The Value of Differing Points of View: Evidence from Financial Analysts' Geographic Diversity

William C. Gerken*

Department of Finance & Quantitative Methods, University of Kentucky, Lexington, KY 40506

Marcus O. Painter

Department of Finance, Saint Louis University, Saint Louis, MO 63103

Abstract

We show that analysts incorporate geographically dispersed information about firms into individual forecasts and that limited analyst geographic diversity adversely affects consensus forecasts and firm liquidity. Using satellite imagery of U.S. retailers' parking lots, we find analysts shade their own forecast in the direction of local car counts relative to other analysts covering the same firm at the same time but from different locations. Examining all industries, we find firms with more geographically concentrated analyst coverage have higher consensus forecast errors and are less liquid. Evidence from shocks in geographic coverage due to brokerage closures suggest these relations are causal.

Keywords: Analysts, Fintech, Geography, Diversity, Forecasts, Liquidity
JEL: G14, G24, O3, M41

☆We thank Tony Cookson, Russell Jame, Maximilian Rohrer, and seminar and conference participants at the Inaugural New Technologies in Finance Conference (Columbia University), the 2nd Future of Financial Information Conference (Stockholm University), the 2nd QES Virtual NLP and Machine Learning in Investment Management Conference, University of Kentucky, University of Missouri, and Miami University. This paper was previously circulated under the title "Are Crowded Crowds Still Wise? Evidence from Financial Analysts' Geographic Diversity."

*Corresponding Author. Tel: +1 (859) 257-2774 Fax: +1 (859) 257-9688

Email addresses: will.gerken@uky.edu (William C. Gerken*), marc.painter@slu.edu (Marcus O. Painter)

Thursday 19th November, 2020

I. Introduction

Well functioning markets require participants to deliver a diverse set of information about assets in that market (Goldstein and Yang, 2015). While some types of information in financial markets are disseminated widely (e.g., annual reports and conference calls), other types of information may be geographically dispersed and not equally accessible to all market participants (Van Nieuwerburgh and Veldkamp, 2009; Bernile, Kumar, and Sulaeman, 2015). However, for dispersed agents observing local information in real time, it is difficult to ascertain how much of their local realization is signal versus noise.

Take, for example, the setting of retail sell-side analysts, where the problem is particularly acute. Analysts, and especially analysts covering retail firms, often highlight store visits in their primary research.[1] A research analyst forming expectations of a firm's earnings may find local firm conditions particularly salient. This practice is also encouraged by asset managers. For example, fund manager Peter Lynch once wrote "visiting stores and testing products is one of the critical elements of the analyst's job" (Lynch and Rothchild, 2000). The busyness of retail stores may be an important signal, but it is difficult for any single analyst to determine how representative any observation of a particular store may be of the firm's overall trend. Generalization from small amounts of personally observed data can often lead individuals to fall victim to the law of small numbers (Tversky and Kahneman, 1971). They may believe their own observation is representative and underestimate the role of chance even in the face of other pieces of data when making their forecast.

Typically, the collective opinion of the crowds of forecasters is often more accurate than individual experts. Yet, for crowds to be wise, they must have diversity of opinion and independence for the individuals' errors to be averaged away. If the crowd of forecasters rely on the same source information (e.g., local retail store busyness for a particular geographic area) and lack a diversity of viewpoints, idiosyncratic noise will not cancel out.

[1] For an example of such behavior, see "Back to school shopping observations from the retail industry." Business Insider, 2019.

The equity research industry in the U.S. is quite geographically concentrated. Over half of research analysts work on or near Wall Street, while the remainder are dispersed among other metropolitan areas. In this paper, we document that retail firm analysts incorporate local firm information (i.e. firm performance for stores near the analyst) into their earnings estimates. We then turn to the implications for the crowd (the consensus) forecast and firm liquidity. If analysts systematically incorporate noise from local information into their forecasts, then we would expect to see the crowd forecast to average away these errors when the consensus is made up of a geographically diverse population – and correlated errors to remain when they are not. This is precisely what we find. Firms with more geographically concentrated analysts have higher consensus forecast errors. These firms also have higher bid-ask spreads around earnings announcements, suggesting that a lack of analyst geographic diversity worsens the information environment for investors.

A major challenge to testing whether analysts generalize from local information is that firm-specific local information is typically hard to observe empirically. To overcome this challenge, we take advantage of satellite image data that provides us with parking lot car counts of retail firms across different metropolitan areas over time. We combine these data with information on the location of sell-side analysts from a series of Freedom of Information Act (FOIA) requests for work histories of brokerage employees to provide us with a measure of firm-specific, time-varying local parking lot information that analysts may be exposed to.

The setting of analysts' quarterly forecasts makes a near ideal setting to study the incorporation of local information for several reasons. First, we confirm the aggregated satellite data contains valuable information about overall firm performance, but as expected, individual local car counts do not contain incremental information beyond the aggregate number. While not experimental randomization, each analyst is presented with a signal that contains valuable information, but also idiosyncratic variation that is difficult to disentangle in real time. This implies that local car counts are a noisy signal, which we can exploit to examine whether analyst follow these signals and if so what are the consequences for the consensus. Second, our setting is useful for identification as we observe many agents making estimates of the same event at the same time but with different exposures to local information. Third, because analysts typically cover multiple firms in the same area, we can separate firm-specific information from common general information about the local and macroeconomy. Finally, analysts are important financial intermediaries and market participants rely on analysts' forecasts as a significant source of cash flow information (Kothari, So, and Verdi, 2016). Therefore, documenting that the geographic diversity of analysts affects the accuracy of consensus forecasts has important implications for the efficiency of asset prices.

As an example of our setting, consider financial analysts working in New York, Minneapolis, and Cleveland who each make a quarterly earnings forecast for Home Depot in Q1 of 2013. While all analysts could access the same SEC filings and listen to the same conference calls, an analyst in Minneapolis may observe full parking lots at local Home Depot locations and infer that the firm is doing well overall, while the analyst in Cleveland may observe mostly empty parking lots and make the opposite inference. Because analysts are located in different locations and the satellite data allows us to observe local firm conditions at these locations at the same point in time, we can control for common information available to all analysts during the quarter. Moreover, because analysts typically cover multiple firms and the satellite data is available for each firm in the locale, we can remove any common "Minneapolis" effect for the analyst even if the local effect varies over time.

We find evidence that analysts shade their quarterly earnings forecast toward their locally observed parking lot car counts. Using a series of fixed effects specifications, we are able to rule out many competing stories that arise due to common information such as company filings, matching between analysts characteristics and the firms that they cover, differences in MSA characteristics, and time-varying analyst specific characteristics like mood or their general predictions about the macroeconomy. In terms of statistical magnitude, our baseline specification suggests that a one standard deviation increase in a firm's local car counts relative to the national average is associated with a 47 bps increase in forecast error, a 6% deviation from the mean. We document that these effects are larger for less experienced analysts and when other analysts are concentrated in the same metropolitan area. We also show that the effect is attenuated when satellite data becomes available for purchase, suggesting that analysts rely on noisy local information primarily when there is no better aggregate alternative available.

The influence of local firm performance on geographically dispersed individual analysts' forecasts has important implications for the formation of consensus forecasts. When analysts are clustered within a single locale, their individual forecasts tend to be shaded in the same direction and result in a relatively higher forecast error. In contrast, when analysts are geographically diverse, the influence of these varied exposures offset, reducing consensus forecast error. Using a broader sample of firms in the IBES database, we find evidence that consensus forecast errors are larger when analysts are more concentrated in fewer metropolitan areas (i.e., less geographically diverse).

We employ within-firm panel regressions to show that increased geographic diversity is associated with lower consensus forecast errors. While there are not obvious reasons to believe that endogeneity would be an important concern in this setting, as brokerage firms are unlikely to shift their entire research staffs' operations in response to their relative geographic dispersion for a single stock, we nonetheless exploit the exogenous shift in geographic coverage due to brokerage closures and mergers. We expand on the approach in prior work such as Hong and Kacperczyk (2010) and Kelly and Ljungqvist (2012) by looking within firms that lose coverage due to the exogenous reduction in analyst coverage. In these tests, we find larger increases in consensus forecasts errors in firms that see reduced geographic coverage (e.g., a firm loses its only analyst who is located in Minneapolis) relative to firms that also lose analyst coverage but retain the same geographic diversity (e.g., a firm with three analysts in the New York city metropolitan area loses one of these three New York analysts). These findings suggest the increased forecast errors that result from brokerage closures stem from not just the reduction in competition between analysts, but also the loss of geographic diversity.

Because analysts are vital for the production of market-related information, their lack of geographic diversity may have spillover effects in financial markets. In particular, the information environment may be worse for firms with less geographic coverage, which could lead to a decrease in stock liquidity (Harford et al., 2019). Consistent with this conjecture, we find bid-ask spreads around earnings announcements are larger for firms with less geographically diverse analysts. This finding is robust to the inclusion of $Firm$ fixed effects and $Number\ of\ Analyst$ fixed effects. Further, using the exogenous reduction in analyst coverage from brokerage closures and mergers setting, we find higher spreads for firms that experience a loss in geographic coverage, suggesting the results are causal.[2]

A firm's information environment is influenced by a variety of sources, including public news (Boudoukh et al., 2019), modern information technologies (Gao and Huang, 2020), analysts' access to management (Green et al., 2014), and analysts' effort allocation (Harford et al., 2019), among others. Our study specifically contributes to a growing literature on the importance of geographically dispersed information about firms. Garcia and Norli (2012) extract state name counts from 10-K filings and show important differences in information environment among firms with business operations concentrated in a few states from firms with operations in multiple states. Addoum, Kumar, and Law (2017) follow a similar approach and provide evidence that the market is slow in aggregating state-level information in diverse firms. Chen (2017) shows that managers overweight observations of local economic conditions at firm headquarters (HQ) when forming their macroeconomic expectations. One key advantage of our setting is that we observe the same economic agent (a financial analyst) who is provided with various signals of local information (car counts of various firms) and provides multiple outputs which other agents who are receiving different signals are also producing. This allows us to use high-dimensional fixed effects to control for unobservable

[2] Kelly and Ljungqvist (2012) use a similar setting to show that bid-ask spreads increase amidst an increase in information asymmetry. Thus, our findings suggest one channel - changes in the geographic diversity of analysts - that contribute to the increase in information asymmetry.

factors that confound analysis at a more aggregated level.

