**Financial Data Providers and the Diversity of Market Opinions**

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

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

**March 2024**

Financial data providers play a critical role in modern capital markets, with many market participants relying on the same common data providers. In this paper, we investigate whether access to similar data providers affects the diversity of market opinions. We study this through the lens of sell-side analysts and find that when two analysts subscribe to similar data feeds, their (1) forecast values, (2) forecast timing, and (3) forecast boldness all tend to converge, consistent with data subscriptions affecting the diversity and timing of analyst opinions. Further, when access to similar data providers changes intertemporally between two analysts, either because of changes in brokerage subscriptions or because of changes in analyst employment, we observe similar effects. Moreover, our findings are stronger when data subscriptions contain more proprietary information and are weaker for analysts who tend to have access to soft information (i.e., All-Star analysts). Finally, consistent with the wisdom of crowds theory, we find that consensus estimates that exhibit more data provider diversity in the underlying forecasts tend to be more accurate.

**Keywords:** Data Sources; Diversity of Opinion; Sell-side Analysts; 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, Mike Drake, Jeremiah Green, Xu Jiang, Jared Jennings, Stan Markov, Jeff McMullin, Mark Piorkowski, Georg Rickmann (Discussant), Richard Saouma, Roy Schmardebeck, Carol Seregni, Brady Twedt, and workshop participants at the BYU Accounting Symposium and the Washington University in St. Louis Accounting Research Conference. We are grateful to the University of Georgia Terry College of Business and the BYU Marriott School of Business for financial support. We also acknowledge the research support of Vanessa Johnson and Ham Christensen. All errors are our own.

## 1. Introduction

Despite living in the information age, we know relatively little about how financial data providers (FDPs, henceforth) affect market participants. This represents an important gap in the literature, as a staggering 45.9% of the market share in the FDP industry is concentrated among five companies (Al Bari, 2023). Accordingly, the objective of our study is to provide insight into the data providers that financial market participants use and whether using similar FDPs affects the diversity of market participants' opinions. We study this through the lens of a prominent financial market intermediary, sell-side analysts, who often report the "sources" of their market research.

FDPs have become a cornerstone of modern capital markets, serving as the primary channels through which market participants receive information. Specifically, FDPs create value by aggregating and disseminating financial information, thereby reducing acquisition costs for subscribing clients (e.g., Akbas et al., 2018). Illustrating the persistent role of FDPs, over 70% of market participants anticipate that data providers will become *more* important to their investment process in the coming years (Refinitiv, 2019a). Yet, despite the increasing importance of data providers, our understanding of their use and corresponding effects on market participants remains limited. Indeed, a long-standing challenge to those interested in studying FDPs is that the data sources market participants have access to are traditionally unobservable. We overcome this empirical hurdle by constructing a comprehensive measure of brokerage-level data subscriptions, based on data source references in sell-side equity analyst research reports.

Sell-side analysts play a pivotal role in shaping investor expectations and directing capital flows (Womack, 1996; Francis and Soffer, 1997; Howe et al., 2009). However, if analysts draw from similar pools of data sources, it raises questions about the diversity and independence of their opinions. A concentration of data subscriptions, as is observed in the FDP industry, could inadvertently lead to homogenized market views. As such, understanding these effects is important, not only for market participants who rely on these information intermediaries, but also for regulators who aim to better understand the operation of financial markets. In this context, our study seeks to shed light on the FDPs that sell-side analysts subscribe to and the impact such subscriptions have on the diversity of their research.

Ex ante, the relationship between data subscriptions and analysts' opinions is unclear. In our conversations with various practitioners and academics, a prominent belief is that sharing similar data feeds should have no systematic or directional effect on analyst opinions. This is because FDPs generally rely on the *same* underlying source data for the bulk of their product offerings. For example, to provide information on firm fundamentals, FDPs typically extract, verify, and aggregate financial information from the same source: EDGAR filings. In this light, market participants view FDPs as offering a one-stop-shop for accessing commodity-like data, at scale, to compare and evaluate investment opportunities. To the extent that this common perspective is accurate, we may not observe an association between shared data subscriptions and the diversity of analyst opinions.

Alternatively, relying on similar data providers potentially leads to a convergence in analyst opinions and actions. This could occur for several reasons. First, some analysts might simply use data in a "plug and chug" manner without adjusting it or forming their own opinions about it. This approach, in essence, defers the opinion of the analyst to that of the data provider, which aggregates and constructs market metrics. As emerging evidence suggests that the measurement of some firm and market metrics can vary considerably across FDPs (Bochkay et al., 2022; Larocque et al., 2023), subscribing to the same data provider could lead to more correlated investment opinions. Second, to strengthen market position, FDPs increasingly offer unique, proprietary datasets in addition to their traditional offerings (Bloomberg, 2017; Refinitiv, 2019b; Bloomberg, 2023). If analysts use and anchor their opinions on such proprietary information, subscribing to the same data provider could lead to similar investment decisions. Third, training sessions that data providers offer potentially standardize how analysts interpret and use the data, encouraging uniformity.

A final possibility is that sharing similar FDPs might actually *increase* the diversity of analyst opinions. If analysts know that their industry peers rely on similar data, this could encourage increased focus on output differentiation, as analysts often feel a need to demonstrate their unique "edge" to their clients and superiors (Brown et al., 2015).[1] Supporting this argument, some scholars suggest that anti-herding behavior is quite common among analysts (e.g., Bernhardt et al., 2006). As such, knowledge of lacking data independence might induce diverse opinions, despite sharing similar data providers.[2]

To investigate this research question, we construct a novel data set containing the financial data "sources" referenced in a large sample of analyst reports. Our sample consists of 595,642 analyst reports, written by 3,596 distinct analysts, and issued from approximately 285 brokerages during the years 2008–2017. To compile the list of sources referenced by analysts, we rely on the common convention of analysts referencing "source:" followed by a list of data providers used in

[1] In our conversations with current and former analysts, this point was mentioned several times.
[2] Relatedly, information theory highlights that sometimes sharing the same public information signal can cause market participants to disagree even more (akin to Kondor, 2012 and Armstrong et al., 2023). In the case of analysts, if they initially disagree about a firm's fundamental value and receive new, uncertain information from a shared FDP, their disagreement could *increase*.

the analyst report.[3] If an analyst within a brokerage references a given source, we consider that source "subscribed" to by the brokerage for the quarter before and after the analyst reference.[4]

As expected, we observe a high percentage of reports referencing Bloomberg, Factset, Thomson Reuters, and S&P Capital IQ. Interestingly, although analysts often cite these well-known FDPs, we find that the largest brokerages in our sample tend to have *different* top data providers. For example, JPMorgan most often references Bloomberg, while Morgan Stanley most commonly cites data provided by Thomson Reuters. We also find that, conditional on citing a source, the average analyst report cites 2.89 unique sources per report. The trend in source citations appears to be relatively constant throughout our sample period, with little temporal variation in the quantity of sources cited by analysts.

To evaluate whether sharing data providers affects analyst forecasting behavior, we construct a panel of analyst pairs. To control for selection concerns related to the covered firm, we constrain analyst pairing to only those analysts forecasting for the same firm in a given forecast period (i.e., annual forecasts), and we include *firm x year* fixed effects to control for firm and time-related attributes that may affect an analyst's forecast. These steps result in a panel of 1,362,220 analyst pair observations. As similarities in analyst experience, brokerage resources, and analyst busyness might also affect the similarity of analyst forecasting behavior, we control for those similarities in each empirical design.

We examine the diversity of analyst opinions by measuring how similar analyst forecasts are to one another. We find that sharing FDPs increases similarity in analyst forecasts, both in terms of point estimates and forecast boldness. These results hold with firm x year fixed effects,

[3] Appendix A provides three examples of analyst reports in our sample that include source references.
[4] A unique benefit of studying *brokerage-level* subscriptions is that it mitigates potential disclosure selection concerns at the analyst-report level.

as described previously, and are consistent with shared financial data feeds affecting the diversity of analyst opinions. In terms of economic magnitude, a one standard deviation increase in FDP similarity between analyst pairs is correlated with a 14.67% increase in point forecast similarity and a 3.00% increase in the probability that both analysts issue forecasts with similar boldness.

We also study the similarity in forecast timing to explore the role of information processing explanations. If FDPs affect the speed of information processing, then analysts with similar data subscriptions should exhibit more aligned forecast timing. We find evidence consistent with such similarity in process. That is, when analysts share similar FDPs, their reports tend to be disclosed more proximately in event time. Accordingly, these results suggest that sharing information may not be the only driver of forecast similarity; rather, there also appears to be a mechanical aspect in the *processing* of information that occurs when sharing data providers.

A unique feature of the analyst pairwise design is that it mitigates a variety of endogeneity concerns. When a brokerage subscribes to a new FDP, it affects data provider similarity with both subscribing *and* non-subscribing peer brokerages. Hence, selection concerns are at the pairwise level, rather than at the individual level. Further, incentives for data provider herding are likely to be minimal in this case, as anchoring on other analysts' forecasts is a low-cost solution to purchasing expensive data subscriptions. That being said, we include a specification that includes brokerage-pair fixed effects. Doing so allows us to hold constant the fixed similarities across brokerage pairs, thereby exploiting inter-temporal variation in FDP similarity that results from changes in subscriptions over time. Under this more stringent specification, we continue to find that increases in FDP similarity are significantly positively associated with similarity in forecast point estimates, boldness, and timing.

In an additional robustness test, we also exploit *across* brokerage changes in employment by analysts as another plausibly exogenous source of variation in data provider similarity within the analyst pair. A benefit of studying employment changes is that FDPs are subscribed to at the brokerage level, and data subscriptions are unlikely to change systematically with new analyst hires.[5] We find that variation in FDP similarity from employment changes leads to consistent results.

To better understand the mechanisms underlying our results, we conduct several tests. First, we examine the nature of information offered by FDPs. If the effects of using similar sources are more pronounced with exclusive or proprietary data, then this would be consistent with analysts anchoring their decisions in part on the unique information the data providers offer. Our findings support this idea. Analysts who share FDPs that require paid subscriptions show a stronger tendency towards similar forecasting than sharing sources that mainly provide public data. Interestingly, even when brokerages use similar *public* data sources, there is still an effect on the similarity of forecasts, though it is less pronounced.[6] One interpretation of this latter result is that analysts potentially use data directly without adding much of their own interpretation, consistent with minimal diversity in information processing contributing to the effects we observe.

Second, given the large concentration of market share by relatively few data providers, we examine whether the effects we document apply broadly to both major and minor paid subscription providers. We define "major" FDPs as S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar based on their status as the top five companies in terms of market share in the

---

[5] Our conversations with analysts suggest that new brokerage hires have little sway over the brokerage's data subscriptions. While this assumption appears valid, it should be particularly true for large brokerages. When running our analysis among large brokerages, we find similar results.

[6] Public data sources include company filings, which are accessible via EDGAR or firms' websites, as well as information from conference calls, which is generally accessible to the public.

FDP industry (Al Bari, 2023). We find that major and minor data providers affect analyst forecasting behavior with generally similar effect sizes, suggesting that both play an important role in explaining our results.

Third, while explicit data provider references are the subject of study in this paper, it is likely that some analysts also have access to unreferenced soft information (e.g., via a relationship with management). Accordingly, we examine how soft information might affect analysts' anchoring on hard data from FDPs. Using All-Star analyst status as a proxy for soft information access (Mayew, 2008; Green et al., 2014), we find that the effects of data provider similarity are reduced for analysts who are more likely to have access to soft information. In additional tests, we also investigate how variation in the number of FDP subscriptions affects our results. We find that the results attenuate for analysts with enhanced data subscription access, consistent with their greater flexibility in choosing which data providers to rely on. Finally, we find that our results hold across annual forecasts issued at various times during the year, suggesting that data providers influence both short- and longer-horizon forecasting behaviors.

