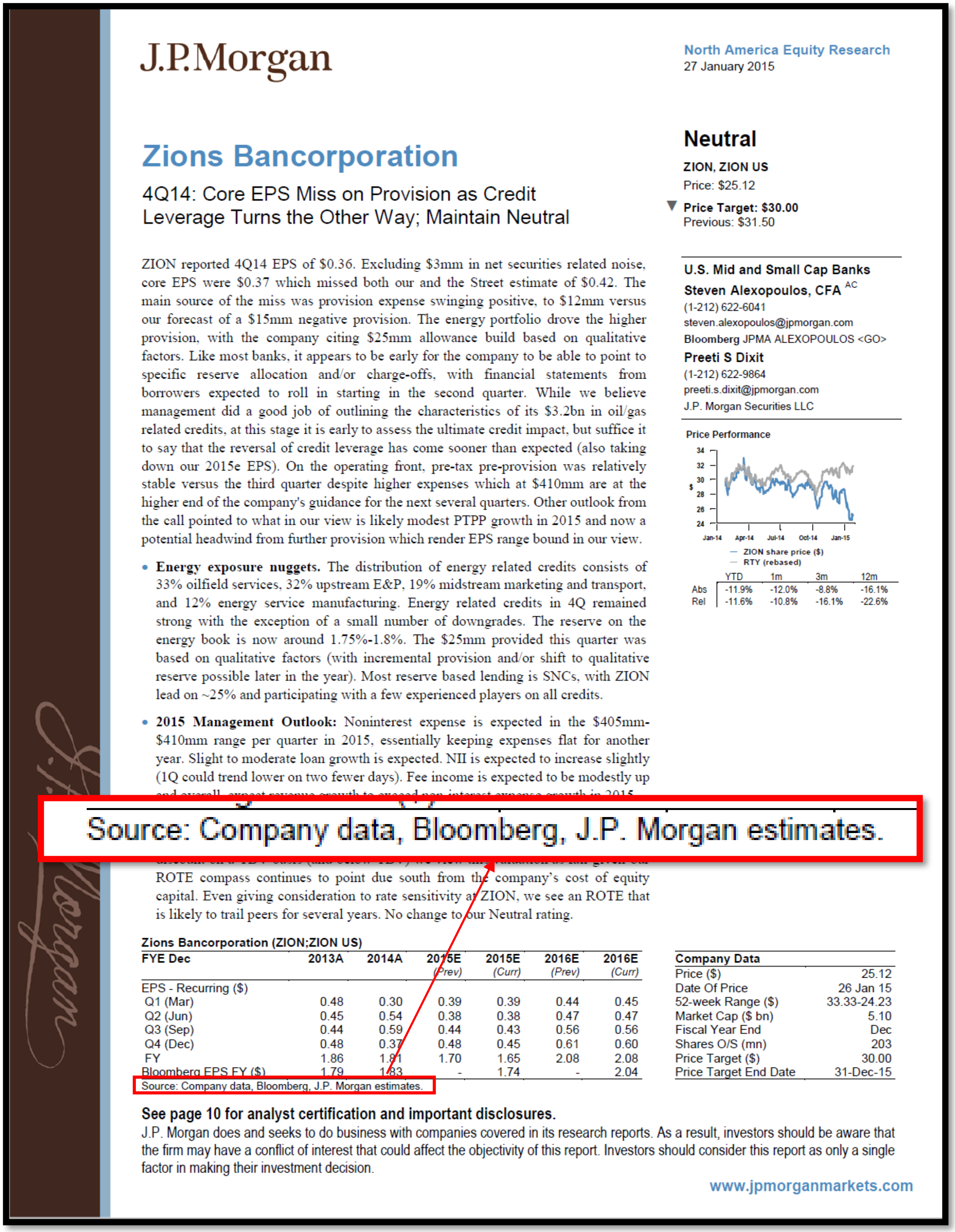
Simpson, A. V. (2010). Analysts’ use of non-financial information disclosures. Contemporary accounting research, 27(1).

Surowiecki, J. (2005). The Wisdom of Crowds. New York: Anchor Books.

Whited, R., Q. Swanquist, J. Shipman, & J. Moon Jr. (2022). Out of control: The (over) use of controls in accounting research. The Accounting Review 97, (3), 395-413.

Appendix A – Example of Subscription References in Analyst Reports

The image below contains a page from an equity research report that is representative of the data in our analyses. Subscription references mentioned in the report are highlighted in red and magnified for clarity.