We also contribute to the literature on how analysts form forecast predictions. Brown, Call, Clement, and Sharp (2015) provide survey evidence that attempts to understand the context within which analysts make their decisions. Malloy (2005) finds that an analyst's proximity to the firm's headquarters is related to forecasting ability. Renjiey (2018) provides evidence that analysts' beliefs are influenced by the performance of other industries that they cover. Hirshleifer, Levi, Lourie, and Teoh (2019) show that analysts that cover multiple firms and make several forecasts in a single day suffer from decision fatigue. Gibbons, Iliev, and Kalodimos (2019) exploit EDGAR logs to show differentials based on its usage by analysts. Merkley, Michaely, and Pacelli (2019) show that ethnic diversity among analysts impacts the format of consensus. Chen, Mayew, and Yan (2018) examine how social interaction between analysts influences forecast quality and find that the analysts who work in the same office transmit important information to each other when these analysts cover different firms that are headquartered in the same location. Cen, Chang, and Dasgupta (2020) also study the analyst location setting but focus on trying to infer whether analysts appear to learn from other geographically dispersed forecasts. While similar in the general spirit of this research, we have a unique setting where we observe time-varying local information signals that are specific to each analyst-firm pairing (across all pairs that share the same MSA). The union of satellite data and analyst location data enables us to use fine-grained fixed effects to isolate the effects of local information in a way that is typically not possible.

Our findings also provide a more nuanced take on the wisdom of crowds.[3] While prior research has shown that group forecasts are less accurate when individuals in the group rely too heavily on public information (Da and Huang, 2019), we document a cost of relying too heavily on correlated private information and a lack of diversity among forecasters. Further, by identifying local firm performance our findings are among the first to point out an actual

[3]The majority of this literature focuses on comparing the forecast accuracy of groups versus individuals (e.g., Charness and Sutter (2012) and Jame, Johnston, Markov, and Wolfe (2016).

source of private information used in forecasting.

Finally, our paper contributes to the burgeoning literature that uses big and alternative data to study questions in finance.[4] Froot, Kang, Ozik, and Sadka (2017) use proprietary real-time measures of consumer activity and show that managers downplay current performance to take advantage of insider trading opportunities. Zhu (2019) studies the release of credit card transaction and satellite imagery data to investors and finds the dissemination of this data disciplines firms' long-term investment decisions. Katona, Painter, Patatoukas, and Zeng (2019) also study the dissemination of satellite imagery but focus on the playing field between investors and show that sophisticated investors become better off while individual investors are hindered. Kang, Stice-Lawrence, and Wong (2019) show that local institutional investors appear to trade in the same direction as local parking lot counts. We contribute to this literature by demonstrating how alternative data can be used to create granular local information measures that allow for the study of open questions in finance and economics.

II. Data

A. Analyst Locations

Our data on financial analyst locations comes from historical filings of the Uniform Application for Securities Industry Registration or Transfer (Form U4), which provides detailed accounts of analysts registrations and work histories including the street address of office location. The Form U4 is filed by an employer when the analyst joins the firm and must be updated upon material changes such as changing jobs. The Form U4 data are aggregated in a database called the Central Registration Depository (CRD), which is jointly operated by FINRA and state securities regulators.

We obtain Form U4 data from a series of Freedom of Information Act requests to state regulators. Our universe of financial analysts consists of those registered in any of the states that respond to our requests during some point of their career. Analysts may register in

[4]See Goldstein et al. (2019) for an early perspective on this literature.

multiple states, so we have data for many analysts in the states that do not supply information. We do not observe location data only for those analysts who never register in a reporting state. Although our sample of office locations is not comprehensive, it covers all major financial centers. As the selection mechanism for inclusion into the sample is a state regulator's interpretation of The Privacy Act of 1974 as it relates to the FOIA request, it is unlikely selection would systematically bias the correlation between our variables of interest which are time-varying within MSAs.

To match the analyst locations to earnings recommendations, we link the the location data to IBES using a combination of name and career history matching. First, we match firm identifiers between the CRD database and IBES database. We then match individuals using a combination of first initial, last name, and career history. We drop observations with ambiguous information in the name fields such as missing or multiple names. Because we have career history in both databases, we can take advantages of the timing of job switches to make matches when name strings are ambiguous (e.g., while there are many "L. Smith" entries in both databases, only one in each database moves from Citigroup to Sutro & Co. in 1998 allowing us to make a distinct match). Further we can verify the quality of the matches, by checking their career start and end dates and FINRA licensing required of analysts. Research analysts must pass either the Series 86 or Series 87 exam to register as a research analyst, which allows the analyst to prepare written or electronic communications that analyze equity securities, companies, and industry sectors.

Consistent with prior studies (e.g., Malloy, 2005), we find that the slight majority (57.5%) of analysts are located in the New York metropolitan area. Analysts typically remain in the same MSA throughout our sample period with only 5% switching MSAs. In panel A of Figure 1, we display the geographic dispersion of analysts in the retail sample, along with the store locations of the retail firms. The size of the circle is proportional to the number of unique analysts in the sample. Outside of New York, there are several geographically dispersed metropolitan areas that each contain a sizable count of analysts such as Chicago, Minneapolis, and San Francisco. Further, there are numerous metropolitans like Cleveland, Atlanta, Nashville, Boston, Dallas, and Washington that have a meaningful headcount of analysts. We also show the location of analysts and stores for an example firm, Home Depot, in panel B of Figure 1. Home Depot's analyst coverage highlights the importance of geographic diversification if analysts look to incorporate local information. Though Home Depot's analysts are relatively dispersed, there are still many stores with no local analyst coverage.

In Table 1, we report that the vast majority (69.3%) of firms have analysts covering the firm from multiple MSAs. However, the reverse is not as true, most brokerage firms concentrate their analysts in a single location and only a handful have large staffs at multiple MSAs. This agglomeration pattern is consistent with the financial research industry having significant localized industry spillovers that dominate the potential information advantages of having dispersed analysts (Ellison and Glaeser, 1999).

B. Parking Lot Data

To measure local firm performance, we use satellite imagery of daily parking lot car counts for major U.S. retail firms. We obtain data on parking lot car counts from Orbital Insight, a leading image processing company that uses machine learning to convert satellite images into quantitative data.[5] The data include the company name and ticker, a unique id for each store location, the latitude and longitude of the store, the date the image was taken, and a count of the number of cars in the parking lot when the image was taken. The measure from Orbital Insight normalizes the car count data to account for the day of the week and time of day the satellite image is taken. Our sample for daily car counts begins in January 2009 and ends in December 2015.

The daily car count data include 5.4 million parking lot observations covering 176,935 unique store locations for the 162 retailers covered by Orbital Insight. From the daily data,

[5]See Katona, Painter, Patatoukas, and Zeng (2019) for a detailed description and background of satellite data.

we create a measure of local firm performance by taking the average car count of a firm's parking lots located in the same MSA during a given quarter. We then create the variable *Car Count Difference*, measured as

$$Car\ Count\ Difference_{i,m,q} = \frac{Local\ car\ count_{i,m,q} - National\ car\ count_{i,q}}{National\ car\ count_{i,q}} \tag{1}$$

where $Local\ car\ count_{i,m,q}$ is the average daily car count for firm i in MSA m during quarter q and $National\ car\ count_{i,q}$ is the average daily car count for firm i during quarter q across all observed store locations. *Car Count Difference* therefore measures how well a firm performed in a given MSA relative to how the firm performed overall during a given quarter. Previous studies (e.g., Katona, Painter, Patatoukas, and Zeng (2019) and Zhu (2019)) find that aggregate satellite data measures like $National\ car\ count_{i,q}$ strongly predict firms' future earnings per share. In contrast, $Local\ car\ count_{i,m,q}$ should have no predictive power beyond the aggregate measure. By looking at the abnormal number of cars in a local MSA relative to the national average, we are able to isolate the noisy local information from the relevant aggregate information.

We then match the satellite data to the analyst location data at the firm, analyst-MSA, and quarter level. The unison of these two data sets allows us to gauge the relative performance of a firm that an analyst is exposed to in a given quarter, which will vary both over time and for analysts in different MSAs.

We also merge the data with CRSP and Compustat to obtain stock return and firm accounting data. Fourteen firms are dropped due to not having analyst coverage or not having stock return and accounting information in CRSP or Compustat. The final sample contains 18,679 quarterly observations for 402 analysts covering 138 firms.

To test whether analysts are influenced by the local performance of a firm, we analyze the relationship between a firm's *Car Count Difference* in an analyst's MSA and that analyst's forecast error. We measure forecast error as

$$Error_{i,j,q} = \frac{Forecast_{i,j,q} - Actual_{j,q}}{Price_{j,q-4}} * 100 \tag{2}$$

Where $Forecast_{i,j,q}$ is analyst i's most recent forecast for firm j's earnings per share in quarter q, $Actual_{j,q}$ is firm j's actual earnings per share in quarter q, and $Price_{j,q-4}$ is firm j's stock price at the end of quarter $q-4$.