Our results thus far suggest that when two analysts are employed by brokerages with similar data subscriptions, the analysts' forecast values, boldness, and timing all tend to converge. In our final analysis, we aggregate our results to the consensus level. Consensus analyst estimates represent a widely used, critical benchmark in capital markets. Therefore, determining whether data provider diversity has implications for consensus estimates is an important question given our main findings. We rely on the wisdom of crowds theory, which predicts that as opinions are more diverse, the crowd becomes more "wise," leading to accuracy improvements in the crowd forecast (Surowiecki, 2005). We test this prediction in our setting and find that as analysts' consensus forecasts are based on more diverse data sources, the accuracy of the consensus tends to improve.

While we hesitate to draw normative conclusions, the collective analyses suggest that sharing data providers affects the similarity of analyst forecasting behavior and deteriorates the quality of consensus forecasts.

Our results contribute to several literatures. Most directly, we provide evidence on the diversity of data subscriptions and its effects on market participant behavior. In an era marked by large consolidations and separations of major FDPs (e.g., Thomson and Reuters, Refinitiv, etc.), our findings underscore the critical role that data providers play in preserving the diversity of opinions and the usefulness of consensus estimates. While it appears that the high level of market concentration among FDPs will continue, the effects of such concentration are still being debated. To our knowledge, we are among the first to provide evidence on how access to similar data subscriptions affects the diversity of market opinions and its implications for consumers of market research.

This study also contributes to an emerging stream of literature examining the role of FDPs in modern capital markets (Akbas et al., 2018; Bochkay et al., 2022; Larocque et al., 2023). While prior and contemporaneous studies examine institutional details of common data providers, we document the effects that data providers have on the diversity of market opinions, a heretofore underexplored issue. In particular, our study offers unique insights into the landscape of data consumption within the analyst community, emphasizing the importance of FDP diversity.

More broadly, our results contribute to the literature on the "wisdom of the crowds." While the foundational premise of this theory suggests that the aggregated opinion from diverse and independent viewpoints often results in decisions that are superior to those of any single individual (Surowiecki, 2005), the role of data source independence in shaping this wisdom is less understood.[7] Our research expands upon this prior work by demonstrating how subscriptions to similar data providers not only affect the diversity of analyst opinions, but also influence the usefulness of consensus forecasts.

Finally, by exploiting a unique data set containing the financial data feeds that analysts have access to, our findings take an important step toward piercing the "black box" of analyst research (Bradshaw, 2011; Brown et al., 2015). While we caveat that we cannot directly observe how analysts input data into their models, we believe that our novel identification of financial data providers extends the literature and opens promising avenues for future research.

## 2. Related Literature and Hypothesis Development

The nature of data that financial analysts use has been a focal point of research since the late 1980s (Barry and Brown, 1985). Generally, information used by financial analysts is grouped by the private vs. public nature of the information. Several studies, both pre- and post-Regulation Fair Disclosure, suggest that analysts enhance their forecasts by privately accessing management insights (Bowen et al., 2002; Mayew et al., 2013; Green et al., 2014). Beyond access to management, other studies have shown that analysts rely on site visits (Cheng et al., 2016) and even FOIA requests of government agencies (Klein et al., 2020) to improve their forecasting activity. In more recent work, Chi et al. (2022) examines the variety of private data sources that analysts reference in their reports and how these sources influence individual forecast accuracy.

From a public information perspective, research shows that analysts often incorporate regulatory changes and other public information into their forecasts (Plumlee, 2003). Simpson

[7] Specifically, prior research in finance and accounting primarily focuses on the general prediction that consensus forecasts are often powerful predictors of future outcomes (Chen et al., 2014; Jame et al., 2016; Bartov et al., 2018; Green et al., 2019; Huang et al., 2020).

(2010) finds that analysts use public, non-financial information in their forecasting activities, and Gibbons et al. (2021) shows that analysts who access public SEC filings via EDGAR tend to produce more accurate forecasts. Such analysts also tend to offer more in-depth and consistent analyses of the companies they cover.

Despite these findings from prior work, the effect of data subscription similarity on analysts' research remains somewhat ambiguous. Various viewpoints exist regarding how data source overlap among analysts might affect the similarity and timing of their financial projections.

(1) *Data as a Commodity*: To the extent that financial data platforms rely on, and disseminate, the same underlying firm data (e.g., firm fundamentals from EDGAR and real-time market data from stock exchanges), then financial data is essentially a commodity good. In this case, sharing similar data sources will not materially alter the underlying information signals, and as a result, would not affect the diversity of analyst actions. Interestingly, this argument has been widely supported in our conversations with practitioners and academics.

(2) *Mechanical Data Processing*: Some analysts might simply use data "as is", without adjusting it or forming their own opinions about it. Essentially, this approach defers the opinion of the analyst to that of the data provider, which aggregates and constructs market metrics. Emerging evidence suggests that the measurement of firm performance (e.g., street earnings) can vary across FDPs, which may lead to differences in market participants' valuation estimates across platforms (e.g., Bochkay et al., 2022, Larocque et al., 2023). As a result, subscribing to the same data provider could lead to more correlated investment

opinions. Moreover, training sessions that data providers offer potentially standardize how analysts interpret and use the data, encouraging uniformity in analyst forecasts.[8]

(3) *Anchoring on Proprietary Information*: To strengthen market position, FDPs increasingly offer unique, proprietary datasets in addition to their traditional offerings (Bloomberg, 2017; Refinitiv, 2019b; Bloomberg, 2023). If analysts use and anchor their opinions on such proprietary information, subscribing to the same data provider could lead to similar investment decisions. For example, if two analysts use the same satellite data for revenue projections, their forecasts might be very similar.

(4) *Output Differentiation*: If analysts know that peers covering the same firm rely on similar data feeds, this could encourage an increased focus on output differentiation, as analysts often feel a need to demonstrate their unique "edge" to their clients and superiors (Brown et al., 2015). In our conversations with current and former analysts, this perspective was mentioned several times. Supporting this argument, some scholars suggest that anti-herding behavior is quite common among analysts (e.g., Bernhardt et al., 2006). As such, knowledge of lacking data independence might induce diverse opinions, despite sharing similar data providers.

(5) *Shared Economic Signals and Divergent Beliefs:* Information theory highlights that sometimes sharing the same public information signal can cause market participants to disagree even more (akin to Kondor, 2012 and Armstrong et al., 2023). In the case of analysts, if they initially disagree about a firm's fundamental value and receive new, uncertain information from a shared FDP, their disagreement could *increase*.

[8] This intuition is related to Bochkay et al. (2022), which studies the reporting practices at Thomson Reuters and finds that when the company changed their definition of street earnings, analysts aligned their reporting practices accordingly.

Given these contrasting perspectives, the effect that data subscriptions have on the diversity of analyst opinions remains unclear ex ante. Accordingly, our first hypothesis is in the null form:

*H1: The similarity in financial data provider subscriptions is unrelated to the similarity of sell-side analysts' forecasting behavior.*

We next consider whether data provider diversity ultimately affects consensus forecast accuracy. Building on the premise that diverse opinions tend to enhance the wisdom of crowds (Surowiecki, 2005), if data diversity affects the diversity of analyst opinions (i.e., a rejection of H1), we would also expect it to affect the accuracy of consensus forecasts. In other words, increased data diversity could lead to improved diversity of analyst opinions and thereby improved forecast accuracy at the consensus level. Alternatively, it is plausible that even if data diversity affects analyst opinions, such variation may not correlate with improvements in consensus forecast accuracy. This could happen if data diversity increases the noise in consensus analyst estimates, while having a limited effect on improving signal quality (i.e., the signal-to-noise ratio decreases). Accordingly, our second hypothesize, in the null form, is as follows:

*H2: The diversity in financial data providers among analysts contributing to the consensus is unrelated to consensus forecast accuracy.*

## 3. Research Design

### *3.1 Data and Sample*

We begin our sample construction by extracting the data source references from a sample of approximately 595,642 analyst reports obtained from Thomson ONE, issued during the years 2008-2017. 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.[9,10] Accordingly, 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.[11] Next, we link the analyst reports and resulting source information to the I/B/E/S detail file. We retain only those sources that are mentioned by five brokerages or more to aid in removing analyst references to internal data sources. Using the identified data source references, we construct a panel of brokerage months that includes the active sources within each brokerage for a given month. Since data subscriptions usually last for several months, 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.[12] Our focus is on brokerage-level data subscriptions for three key reasons: (1) these subscriptions constitute the data analysts have access to; (2) our conversations with analysts indicated that data subscriptions are a brokerage-level decision; and (3) focusing on brokerage-level subscriptions has the added advantage of

[9] We obtained these inferences by randomly sampling 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 report 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 sources."

[10] There are various reasons why analysts cite the data sources they use when preparing research reports. First, there is a legal basis for citing financial data sources, as data providers often have licensing agreements that require source attribution (e.g., Thomson Reuters' General Terms and Conditions lists this requirement). Second, ethical guidelines from the CFA Institute and analyst employers encourage transparency and credibility in reporting. Finally, anecdotal evidence from discussions with a prior UBS equity analyst suggests that analysts also reference the sources they use to increase clients' confidence in the report content, consistent with a credibility motive. Brokerage reports that do not follow this referencing convention are excluded from our analysis to mitigate source disclosure selection concerns.

[11] 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 therein. We then identified the most common sources referenced among the random sample. While adding additional sources to our list might reduce measurement error in *SourceSimilarity*, we are unaware of a reason focusing on common sources would induce bias in our results.

[12] If we adjust this assumption and instead assume that a brokerage's subscription begins in the month of the analyst's reference and continues for six months, we observe similar inferences.

mitigating analyst-report-level disclosure selection concerns. Table 1 lists the top 20 data sources in our sample based on the number of unique brokerages mentioning the source. Many of the sources are well-known and include data providers such as Bloomberg, FactSet, and Thomson Reuters.

For brokerages with non-missing data source information, 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 1,362,220 analyst forecast pairs.

*3.2 Empirical Model*

We investigate whether access to similar data providers affects the attributes of analysts' forecasts. To do so, 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*. Figure 1 presents an illustration of this analyst pair research design. After forming the analyst pairs, we consider three distinct attributes of the forecasts: (1) forecast similarity, (2) forecast timing, and (3) forecast boldness. We use the following model to examine whether source similarity is associated with the aforementioned attributes:

$$SimilarForecast_{p,f,t} \ / \ SimilarTiming_{p,f,t} \ / \ SimilarBoldness_{p,f,t} = \alpha_1 SourceSimilarity_{p,t} + \alpha Controls_{p,f,t} + \beta Fixed\ Effects_{f,t} + \varepsilon_{p,f,t} \quad (1)$$

In the above model, *p* indexes unique analyst forecast pairs, *f* indexes the covered firm, and *t* indexes the year. Our primary independent variable of interest is *SourceSimilarity,* which is the number of data providers that both analysts in the pair have access to at their respective brokerages (i.e., the number of overlapping sources), scaled by the number of all possible data providers.[13] We decile rank this variable each year. Thus, higher values of *SourceSimilarity* indicate more source overlap for both analysts in the pair.

We consider three dependent variables that represent important attributes of the analysts' forecasts. First, *SimilarForecast* is the absolute value of the difference between the two forecasts in each unique analyst forecast pair. We scale this difference by the firm's stock price measured two trading days prior to the first analyst's forecast issuance date in the analyst pair and multiply this value by negative one. We decile rank the resulting value each year. Thus, higher values of *SimilarForecast* indicate more similar forecasts between the two analysts in the pair. Moreover, a positive coefficient on *SourceSimilarity* ($\alpha_1$) would be consistent with analysts' earnings point estimates becoming more similar as the analysts share more data providers.