J.P.Morgan

North America Equity Research
27 January 2015

Zions Bancorporation

4Q14: Core EPS Miss on Provision as Credit Leverage Turns the Other Way; Maintain Neutral

Neutral

ZION, ZION US
Price: $25.12
▼ **Price Target: $30.00**
Previous: $31.50

U.S. Mid and Small Cap Banks
Steven Alexopoulos, CFA AC
(1-212) 622-6041
steven.alexopoulos@jpmorgan.com
Bloomberg JPMA ALEXOPOULOS <GO>
Preeti S Dixit
(1-212) 622-9864
preeti.s.dixit@jpmorgan.com
J.P. Morgan Securities LLC

ZION reported 4Q14 EPS of $0.36. Excluding $3mm in net securities related noise, core EPS were $0.37 which missed both our and the Street estimate of $0.42. The main source of the miss was provision expense swinging positive, to $12mm versus our forecast of a $15mm negative provision. The energy portfolio drove the higher provision, with the company citing $25mm allowance build based on qualitative factors. Like most banks, it appears to be early for the company to be able to point to specific reserve allocation and/or charge-offs, with financial statements from borrowers expected to roll in starting in the second quarter. While we believe management did a good job of outlining the characteristics of its $3.2bn in oil/gas related credits, at this stage it is early to assess the ultimate credit impact, but suffice it to say that the reversal of credit leverage has come sooner than expected (also taking down our 2015e EPS). On the operating front, pre-tax pre-provision was relatively stable versus the third quarter despite higher expenses which at $410mm are at the higher end of the company's guidance for the next several quarters. Other outlook from the call pointed to what in our view is likely modest PTPP growth in 2015 and now a potential headwind from further provision which render EPS range bound in our view.

- **Energy exposure nuggets.** The distribution of energy related credits consists of 33% oilfield services, 32% upstream E&P, 19% midstream marketing and transport, and 12% energy service manufacturing. Energy related credits in 4Q remained strong with the exception of a small number of downgrades. The reserve on the energy book is now around 1.75%-1.8%. The $25mm provided this quarter was based on qualitative factors (with incremental provision and/or shift to qualitative reserve possible later in the year). Most reserve based lending is SNCs, with ZION lead on ~25% and participating with a few experienced players on all credits.
- **2015 Management Outlook:** Noninterest expense is expected in the $405mm-$410mm range per quarter in 2015, essentially keeping expenses flat for another year. Slight to moderate loan growth is expected. NII is expected to increase slightly (1Q could trend lower on two fewer days). Fee income is expected to be modestly up

Source: Company data, Bloomberg, J.P. Morgan estimates.

ROTE compass continues to point due south from the company's cost of equity capital. Even giving consideration to rate sensitivity at ZION, we see an ROTE that is likely to trail peers for several years. No change to our Neutral rating.

Price Performance

	YTD	1m	3m	12m
Abs	-11.9%	-12.0%	-8.8%	-16.1%
Rel	-11.6%	-10.8%	-16.1%	-22.6%

Zions Bancorporation (ZION;ZION US)

FYE Dec	2013A	2014A	2015E (Prev)	2015E (Curr)	2016E (Prev)	2016E (Curr)
EPS - Recurring ($)						
Q1 (Mar)	0.48	0.30	0.39	0.39	0.44	0.45
Q2 (Jun)	0.45	0.54	0.38	0.38	0.47	0.47
Q3 (Sep)	0.44	0.59	0.44	0.43	0.56	0.56
Q4 (Dec)	0.48	0.37	0.48	0.45	0.61	0.60
FY	1.86	1.81	1.70	1.65	2.08	2.08
Bloomberg EPS FY ($)	1.79	1.83	-	1.74	-	2.04

Source: Company data, Bloomberg, J.P. Morgan estimates.

Company Data	
Price ($)	25.12
Date Of Price	26 Jan 15
52-week Range ($)	33.33-24.23
Market Cap ($ bn)	5.10
Fiscal Year End	Dec
Shares O/S (mn)	203
Price Target ($)	30.00
Price Target End Date	31-Dec-15

See page 10 for analyst certification and important disclosures.

J.P. Morgan does and seeks to do business with companies covered in its research reports. As a result, investors should be aware that the firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision.