We examine summary statistics for the merged sample in the first panel of Table 1. Analyst forecasts tend to be pessimistic, as the average error is -.08. The average *Car Count Difference* is 17%, suggesting stores of firms located near analysts tend to outperform the firm's national average. There is large variation in *Car Count Difference*, with a standard deviation of 47%. There continues to be variation in *Car Count Difference* within a firm-quarter, which we exploit in the design of our tests. This variation can arise due to several factors, including changes in local competition, weather, local advertising campaigns, and variation in regional management quality of a firm. For this study, we are focused less on how this variation emerges and more on how the variation influences financial analysts' information set.

The stores of firms in our sample are widespread, as the average firm has store locations in 58% of MSAs. In a given quarter, forecasts are issued by 6.4 analysts for the average firm and those analysts are located in an average of 3.77 MSAs. 14% of forecasts in our sample are made by analysts who do not live near a store of the firm they are forecasting.[6] The average analyst has approximately ten years of overall experience, 4.25 years of firm-specific experience, and issues forecasts two months before the earnings announcement. Firms in our sample have an average share turnover over the prior year of 0.27, market equity of $9.9 billion, and a book to market ratio of 0.42.

In panel B of Table 1, we examine the summary statistics for the broader sample of all IBES

[6] In these cases, we set *Car Count Difference* equal to zero and include an indicator variable for a "no local store" analyst.

analysts. To test whether analysts' generalization of local information has implications for the accuracy of analyst's consensus forecasts, we analyze the relation between the geographic concentration of analysts covering a specific firm and the consensus forecast error for that firm. We construct the consensus analyst forecast error as the price-scaled absolute error between the consensus forecast and the actual. We calculate the MSA Herfindahl Index using the count of analysts in each MSA in each firm-quarter to a market share by MSA which is then squared and summed across all MSAs in the firm-quarter. We multiply this index by negative one so that it may be interpreted as a dispersion index. The average consensus forecast error is 0.28 and the average MSA dispersion index is -0.6.

When comparing the analyst and firm characteristics between the satellite data sample and the full IBES sample, we find that firms in the satellite data sample generally have more experienced analysts, higher turnover, higher market value, and lower book to market values. The difference in characteristics between the two samples should not affect the validity of our findings for three reasons. First, our tests examining the satellite data sample are done at the $Firm \times Year \times Quarter$ level, so all firm-level characteristics are subsumed. Second, we also use $Analyst \times Firm$ and $Analyst \times Year \times Quarter$ fixed effects to show that the same analyst's forecast for the same firm will be influenced based on whether that firm's local stores perform better or worse than the firm's national performance in a given quarter. Third, we find consistent results in both the satellite and full IBES samples, suggesting differences in firm and analyst characteristics do not bias our results.

We delve further into the satellite data in panels C and D of Table 1. Panel C reports the top five retail firms in the satellite sample by market equity, the percentage of MSAs where these firms have store locations, the number of unique MSAs that analysts covering the firm are located over the sample period, the number of analysts covering the firm over the sample period, and the average number of cars in each firm's parking lots. Three main inferences can be drawn from this panel. First, the high percentage of MSAs with store locations for these firms shows that they are geographically diverse and suggests that it would be difficult to forecast the performance of these stores based on the local performance in one MSA. Second, despite the geographic diversity of firms, analysts tend to cluster in a few MSAs. For example, though there are Walmart stores in 94% of the MSAs in the US, the 31 analysts who cover Walmart are spread across 1.3% of those MSAs. Third, the geographic concentration of analysts varies even within the largest firms in our sample as seen by contrasting the five unique locations of Walmart analysts with the eight unique locations of Home Depot analysts.

In panel D, we examine the relationship between analyst geographic dispersion and forecast accuracy in a univariate setting. Contrary to the perception that all analysts are on Wall Street, this table shows that less than a third (30.7%) of firms have analysts that are all located in the same MSA. The majority of firms (69.3%) have analysts in two or more MSAs and 17.5% of firms have analysts in five or more MSAs.

Additionally, panel D provides univariate evidence consistent with the idea that forecast accuracy is positively related to the geographic dispersion of analysts. The average forecast error for firms with all analysts located in the same MSA is -0.088. This error monotonically trends toward zero (i.e., becomes more accurate) as analysts spread across more MSAs, with an average forecast error for firms with analysts in five or more MSAs of -0.063.

III. Predicting returns with retail store car counts

Our first tests establish that car count information (at least in aggregate) contains value relevant information. To test this idea, we test its predictive power, examining the return predictability of car counts. Local car counts for a particular firm are highly correlated but much noisier than the firm's aggregated national car count. When a firm has high car counts nation-wide, it obviously in aggregate has higher counts across the various local stores. However, idiosyncratic local conditions such as weather, local management, local competition, or local economic conditions, can affect individual localities at any point in time. These idiosyncratic local components average out across the national measures. To illustrate this idea, we plot the relationship between local car counts and national car counts for an example firm, Home Depot, in Figure 2. Panel A demonstrates that local car counts are quite varied, suggesting it would be difficult to accurately predict national performance using only one store's car count. However, panel B shows that on average, when local car counts are higher, aggregate national car counts are higher.

During the majority of our sample period, access to real-time car count data directly was limited. Instead, analysts likely observe local realizations of store car counts along with national level information which is available to all analysts. We examine return predictability using the following model:

$$Announcement\ Return_{j,q} = \beta_1 * National\ Car\ Count_{j,q,m} + \beta_2 * Firm\ controls_{j,q} + \gamma' * FE + \epsilon_{j,q,m} \tag{3}$$

Where $Announcement\ Return_{j,q}$ is firm $j's$ stock return from the day of an earnings announcement until three days after.[7] $National\ Car\ Count_{j,q,m}$ is firm $j's$ national car count during quarter q in MSA m. We also examine the difference between local and national car counts. We include firm controls for turnover (log of the number of shares traded divided by the number of shares outstanding over the prior 12 months), size (the log of the market capitalization of the firm), book to market (the log of book value of equity minus the log market value of equity), and past return (the return of firm j from 30 days prior to announcement until 3 days prior to the announcement). The unit of observation is an MSA-firm-quarter. We also include both $Firm$ and $Quarter$ fixed effects and cluster standard errors at the MSA level.

Table 2 shows the results for our return prediction tests. First, column 1 shows a significantly strong positive relationship between national car counts and announcements returns. In column 2, we include the difference between the local measure and the national measure. In this model, we find that the national car count measure is still significant while the local differences are not. The key implication from Figure 2 and Table 2 is that local car

[7]Results are robust to measuring announcement returns using various windows.

counts contain relevant information, but primarily because they are correlated with national car counts. The local differences are not in themselves value relevant.

Our findings are consistent with those found in other papers documenting the investment value of satellite data (e.g., Katona, Painter, Patatoukas, and Zeng (2019); Zhu (2019)). Together, these results show that car counts contain value relevant information, but analysts who make predictions that generalize from local car counts will be exposed to substantial noise.

IV. Evidence of generalization of local information in analyst forecasts

A. Main Results

In this section, we conduct a series of tests to examine how the local performance of a firm affects individual analysts' forecasts. We employ a variety of fixed effect specifications which allows us to rule out any plausible alternative explanations. Our initial regression tests examine the satellite data sample using the following specification:

$$\begin{aligned} Error_{i,j,q} = \beta_1 * Car\ Count\ Difference_{i,j,q} + \beta_2 * Analyst\ controls_{i,j,q} \\ + \beta_3 * Firm\ controls_{i,q} + \gamma' * FE + \epsilon_{i,j,q} \end{aligned} \tag{4}$$

where *Error* is the price-scaled forecast error of analyst i for firm j in quarter q and *Car Count Difference* is the relative local performance of firm j in the MSA of analyst i during quarter q. We follow the analyst forecasting literature and include controls for the log of one plus overall experience of an analyst (measured in quarters), the log of one plus the analyst's firm-specific experience, and the log of the age of the analyst forecast. Additionally, we include firm level controls for the log of the number of shares traded divided by the number of shares outstanding over the prior 12 months (Turnover), the log of the market capitalization of the firm (Size), the log of book value of equity minus the log market value of equity (Book to market), and the log of the number of analysts covering a firm (Num Analysts). We also control for whether an analyst is located near the firm's headquarters and whether the analyst is in an area with no local stores of the covered firm. We cluster standard errors at the MSA level.

We present the results of the effect of local performance on individual analyst forecast error in Table 3. In column 1, we include $Firm$ and $Year \times Quarter$ fixed effects to control for the possibility that some firms are more difficult to forecast than others as well as any potential time trends. We find that local firm performance is significantly associated with analyst forecast error. Specifically, a one standard deviation increase in car count difference (0.47) is associated with a 31 bps increase in forecast error, representing a 3.8% deviation from the mean.

In column 2, we introduce $Firm \times Year \times Quarter$ fixed effects to control for any information about a firm that all analysts can access (e.g., company filings, earnings guidance calls with management, etc.) in a given quarter. All variation in firm specific characteristics (turnover, size, book to market, and number of analysts) is subsumed under this model. The remaining specifications in Table 3 are thus identified by variation in the local signal that analysts receive, as proxied by *Car Count Difference*. The introduction of $Firm \times Year \times Quarter$ fixed effects strengthens the relation between forecast error and *Car Count Difference*, as the coefficient on *Car Count Difference* in column 2 increases to 0.009. In column 3, we introduce MSA fixed effects to examine variation in car count difference in a particular MSA for a firm. This model will control for any time-invariant measurement error that may arise from the satellite imagery in a certain location (e.g., incomplete coverage due to underground parking garages). The coefficient remains significant and similar in magnitude under this specification. In terms of economic magnitude, a one standard deviation increase in *Car Count Difference* is associated with a 47 bps increase in forecast error, representing a 6% deviation from the mean. For reference, in Hirshleifer et al. (2020) analyst forecasts are 3% higher than the mean after experiencing a better first impression of a firm (see section 4.1 of Hirshleifer et al. (2020)) and in Hong and Kacperczyk (2010) the loss of an analyst due to a broker merger or closure increases bias by about 5.6%

of the mean long-term bias (see Table 5 of Hong and Kacperczyk (2010)).