Second, we examine *SimilarTiming,* which measures how clustered analysts' forecasts are in event time. To construct this measure, 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, and zero otherwise. Forecast horizon is the number of days between the covered firm's fiscal period end date and the forecast issuance date. A positive coefficient on *SourceSimilarity* ($\alpha_1$) would be consistent with analysts' forecast horizons becoming more similar as the analysts share more data providers.[14]

Third, we examine *SimilarBoldness*, which is set 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), and zero otherwise. We follow Clement and Tse (2005) in calculating forecast boldness,

[13] An alternative scaler is the total number of sources available to either brokerage in the pair. We observe similar inferences using this alternative approach.
[14] An alternative measurement approach is to decile rank the difference in forecast issuance days in the analyst pair. Using this alternative approach, we observe similar inferences.

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.[15] A positive coefficient on *SourceSimilarity* ($\alpha_1$) would be consistent with analysts' forecast boldness becoming more similar as the analysts share more data providers.

We include a number of fixed effects and control variables in our models to better isolate the relationship between shared data sources and analyst forecast attributes. First, we include time-varying covered firm control variables such as *BTM* (book-to-market ratio), *MVE* (market value of equity), and *ROA* (return on assets). In additional specifications, we introduce firm-year fixed effects. This augmented research design mitigates the impact of generally stable or time-invariant characteristics of the covered firms. Additionally, because this specification includes a unique fixed effect for each firm-year in our panel, it effectively neutralizes time-varying characteristics of the firms, 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, 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. Second, we control for whether the analysts are employed by brokerages with similar resources.

[15] In the relatively infrequent 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.

*SimilarResources* is 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. 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, 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.[16] We cluster standard errors at the firm-year level in each of our estimations.[17]

## 4. Results

### *4.1 Descriptive Statistics*

Table 1 Panel A provides descriptive evidence on the most commonly cited data providers in our sample. Specifically, we report the top 20 sources based on the total number of citing brokerages. We find various well-known FDPs such as Bloomberg, FactSet, S&P Capital IQ, Thomson Reuters, and others to be on this list. Additionally, as expected, we find that publicly available information sources, such as conference calls or company information on the covered firms, rank very highly on the list (with company information being the most cited source overall).[18] Further, Table 1 Panel B reports the top two non-brokerage sources most commonly cited by the 20 largest brokerages in our sample (based on report volume). Here, we observe variation in the top data providers both across and within specific brokerages. For instance,

[16] Our inferences remain unchanged if we decile rank the absolute differences in analysts' experience, brokerage size, and busyness within each analyst pair.

[17] Our inferences remain unchanged if we cluster at the brokerage-pair level.

[18] Our results are not sensitive to the exclusion of conference calls or company information as a data source, as highlighted in Table 5.

JPMorgan's top referenced source is Bloomberg, UBS and Credit Suisse rely most on products produced by Thomson Reuters, and other smaller brokerages rely more on sources such as FactSet or SNL, highlighting the data source variation in our sample. Table 1 Panel C reports a data source transition matrix. Conditional on subscribing (not subscribing) to a data source, brokerages have an 85.44% (96.38%) likelihood of subscribing (not subscribing) to that dataset the following year.

Figure 2 presents our assessment of (1) subscription features across common FDPs and (2) how similar financial data is presented across FDPs. Panel A presents the typical features that financial data platforms offer. While most data providers offer access to firm filings, market news, and include charting and data visualization tools, meaningful differences emerge when considering whether the providers offer in-house research and proprietary data, in-house news desks, messaging services, and the ability to trade in-platform. Panel B highlights differences in reporting across common financial platforms while holding the covered firm constant. Specifically, we use Ryanair's 2022 fiscal year end (March 31, 2023 report date) as an example.[19] We focus on the platforms' reported Gross Profit for simple illustrative purposes. In this example, S&P Capital IQ provides the most disaggregated information, which is different from the disaggregation used by Refinitiv Eikon and Morningstar. Bloomberg does not disaggregate related expenses in this case, and instead reports them under "Other Operating Expenses." Collectively, we observe that, even when data is drawn from the same source (firm financial reports), the aggregation and reporting *across* platforms can vary considerably. This intuition is akin to Du et al. (2023), which recently documented that Compustat and FactSet data items contain large and frequent discrepancies from as-filed data. Thus, Figure 2 highlights that while there are differences in major features across

[19] This illustration was inspired by the Harvard business case on Ryanair (Bradshaw, 2006).

platforms, there are even prominent differences when it comes to more basic tasks (e.g., reporting a firm's most recent earnings).

Table 2 provides basic descriptive statistics on the primary variables used in our models. Regarding the forecast pairs in our main sample, we find that approximately 11% are issued by analysts with similar experience, 9% have similar brokerage resources, and 13% have similarly sized analyst portfolios. Further, approximately 38% of the forecasts have similar timing and 56% have similar boldness. All remaining variables are reported in Table 2.

*4.2 Source Similarity and Forecast Similarity*

Figure 3 presents the univariate illustration of correlations between *SourceSimilarity* and (1) *SimilarForecast*, (2) *SimilarTiming*, and (3) *SimilarBoldness*. In each case, we observe a positive correlation with a generally monotonic increase across deciles of *SourceSimilarity*. This univariate observation is consistent with the inference that, as data subscription overlap increases, so too does similarity in point forecasts, forecast timing, and forecast boldness.

Table 3 presents our main empirical result examining the association between provider similarity and forecast similarity. We examine three separate dependent measures that capture unique attributes of forecast similarity based on point estimates (Panel A), forecast timing (Panel B), and forecast boldness (Panel C). In each panel, we estimate our main analysis in three ways: (1) without controls or fixed effects (Column 1) (Whited et al., 2022); (2) with controls, but no fixed effects (Column 2) (Jennings et al., 2023); and (3) with controls and fixed effects. As mentioned previously, because our third column includes firm x year fixed effects, it makes firm-year controls redundant. Thus, the variables *BTM, MVE,* and *ROA* are dropped from this column.

Table 3 Panel A reports our results when examining point forecast similarity. Across each column, we find a positive and statistically significant coefficient on *SourceSimilarity*. This suggests that forecast point estimates become more similar as analysts increasingly share the same financial data providers.[20] In terms of economic magnitude, a one standard deviation increase in source similarity equates to about a 14.67% increase in forecast similarity, relative to the mean.[21] Panel B reports our results when examining forecast timing. We continue to find a positive and statistically significant loading on *SourceSimilarity,* suggesting that sharing similar data providers not only influences forecast point estimates but can influence the timing of these estimates. Accordingly, these results suggest that sharing proprietary information is not the only driver of forecast similarity; rather, there also appears to be something mechanical in the *processing* of information that occurs when sharing similar data sources. In terms of economic magnitude, a one standard deviation increase in source similarity equates to a 3.33% increase in the probability of sharing a similar horizon decile.[22] Panel C reports our results that examine forecast boldness. Across each column, we find a positive and statistically significant coefficient on *SourceSimilarity*. In terms of economic magnitude, a one standard deviation increase in source similarity equates to a 3% increase in the probability that both analysts are similar in the boldness of their forecasts.[23] Overall, our collective evidence is consistent with data provider similarity influencing the similarity of analysts' forecasts in terms of point estimates, timing, and boldness.

*4.3 Source Similarity and Forecast Similarity – Robustness*

While our pairwise research design and firm x year fixed effects structure alleviate various concerns with endogeneity, in additional analyses we address potential alternative explanations that remain. Specifically, although we explicitly control for brokerage resource similarity in each

[20] While we consider forecast horizon to be an outcome of interest (*SimilarTiming*), results in Table 3 Panels A and C are robust to constraining to analyst pairs that release their forecasts on the exact same day.

[21] $0.051 \cdot 3.33 \cdot (0.0095/0.0110) = 14.67\%$; 0.0097 and 0.0110 are the mean and average decile change in forecast similarity (unranked), respectively.

[22] $0.010 \cdot 3.33 = 3.33\%$;

[23] $0.009 \cdot 3.33 = 3.00\%$;

of our main tests, we acknowledge that other similarities across brokerages may correlate with both data provider similarity and forecast similarity. Therefore, in a subsequent test, we seek to hold the brokerage-pairs constant and exploit inter-temporal variation in *SourceSimilarity.* To the extent that correlated omitted variables are fixed between brokerage pairs, or are uncorrelated with *changes* in data provider similarity, then such inter-temporal variation in source similarity can help rule out these alternative explanations. Accordingly, we estimate a specification of our model that includes brokerage pairwise fixed effects. Specifically, we create a distinct fixed effect for each brokerage pair in our sample of forecasts. Table 4 Panel A reports this finding. We continue to find consistent results, suggesting that fixed pairwise attributes between brokerages are not driving our main empirical findings.

To refine these inferences, we note that changes in data provider similarity, holding brokerage pairs constant, can come from two sources: (1) brokerages changing their data provider subscriptions over time and (2) analysts changing their employment (i.e., changing brokerages) over time. As such, we provide inference on each of these points in the following tests. To investigate the effects of intertemporal changes in FDP subscriptions 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. A unique benefit of adding analyst pair fixed effects is that it also helps mitigate concerns that fixed similarities between *analysts* are driving our results. Table 4 Panel B reports this finding. In summary, we find that when brokerages change their subscriptions intertemporally, the results are consistent with data provider similarity influencing forecast similarity.

Finally, to capture the effects of *employment change* on source similarity, we control for what FDP similarity is at analysts' former employers in the current time period

(*OldSourceSimilarity*). To the extent that current access to data providers is what drives our inferences, we would expect to see effects from *SourceSimilarity* and not from *OldSourceSimilarity*. A unique benefit from this employment change analysis is that source similarity *between* brokerages is unlikely to change systematically when hiring new analysts. Table 4 Panel C reports results from this analysis. We continue to find a positive and significant coefficient on *SourceSimilarity,* while the coefficient on *OldSourceSimilarity* is insignificantly different from zero. We also observe that *SourceSimilarity* is statistically different from *OldSourceSimilarity*. Overall, this reinforces our main result, suggesting that the data providers to which brokerages subscribe influence analyst forecasting behavior, and helps mitigate a variety of omitted variable bias concerns.

*4.4 Public versus Paid Subscription Sources*

To better understand the mechanisms by which FDP similarity influences forecasting behavior, we next estimate a cross-sectional test that exploits variation in the underlying nature of the data sources in our sample. In particular, we examine whether data sources that contain primarily proprietary information (i.e., paid subscription sources) lead to a more pronounced effect than data sources containing information that is generally in the public domain (i.e., public sources).

To do so, we estimate the following model:

$$SimilarForecast/SimilarTiming/SimilarBoldness_{p,f,t} = \alpha_1 PaidSourceSimilarity_{p,t} + \alpha_2 PublicSourceSimilarity_{p,t} + \alpha Controls_{p,f,t} + \beta Fixed\ Effects_{f,t} + \varepsilon_{p,f,t} \quad (2)$$

We include all controls and fixed effects as in Model (1). However, we partition *SourceSimilarity* into both a paid and public source component. Specifically, for each forecast pair, *PaidSourceSimilarity* measures the brokerages' similarity between private FDPs, while

*PublicSourceSimilarity* measures the brokerages' similarity between public FDPs. Similar to how we construct *SourceSimilarity,* we decile rank each similarity measure each year. If private information from the data providers contributes to our result, we would expect a larger coefficient on $\alpha_1$ as compared to $\alpha_2$.

Table 5 reports this result. For brevity, we tabulate only our strictest specification that includes firm x year fixed effects. Across each of our three main dependent variables (*SimilarForecast, SimilarTiming,* and *SimilarBoldness*), we find that the *PaidSourceSimilarity* coefficient is significantly larger than *PublicSourceSimilarity*. Overall, this suggests that similarity in private information from data providers is a primary mechanism for our result.[24,25] Interestingly, even when using shared *public* data sources (*PublicSourceSimilarity*), there's still an effect on the similarity of forecasts, though it's less pronounced. One interpretation of this latter result is that analysts are potentially using data directly without adding much of their own interpretation, consistent with minimal diversity in information processing contributing, in part, to the effects of data source similarity.