www.jpmorganmarkets.com

Appendix B – Variable Definitions

Dependent Variables:	Definition:
Accuracy	is the absolute value of the analyst's forecast minus the covered firm's actual earnings, scaled by stock price measured two trading days prior to the forecast issuance date, multiplied by negative one, and decile ranked by year.
SimilarForecast	is the absolute value of the difference between the two forecasts in each unique analyst pair, scaled by the firm's stock price measured two trading days prior to the first analyst's forecast issuance date in the analyst pair, multiplied by negative one, and decile ranked by year.
SimilarTiming	is an indicator variable set to one if the analysts in the pair share the same decile rank of forecast horizon, where forecast horizon is the number of days between the covered firm's fiscal period end date and the forecast issuance date. We decile rank horizon each year.
SimilarBoldness	is an indicator variable set to one if both forecasts in the analyst pair are similar in terms of boldness (i.e., both analysts are bold or both analysts are not bold), and zero otherwise. We follow Clement and Tse (2005) in calculating forecast boldness, where bold forecasts are those with forecast values that exceed (or are below) both the analyst's prior forecast for the firm and the prevailing consensus forecast at the time; all remaining forecasts are classified as nonbold. If a forecast's boldness cannot be calculated (e.g., there is no prior forecast to reference), *SimilarBoldness* is set equal to zero.
Similar Recommendations	is an indicator variable set equal to one if both analysts in the pair have the same outstanding recommendation at the time of the analysts' last annual forecasts, issued at least 30 days prior to the fiscal year end for each firm, and zero otherwise.
SimilarReports	is the cosine similarity between the analyst reports in the pair, decile ranked by year. We constrain to reports that are issued surrounding the analysts' last annual forecasts, issued at least 30 days prior to the fiscal year end for each firm.
SimilarError	is an indicator variable set equal to one if both analysts in the pair have the same directional earnings forecast error (i.e., both analysts over- or under-forecast actual earnings for the firm) and zero otherwise.
SimilarError Magnitude	is an indicator variable set equal to one if both analysts in the pair have an absolute forecast error, scaled by actual announced earnings, that is in the same yearly decile rank and zero otherwise.
SimilarError& Magnitude	is an indicator variable set equal to one if both analysts in the pair have the same directional earnings forecast error (i.e., *SimilarError* = 1) and the magnitude of the error is in the same decile rank (i.e., *SimilarErrorMagnitude* = 1) and zero otherwise.
ConsensusAccuracy	is the absolute value of the difference between the covered firm's reported earnings and the median analyst consensus forecast, scaled by the firm's stock price from the most recent quarter, multiplied by negative one, and decile ranked by year. The median consensus forecast is calculated using the most recent analyst forecasts issued thirty days before the firm's earnings announcement date.
Independent Variables:	Definition:
NumSubscriptions	is the number of data providers that a brokerage subscribes to at the time the analyst's forecast is issued, decile ranked by year.
Horizon	is the number of days between the covered firm's fiscal period end date and the forecast issuance date, decile ranked by year.
Experience	is the analyst's total number of years forecasting on I/B/E/S as of the prior year, decile ranked by year.
BrokerageSize	is the number of analysts employed at the brokerage as of the prior year, decile ranked by year.
PortfolioSize	is the number of firms the analyst covers as of the prior year, decile ranked by year.
MajorProvider	is the number of data subscriptions the brokerage subscribes to among the five major data providers (S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar), decile ranked by year.
MinorProvider	is the number of data subscriptions the brokerage subscribes to that are not among the five major data providers (S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar), decile ranked by year.
PublicSources	is set equal to one if there are brokerage references to company filings (e.g., EDGAR) or source citations of conference calls. The variable is set equal to two if both company filings and conference calls are mentioned, and zero if neither are referenced.
AllStar	is an indicator variable set equal to one if the analyst received All-Star designation during the year, and zero otherwise.

LowExperience	is an indicator variable set equal to one if the analyst is in the lowest two deciles of general experience, and zero otherwise.
HighBusyness	is an indicator variable set equal to one if the number of the firms the analyst covers is in the highest two deciles of portfolio size, and zero otherwise.
LongHorizon	is an indicator variable set equal to one if the forecast horizon is in the highest two deciles, and zero otherwise.
SubscriptionSimilarity	is the number of subscriptions that both analysts in the pair have access to at their respective brokerages, scaled by the number of all possible data subscriptions, and decile ranked by year.
SimilarExperience	is an indicator variable set equal to one if both analysts in the pair have a similar number of years of experience forecasting on I/B/E/S, and zero otherwise. Analysts are determined to have similar forecasting experience if both are in the same experience decile rank, based on the total years forecasting on I/B/E/S as of the prior year, calculated annually.
SimilarResources	is an indicator variable set equal to one if both analysts in the pair are employed by brokerages with similar resources, and zero otherwise. Brokerages are determined to have similar resources if each brokerage is in the same decile rank, based on the number of analysts employed at the brokerage as of the prior year, calculated annually.
SimilarBusyness	is an indicator variable set equal to one if both analysts in the pair cover a similar number of firms on I/B/E/S, and zero otherwise. Analysts are determined to cover a similar number of firms if both are in the same decile rank, based on the number of covered firms as of the prior year, calculated annually.
BTM	is the covered firm's book-to-market ratio as of the most recently reported quarter, decile ranked by year.
MVE	is the market value of equity as of the most recently reported quarter, decile ranked by year.
ROA	is the covered firm's return on assets ratio as of the most recently reported quarter, decile ranked by year.
AllStars	is an indicator variable set equal to one if both analysts in the pair received All-Star designation during the year, and zero otherwise.
OldSubscription Similarity	is the *SubscriptionSimilarity* between an analyst's prior brokerage and the brokerage of the paired analyst, in the concurrent period.
MajorSubscription Similarity	is the percentage of major sources that both analysts in the pair have access to at their respective brokerages, decile ranked by year. Major sources are defined as S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar.
MinorSubscription Similarity	is the percentage of non-major, paid sources that both analysts in the pair have access to at their respective brokerages, decile ranked by year.
HighSubscription Access	is an indicator variable set equal to one if both analysts in the pair are employed by brokerages with a high number of data subscriptions, and zero otherwise. Brokerages are determined to have a high number of data subscriptions if they are in the upper 50^{th} percentile based on the number of data sources that each brokerage reports, calculated yearly.
AvgSubscription Similarity	is the average value of *SubscriptionSimilarity,* prior to its decile ranking, calculated at the firm-year level. The variable is then decile ranked by year.
NumConsensus Subscriptions	is the number of unique data subscriptions available to consensus analysts, decile ranked by year.
AvgExperience	is the average experience of the analysts contributing to the consensus forecast, calculated at the firm-year level, and decile ranked by year.
AvgBrokerageSize	is the average size of the analysts' brokerages contributing to the consensus forecast, calculated at the firm-year level, and decile ranked by year.
AvgHorizon	is the average horizon of the analysts' forecasts contributing to the consensus forecast, calculated at the firm-year level, and decile ranked by year.
Loss	is set equal to one if the covered firm's earnings are negative, and zero otherwise.