B. Heterogeneity

We next examine the cross-section of analysts to investigate which characteristics help to reduce the generalization of local information. Specifically, we test the prediction from Tversky and Kahneman (1971) that agents can learn to recognize and overcome the law of small numbers bias. Motivated by Clement (1999), who finds analysts with more experience generally produce more accurate forecasts, we hypothesize that analysts will be less influenced by local information as they gain more experience. We also examine whether analysts are more or less prone to generalize local information based on the number of analysts covering the firm and the number of unique MSAs in which analysts covering the firm are located. A significant decrease in local information weighting based on increased analyst following would suggest that analysts are able to reduce the influence of local information by learning from other analysts. A significant decrease in local information weighting based on the number of unique MSAs would suggest that analysts reduce the influence of local information by learning from analysts *who are located in other MSAs* and therefore not exposed to the same local information. Finally, we test whether the availability of satellite data - which gives analysts the ability to see all parking lots - reduces analysts' weighting on local information.

We show results for the heterogeneity of local information weighting in Table 4. In column 1, we interact *Car Count Difference* with analyst experience. In column 2, we interact *Car Count Difference* with number of analysts. In column 3, we interact *Car Count Difference* with the number of MSAs. Lastly, in column 4 we interact *Car Count Difference* with an indicator for whether a satellite data is available for a firm in that quarter. For ease of interpretation, we z-score each continuous interaction term. We find evidence that the effect is stronger for less experienced analysts. This suggests that part of the reason that more experienced analysts are more accurate is because analysts become better able to recognize signals vs. noise in local information over time. The coefficient on the interaction term suggests that a one standard deviation increase in experience is associated with a 28%

(0.0028/0.01) decrease in local information generalization relative to an analyst with average experience. We find no evidence that the raw number of analysts modulates the magnitude of the effect (column 2). However, we find the effect is mitigated when there are more analysts in geographically diverse locations (column 3). This suggests that analysts only reduce the weighting of local information when there are other analysts covering the firm *that are not exposed to the same local information*. One possibility is that when analysts are in the same MSA, the analyst can observe the other analysts' forecasts that include the same local information. When their own local signal differs from analysts in other MSAs, their own local signal may conflict with the implied signal in the other forecasts, causing the analyst to put less weight on the signal. We explore the implications of this on the consensus forecast in Section V.

One potential reason analysts tend to generalize from local parking lots is that local lots contain a noisy signal and the more informative signal from nationwide parking lots has historically been unavailable. If this were the case, we would likely see analysts rely less on local parking lots when access to satellite data becomes available. The majority of our sample (78% of observations) covers the period before satellite data was disseminated but the data for some companies did become available starting in mid-2014. Therefore we are able to test whether the availability of satellite data affects local information weighting by interacting *Car Count Difference* with an indicator variable identifying whether a firm's satellite data is available from a satellite vendor.[8] The staggered introduction of the satellite data allows us to use firms that either had not had their data disseminated yet or never have their data released during our sample as a control group. Consistent with our conjecture, we find that analysts substantially reduce the generalization of information from local parking lots for firms that have satellite data available. This finding is consistent with Katona et al. (2019), who find that analyst forecast revisions correlate with parking lot traffic growth.

[8] We collect the release dates of firm's satellite data for both Orbital Insight and RS Metrics, the two leading satellite data vendors in the U.S. The Orbital Insight dates are sourced from Orbital Insight and the RS Metrics release dates are from Katona et al. (2019).

C. Robustness

Our next set of tests are robustness checks to ensure the validity of our results. O'Brien and Tan (2015) find that analysts are more likely to cover local firms than non-local ones and may provide more accurate forecasts (Bae, Stulz, and Tan, 2008). To control for potential selection between analysts and the firms they cover, we include $Analyst \times Firm$ fixed effects in Table 5, column 1. These effects also control for time-invariant analyst characteristics such as analyst proximity to a firm's headquarters, an analyst's personal connections (e.g., school ties) with firm management, analyst skill, and whether an analyst has industry-relevant experience.[9] We continue to find a stronger effect under this model, with a coefficient on *Car Count Difference* of 0.01.

Jennings, Lee, and Matsumoto (2017) find that managers frequently refer to other firms in the same geographic location in earnings conference calls, suggesting the relevance of local information for analysts. In the second column of Table 5, we add $Analyst \times Year \times Quarter$ fixed effects to examine how local firm performance affects forecast error for the same analyst in the same quarter. This specification rules out the possibility that analysts may be biased by any other local effects, including the performance of the local economy, weather conditions (Dehaan et al., 2017), local sports teams, etc. The inclusion of $Analyst \times Year \times Quarter$ fixed effects allows us to isolate local firm performance by exploiting variation in *Car Count Difference* for the different firms that an analyst covers. The coefficient for *Car Count Difference* is largest under this specification, with a coefficient of 0.0136 and significant at the 1% level. The overall increase in the size of the coefficient as our identification becomes more conservative is consistent with the idea that the fixed effects are controlling for unobserved variables that would bias the results towards zero. For example, including $Analyst \times Firm$ fixed effects controls for the possibility that some analysts are

[9]Malloy (2005) shows that analysts near firm headquarters are more accurate, Bradley et al. (2020) find that analysts with professional connections to management issue more informative recommendations, Crane and Crotty (2020) document that there is persistence in analyst skill, and Bradley et al. (2015) show that prior industry experience can help improve forecast accuracy.

located near a firm's headquarters, which could potentially allow those analysts to overcome the generalization of local information by speaking with management and acquiring a more complete information set about the firm's current performance. By controlling for proximity to headquarters and other unobserved variables, we are able to isolate the effect of local information and find a stronger effect.

Another concern is that given the industry concentration in New York, the results may be driven by some difference between New York and other metropolitan areas. To address this concern, in column 3 we estimate the model excluding New York-based analysts. We find consistent results, with a significant coefficient on *Car Count Difference* of 0.008. Another concern is that our data on parking lot car counts is not frequent enough to be representative of how a firm is actually performing in an MSA. Because the satellites do not cover every MSA with the same frequency, there exists a possibility that the low number of parking lot observations in some MSAs during a quarter could create a spurious correlation. We address this potential concern in column 4 by excluding MSA-firm-quarters with fewer than five satellite observations and again find consistent results. Our last robustness check examines whether there is any asymmetry in the influence of local performance. Prior research on financial analysts has shown that negative signals can be more influential than positive signals (Hirshleifer et al., 2020). To examine whether this is the case regarding local information, we include in our regression the interaction of *Car Count Difference* with indicators for whether the local car count is above or below the median car count for that firm in our sample. In column 5 of Table 5 we find that negative and positive local signals influence analysts equally, as the coefficients on the interactions are both statistically significant and similar in magnitude (0.0103 for high car counts and 0.0099 for low car counts), suggesting there is no asymmetry in influence of positive versus negative local performance.

V. Firm-level forecast distribution and consensus error

In this section, we examine whether local information affects the range of analyst forecasts, the consensus forecast error, and firm liquidity. In a frictionless market, we would expect brokerage firms to address local information generalization by locating a new analyst in an area that does not have an existing analyst covering the same firm. In this frictionless market, noise from local information would cancel out and we would see no effect on consensus forecast errors. However, several frictions prevent brokerage firms from adequately dispersing analysts. These frictions include the need to keep employees local in order to monitor them, the concentration of analyst talent, and the fact that brokerage firms typically only have one analyst cover each firm. Together, these frictions create an environment where firms have varying degrees of analyst geographic dispersion.

A. Car count distribution and analyst forecast distribution: satellite-retail sample

As we can observe local information in the form of car counts over time for each specific firm, we can observe whether time-varying within firm changes in the distribution of local information are related to changes in the distribution of geographically dispersed analysts. If analysts incorporate local information into their forecasts, we would expect to see a wide range of forecasts when car counts are more varied for a firm in a quarter, and a narrow range when car counts are more consistent across locations. To examine how local information affects the distribution of forecasts, we next examine the distribution of individual forecasts within each firm-quarter. The key independent variable is the range of *Car Count Difference* across all analysts that cover the firm during a quarter. We also control for turnover, size, book to market, and number of analysts.

In column 1 of Table 6, we find that firm-quarters with higher dispersion of car counts among covering analysts are related to a wider distribution in covering analysts forecasts. In column 2, we also include *Firm* fixed effects and still find a positive and significant relation, suggesting that even within the same firm during quarters where the geographic dispersion of car counts is higher covering analysts produce a wider range of forecast estimates. These results are consistent with the individual-level results in the prior section and that local difference in information manifest in the aggregate firm-level forecast.

B. Geographic diversity, consensus earnings estimates, and firm liquidity: satellite-retail and IBES sample

If forecasts are clustered in geographic areas based on the same local information then we would expect the forecast errors to aggregate when analysts are clustered in the same MSA and to cancel out when they are disperse. Consequently, investors may experience a worse information environment regarding stocks that have less geographic coverage. We next turn to examining how the individual forecast errors interact at the consensus forecast level. We also test whether analysts' geographic diversity is relevant for a stock's bid-ask spread, a common proxy for the quality of a stock's information environment. Since we can test these implications without the use of satellite data, in this section, we expand the sample to include all firms as we can measure the geographic dispersion using solely the intersection of the IBES and CRD databases.