Within paid subscription services, there is a high concentration of market share among five major financial data platforms: S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar (Al Bari, 2023). To provide insight into the effects of these data providers, as well as assess the generalizability of the results, we next study the effects of sharing subscriptions to a major data provider versus sharing subscriptions to a minor data provider. Specifically, for each

[24] Several paid data providers offer their clients data on consensus analyst forecasts, which might encourage analyst herding behavior. However, when we remove such data providers from our analysis (per the listing in Larocque et al., 2023), we find similar inferences. While we cannot say whether this practice contributes to our results, this analysis (untabulated) suggests that providing consensus data is not the sole factor driving our results.

[25] Cross-sectional variation in the type of data subscriptions (Table 5) and the quantity of data subscriptions (Table 8) help mitigate several correlated omitted variable concerns. Remaining correlated omitted variables would need to describe why there are cross-sectional differences across features of the underlying data subscriptions, and also explain why such variation is not consistent with a direct effect of source similarity on the outcome variables of interest.

forecast pair, *MajorSourceSimilarity* measures the brokerages' similarity in subscriptions among the five major data providers, while *MinorSourceSimilarity* measures the brokerages' similarity in subscriptions among the minor private data providers. We assess these effects by modifying equation (2), replacing *PaidSourceSimilarity* and *PublicSourceSimilarity* with *MajorSourceSimilarity* and *MinorSourceSimilarity*.

Table 6 presents these results. As before, for brevity, we tabulate only our strictest specification that includes firm x year fixed effects. The first observation from Table 6 is that the effect of subscription similarity on forecasting behavior similarity is evident in both major and minor paid subscription sources. Additionally, while the effects are generally quite similar across both major and minor paid subscription sources, the effects on *SimilarForecast* and *SimilarTiming* are more pronounced for *MajorSourceSimilarity*. Collectively, these results highlight two key insights. First, subscriptions to major data providers appear to play an important role in shaping the documented effects on forecasting convergence, particularly in light of their substantial market share. Second, subscription similarity effects appear to generalize across both major and minor data providers, suggesting the results are not simply a byproduct of access to less conventional data sources.

*4.5 Soft and Hard Information Access*

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

$$SimilarForecast/SimilarTiming/SimilarBoldness_{p,f,t} = \alpha_1 SourceSimilarity_{p,t} \cdot AllStars_{p,t} + \alpha_2 SourceSimilarity_{p,t} + \alpha_3 AllStars_{p,t} + \alpha Controls_{p,f,t} + \beta Fixed\ Effects_{f,t} + \varepsilon_{p,f,t} \quad (3)$$

We include all controls and fixed effects as in Model (1). In this model, we use analysts' All-Star status as a proxy for soft information access (Mayew, 2008; Green et al., 2014). Specifically, we interact *SourceSimilarity* with an indicator variable, *AllStars,* that is set equal to one if both analysts in the pair receive the All-Star award designation during the year, and zero otherwise. Because All-Star analysts tend to have greater access to soft information, our effect may be attenuated for these analysts. In this case, the coefficient on *SourceSimilarity · AllStars* ($\alpha_1$) would be negative.

Table 7 reports this result. As before, we tabulate only our strictest specification that includes firm x year fixed effects. Across each of our three main dependent variables (*SimilarForecast, SimilarTiming,* and *SimilarBoldness*), we find a negative and significant coefficient on *SourceSimilarity · 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 their brokerages' data feeds).

In contrast to soft information, brokerages also have varying levels of hard information resources, with the importance of data feeds likely varying substantially across brokerages. For instance, analysts at brokerages with more data sources have relatively more flexibility in their choice of data to rely on. Following this intuition, we expect the effect of sharing data to be weaker for analysts with access to more data subscriptions. To examine this in greater depth, we exploit variation in our sample by measuring the number of data sources each brokerage reports. We then rank the brokerages each year based on the number of available data sources. If the analysts in a given pair are both employed by brokerages in the upper $50^{th}$ percentile of data source subscriptions, we set the variable *HighSourceAccess* equal to one and zero otherwise. We then interact *HighSourceAccess* with *SourceSimilarity* and include these variables in a modified version of Model (3) from above. If we observe a negative and significant coefficient on *SourceSimilarity* · *HighSourceAccess*, this would be consistent with our results being attenuated when analysts have access to a greater number of data sources.

Table 8 reports this result. As before, we tabulate only our strictest specification that includes firm x year fixed effects. Across each of our three main dependent variables (*SimilarForecast, SimilarTiming,* and *SimilarBoldness*), we find a negative and significant coefficient on *SourceSimilarity* · *HighSourceAccess.* Overall, this suggests that data provider similarity is less potent for analysts with access to a greater number of hard financial data sources.[26]

*4.6 Longer Horizon Forecasts*

Our primary sample consists of analysts' most recent (i.e., last) one-year-ahead annual earnings forecasts issued at least a month before the covered firm's fiscal year-end date. Accordingly, we next assess the generalizability of our inferences to longer forecast horizons. Specifically, we retain analysts' *first* annual forecasts issued for a given firm's fiscal year-end, and we require that the forecasts be issued during the first two quarters of the year.

Table 9 reports this result. We tabulate only our strictest specification that includes firm x year fixed effects. Across the dependent variables *SimilarForecast* and *SimilarTiming,* we find a positive and significant coefficient on *SourceSimilarity,* which is consistent with our main findings

[26] Because larger brokerages generally have access to more data subscriptions than smaller brokerages, in untabulated analyses we examine whether our results are robust to examining forecasts issued by analysts *only* at large brokerages. We find similar inferences and effect sizes as our main results presented in Table 3. 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.

reported in Table 3 (based on a sample of analysts' last annual forecasts).[27] Interestingly, coefficient magnitudes are somewhat attenuated over longer horizons, indicating that data subscriptions may be more important the shorter the forecast horizon. Overall, these results suggest that our inferences are not overly sensitive to the forecast horizon or the sample of analyst forecasts we employ.

*4.7 Source Independence and Consensus Forecast Accuracy*

The wisdom of crowds theory suggests that, as opinions are more diverse, the crowd becomes more "wise," resulting in improvements to the accuracy of the crowd forecast (Surowiecki, 2005). To examine how data independence relates to this intuition, we aggregate our results to the consensus level. To do so, we estimate the following model:

$$ConsensusAccuracy_{f,t} = \alpha_1 AvgSourceIndependence_{f,t} + \alpha Controls + \beta Fixed\ Effects_{f,t} + \varepsilon_{f,t} \tag{4}$$

In the above model, *f* indexes firms and *t* indexes year. *AvgSourceIndependence* is our key independent variable of interest, and is defined as the average value of *SourceSimilarity,* prior to its decile ranking, calculated at the firm-year level. The variable is then decile ranked by year and multiplied by negative one. As a result, increases in *AvgSourceIndependence* reflect more data source diversity. Our dependent measure is *ConsensusAccuracy,* which is defined as the absolute value of the difference between the covered firm's reported earnings and the consensus forecast, scaled by the firm's stock price from the most recent quarter, multiplied by negative one, and decile ranked by year. Thus, higher values of *ConsensusAccuracy* indicate a more accurate consensus forecast. Overall, if data provider diversity leads to a more accurate consensus forecast, we would observe a positive coefficient on *AvgSourceIndependence* ($\alpha_1$).

[27] We are unable to examine the outcome variable *SimilarBoldness* in this test because we examine analysts' *first* forecast. As highlighted in Appendix B, calculating forecast boldness requires at least two forecasts from the same analyst.

Given that this model includes only one unique observation at the firm-year level, we are unable to include firm-year fixed effects. However, we do 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*), the average horizon for each forecast forming the consensus (*AvgHorizon*), and the number of unique analysts contributing to the consensus forecast. We also include several additional control variables related to the covered firms that are associated with consensus forecast accuracy. In particular, we include the firms' book-to-market ratio (*BTM*), size (market value of equity, *MVE*), profitability (*ROA*), and an indicator for whether the firm reports a loss (*Loss*).

Table 10 reports this result. Panel A shows results using the mean consensus value, while Panel B uses the median. In Column 1, we include firm fixed effects and the control variables described above. In Column 2, we include firm and year fixed effects, along with the control variables. Across both panels and both columns, we find a positive and significant coefficient on *AvgSourceIndependence*. Overall, this suggests that, as analysts' consensus forecasts are based on more diverse data providers, the accuracy of the consensus forecast tends to improve. This finding is consistent with the wisdom of crowds theory and highlights the importance of financial data diversity when forming consensus opinions.

## 5. Conclusion

Our study investigates whether using similar financial data providers affects the diversity of sell-side analysts' forecasting behavior. Using a comprehensive and novel dataset of 595,642 equity research reports to identify the "sources" analysts rely on, we find compelling evidence that sharing similar data providers significantly influences analysts' forecasting behavior. Specifically, analysts who subscribe to similar data providers tend to exhibit greater similarity in their forecast values, timing, and boldness. These effects are more pronounced when the data providers offer proprietary information and are attenuated when analysts have access to soft information or have access to a greater number of unique financial data providers. Furthermore, our results suggest that both large and small data providers (in terms of market share) influence analyst forecasting behavior. Finally, consistent with the wisdom of crowds theory, we also find that consensus estimates that exhibit more data source diversity in the underlying forecasts tend to be more accurate.

While our findings provide valuable insights into how data provider similarity shapes market opinions, certain caveats are worth noting. First, our study focuses on sell-side analysts. As such, the dynamics we observe may not directly translate to other market participants, such as investors, market makers, etc. Second, we base our inferences on analyst reports that reveal the data providers that brokerages subscribe to. While we do not anticipate omitted brokerages to bias inferences in a systematic way given our analyst-pairwise research design, we cannot rule this out definitively. Lastly, while we've attempted to rule out alternative explanations for our findings using an analyst pairwise research design, brokerage pair fixed effects, changes in analysts' employment, and cross-sectional analyses, we acknowledge that other omitted explanations may persist. However, we note that any such alternative explanations would need to align with the totality of our results. Despite these caveats, we believe our findings provide novel insights into the role of data independence and its effects on analyst research.

We also believe that our novel identification of sources that analysts use opens several avenues for future research. For instance, researchers might examine the role of different data providers in forming market opinions, and how that may depend on specific market conditions, regulatory shifts, or technological innovations. Research might also study how the adoption of data subscriptions affects analyst activity intertemporally. Overall, we believe that studying the use of financial data by market participants will be a fruitful area of research in the years to come.

During the preparation of this work the authors used Grammarly and ChatGPT to help with editorial and programming suggestions. After using these services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

**Bibliography**


Akbas, F., Markov, S., Subasi, M., & Weisbrod, E. (2018). Determinants and consequences of information processing delay: Evidence from the Thomson Reuters Institutional Brokers' Estimate System. *Journal of Financial Economics*, 127(2), 366-388.

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* (Forthcoming).

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

Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review*, 93(3), 25-57.

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. (2006). Ryanair Holdings Plc. *Harvard Business School Publishing*, Case No. 106-003.

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.

Chen, H., De, P., Hu, Y., & Hwang, B. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5), 1367-1403.

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., & Zheng, Y. (2022). The Use and Usefulness of Big Data in Finance. *Working Paper.*

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.

Du, K., Huddart, S., & Jiang, X. D. (2023). Lost in standardization: Effects of financial statement database discrepancies on inference. *Journal of Accounting and Economics*, *76*(1), 101573.

Francis, J., & Soffer, L. (1997). The relative informativeness of analysts’ stock recommendations and earnings forecast revisions. *Journal of Accounting Research*, 35(2), 193-211.