Table 1 – Data Subscription Descriptive Statistics
This table provides descriptive evidence on the data subscriptions that analysts cite in our sample of reports. Panel A lists the top 15 cited subscriptions.

Panel A: Top 15 Cited Data Subscriptions

Subscription	# of Brokerages Citing the Subscription	% of Brokerages
Bloomberg	188	66%
FactSet	134	47%
S&P Capital IQ	129	45%
Thomson	123	43%
Reuters	112	39%
Thomson Reuters	90	32%
NASDAQ	88	31%
EIA	69	24%
First Call	66	23%
SNL	62	22%
IDC	60	21%
IHS	59	21%
Nielsen	55	19%
Street Account	55	19%
IMS	51	18%

Table 1 – Data Subscription Descriptive Statistics, Continued

Panel B lists the probability that an analyst's peer at the same brokerage will also cite the same referenced data source as the given analyst at least once during the next six months. This is compared to a sample of analysts who are randomly assigned from different brokerages. Panel C reports a data subscription transition matrix, illustrating the likelihood that a data source subscribed to in year t will continue to be used in year t+1.

Panel B: Probability a Peer Analyst at the Brokerage Uses the Dataset

	Actual Brokerage	Randomly Assigned Brokerage
Bloomberg	94.9%	36.0%
Thomson Reuters	95.3%	16.7%
Factset	97.3%	25.7%
S&P Capital IQ	91.8%	23.8%
Median Dataset	89.4%	1.9%
Mean Dataset	80.5%	5.8%

Panel C: Data Subscription Retention/Transition Matrix

	Subscribe$_{t+1}$	***Unsubscribe$_{t+1}$***
Subscribe$_t$	*85.44%*	*14.56%*
Unsubscribe$_t$	*3.62%*	*96.38%*

Table 2 – Sample Descriptive Statistics

Panel A provides descriptive statistics for the individual analyst accuracy sample. Panel B provides descriptive statistics for the pairwise sample. All continuous variables are decile ranked by year. Variable definitions are provided in the appendix.

Panel A – Individual Analyst Accuracy Sample

Variable	N	Mean	Std. Dev	25th	Median	75th
Accuracy	213,523	4.58	2.85	2.00	5.00	7.00
NumSubscriptions (raw value)	213,523	9.70	7.85	3.00	7.00	15.00
NumSubscriptions (decile ranked)	213,523	4.54	2.85	2.00	5.00	7.00
Horizon	213,523	4.43	2.79	2.00	4.00	7.00
Experience	213,523	4.59	2.86	2.00	5.00	7.00
BrokerageSize	213,523	4.50	2.85	2.00	5.00	7.00
PortfolioSize	213,523	4.65	2.81	2.00	5.00	7.00
PublicSources	213,523	1.52	0.59	1.00	2.00	2.00
MajorProvider	213,523	4.49	2.85	2.00	4.00	7.00
MinorProvider	213,523	4.52	2.79	2.00	5.00	7.00
AllStar	213,523	0.13	0.34	0.00	0.00	0.00
LowExperience	213,523	0.18	0.39	0.00	0.00	0.00
HighBusyness	213,523	0.20	0.40	0.00	0.00	0.00
LongHorizon	213,523	0.18	0.38	0.00	0.00	0.00
BTM	213,523	4.50	2.87	2.00	4.00	7.00
MVE	213,523	4.50	2.87	2.00	5.00	7.00
ROA	213,523	4.50	2.87	2.00	4.00	7.00

Panel B – Pairwise Sample

Variable	N	Mean	Std. Dev	25th	Median	75th
SimilarForecast	1,337,709	4.56	2.86	2.00	5.00	7.00
SimilarTiming	1,337,709	0.38	0.49	0.00	0.00	1.00
SimilarBoldness	1,337,709	0.54	0.50	0.00	1.00	1.00
SubscriptionSimilarity (raw value)	1,337,709	0.10	0.10	0.02	0.07	0.16
SubscriptionSimilarity	1,337,709	4.53	2.83	2.00	4.00	7.00
SimilarExperience	1,337,709	0.11	0.31	0.00	0.00	0.00
SimilarResources	1,337,709	0.09	0.29	0.00	0.00	0.00
SimilarBusyness	1,337,709	0.13	0.33	0.00	0.00	0.00
AllStars	1,337,709	0.02	0.15	0.00	0.00	0.00
HighSubscriptionAccess	1,337,709	0.29	0.45	0.00	0.00	1.00
MajorSubscriptionSimilarity	1,337,709	4.53	2.85	2.00	5.00	7.00
MinorSubscriptionSimilarity	1,337,709	4.52	2.75	2.00	4.00	7.00
BTM	1,337,709	4.49	2.87	2.00	4.00	7.00
MVE	1,337,709	4.50	2.87	2.00	4.00	7.00
ROA	1,337,709	4.51	2.87	2.00	5.00	7.00