As a simple test, we plot the relation between consensus forecast error (panel A) and bid-ask spread (panel B) by number of distinct MSAs covering a firm in Figure 3. Panel A provides suggestive evidence that when a firm is covered by analysts from more distinct MSAs the individual weightings on local information cancel out, producing lower consensus forecast errors. Likewise, panel B suggests that firm liquidity improves as analysts become more geographically dispersed, as a firm's bid-ask spread around earnings announcements is lower when analysts are located in more distinct MSAs. However, a number of factors could confound such interpretations, so we proceed to estimate regression models to attempt to rule these potential confounders out.

To more formally test the relationship between consensus forecast error and analyst geographic dispersion, we estimate the following regression model:

$$Consensus\ Abs(Error)_{j,q} = \beta_1 * Y_{j,q} + \beta_2 * Controls_{j,q} + \gamma' * FE + \epsilon_{j,q}. \tag{5}$$

We consider two measures of geographic dispersion. First, we create a Herfindahl index of the concentration of analysts in MSAs using the number of analysts covering a firm in each MSA. We multiply this Herfindahl index by negative one so that it may be interpreted as a dispersion index. Second, we use the number of distinct MSAs for which analysts covering the firm are located. If increased geographic dispersion is associated with lower consensus forecast errors, we would expect a positive β_1 on both the dispersion index and on the number of MSAs. The specifications include controls for *Turnover*, *Book to Market*, and *Number of Analysts*. Standard errors are clustered by MSA.

Table 7 reports tests of geographic dispersion on consensus analyst forecast error. We report results for the Orbital Insight sample in panel A and for the IBES-CRD sample in panel B. In both samples, we find evidence that consensus analyst forecast errors are lower when analysts are more dispersed (column 1) or in more MSAs (column 2). Importantly, column 2 also includes *Number of Analysts* fixed effects, holding constant the effect of the number of analysts on consensus forecast error, which provides compelling evidence that it is the location and not the amount of coverage that is driving the observed relationship.

We next examine whether there are diminishing effects of geographic dispersion by including indicator variables in equation (5) that identify the exact number of MSAs for each firm. We use firms whose analysts are all in one MSA as the baseline so that coefficients are interpreted relative to firms with the least geographically diverse analysts. We include the same controls as before as well as fixed effects for *Industry*, *Number of Analysts*, and the *Year-Quarter* of the earnings announcement. We report the resulting coefficients and 95% confidence intervals in panel A of Figure 4. We find no evidence of diminishing effects of geographic dispersion, as the consensus forecast error monotonically declines as the number of MSAs increases.

We next test another financial market implication of analyst geographic diversity: information quality. Because analysts are one of the primary information intermediaries for stock valuation, shortcomings in their ability to convey all information about a firm could decrease the quality of the information available to investors. We test this idea using the following model:

$$Spread_{j,q} = \beta_1 * Y_{j,q} + \beta_2 * Controls_{j,q} + \gamma' * FE + \epsilon_{j,q}. \tag{6}$$

Where $Spread_{j,q}$ is the closing bid-ask spread scaled by the midpoint for the three days around the earnings announcement of firm j in quarter q. We again use the MSA Dispersion index and the number of MSAs as our variables of interest. The results, shown in columns 3 and 4 of Table 7, support the argument that analyst geographic diversity is positively related to an improvement in the information environment. We find spreads are lower when analysts are more dispersed as measured by the MSA Dispersion index. We also find a reduction in spreads for firms that have analysts located in more MSAs, after controlling for the number of analysts. Specifically, a firm with one more MSA covered has a spread that is 1.74bps (IBES sample) lower than an otherwise similar firm, representing a 14.5% reduction from the median spread of 12bps. Panel B of Figure 4 shows that spreads reduce monotonically as the number of covered MSAs increase. We note that results across the Orbital Insight and IBES samples are remarkably similar, further supporting our claim that any differences between the samples are unlikely to diminish the generalizability of our results.

Another compelling result is shown in Table 8, columns 1 and 3, where we include $Firm$ fixed effects and continue to find a negative β_1 on the MSA Dispersion index even within the same firm. Because the variation is within-firm, the effect is being estimated off changes in geographic coverage decisions of brokerage houses. While the choice of an individual analyst to cover a firm could plausibly be affected by the information environment of the firm, it is less clear why the geographic dispersion of coverage would be similarly affected (controlling for the number of analysts).

There are no obvious reasons to believe that endogeneity would be an important concern in this setting, as brokerage firms are unlikely to shift their entire research staffs' operations in response to their relative geographic dispersion for a single stock. Nonetheless, we next exploit the exogenous shift in geographic coverage due to brokerage closures and mergers in order to rule out any uncertainty regarding endogeneity. We follow Hong and Kacperczyk (2010) and Kelly and Ljungqvist (2012) to identify exogenous drops in coverage and look within firms that lose coverage due to the exogenous reduction in analyst coverage. The reductions in analyst coverage are either the result of the closure of a brokerage's entire research operation or due to a brokerage-firm merger where both firms employed analysts covering the same firm before the merger and only one analyst covers the firm after the merger. While prior research has used this setting as a shock to competition and information asymmetry, we are interested in the shift in analyst geographic diversity due to the brokerage closures and mergers. For example, a stock that loses coverage from its only analyst located in Minneapolis will lose geographic diversity while a stock that loses coverage from one of three analysts located in New York will retain the same level of geographic diversity. Further, this exogenous shift will only affect analyst forecast accuracy and the information environment through the change in geographic dispersion and will be uninformative about stocks' future performance. An increase in consensus forecast error for firms that lose geographic diversity relative to those that retain the same geographic diversity would lend even more support to the idea that analyst geographic dispersion is an important determinant of analyst forecast accuracy. An increase in the bid-ask spread for firms that lose a geographically diverse analyst would suggest geographic diversity is one channel through which analysts can improve a stock's information environment.

We show results for the exogenous change in analyst geographic dispersion in columns 2 and 4 of Table 8. We find larger effects on both consensus error and spread in those firms that see reduced geographic coverage (e.g., a firm loses its only analyst who is located in Minneapolis) relative to firms that also lose analyst coverage but retain the same geographic diversity (e.g., a firm with three analysts in the New York city metropolitan area loses one of these three New York analysts). This finding provides causal evidence that analyst geographic dispersion increases consensus forecast quality and improves the information environment.

VI. Conclusion

Exploiting novel data from satellite images of parking lots of U.S. retailers, we demonstrate that analysts shade their forecast in the direction of the local car counts relative to other analysts covering the same firm from different locations. Research analysts appear to rely on local information (and generalize the representative nature of this information) in the production of their forecasts. The influence of local information and diversity of local sources affects the distribution of individual analyst forecasts, the overall consensus forecast error, and firm liquidity. When firms have more geographically diverse coverage the consensus forecast error is lower. Bid-ask spreads are also reduced for firms with dispersed analysts, suggesting that analyst geographic diversity has important implications for market liquidity. We provide evidence that these effects are causal using within-firm analyses and exogenous shocks to coverage due to brokerage closures.

Research on geography in financial markets has primarily focused on the importance of being close to a firm's headquarters for access to information. We document a rich relationship between geographic diversity of analysts in the production of equity research. Our findings also highlight a key driver regarding the wisdom of crowds. By identifying local firm performance as a source of private information, we show that a systematic reliance on the same private information can lead to a reduction in accuracy. In other words, there is value in differing points of view.

References

Addoum, Jawad M., Alok Kumar, Kelvin K. F. Law. 2017. Slow diffusion of state-level information and return predictability. *Working Paper, Cornell University*.

Bae, Kee-Hong, Rene M. Stulz, Hongping Tan. 2008. Do local analysts know more? a cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics*, **88** 581–606.

Bernile, Gennaro, Alok Kumar, Johan Sulaeman. 2015. Home away from home: Geography of information and local investors. *Review of Financial Studies*, **28** 2009–2049.

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, Matthew Richardson. 2019. Information, trading, and volatility: Evidence from firm-specific news. *The Review of Financial Studies*, **32**(3) 992–1033.

Bradley, Daniel, Sinan Gokkaya, Xi Liu. 2015. Before an analyst becomes an analyst: Does industry expertise matter. *Journal of Finance*.

Bradley, Daniel, Sinan Gokkaya, Xi Liu. 2020. Ties that bind: The value of professional connections to sell-side analysts. *Management Science*.

Brown, Lawrence D., Andrew C. Call, Michael B. Clement, Nathan Y. Sharp. 2015. Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research*, **53** 1–47.

Cen, Ling, Yuk Ying Chang, Sudipto Dasgupta. 2020. Do analysts learn from each other? evidence from analysts' location diversity. *Working Paper, Chinese University of Hong Kong*.

Charness, Gary, Matthias Sutter. 2012. Groups make better self-interested decisions. *Journal of Economic Perspectives*, **26**(3) 157–76.

Chen, Brian S. 2017. Seeing is believing: The impact of local economic conditions on firm expectations, employment and investment. *Working Paper, Harvard University*.

Chen, Qi, William J Mayew, Huihao Yan. 2018. Do social interactions communicate or garble information? evidence from equity analysts. *Working Paper, Duke University*.

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

Crane, Alan, Kevin Crotty. 2020. How skilled are security analysts? *The Journal of Finance*.

Da, Zhi, Xing Huang. 2019. Harnessing the wisdom of crowds. *Management Science*.

Dehaan, Ed, Joshua Madsen, Joseph D Piotroski. 2017. Do weather-induced moods affect the processing of earnings news? *Journal of Accounting Research*, **55**(3) 509–550.

Ellison, Glenn, Edward L. Glaeser. 1999. The geographic concentration of industry: Does natural advantage explain agglomeration? *American Economic Review*, **89** 311–316.

Froot, Kenneth, Namho Kang, Gideon Ozik, Ronnie Sadka. 2017. What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? *Journal of Financial Economics*, **125**(1) 143–162.

Gao, Meng, Jiekun Huang. 2020. Informing the market: The effect of modern information technologies on information production. *The Review of Financial Studies*, **33**(4) 1367–1411.