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

Green, T. C., Huang, R., Wen, Q., & Zhou, D. (2019). Crowdsourced employer reviews and stock returns. *Journal of Financial Economics*, 134(1), 236-251.

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.

Howe, J., Unlu, E., & Yan, X. (2009). The predictive content of aggregate analyst recommendations. *Journal of Accounting Research*, 47(3), 799-821.

Huang, K., Li, M., & Markov, S. (2020). What do employees know? Evidence from a social media platform. *The Accounting Review*, *95*(2), 199-226.

Jame, R., Johnston, R., Markov, S., & Wolfe, M. (2016). The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, *54*(4), 1077-1110.

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.

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

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 (2019a). The rise of data analytics in trading | Future of trading. Retrieved November 8, 2023, from https://www.refinitiv.com/en/resources/special-report/data-analytics-in-trading

Refinitiv (2019b). 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.

Trueman, B. (1994). Analyst forecasts and herding behavior. *The Review of Financial Studies*, *7*(1), 97-124.

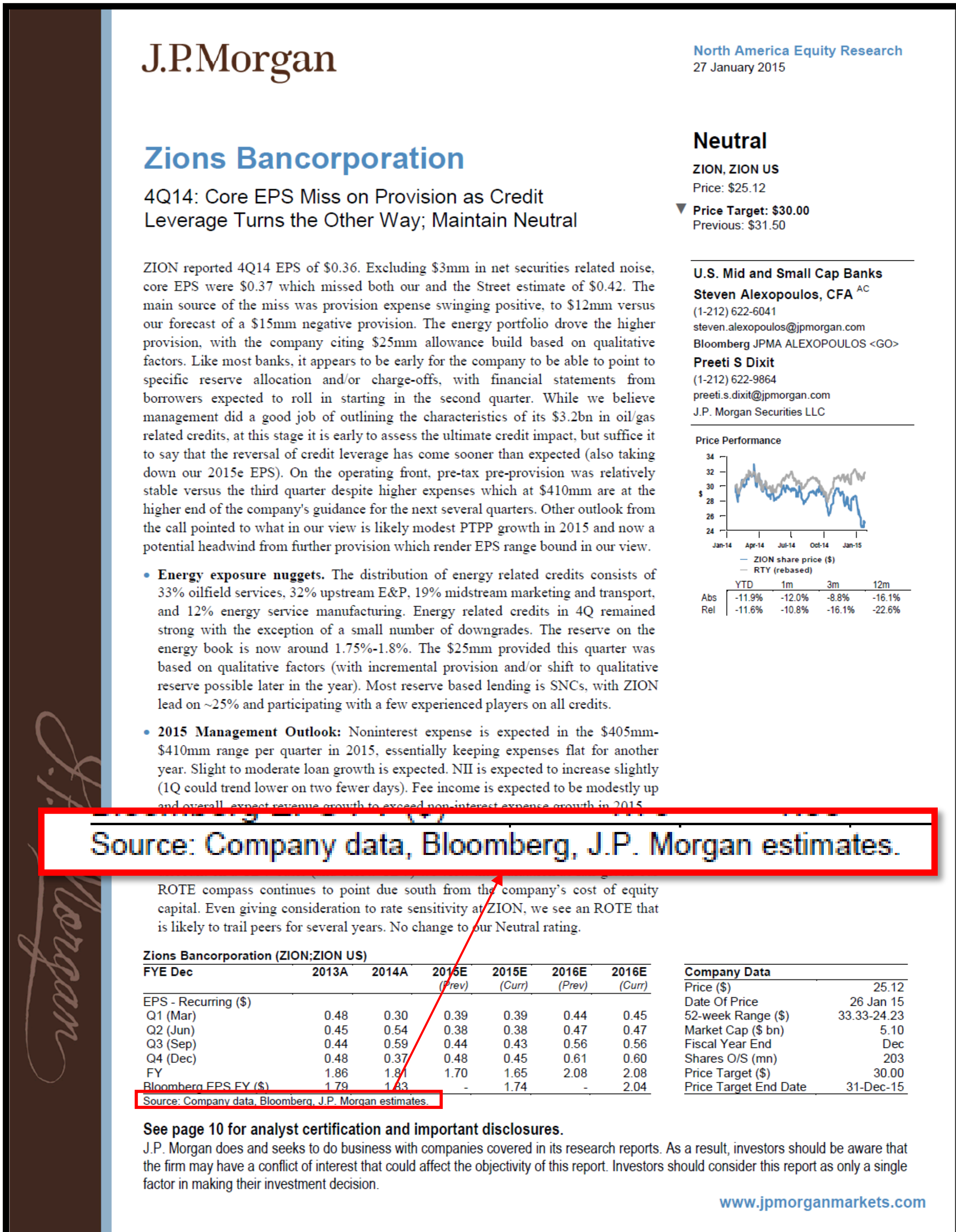
Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics*, *58*(3), 369-396.

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.

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance*, 51(1), 137-167.

## Appendix A – Examples of Source References in Analyst Reports

The images below contain pages from three different equity research reports, and are representative of the data in our analyses. Source references mentioned in these reports 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 and overall, expect revenue growth to exceed non-interest expense growth in 2015.

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 A – Examples of Source References in Analyst Reports, Continued

UBS | Global Research | 18 December 2013

### Initiation of Coverage

# American Airlines Group

## Merger Integration Risk and Weak Cash Mask Improving Industry Fundamentals

**Equities**

Americas
Airlines

| | |
|---|---|
| **12-month rating** | **Neutral** |
| | *Prior: Not Rated* |
| **12m price target** | **US$29.00** |
| | *Prior: -* |
| **Price** | **US$26.10** |
| **RIC:** AAL.O **BBG:** AAL US | |

**Trading data and key metrics**

| | |
|---|---|
| **52-wk range** | US$26.61-21.87 |
| **Market cap.** | US$4.24bn |
| **Shares o/s** | 162m (COM) |
| **Free float** | 99% |
| **Avg. daily volume ('000)** | 24,062 |
| **Avg. daily value (m)** | US$593.3 |
| **Common s/h equity (12/13E)** | US$14.3bn |
| **P/BV (12/13E)** | 0.3x |
| **Net debt / EBITDA (12/13E)** | 4.1x |

**EPS (UBS, diluted) (US$)**

| | 12/13E | | | |
|---|---|---|---|---|
| | From | To | % ch | Cons. |
| **Q1** | - | 0.27 | - | 0.02 |
| **Q2** | - | 1.56 | - | 0.92 |
| **Q3** | - | 1.16 | - | 1.38 |
| **Q4E** | - | 0.84 | - | 0.50 |
| **12/13E** | - | 3.75 | - | 2.51 |
| **12/14E** | - | 1.95 | - | 3.71 |
| **12/15E** | - | 2.80 | - | 4.50 |

Darryl Genovesi
Analyst
darryl.genovesi@ubs.com
+1-212-713 4016

**Start at Neutral**

Post restructuring, AAL offers what we think could be the best earnings growth outlook among the airlines we cover. However, we start it at Neutral as we think its financial forecast appears aggressive relative to prior mergers while cash generation will likely be weak for several years and technical/liquidity risk and remaining bankruptcy claims are wildcards. AAL looks cheap on earnings, but more expensive on EBITDAR and cash given higher leverage compared to the group and pending ramp up in new aircraft deliveries.

**Cheap on earnings, but no cash**

While AAL trades at a discount on earnings, which we expect should grow meaningfully, we don't foresee it generating much FCF over the next few years as it spends nearly $5B annually on merger integration and new airplanes. We also see risk to AAL's five-year revenue and EBITDAR forecasts from its restructuring plan. While our estimates are not much different from the few published estimates out there, a consensus is still forming, and we think the stock could lag ahead of what we see as a likely cut to guidance.

**Potential selling pressure as new shares are issued**

Natural sellers, principally AMR debt holders and labor, will be issued more than 300 million shares of AAL stock over the next four months, representing roughly 40% of all AAL stock. Given our view on AAL's financial forecasts and cash flow outlook, we will wait for this liquidity risk to subside, updated proforma financials to be issued, and the integration plan to be read before revisiting our Neutral stance.

**Valuation: $29 price target**

Our price target reflects 6x our 2015 EBITDAR estimate.

Source: Company accounts, Thomson Reuters, UBS estimates.

| Highlights (US$m) | 12/10 | 12/11 | 12/12 | 12/13E | 12/14E | 12/15E | 12/16E | 12/17E |
|---|---|---|---|---|---|---|---|---|
| **Revenues** | 11,908 | 13,055 | 13,830 | 16,093 | 42,314 | 44,055 | 45,827 | 47,696 |
| **EBIT (UBS)** | 785 | 452 | 893 | 1,452 | 3,329 | 4,312 | 5,065 | 5,889 |
| **Net earnings (UBS)** | 447 | 112 | 536 | 942 | 1,478 | 2,132 | 2,643 | 3,199 |
| **EPS (UBS, diluted) (US$)** | 2.22 | 0.68 | 2.62 | 3.75 | 1.95 | 2.80 | 3.45 | 4.15 |
| **DPS (US$)** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **Net (debt) / cash** | (2,541) | (2,619) | (2,417) | (7,327) | (7,644) | (7,624) | (6,525) | (4,472) |

| Profitability/valuation | 12/10 | 12/11 | 12/12 | 12/13E | 12/14E | 12/15E | 12/16E | 12/17E |
|---|---|---|---|---|---|---|---|---|
| **EBIT margin %** | 6.6 | 3.5 | 6.5 | 9.0 | 7.9 | 9.8 | 11.1 | 12.3 |
| **ROIC (EBIT) %** | 23.5 | 13.8 | 24.8 | 8.8 | 11.1 | 13.8 | 15.6 | 17.7 |
| **EV/EBITDA (core) x** | - | - | - | 9.4 | 4.1 | 3.3 | 2.8 | 2.2 |
| **P/E (UBS, diluted) x** | - | - | - | 7.0 | 13.4 | 9.3 | 7.6 | 6.3 |
| **Equity FCF (UBS) yield %** | - | - | - | (18.7) | (12.2) | (0.2) | 25.2 | 47.7 |
| **Net dividend yield %** | - | - | - | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Source: Company accounts, Thomson Reuters, UBS estimates. Metrics marked as (UBS) have had analyst adjustments applied. Valuations: based on an average share price that year, (E): based on a share price of US$26.10 on 17 Dec 2013 19:45 EST

www.ubs.com/investmentresearch

## Appendix A – Examples of Source References in Analyst Reports, Continued

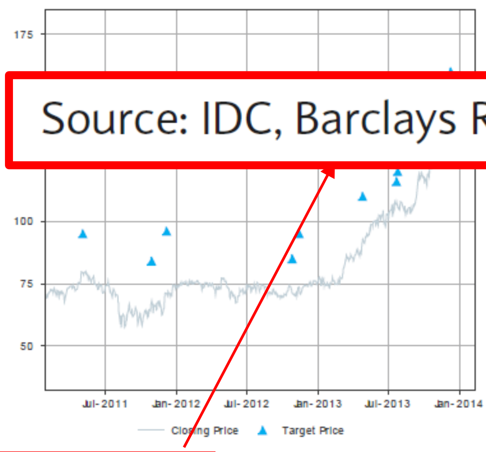
Barclays | Boeing Co.