Table 3 – Individual Analyst Forecast Accuracy

This table provides results from estimating Model (1), in which we investigate the relationship between the number of unique data subscriptions available to an analyst and forecast accuracy. Panel A reports the primary specification while Panel B provides additional robustness analyses. Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Forecast Accuracy

Dependent Variable: *Accuracy*	(1)	(2)
NumSubscriptions	**0.019*****	**0.023*****
	(7.86)	**(13.18)**
Horizon	-0.053***	-0.155***
	(-16.58)	(-71.64)
Experience	0.038***	0.016***
	(16.79)	(11.43)
BrokerageSize	-0.035***	-0.013***
	(-13.94)	(-7.50)
PortfolioSize	0.043***	0.005***
	(15.40)	(3.52)
BTM	-0.184***	
	(-34.80)	
MVE	0.261***	
	(48.76)	
ROA	0.204***	
	(39.23)	
Firm-Year FE	No	Yes
N	213,523	209,033
Adj. R^2	0.23	0.71

Table 3 – Individual Analyst Forecast Accuracy, Continued

Panel B: Forecast Accuracy Robustness

Dependent Variable: *Accuracy*	(1)	(2)	(3)
NumSubscriptions	**0.013*** **	**0.015*** **	**0.018*** **
	(4.21)	**(4.38)**	**(4.82)**
Horizon	-0.142***	-0.151***	-0.138***
	(-64.83)	(-69.72)	(-62.31)
Experience	0.014*	0.007***	0.011
	(1.66)	(4.79)	(1.09)
BrokerageSize	-0.011***	-0.023***	-0.009
	(-3.01)	(-4.64)	(-1.49)
PortfolioSize	-0.004	0.004***	-0.005*
	(-1.51)	(2.64)	(-1.68)
Firm-Year FE	Yes	Yes	Yes
Analyst FE	Yes	No	No
Brokerage FE	No	Yes	No
Analyst-Brokerage FE	No	No	Yes
N	208,153	209,018	207,971
Adj. R^2	0.72	0.71	0.72

Table 4 – Individual Analyst Forecast Accuracy: Cross-sectional Tests

This table provides results from estimating Models (2-3). Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable: *Accuracy*		(1)	(2)	(3)	(4)	(5)	(6)
MajorProvider		**0.010***** **(4.99)**					
MinorProvider		**0.016***** **(7.43)**					
NumSubscriptions			**0.022***** **(10.30)**				
PublicSources			**0.008** **(0.92)**				
AllStar* · *NumSubscriptions				**-0.011**** **(-2.30)**			
LowExperience* · *NumSubscriptions					**0.009***** **(2.70)**		
HighBusyness* · *NumSubscriptions						**0.008**** **(2.54)**	
LongHorizon* · *NumSubscriptions							**0.012***** **(3.04)**
Controls		Yes	Yes	Yes	Yes	Yes	Yes
Firm-Year FE		Yes	Yes	Yes	Yes	Yes	Yes
N		209,033	209,033	209,033	209,033	209,033	209,033
Adj. R^2		0.71	0.71	0.71	0.71	0.71	0.71
Within Regression F-Tests							
MajorProvider = MinorProvider	Diff	-0.006					
	f-stat	2.34					
NumSubscriptions = PublicSources	Diff		0.014				
	f-stat		1.84				

Table 5 – Subscription Similarity and Forecast Similarity

This table provides results from estimating Model (4), in which we investigate the relationship between shared data subscriptions and various analyst forecast attributes. Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable:	*SimilarForecast*		*SimilarTiming*		*SimilarBoldness*	
	(1)	(2)	(3)	(4)	(5)	(6)
SubscriptionSimilarity	**0.054*** **	**0.053*** **	**0.009*** **	**0.011*** **	**0.009*** **	**0.009*** **
	(24.05)	**(33.85)**	**(25.64)**	**(31.45)**	**(23.87)**	**(26.70)**
SimilarExperience	-0.021***	0.025***	0.010***	0.008***	0.002	0.004***
	(-2.70)	(4.28)	(6.68)	(6.06)	(1.05)	(2.69)
SimilarResources	-0.195***	-0.115***	-0.036***	-0.029***	-0.039***	-0.030***
	(-25.93)	(-20.40)	(-27.43)	(-23.70)	(-28.03)	(-23.16)
SimilarBusyness	-0.133***	0.020***	-0.002	0.007***	-0.005***	0.005***
	(-11.79)	(3.38)	(-1.21)	(5.26)	(-2.96)	(3.85)
BTM	-0.227***		-0.011***		-0.003***	
	(-32.44)		(-16.07)		(-4.53)	
MVE	0.166***		-0.004***		0.002***	
	(26.78)		(-6.29)		(3.63)	
ROA	0.162***		0.005***		0.004***	
	(23.26)		(6.69)		(5.60)	
Firm-Year FE	No	Yes	No	Yes	No	Yes
N	1,337,709	1,333,013	1,337,709	1,333,013	1,337,709	1,333,013
Adj. R^2	0.17	0.51	0.01	0.15	0.00	0.15