Garcia, Diego, Oyvind Norli. 2012. Geographic dispersion and stock returns. *Journal of Financial Economics*, **106** 547–565.

Gibbons, Brian, Peter Iliev, Jonathan Kalodimos. 2019. Analyst information acquisition via edgar. *Working Paper, Pennsylvania State University*.

Goldstein, Itay, Wei Jiang, G Andrew Karolyi. 2019. To fintech and beyond. *The Review of Financial Studies*, **32**(5) 1647–1661.

Goldstein, Itay, Liyan Yang. 2015. Information diversity and complementarities in trading and information acquisition. *The Journal of Finance*, **70**(4) 1723–1765.

Green, T Clifton, Russell Jame, Stanimir Markov, Musa Subasi. 2014. Access to management and the informativeness of analyst research. *Journal of Financial Economics*, **114**(2) 239–255.

Harford, Jarrad, Feng Jiang, Rong Wang, Fei Xie. 2019. Analyst career concerns, effort allocation, and firms' information environment. *The Review of Financial Studies*, **32**(6) 2179–2224.

Hirshleifer, David, Yaron Levi, Ben Lourie, Siew Teoh. 2019. Decision fatigue and heuristic analyst forecasts. *Journal of Financial Economics*, **forthcoming**.

Hirshleifer, David A, Ben Lourie, Thomas Ruchti, Phong Truong. 2020. First impression bias: Evidence from analyst forecasts. *Review of Finance, Forthcoming*.

Hong, Harrison, Marcin Kacperczyk. 2010. Competition and bias. *Quarterly Journal of Economics*, **125** 1683–1725.

Jame, Russell, Rick Johnston, Stanimir Markov, Michael C Wolfe. 2016. The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, **54**(4) 1077–1110.

Jennings, Jared, Joshua Lee, Dawn A. Matsumoto. 2017. The effect of industry co-location on analysts' information acquisition costs. *Accounting Review*, **92** 103–127.

Kang, Jung Koo, Lorien Stice-Lawrence, Forester Wong. 2019. Attention, acquisition costs, or executive information? using big data to uncover the sources of local information. *Working Paper, University of Southern California*.

Katona, Zsolt, Marcus Painter, Panos Patatoukas, Jean Zeng. 2019. On the capital market consequences of alternative data: Evidence from outer space. *Working Paper*.

Kelly, Bryan, Alexander Ljungqvist. 2012. Testing asymmetric-information asset pricing models. *Review of Financial Studies*, **25** 1366–1413.

Kothari, SP, Eric So, Rodrigo Verdi. 2016. Analysts' forecasts and asset pricing: A survey. *Annual Review of Financial Economics*, **8** 197–219.

Lynch, Peter S, John Rothchild. 2000. *One up on Wall Street: how to use what you already know to make money in the market*. Simon and Schuster.

Malloy, Christopher J. 2005. The geography of equity analysis. *Journal of Finance*, **60** 719–755.

Merkley, Kenneth, Roni Michaely, Joseph Pacelli. 2019. Cultural diversity on Wall Street: Evidence from sell-side analysts' forecasts. *Working Paper, Indiana University*.

O'Brien, Patricia, Hongping Tan. 2015. Geographic proximity and analyst coverage decisions: Evidence from ipos. *Journal of Accounting and Economics*, **59** 41–59.

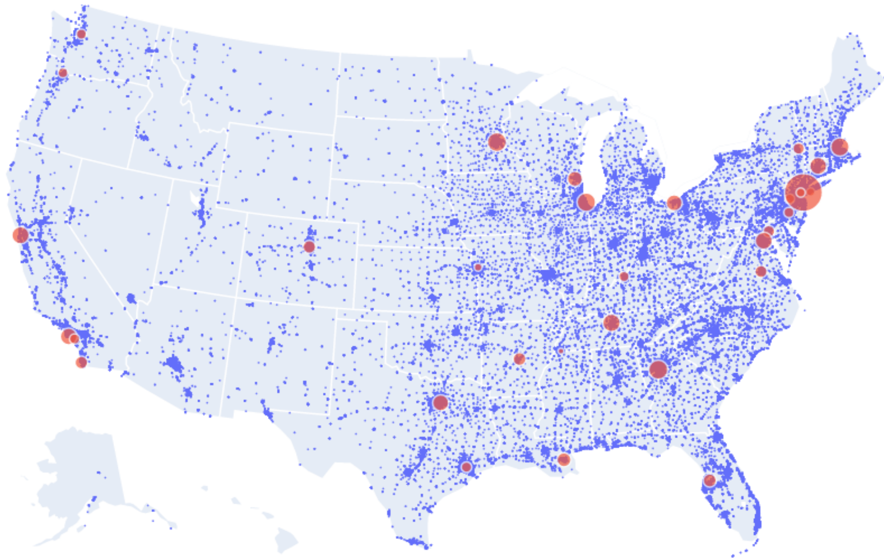
Renjiey, Rex Wang. 2018. Noise from other industries: Overgeneralization and analyst beliefs. *Working Paper, Erasmus School of Economics*.

Tversky, Amos, Daniel Kahneman. 1971. Belief in the law of small numbers. *Psychological bulletin*, **76**(2) 105.

Van Nieuwerburgh, Stijn, Laura Veldkamp. 2009. Information immobility and the home bias puzzle. *Journal of Finance*, **64** 1187–1215.

Zhu, Christina. 2019. Big data as a governance mechanism. *The Review of Financial Studies*, **32**(5) 2021–2061.

(a) All Analyst and Store Locations



(b) Home Depot Analyst and Store Locations

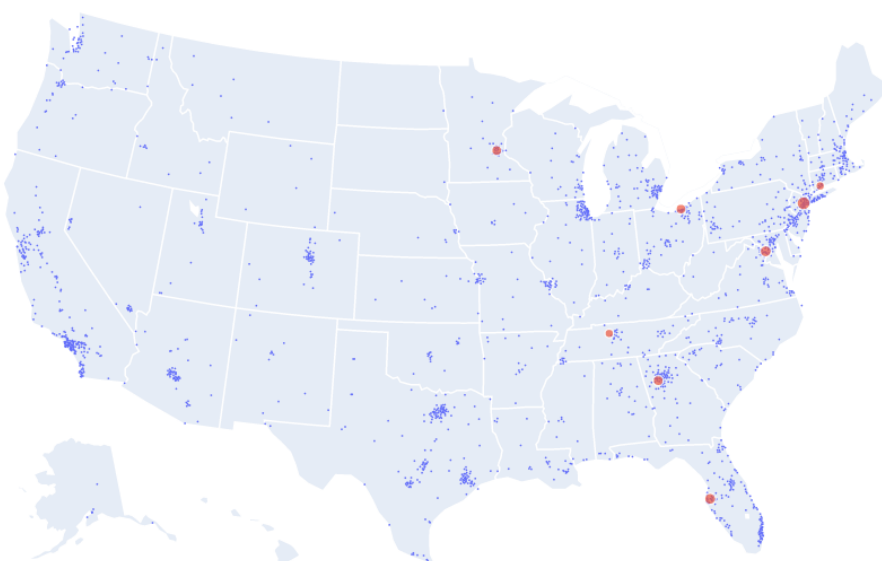
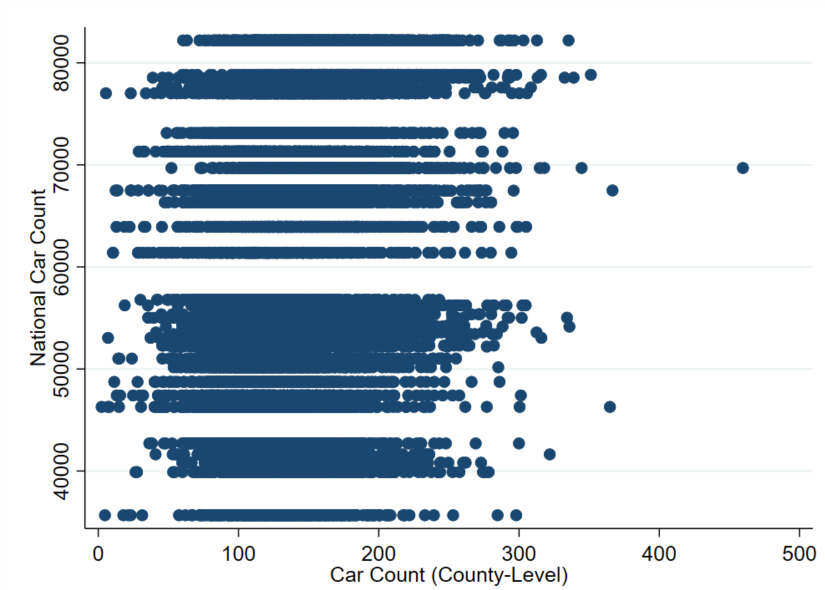


Fig. 1
Location of Financial Analysts
This figure presents a graphical representation of the location of the store locations in our Orbital Insight sample. We also indicate the location of financial analysts who cover the retail firm sample. Panel B shows the analyst and store locations for an example firm in our sample, Home Depot. The area of the red circle is proportional to the relative number of analysts in each MSA.