**IMPORTANT DISCLOSURES CONTINUED**

**Boeing Co. (BA)** | Stock Rating | Industry View

**USD 136.65 (24-Jan-2014)** | **OVERWEIGHT** | **POSITIVE**

**Rating and Price Target Chart - USD (as of 24-Jan-2014)** | **Currency=USD**




Source: IDC, Barclays Research

| Date | Closing Price | Rating * | Adjusted Price Target |
|---|---|---|---|
| 09-Dec-2013 | 134.68 | | 160.00 |
| 24-Oct-2013 | 128.98 | | 145.00 |
| [illegible] | [illegible]70 | | 120.00 |
| [illegible] | [illegible]36 | | 116.00 |
| [illegible] | [illegible]7 | | 110.00 |
| 12-Nov-2012 | 73.69 | | 95.00 |
| 25-Oct-2012 | 71.54 | | 85.00 |
| 05-Dec-2011 | 71.09 | | 96.00 |
| 27-Oct-2011 | 67.49 | | 84.00 |
| 02-May-2011 | 79.53 | | 95.00 |

Source: Thomson Reuters, Barclays Research

Historical stock prices and price targets may have been adjusted for stock splits and dividends.

*The rating for this security remained Overweight during the relevant period.

Source: IDC, Barclays Research

*Link to Barclays Live for interactive charting*

A: Barclays Bank PLC and/or an affiliate has been lead manager or co-lead manager of a publicly disclosed offer of securities of Boeing Co. in the previous 12 months.

C: Barclays Bank PLC and/or an affiliate is a market-maker and/or liquidity provider in equity securities issued by Boeing Co. or one of its affiliates.

D: Barclays Bank PLC and/or an affiliate has received compensation for investment banking services from Boeing Co. in the past 12 months.

J: Barclays Bank PLC and/or an affiliate trades regularly in the securities of Boeing Co..

K: Barclays Bank PLC and/or an affiliate has received non-investment banking related compensation from Boeing Co. within the past 12 months.

L: Boeing Co. is, or during the past 12 months has been, an investment banking client of Barclays Bank PLC and/or an affiliate.

M: Boeing Co. is, or during the past 12 months has been, a non-investment banking client (securities related services) of Barclays Bank PLC and/or an affiliate.

N: Boeing Co. is, or during the past 12 months has been, a non-investment banking client (non-securities related services) of Barclays Bank PLC and/or an affiliate.

**Valuation Methodology:** Our price target of $160 is based on our view that within the next 12 months, BA shares can achieve a ~18.0x P/E on our 2015 "Core" EPS estimate of $8.85.

**Risks which May Impede the Achievement of the Barclays Research Price Target:** Boeing's served markets are military and commercial aerospace markets, which are subject to large cyclical movements. The commercial aerospace industry is currently in its worst downturn ever and a slow recovery in demand for air travel or increases in airline bankruptcies would have major negative implications for the company. Boeing realizes substantial tax benefits from favorable tax treatment of export sales. The future of these tax benefits are uncertain and their repeal would have a material negative impact on our forecasts.

27 January 2014 | 8

**Appendix B – Variable Definitions**

| **Dependent Variables:** | **Definition:** |
|---|---|
| *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. |
| *ConsensusAccuracy* | is the absolute value of the difference between the covered firm's reported earnings and the average consensus forecast, scaled by the firm's stock price from the most recent quarter, multiplied by negative one, and decile ranked by year. The consensus forecast is calculated using the most recent analyst forecasts issued thirty days before the firm's earnings announcement date. |
| **Independent Variables:** | **Definition:** |
| *SourceSimilarity* | is the number of sources that both analysts in the pair have access to at their respective brokerages, scaled by the number of all possible data sources, 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. |
| *OldSourceSimilarity* | is the *SourceSimilarity* between an analyst's prior brokerage and the brokerage of the paired analyst, in the concurrent period. |
| *PaidSourceSimilarity* | is the percentage of private sources that both analysts in the pair have access to at their respective brokerages, decile ranked by year. |
| *PublicSourceSimilarity* | is the percentage of public sources that both analysts in the pair have access to at their respective brokerages, decile ranked by year. |
| *MajorSourceSimilarity* | is the percentage of major, paid sources that both analysts in the pair have access to at their respective brokerages, decile ranked by year. Major, paid sources are defined as S&P Capital IQ, FactSet, Bloomberg, Thomson Reuters, and Morningstar. |
| *MinorSourceSimilarity* | 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. |

| | |
|---|---|
| *HighSourceAccess* | 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. |
| *AvgSourceIndependence* | is the average value of *SourceSimilarity,* prior to its decile ranking, calculated at the firm-year level. The variable is then decile ranked by year and multiplied by negative one. |
| *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. |
| *AnalystCount* | is the number of analysts 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. |

**Figure 1 – Visual Depiction of Analyst Pair Comparisons**

This figure illustrates the pairwise comparison made in our main analyses between analysts regarding their forecast and data source similarity. In this design, we match each analyst forecasting for firm *f* (e.g., Ford) with fiscal period end date *t* to all other analysts forecasting for the same firm (e.g., Ford) and the same fiscal period end date *t*. We retain one unique pairing between each analyst forecasting for firm *f* with fiscal period end date *t*.

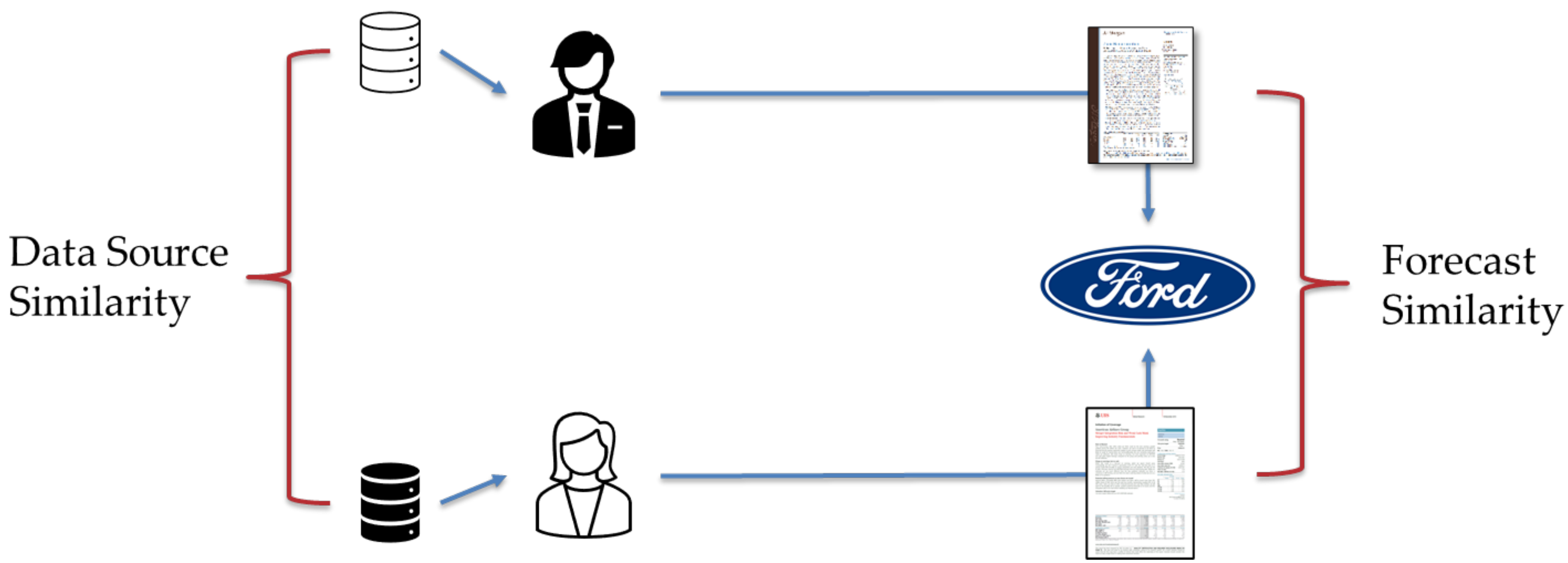


**Figure 2 – Platform Comparisons**

This figure compares features and reporting differences across common financial data providers. Panel A compares the prominent subscription features across financial data platforms. Panel B highlights differences in reporting across financial platforms, using Ryanair's 2022 fiscal year end (March 31, 2023 report date) as a demonstration.

*Panel A: Subscription Features across Financial Data Platforms*

| *Feature List* | *S&P Capital IQ* | *Bloomberg* | *FactSet* | *Refinitiv Eikon* | *Morningstar* | *Yahoo! Finance* | *Compustat* | *EDGAR* |
|---|---|---|---|---|---|---|---|---|
| In-House News Desk | No | Yes | No | Yes | No | Yes | No | No |
| Proprietary Research | No | Yes | No | Yes | Yes | No | No | No |
| Access to Market News | Yes | Yes | Yes | Yes | Yes | Yes | No | No |
| Messaging Service | No | Yes | No | Yes | No | No | No | No |
| Charting, Data Viz, & Analytics | Yes | Yes | Yes | Yes | Yes | Yes | No | No |
| Access to Company Filings | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Filing Alerts and Monitoring | Yes | Yes | Yes | Yes | Yes | No | No | Yes |
| Trade Execution | No | Yes | No | Yes | No | No | No | No |

**Figure 2 – Platform Comparisons, Continued**

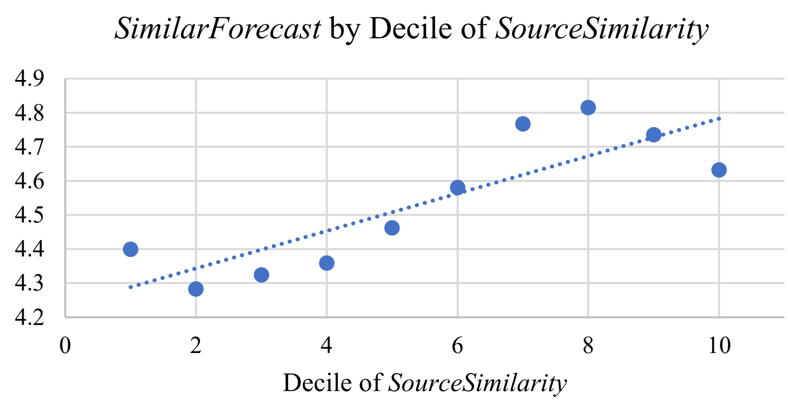
*Panel B: Gross Profit across Financial Data Platforms*

| | Ryanair Gross Profit, 2022 | | | | | | |
|---|---|---|---|---|---|---|---|
| | *S&P Capital IQ* | *Bloomberg*[*] | *Refinitiv Eikon* | *Morningstar* | *Yahoo! Finance* | *Compustat* | *EDGAR*[*] |
| **Revenues** | 10,775.20 | 10,775.20 | 10,775.20 | 10,775.20 | 10,775.20 | 11,706.33 | 10,775.20 |
| Scheduled Revenues | 6,930.30 | N/A | N/A | N/A | N/A | N/A | 6,930.30 |
| Ancillary Revenues | 3,844.90 | N/A | N/A | N/A | N/A | N/A | 3,844.90 |
| | | | | | | | |
| **Cost of Revenues** | 7,735.00 | 7,604.50 | 7,466.8 | 8,552.20 | 8,658.20 | 8,403.41 | 9,332.60 |
| Fuel & Oil | 4,025.70 | N/A | N/A | 4,025.70 | N/A | N/A | 4,025.70 |
| Route Charges | 903.70 | N/A | N/A | N/A | N/A | N/A | 903.70 |
| Staff Cost | 1,191.40 | N/A | N/A | 1,085.40 | N/A | N/A | 1,191.40 |
| Airport & Handling Charges | 1,240.50 | N/A | N/A | N/A | N/A | N/A | 1,240.50 |
| Mainten., Materials & Repairs | 373.70 | N/A | N/A | 373.70 | N/A | N/A | 373.70 |
| Depr. and Amort. | N/A | N/A | 923.20 | 923.20 | N/A | N/A | 923.20 |
| Mrkting, Distr., & Other | N/A | N/A | N/A | N/A | N/A | N/A | 674.40 |
| Cost of Revenues, Other | N/A | N/A | 6,543.6 | 2,144.20 | N/A | N/A | N/A |
| | | | | | | | |
| **Gross Profit** | 3,040.20 | 3,170.70 | 3,308.40 | 2,223.00 | 2,117.00 | 3,302.92 | 1,442.60 |

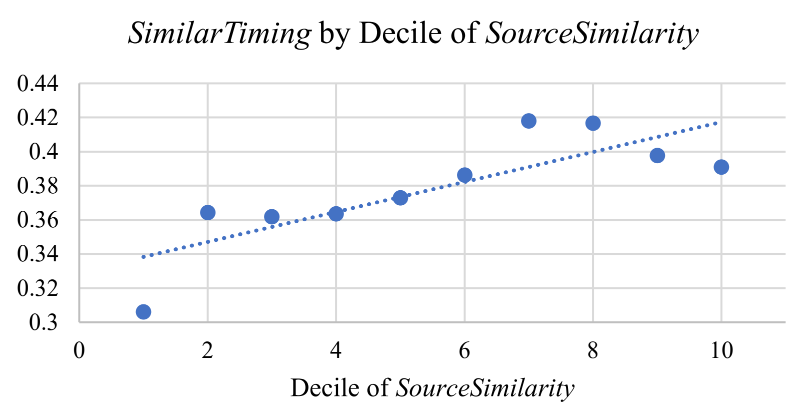
[*] Bloomberg and EDGAR do not report a cost of revenue number for Ryanair (instead, all expenses are listed as operating expenses). Since Bloomberg's income statement does not disaggregate any common Cost of Revenue charges, we list their "Other Operating Expense" number as "Cost of Revenues" to be as consistent as possible with the other platforms. EDGAR does disaggregate such items on the presented income statement, and as such, we list all such referenced expenses, if available.