Table 6 – Subscription Similarity and Forecast Similarity: Robustness

This table provides results from estimating variations of Model (4) with augmented fixed effect designs and additional control variables. Panel B reports fewer observations, as this sample constitutes the analyst pairs where one of the analysts moved brokerages. Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Brokerage Pairwise Fixed Effects

Dependent Variable:	*SimilarForecast*		*SimilarTiming*		*SimilarBoldness*	
	(1)	(2)	(3)	(4)	(5)	(6)
SubscriptionSimilarity	**0.054*** **	**0.042*** **	**0.012*** **	**0.009*** **	**0.010*** **	**0.006*** **
	(15.29)	**(11.58)**	**(16.73)**	**(11.30)**	**(12.86)**	**(7.33)**
SimilarExperience	0.011*	-0.007	0.005***	0.002	0.002	-0.005*
	(1.93)	(-0.56)	(3.49)	(0.53)	(1.12)	(-1.67)
SimilarResources	-0.001	0.006	0.001	0.001	-0.002	-0.003
	(-0.08)	(0.63)	(0.77)	(0.39)	(-1.01)	(-1.44)
SimilarBusyness	0.007	-0.002	0.003***	0.000	0.004***	0.001
	(1.20)	(-0.29)	(2.80)	(-0.08)	(3.12)	(0.36)
Brokerage Pairwise FE	Yes	No	Yes	No	Yes	No
Analyst Pair-Brokerage Pair FE	No	Yes	No	Yes	No	Yes
Firm-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,329,430	1,181,869	1,329,430	1,181,869	1,329,430	1,181,869
Adj. R^2	0.53	0.58	0.22	0.32	0.17	0.21

Table 6 – Subscription Similarity and Forecast Similarity: Robustness, Continued

Panel B: Analyst Employment Changes

Dependent Variable:		*SimilarForecast*	*SimilarTiming*	*SimilarBoldness*
		(1)	(2)	(3)
SubscriptionSimilarity		**0.042*****	**0.011*****	**0.007*****
		(6.11)	**(5.94)**	**(3.52)**
OldSubscriptionSimilarity		**0.006**	**-0.001**	**0.000**
		(0.80)	**(-0.65)**	**(0.01)**
SimilarExperience		0.005	0.003	0.000
		(0.46)	(1.01)	(0.08)
SimilarResources		-0.057***	-0.015***	-0.014***
		(-4.49)	(-5.30)	(-4.87)
SimilarBusyness		0.015	0.006**	0.004
		(1.30)	(2.36)	(1.46)
Firm-Year FE		Yes	Yes	Yes
N		301,871	301,871	301,871
Adj. R^2		0.51	0.19	0.19
Within Regression F-Tests				
SubscriptionSimilarity =	Diff	0.036***	0.012***	0.007*
OldSubscriptionSimilarity	f-stat	7.28	11.13	3.15

Table 7 – Subscription Similarity and Forecast Similarity: Cross-sectional Tests

This table provides cross-sectional results from estimating Models (5-6). Panel A presents results for the similarity in point forecasts. Panel B presents results for the similarity in forecast timing. Panel C presents results for the similarity in forecast boldness. Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Similarity in Point Forecasts

Dependent Variable: *SimilarForecast*		(1)	(2)	(3)
MajorSubscriptionSimilarity		**0.034***** **(17.11)**		
MinorSubscriptionSimilarity		**0.023***** **(12.92)**		
SubscriptionSimilarity · AllStars			**-0.035***** **(-5.69)**	
SubscriptionSimilarity · HighSubscriptionAccess				**-0.026***** **(-4.83)**
Controls		Yes	Yes	Yes
Firm-Year FE		Yes	Yes	Yes
N		1,333,013	1,333,013	1,333,013
Adj. R^2		0.51	0.51	0.51
Within Regression F-Tests				
MajorSubscriptionSimilarity =	Diff	0.011***		
MinorSubscriptionSimilarity	f-stat	11.10		

Panel B: Similarity in Forecast Timing

Dependent Variable: *SimilarTiming*		(1)	(2)	(3)
MajorSubscriptionSimilarity		**0.006***** **(15.91)**		
MinorSubscriptionSimilarity		**0.005***** **(12.46)**		
SubscriptionSimilarity · AllStars			**-0.011***** **(-7.71)**	
SubscriptionSimilarity · HighSubscriptionAccess				**-0.006***** **(-5.00)**
Controls		Yes	Yes	Yes
Firm-Year FE		Yes	Yes	Yes
N		1,333,013	1,333,013	1,333,013
Adj. R^2		0.15	0.15	0.15
Within Regression F-Tests				
MajorSubscriptionSimilarity =	Diff	0.001**		
MinorSubscriptionSimilarity	f-stat	4.75		