(a) Scatterplot - Home Depot



(b) Binned Scatterplot - Home Depot

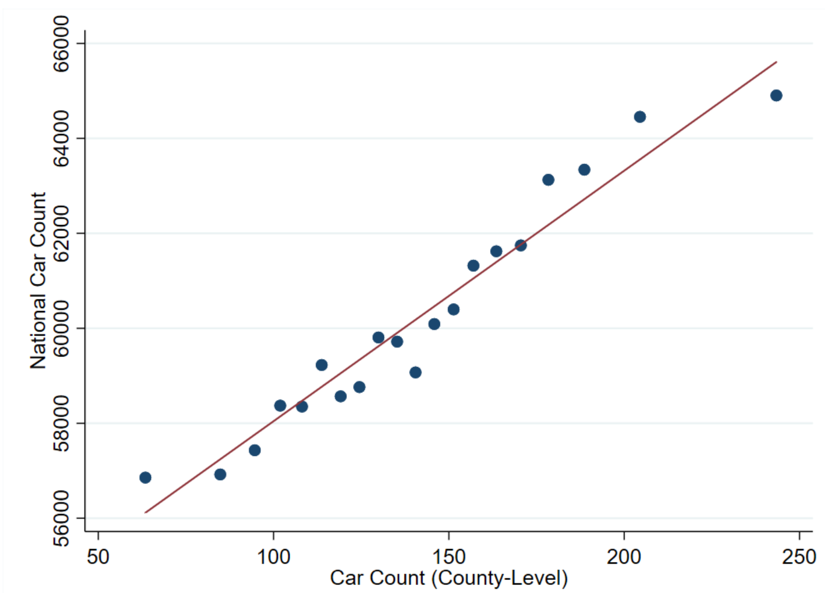
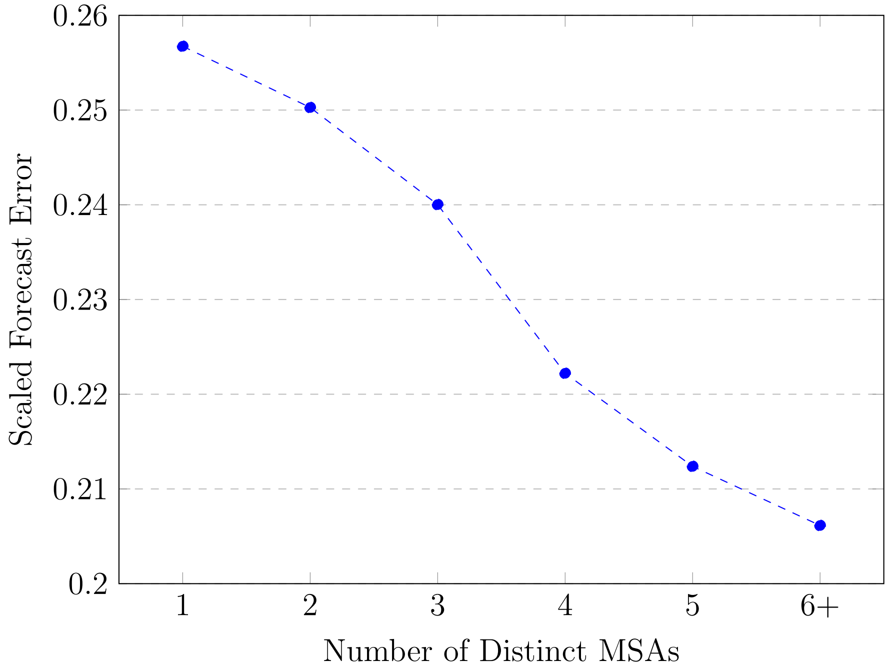


Fig. 2
National versus Local Car Counts
This figure presents scatterplots of national versus local car counts for an example firm in our data. In panel (a), we scatter plot all local car counts versus the aggregate national count within a firm-quarter. In panel (b), we binscatter plot the relative local car count (relative to county average for the firm) versus the relative aggregate national count (relative to the firm's overall average).

(a) Consensus Error by Analyst Geographic Dispersion



(b) Firm Liquidity by Analyst Geographic Dispersion

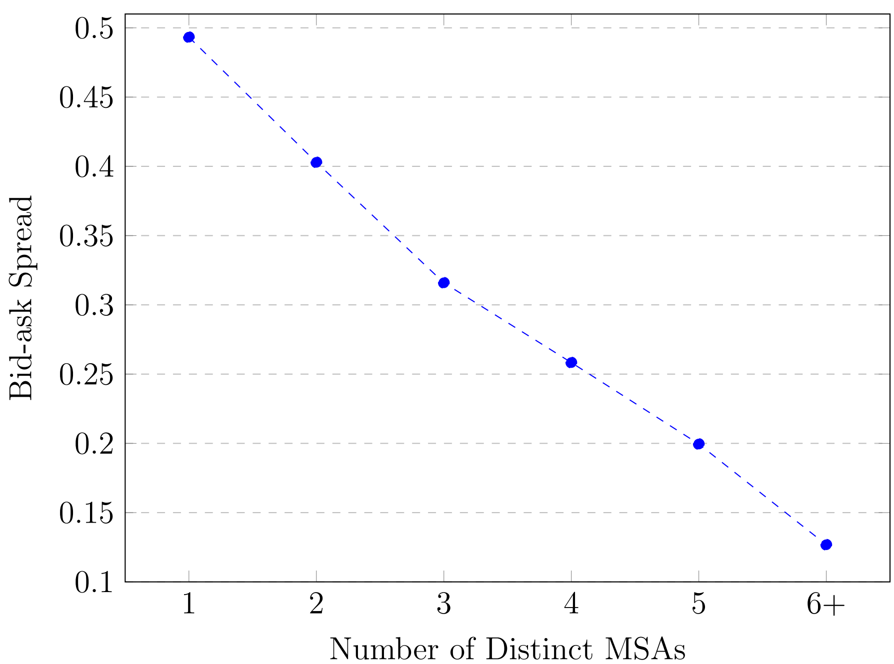
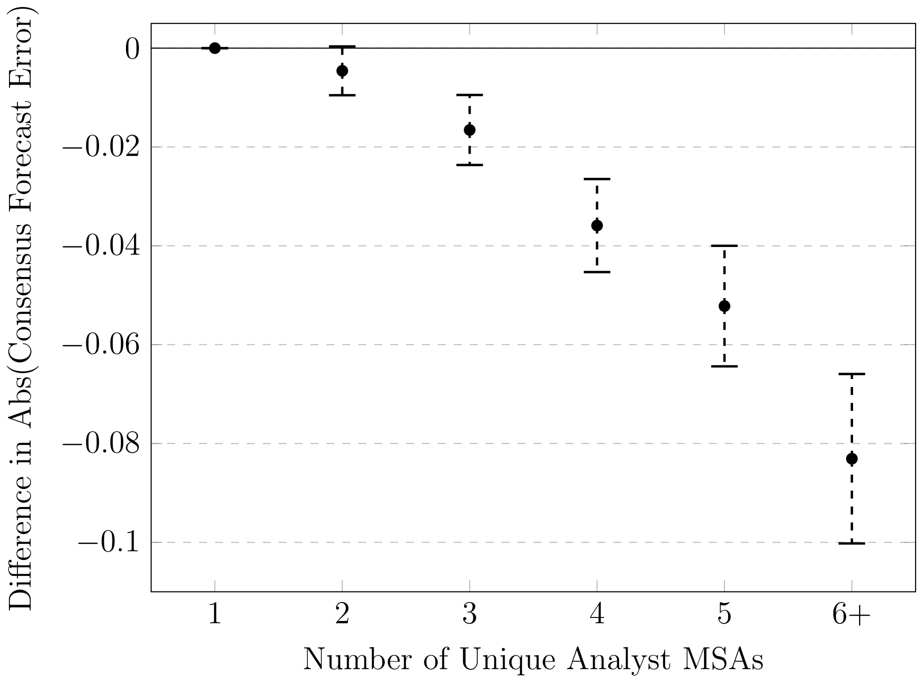


Fig. 3
Analyst Geographic Dispersion, Consensus Forecasts, and Firm Liquidity - Univariate Evidence
This figure presents means of absolute consensus forecast error (panel a) and bid-ask spread (panel b) by number of distinct MSAs from which analysts cover a firm. The sample is the union of the IBES and CRD datasets, described in Section II.

(a) Consensus Error by Analyst Geographic Dispersion



(b) Firm Liquidity by Analyst Geographic Dispersion

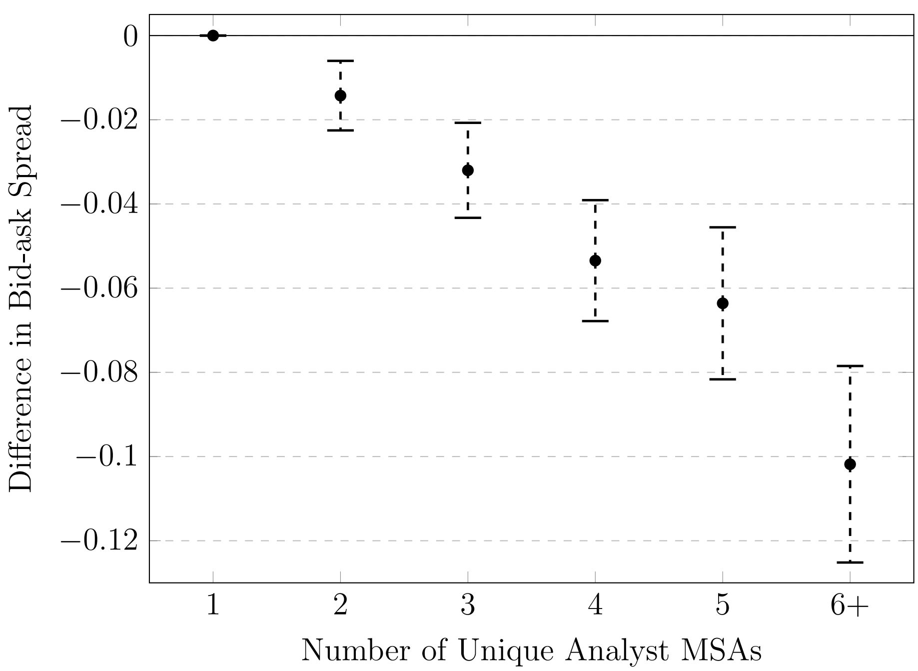


Fig. 4
Analyst Geographic Dispersion, Consensus Forecasts, and Firm Liquidity - Regression Evidence
This figure plots coefficients of indicator variables that identify the number of distinct MSAs from which analysts cover each firm, which we include in regressions of equation (5) and (6). The dependent variable in panel A is the absolute consensus analyst forecast error. The dependent variable in panel B is the bid-ask spread. Regressions include Industry, Number of Analyst, and Year-Quarter fixed effects and controls for Turnover, Book-to-Market, and Size. Error bars, based on standard errors clustered by firm, denote 95% confidence intervals.