**Figure 3 – Source Similarity and Forecast Similarity**
The graphs below plot the values of *SimilarForecast* (Panel A), *SimilarTiming* (Panel B), and *SimilarBoldness* (Panel C) across deciles of *SourceSimilarity*.

*Panel A: Similarity in Point Forecasts*




*Panel B: Similarity in Forecast Timing*




*Panel C: Similarity in Forecast Boldness*

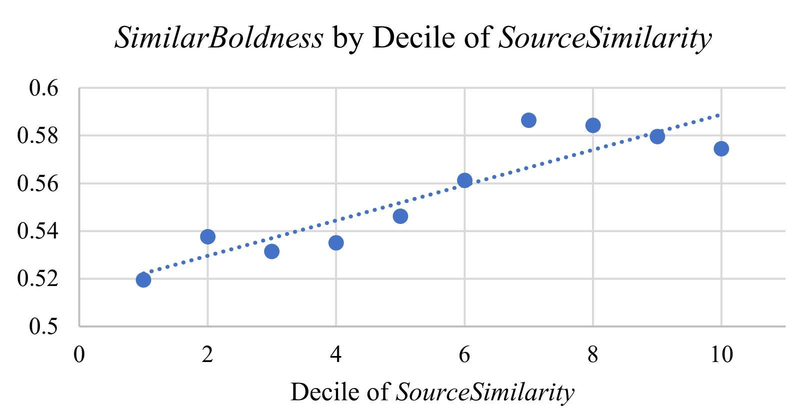


**Table 1 – Data Source Descriptive Statistics**

This table provides descriptive evidence on the sources that analysts cite in our sample of reports. Panel A lists the top 20 cited sources. Panel B lists the top sources for the 20 largest brokerages in our sample, based on the number of reports. For interpretational convenience, we exclude in-house brokerage references.

*Panel A: Top 20 Cited Data Sources*

| Source | # of Brokerages Citing the Source | % of Brokerages |
|---|---|---|
| Company | 260 | 91% |
| Bloomberg | 188 | 66% |
| FactSet | 134 | 47% |
| Conference Call | 132 | 46% |
| 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% |
| IBES | 40 | 14% |
| JP Morgan | 39 | 14% |
| Datastream | 37 | 13% |

**Table 1 – Data Source Descriptive Statistics, Continued**

*Panel B: Top Source References for the 20 Largest Brokerages*

| Brokerage Name | Top Source | Second Top Source |
|---|---|---|
| JPMorgan | Bloomberg | Reuters |
| RBC Capital Markets | Bloomberg | FactSet |
| UBS Research | Thomson | Reuters |
| Credit Suisse | Thomson | Reuters |
| Deutsche Bank | Thomson | FactSet |
| Piper Jaffray | Bloomberg | FactSet |
| Wells Fargo Securities | Reuters | FactSet |
| Jefferies | FactSet | Bloomberg |
| Morgan Stanley | Thomson | Thomson Reuters |
| Suntrust Robinson Humphrey | FactSet | SNL |
| BMO Capital Markets | FactSet | Thomson |
| William Blair & Company | FactSet | Thomson |
| Macquire Research | FactSet | Bloomberg |
| Stephens Inc. | FactSet | SNL |
| Evercore ISI | FactSet | S&P Capital IQ |
| Canaccord Genuity | FactSet | Thomson |
| JMP Securities LLC | Thomson | Reuters |
| Oppenheimer and Co. | Bloomberg | FactSet |
| Wedbush Securities Inc. | Thomson | Reuters |
| Sandler O'Neill & Partners | SNL | FactSet |

*Panel C: Data Source 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**
This table provides descriptive statistics for our primary sample of 1,362,220 pairwise observations. Variable definitions are provided in the appendix.

| Variable | N | Mean | Std. Dev | 25th | Median | 75th |
|---|---|---|---|---|---|---|
| *SimilarForecast* | 1,362,220 | 4.53 | 2.87 | 2.00 | 5.00 | 7.00 |
| *SimilarTiming* | 1,362,220 | 0.38 | 0.49 | 0.00 | 0.00 | 1.00 |
| *SimilarBoldness* | 1,362,220 | 0.54 | 0.50 | 0.00 | 1.00 | 1.00 |
| *SourceSimilarity* (raw value) | 1,362,220 | 0.12 | 0.10 | 0.03 | 0.09 | 0.17 |
| *SourceSimilarity* (decile ranked) | 1,362,220 | 4.54 | 2.84 | 2.00 | 4.00 | 7.00 |
| *SimilarExperience* | 1,362,220 | 0.11 | 0.31 | 0.00 | 0.00 | 0.00 |
| *SimilarResources* | 1,362,220 | 0.09 | 0.29 | 0.00 | 0.00 | 0.00 |
| *SimilarBusyness* | 1,362,220 | 0.13 | 0.33 | 0.00 | 0.00 | 0.00 |
| *AllStars* | 1,362,220 | 0.02 | 0.15 | 0.00 | 0.00 | 0.00 |
| *HighSourceAccess* | 1,362,220 | 0.29 | 0.45 | 0.00 | 0.00 | 1.00 |
| *PaidSourceSimilarity* | 1,362,220 | 4.53 | 2.84 | 2.00 | 4.00 | 7.00 |
| *PublicSourceSimilarity* | 1,362,220 | 4.60 | 2.34 | 2.00 | 5.00 | 6.00 |
| *MajorSourceSimilarity* | 1,362,220 | 4.52 | 2.85 | 2.00 | 5.00 | 7.00 |
| *MinorSourceSimilarity* | 1,362,220 | 4.52 | 2.75 | 2.00 | 4.00 | 7.00 |
| *BTM* | 1,362,220 | 4.50 | 2.87 | 2.00 | 5.00 | 7.00 |
| *MVE* | 1,362,220 | 4.50 | 2.87 | 2.00 | 5.00 | 7.00 |
| *ROA* | 1,362,220 | 4.50 | 2.87 | 2.00 | 4.00 | 7.00 |

**Table 3 – Source Similarity and Forecast Similarity**

This table provides our main results from estimating Model (1), in which we investigate the relationship between shared data sources and various forecast attributes. In Panel A, the dependent variable is *SimilarForecast.* In Panel B, the dependent variable is *SimilarTiming.* In Panel C, the dependent variable is *SimilarBoldness.* 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) |
|---|---|---|---|
| ***SourceSimilarity*** | **0.057***** | **0.052***** | **0.051***** |
| | **(22.64)** | **(23.67)** | **(33.54)** |
| *SimilarExperience* | | -0.022*** | 0.027*** |
| | | (-2.78) | (4.57) |
| *SimilarResources* | | -0.198*** | -0.115*** |
| | | (-26.44) | (-20.59) |
| *SimilarBusyness* | | -0.149*** | 0.020*** |
| | | (-12.99) | (3.54) |
| *BTM* | | -0.227*** | |
| | | (-32.53) | |
| *MVE* | | 0.163*** | |
| | | (26.33) | |
| *ROA* | | 0.163*** | |
| | | (23.33) | |
| Firm-Year FE | No | No | Yes |
| N | 1,362,220 | 1,362,220 | 1,357,254 |
| Adj. $R^2$ | 0.00 | 0.17 | 0.51 |

**Table 3 – Source Similarity and Forecast Similarity, Continued**

*Panel B: Similarity in Forecast Timing*

| Dependent Variable: *SimilarTiming* | (1) | (2) | (3) |
|---|---|---|---|
| ***SourceSimilarity*** | **0.008***** | **0.008***** | **0.010***** |
| | **(24.12)** | **(25.37)** | **(31.18)** |
| *SimilarExperience* | | 0.009*** | 0.008*** |
| | | (6.47) | (5.99) |
| *SimilarResources* | | -0.037*** | -0.030*** |
| | | (-28.02) | (-24.10) |
| *SimilarBusyness* | | -0.003* | 0.007*** |
| | | (-1.87) | (5.14) |
| *BTM* | | -0.011*** | |
| | | (-16.29) | |
| *MVE* | | -0.004*** | |
| | | (-6.53) | |
| *ROA* | | 0.005*** | |
| | | (6.76) | |
| Firm-Year FE | No | No | Yes |
| N | 1,362,220 | 1,362,220 | 1,357,254 |
| Adj. $R^2$ | 0.00 | 0.01 | 0.15 |

**Table 3 – Source Similarity and Forecast Similarity, Continued**

*Panel C: Similarity in Forecast Boldness*

| Dependent Variable: *SimilarBoldness* | (1) | (2) | (3) |
|---|---|---|---|
| ***SourceSimilarity*** | **0.008***** | **0.008***** | **0.009***** |
| | **(23.53)** | **(23.57)** | **(26.56)** |
| *SimilarExperience* | | 0.002 | 0.004*** |
| | | (1.03) | (2.77) |
| *SimilarResources* | | -0.040*** | -0.030*** |
| | | (-28.55) | (-23.42) |
| *SimilarBusyness* | | -0.006*** | 0.005*** |
| | | (-3.65) | (3.68) |
| *BTM* | | -0.003*** | |
| | | (-4.62) | |
| *MVE* | | 0.002*** | |
| | | (3.53) | |
| *ROA* | | 0.004*** | |
| | | (5.63) | |
| Firm-Year FE | No | No | Yes |
| N | 1,362,220 | 1,362,220 | 1,357,254 |
| Adj. $R^2$ | 0.00 | 0.00 | 0.15 |

**Table 4 – Source Similarity and Forecast Similarity: Robustness**

This table provides our main results from estimating variations of Model (1) with augmented fixed effect designs and additional control variables. Panel C 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) |
| ***SourceSimilarity*** | **0.063***** | **0.015***** | **0.011***** |
| | **(17.41)** | **(19.40)** | **(14.76)** |
| *SimilarExperience* | 0.014** | 0.005*** | 0.002* |
| | (2.43) | (3.89) | (1.68) |
| *SimilarResources* | 0.000 | 0.002 | -0.002 |
| | (0.03) | (1.36) | (-1.08) |
| *SimilarBusyness* | 0.009 | 0.003** | 0.004*** |
| | (1.58) | (2.44) | (2.87) |
| Brokerage Pairwise FE | Yes | Yes | Yes |
| Firm-Year FE | Yes | Yes | Yes |
| N | 1,355,706 | 1,355,706 | 1,355,706 |
| Adj. $R^2$ | 0.52 | 0.17 | 0.16 |