Panel C: Similarity in Forecast Boldness

Dependent Variable: *SimilarBoldness*		(1)	(2)	(3)
MajorSubscriptionSimilarity		**0.005*****		
		(12.62)		
MinorSubscriptionSimilarity		**0.005*****		
		(11.50)		
SubscriptionSimilarity · AllStars			**-0.010*****	
			(-7.94)	
SubscriptionSimilarity · HighSubscriptionAccess				**-0.006*****
				(-5.16)
Controls		Yes	Yes	Yes
Firm-Year FE		Yes	Yes	Yes
N		1,333,013	1,333,013	1,333,013
Adj. R^2		0.15	0.15	0.15
Within Regression F-Tests				
MajorSubscriptionSimilarity =	Diff	0.00		
MinorSubscriptionSimilarity	f-stat	0.77		

Table 8 – Subscription Similarity and Alternative Analyst Research Outputs

This table provides results for analyses that examine subscription similarity and alternative analyst research outputs. Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable:	*SimilarRecommendations*	*SimilarReports*
	(1)	(2)
SubscriptionSimilarity	**0.009*** **	**0.029*** **
	(27.65)	**(7.12)**
SimilarExperience	-0.001	-0.020
	(-0.49)	(-1.18)
SimilarResources	0.009***	0.636***
	(4.48)	(31.62)
SimilarBusyness	0.002	0.082***
	(0.99)	(4.58)
Firm-Year FE	Yes	Yes
N	760,390	262,424
Adj. R^2	0.05	0.09

Table 9 – Subscription Similarity and Forecast Errors

This table provides results from analyses that examine the relationship between subscription similarity and analyst forecast errors. Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable:	*SimilarError*	*SimilarErrorMagnitude*	*SimilarError&Magnitude*
	(1)	(2)	(3)
SubscriptionSimilarity	**0.003*** **	**0.003*** **	**0.003*** **
	(10.40)	**(13.99)**	**(13.19)**
SimilarExperience	0.002*	0.000	-0.000
	(1.76)	(0.33)	(-0.04)
SimilarResources	-0.002*	-0.004***	-0.004***
	(-1.77)	(-3.45)	(-3.61)
SimilarBusyness	0.001	0.001	0.002
	(0.78)	(1.00)	(1.50)
Firm-Year FE	Yes	Yes	Yes
N	1,148,721	1,148,205	1,148,205
Adj. R^2	0.29	0.19	0.22

Table 10 – Data Subscriptions and Consensus Forecast Accuracy

This table provides results from estimating Model (7), in which we investigate the relationship between data subscriptions and consensus forecast accuracy. Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable: *ConsensusAccuracy*	(1)	(2)
NumConsensusSubscriptions	**0.052*** **	**0.049*** **
	(3.99)	**(3.79)**
AvgSubscriptionSimilarity	**-0.027*** **	**-0.024*** **
	(-3.40)	**(-3.03)**
AvgHorizon	-0.022***	-0.022***
	(-4.25)	(-4.12)
AvgExperience	-0.003	-0.003
	(-0.36)	(-0.42)
AvgBrokerageSize	-0.033***	-0.032***
	(-3.25)	(-3.15)
BTM	-0.162***	-0.163***
	(-14.96)	(-14.99)
MVE	0.349***	0.339***
	(16.69)	(15.90)
ROA	0.048***	0.044***
	(5.50)	(5.02)
Loss	-0.896***	-0.911***
	(-13.24)	(-13.36)
Firm FE	Yes	Yes
Year FE	No	Yes
N	26,789	26,789
Adj. R^2	0.53	0.53

Internet Appendix:

The Influence of Financial Data Subscriptions on Analyst Research

In this Internet Appendix (IA), we tabulate several additional analyses that we mention in the paper but do not tabulate due to space constraints. First, we report our primary results using analyst sales forecasts, as compared to analyst EPS forecasts, which are used in our main analyses. We find consistent results, which we present in IA Table 1 through IA Table 3 below. Next, we examine whether shared FDS flow through to stock return comovement. Since shared data subscriptions lead to correlated forecast activity *within* firms (Tables 5–9), greater overlap in FDS across analysts covering the same firms may increase return comovement between those firms. We find that higher FDS overlap between firm pairs is associated with greater return comovement, even after controlling for brokerage overlap and including firm-pair and year fixed effects. These findings suggest that as brokerages covering a pair of firms share more FDS, the firms' returns become more closely correlated—consistent with shared FDS influencing market-level outcomes. We present these results in IA Table 4.