**Table 4 – Source Similarity and Forecast Similarity: Robustness, Continued**

*Panel B: Analyst Pair-Brokerage Pair Fixed Effects*

| Dependent Variable: | *SimilarForecast* | *SimilarTiming* | *SimilarBoldness* |
|---|---|---|---|
| | (1) | (2) | (3) |
| ***SourceSimilarity*** | **0.054***** | **0.014***** | **0.008***** |
| | **(15.19)** | **(15.90)** | **(10.07)** |
| *SimilarExperience* | 0.004 | 0.004 | -0.005* |
| | (0.32) | (1.37) | (-1.86) |
| *SimilarResources* | 0.010 | 0.004* | -0.004* |
| | (1.12) | (1.84) | (-1.71) |
| *SimilarBusyness* | 0.004 | -0.001 | -0.001 |
| | (0.48) | (-0.39) | (-0.43) |
| Analyst Pair-Brokerage Pair FE | Yes | Yes | Yes |
| Firm-Year FE | Yes | Yes | Yes |
| N | 1,277,281 | 1,277,281 | 1,277,281 |
| Adj. $R^2$ | 0.57 | 0.25 | 0.20 |

**Table 4 – Source Similarity and Forecast Similarity: Robustness, Continued**

*Panel C: Analyst Employment Changes*

| Dependent Variable: | | *SimilarForecast* | *SimilarTiming* | *SimilarBoldness* |
|---|---|---|---|---|
| | | (1) | (2) | (3) |
| ***SourceSimilarity*** | | **0.043***** | **0.010***** | **0.008***** |
| | | **(6.48)** | **(5.51)** | **(4.65)** |
| ***OldSourceSimilarity*** | | **0.008** | **0.002** | **-0.000** |
| | | **(1.25)** | **(1.26)** | **(-0.09)** |
| *SimilarExperience* | | 0.005 | 0.004 | 0.001 |
| | | (0.41) | (1.46) | (0.22) |
| *SimilarResources* | | -0.056*** | -0.016*** | -0.015*** |
| | | (-4.53) | (-5.95) | (-5.49) |
| *SimilarBusyness* | | 0.018 | 0.007*** | 0.000 |
| | | (1.51) | (2.82) | (0.03) |
| Firm-Year FE | | Yes | Yes | Yes |
| N | | 317,987 | 317,987 | 317,987 |
| Adj. $R^2$ | | 0.52 | 0.18 | 0.19 |
| Within Regression F-Tests | | | | |
| *SourceSimilarity =* | Diff | 0.035*** | 0.008** | 0.008** |
| *OldSourceSimilarity* | f-stat | 7.18 | 4.62 | 5.79 |

**Table 5 – Source Similarity and Forecast Similarity: Public vs. Paid Subscription Sources**

This table provides our main results from estimating Model (2). 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) |
| ***PaidSourceSimilarity*** | | **0.050*****| **0.009***** | **0.008***** |
| | | **(26.10)** | **(23.21)** | **(20.43)** |
| ***PublicSourceSimilarity*** | | **0.005**** | **0.003***** | **0.002***** |
| | | **(2.04)** | **(5.65)** | **(3.50)** |
| *SimilarExperience* | | 0.027*** | 0.008*** | 0.004*** |
| | | (4.55) | (5.99) | (2.76) |
| *SimilarResources* | | -0.115*** | -0.030*** | -0.030*** |
| | | (-20.64) | (-24.09) | (-23.41) |
| *SimilarBusyness* | | 0.020*** | 0.007*** | 0.005*** |
| | | (3.48) | (5.13) | (3.66) |
| Firm-Year FE | | Yes | Yes | Yes |
| N | | 1,357,254 | 1,357,254 | 1,357,254 |
| Adj. $R^2$ | | 0.51 | 0.15 | 0.15 |
| Within Regression F-Tests | | | | |
| *PaidSourceSimilarity =* | Diff | 0.045*** | 0.006*** | 0.006*** |
| *PublicSourceSimilarity* | f-stat | 150.53 | 73.32 | 69.52 |

**Table 6 – Source Similarity and Forecast Similarity: Major vs. Minor Paid Sources**
This table provides our main results from estimating a modification to Model (2) for major and minor paid subscription sources. 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) |
| ***MajorSourceSimilarity*** | | **0.034*****| **0.006***** | **0.005***** |
| | | **(17.21)** | **(15.89)** | **(12.75)** |
| ***MinorSourceSimilarity*** | | **0.023***** | **0.005***** | **0.005***** |
| | | **(13.04)** | **(12.63)** | **(11.56)** |
| *SimilarExperience* | | 0.027*** | 0.008*** | 0.004*** |
| | | (4.59) | (6.00) | (2.77) |
| *SimilarResources* | | -0.115*** | -0.030*** | -0.030*** |
| | | (-20.65) | (-24.11) | (-23.49) |
| *SimilarBusyness* | | 0.020*** | 0.007*** | 0.005*** |
| | | (3.46) | (5.10) | (3.65) |
| Firm-Year FE | | Yes | Yes | Yes |
| N | | 1,357,254 | 1,357,254 | 1,357,254 |
| Adj. $R^2$ | | 0.51 | 0.15 | 0.15 |
| Within Regression F-Tests | | | | |
| *MajorSourceSimilarity* = | Diff | 0.011*** | 0.001** | 0.000 |
| *MinorSourceSimilarity* | f-stat | 11.12 | 4.31 | 0.85 |

**Table 7 – Source Similarity and Forecast Similarity: All-star Analysts**
This table provides our main results from estimating Model (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: | *SimilarForecast* | *SimilarTiming* | *SimilarBoldness* |
|---|---|---|---|
| | (1) | (2) | (3) |
| ***SourceSimilarity · AllStars*** | **-0.034*****| **-0.010***** | **-0.010***** |
| | **(-5.39)** | **(-7.10)** | **(-7.48)** |
| *SourceSimilarity* | 0.050*** | 0.010*** | 0.009*** |
| | (32.51) | (30.74) | (25.72) |
| *AllStars* | 0.452*** | 0.102*** | 0.122*** |
| | (9.11) | (8.99) | (10.99) |
| *SimilarExperience* | 0.026*** | 0.008*** | 0.003*** |
| | (4.40) | (5.89) | (2.60) |
| *SimilarResources* | -0.117*** | -0.030*** | -0.031*** |
| | (-21.00) | (-24.40) | (-23.85) |
| *SimilarBusyness* | 0.018*** | 0.006*** | 0.004*** |
| | (3.13) | (4.90) | (3.28) |
| Firm-Year FE | Yes | Yes | Yes |
| N | 1,357,254 | 1,357,254 | 1,357,254 |
| Adj. $R^2$ | 0.51 | 0.15 | 0.15 |

**Table 8 – Source Similarity and Forecast Similarity: High Source Access**

This table provides our main results from estimating a modification to Model (3) for brokerages with access to a greater number of data sources. 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) |
| ***SourceSimilarity · HighSourceAccess*** | **-0.020*** ** | **-0.006*** ** | **-0.006*** ** |
| | **(-3.73)** | **(-5.09)** | **(-5.33)** |
| *SourceSimilarity* | 0.055*** | 0.013*** | 0.011*** |
| | (25.27) | (27.12) | (21.37) |
| *HighSourceAccess* | 0.133*** | 0.026*** | 0.038*** |
| | (3.29) | (2.99) | (4.34) |
| *SimilarExperience* | 0.027*** | 0.008*** | 0.004*** |
| | (4.55) | (5.93) | (2.73) |
| *SimilarResources* | -0.113*** | -0.029*** | -0.029*** |
| | (-20.06) | (-23.15) | (-22.77) |
| *SimilarBusyness* | 0.020*** | 0.007*** | 0.005*** |
| | (3.56) | (5.17) | (3.71) |
| Firm-Year FE | Yes | Yes | Yes |
| N | 1,357,254 | 1,357,254 | 1,357,254 |
| Adj. $R^2$ | 0.51 | 0.15 | 0.15 |

**Table 9 – Source Similarity and Forecast Similarity: Longer Horizon Forecasts**

This table provides our main results from estimating Model (1) on an alternative sample that consists of the first forecasts that analysts issue for a given firm's annual reporting period. 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* |
|---|---|---|
| | (1) | (2) |
| ***SourceSimilarity*** | **0.024***** | **0.010***** |
| | **(16.27)** | **(24.18)** |
| *SimilarExperience* | 0.008 | 0.010*** |
| | (1.26) | (6.59) |
| *SimilarResources* | 0.015** | -0.013*** |
| | (2.23) | (-8.78) |
| *SimilarBusyness* | 0.018*** | 0.013*** |
| | (2.78) | (8.25) |
| Firm-Year FE | Yes | Yes |
| N | 990,069 | 990,069 |
| Adj. $R^2$ | 0.56 | 0.16 |

**Table 10 – Source Similarity and Consensus Forecast Accuracy**
This table provides our results from estimating Model (4), in which we investigate the relationship between data source independence and consensus forecast accuracy. Panels A and B present results using the mean and median consensus values, respectively. Variable definitions are provided in the appendix. t-statistics are reported in parentheses, and standard errors are clustered by firm and year. All p-values are two-tailed. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

*Panel A: Consensus Accuracy – Mean*

| Dependent Variable: *ConsensusAccuracy* | (1) | (2) |
|---|---|---|
| ***AvgSourceIndependence*** | **0.018**** | **0.015**** |
| | **(2.95)** | **(2.75)** |
| *AvgExperience* | 0.006 | 0.007 |
| | (0.61) | (0.72) |
| *AvgBrokerageSize* | -0.035** | -0.035** |
| | (-3.09) | (-3.08) |
| *AvgHorizon* | -0.060*** | -0.059*** |
| | (-8.62) | (-9.03) |
| *AnalystCount* | -0.000 | 0.000 |
| | (-0.02) | (0.04) |
| *BTM* | -0.172*** | -0.170*** |
| | (-14.65) | (-14.80) |
| *MVE* | 0.343*** | 0.342*** |
| | (11.03) | (11.50) |
| *ROA* | 0.060*** | 0.058*** |
| | (4.41) | (4.15) |
| *Loss* | -0.940*** | -0.941*** |
| | (-9.06) | (-9.19) |
| | | |
| Firm FE | Yes | Yes |
| Year FE | No | Yes |
| N | 27,848 | 27,848 |
| Adj. $R^2$ | 0.55 | 0.55 |

**Table 10 – Source Similarity and Consensus Forecast Accuracy, Continued**

*Panel B: Consensus Accuracy – Median*

| Dependent Variable: *ConsensusAccuracyMedian* | (1) | (2) |
|---|---|---|
| ***AvgSourceIndependence*** | **0.017**** | **0.014**** |
| | **(2.67)** | **(2.47)** |
| *AvgExperience* | 0.001 | 0.002 |
| | (0.15) | (0.28) |
| *AvgBrokerageSize* | -0.024* | -0.024* |
| | (-2.00) | (-2.02) |
| *AvgHorizon* | -0.029*** | -0.029*** |
| | (-4.59) | (-4.73) |
| *AnalystCount* | 0.052*** | 0.053*** |
| | (5.14) | (5.35) |
| *BTM* | -0.162*** | -0.160*** |
| | (-15.02) | (-14.78) |
| *MVE* | 0.333*** | 0.333*** |
| | (10.99) | (11.36) |
| *ROA* | 0.050*** | 0.049*** |
| | (4.51) | (4.27) |
| *Loss* | -0.883*** | -0.884*** |
| | (-8.69) | (-8.76) |
| Firm FE | Yes | Yes |
| Year FE | No | Yes |
| N | 27,848 | 27,848 |
| Adj. $R^2$ | 0.53 | 0.53 |