IA Table 1 – Individual Analyst Sales Forecast Accuracy

This table provides results from investigating the relationship between the number of unique data subscriptions available to an analyst and sales forecast accuracy. We define *SalesForecastAccuracy* as the absolute value of the analyst's sales forecast minus the covered firm's actual sales, scaled by the firm's market value of equity, multiplied by negative one, and decile ranked by year. Other variable definitions are provided in the main appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable: *SalesForecastAccuracy*	(1)	(2)
NumSubscriptions	**0.024***** **(8.06)**	**0.014***** **(6.43)**
Horizon	-0.151*** (-40.57)	-0.244*** (-85.89)
Experience	0.016*** (5.74)	0.009*** (4.64)
BrokerageSize	-0.045*** (-14.45)	0.003 (1.43)
PortfolioSize	0.028*** (8.91)	0.006*** (2.86)
BTM	-0.316*** (-56.08)	
MVE	0.142*** (24.43)	
ROA	0.030*** (5.46)	
Firm-Year FE	No	Yes
N	169,045	163,762
Adj. R^2	0.16	0.64

IA Table 2 – Subscription Similarity and Sales Forecast Similarity

This table provides results from investigating the relationship between shared data subscriptions and similarity in analyst sales forecasts. *SimilarSalesForecast* is the absolute value of the difference between the two sales forecasts in each unique analyst pair, scaled by the firm's market value of equity, multiplied by negative one, and decile ranked by year. Other variable definitions are provided in the main appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm-year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable:	*SimilarSalesForecast*	
	(1)	(2)
SubscriptionSimilarity	**0.041***** **(14.87)**	**0.054***** **(27.40)**
SimilarExperience	-0.003 (-0.34)	0.007 (0.93)
SimilarResources	-0.175*** (-18.94)	-0.112*** (-15.44)
SimilarBusyness	-0.095*** (-8.11)	0.011 (1.51)
BTM	-0.343*** (-47.86)	
MVE	0.088*** (12.64)	
ROA	0.033*** (4.46)	
Firm-Year FE	No	Yes
N	888,994	883,799
Adj. R^2	0.15	0.49

IA Table 3 – Data Subscriptions and Consensus Sales Forecast Accuracy

This table provides results from investigating the relationship between data subscriptions and consensus sales forecast accuracy. *ConsensusSalesAccuracy* is the absolute value of the difference between the covered firm's reported sales and the median analyst consensus sales forecast, scaled by the firm's market value of equity, multiplied by negative one, and decile ranked by year. Other variable definitions are provided in the main appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable: *ConsensusSalesAccuracy*	(1)	(2)
NumConsensusSubscriptions	**0.090***** **(5.62)**	**0.097***** **(6.07)**
AvgSubscriptionSimilarity	**-0.052***** **(-5.09)**	**-0.052***** **(-5.11)**
AvgHorizon	-0.087*** (-13.24)	-0.086*** (-13.01)
AvgExperience	0.007 (0.71)	0.009 (0.91)
AvgBrokerageSize	-0.022* (-1.67)	-0.025* (-1.91)
BTM	-0.176*** (-13.26)	-0.165*** (-12.32)
MVE	0.428*** (17.46)	0.455*** (18.33)
ROA	0.024** (2.29)	0.030*** (2.80)
Loss	-0.285*** (-3.76)	-0.261*** (-3.43)
Firm FE	Yes	Yes
Year FE	No	Yes
N	19,896	19,896
Adj. R^2	0.48	0.48

IA Table 4 – Cross-brokerage Data Subscription Similarity and Future Return Comovement

This table provides results from investigating the relationship between data subscription similarity across brokerages and future return comovement for pairs of covered firms. *FutureReturnComovement* is the Pearson correlation coefficient between a firm's stock return (firm i) and a peer's stock return (firm j) over the one-year period beginning on the firms' matching fiscal period end dates, and decile ranked by year. *CrossBrokerageSubscriptionSimilarity* is the average similarity in financial data subscriptions across the brokerages covering firm i and firm j as of the firms' fiscal period end date, and decile ranked by year. *SimilarROA*, *SimilarBTM*, and *SimilarMVE* are indicator variables set equal to one if the firms are in the same decile of ROA (return on assets), BTM (book-to-market ratio), or MVE (market value of equity), respectively, and zero otherwise. t-statistics are reported in parentheses, and standard errors are clustered by firm-year for each firm in the pair (i.e., firm *i* year and firm *j* year). All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable: *FutureReturnComovement*	(1)	(2)	(3)	(4)
CrossBrokerageSubscriptionSimilarity	**0.025*** **	**0.025*** **	**0.026*** **	**0.020*** **
	(6.40)	**(6.41)**	**(6.79)**	**(2.99)**
SimilarROA		0.040***	0.044***	0.030**
		(5.59)	(6.23)	(2.41)
SimilarBTM		0.032***	0.032***	0.039***
		(6.59)	(6.48)	(3.88)
SimilarMVE		0.033***	0.039***	0.025
		(3.78)	(4.70)	(1.63)
Firm-Pair FE	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes
N	6,619,509	6,619,509	6,619,509	1,119,048
Adj. R^2	0.49	0.49	0.50	0.52