Table 1

Summary Statistics

This table reports summary statistics for the sample of retail firms covered by Orbital Insights satellite data from January 2009 through December 2015. Panel A contains summary statistics for retail firm-quarter observations. Panel B presents summary statistics for the broader coverage of all IBES analysts. Panel C shows geographic coverage statistics for the top five largest firms in the retail sample. Panel D shows the retail sample's geographic coverage by MSAs.

Panel A: Retail sample summary statistics

	Mean	Std Dev	Q1	Median	Q3
Error	-0.08	0.17	-0.15	-0.05	0.0
Car count difference	0.17	0.47	-0.17	0.12	0.45
MSAs with Store Location	0.58	0.26	0.37	0.62	0.80
Number of analyst covering firm	6.40	5.15	2	5	10
Number of MSAs with an analyst	3.77	2.43	2	3	5
Headquarters	0.07	0.25	0	0	0
No Local Stores	0.14	0.35	0	0	0
Analyst total experience (quarters)	40.27	31.89	12	37	61
Analyst firm experience (quarters)	13.55	16.34	3	8	18
Forecast age (days)	59.82	34.52	23	70	90
Turnover	0.27	0.14	0.16	0.25	0.35
Market equity ($ millions)	9905.20	25962.47	810.12	2359.27	7967.94
Book to market	0.42	0.27	0.23	0.37	0.58

Panel B: IBES sample summary statistics

	Mean	Std Dev	Q1	Median	Q3
Consensus Abs(Error)	0.28	0.28	0.06	0.16	0.42
Bid-ask Spread	0.39	0.59	0.05	0.12	0.40
MSA Dispersion Index	-0.60	-0.29	-0.33	-0.50	-1.00
Number of analysts covering firm	3.90	3.13	2	3	5
Number of MSAs with an analyst	2.06	1.37	1	2	3
Analyst total experience (quarters)	34.50	25.90	13	29	50
Analyst firm experience (quarters)	10.46	12.96	2	6	14
Forecast age (days)	63.46	40.12	31	72	90
Turnover	0.19	0.20	0.08	0.14	0.24
Market equity ($ millions)	5181.18	19720.40	268.63	849.75	2930.39
Book to market	0.68	3.40	0.28	0.49	0.79

Panel C: Geographic coverage of five largest firms in retail sample						
Company Name	SIC	Market Equity ($ millions)	% MSAs have Store	# MSAs have Analyst	# Analysts covering firm	Average car count
Walmart	5331	221,022.7	94%	5	31	245.77
Home Depot	5211	90,380.6	87%	8	29	131.81
CVS Health	5912	68,711.7	67%	6	29	16.07
Walgreens	5912	45,637.8	90%	7	28	17.90
Lowe's	5211	42,866.6	88%	8	29	116.80

Panel D: Retail sample geographic coverage by MSA			
# MSAs	%	Avg. # Analysts	Avg. Error
1	30.7%	2.4	-0.088
2	23.0%	4.7	-0.080
3	16.6%	7.6	-0.078
4	12.2%	10.2	-0.067
5 or more	17.5%	11.9	-0.063

Table 2

Predicting Retail Firm Returns with Parking Lot Data
This table reports tests of stock return predictability using parking lot data from Orbital Insight. The unit of observation is the MSA-Firm-Quarter. The dependent variable is a firm's return from the day of an earnings announcement until three days after the announcement. *Local (National) Car Count* is the logarithm of the average number of cars for a firm in an MSA (nationally) during a given quarter. *Car Count Difference (Local-National)* , is different between the local and nation car count. Controls are included for a firm's share turnover, size, book-to-market, and stock return from 30 days before the earnings announcement until three days before. Standard errors are clustered by MSA.

	Earnings Announcement Return	
	(2)	(3)
National Car Count	0.0651***	0.0657***
	(0.0156)	(0.0157)
Car Count Difference (Local-National)		0.0010
		(0.0016)
Controls	Yes	Yes
Firm FE	Yes	Yes
Quarter FE	Yes	Yes
R^2	0.101	0.101
Observations	7,249	7,249

Table 3

Local Information and Individual Analyst Forecast Error - Retail Sample

This table reports tests of local information's effect on individual analyst forecast error. The sample is 18,679 retail firm analyst-quarters. The dependent variable is individual analyst forecast error, which is the price scaled forecast error of analyst i for firm j in quarter q. The key independent variable is *Car Count Difference*, which is the relative local performance of firm j in the MSA of analyst i during quarter q. *No Local Stores* is equal to one for an analyst that lives in an MSA with no retail stores of the forecasted firm. *Headquarters* is equal to one for an analyst that lives in the MSA of the forecasted firm's headquarters. *Turnover* is the log of the number of shares traded by shares outstanding over the prior 12 months. *Size* is the log of the market capitalization of the firm, calculated as the share price multiplied by the number of shares outstanding. *Book to Market* is the log book value of equity minus the log market value of equity. *Number of Analysts* is the number of analysts covering firm j in quarter q. *Analyst Experience* is measured as number of quarters the analyst has worked in industry *Overall* or covering the *Firm*. *Forecast Age* is the number of days between forecast and actual date. Standard errors are clustered by MSA.

		Error	
	(1)	(2)	(3)
Car Count Difference	0.0066***	0.0090**	0.0100***
	(0.0020)	(0.0033)	(0.0032)
No Local Stores	0.0200**	0.0016	0.0021
	(0.0078)	(0.0052)	(0.0052)
Headquarters	-0.0089	-0.0048	-0.0063
	(0.0054)	(0.0038)	(0.0043)
Turnover	-0.0895**		
	(0.0367)		
Size	0.0296***		
	(0.0063)		
Book to Market	0.0673***		
	(0.0197)		
log(Num Analysts)	0.0412***		
	(0.0056)		
Analyst Experience	-0.0012	-0.0002	-0.0003
	(0.0008)	(0.0005)	(0.0004)
Analyst Experience Covering Firm	0.0007	-0.0002	-0.0000
	(0.0011)	(0.0004)	(0.0004)
log(Forecast Age)	-0.0030***	-0.0015**	-0.0010**
	(0.0007)	(0.0007)	(0.0005)
Firm FE	Yes	No	No
Year×Quarter FE	Yes	No	No
Firm×Year×Quarter FE	No	Yes	Yes
MSA FE	No	No	Yes
R^2	0.164	0.815	0.816
Observations	18,679	18,190	18,189

Table 4

Heterogeneity in Local Information and Individual Analyst Forecast Error - Retail Sample

This table reports tests of local information's effect on individual analyst forecast error. The sample is 18,679 retail firm analyst-quarters. The dependent variable is individual analyst forecast error, which is the price scaled forecast error of analyst i for firm j in quarter q. The key independent variable is *Car Count Difference*, which is the relative local performance of firm j in the MSA of analyst i during quarter q. In column (1), we interact *Car Count Difference* with *Experience*. In column (2), we interact *Car Count Difference* with *Number of Analysts*. In column (3), we interact *Car Count Difference* with *Number of MSAs*. In column (4), we interact *Car Count Difference* with an indicator identifying whether a firm's satellite data is available from a satellite vendor. The specifications include, but we do not report, coefficients for *No Local Stores*, *Headquarters*, *Turnover*, *Size*, *Book to Market*, *Number of Analysts*, *Analyst Experience*, and *Forecast Age*. Standard errors are clustered by MSA.

	Error			
	(1)	(2)	(3)	(4)
Car Count Difference	0.0100***	0.0098***	0.0120***	0.0118***
	(0.0032)	(0.0032)	(0.0036)	(0.0037)
Car Count Difference × Experience	-0.0028*			
	(0.0015)			
Car Count Difference × # Analysts		0.0015		
		(0.0031)		
Car Count Difference × # MSAs			-0.0054**	
			(0.0024)	
Car Count Difference × Satellite Data Release				-0.0113**
				(0.0042)
Controls	Yes	Yes	Yes	Yes
Firm×Year×Quarter FE	Yes	Yes	Yes	Yes
MSA FE	Yes	Yes	Yes	Yes
R^2	0.816	0.816	0.816	0.816
Observations	18,189	18,189	18,189	18,189

Table 5

Robustness of Local Information and Individual Analyst Forecast Error - Retail Sample

This table reports tests of local information's effect on individual analyst forecast error. The sample is 18,679 retail firm analyst-quarters. We include *Analyst* × *Firm* fixed effects in column (1). We include $Analyst \times Year \times Quarter$ fixed effects in column (2). In column (3), we exclude analysts located in the New York metropolitan area. In column (4), we exclude analysts from MSA-firm-quarters with fewer than five satellite pictures. The dependent variable is individual analyst forecast error, which is the price scaled forecast error of analyst i for firm j in quarter q. The key independent variable is *Car Count Difference*, which is the relative local performance of firm j in the MSA of analyst i during quarter q. We create a dummy *Low Car Count* which equals one if the relative local performance of firm is in the lowest quartile, and create a dummy *High Car Count* which equals one if the relative local performance of firm is in the highest quartile (column 5). The specifications include, but we do not report, coefficients for *No Local Stores*, *Headquarters*, *Turnover*, *Size*, *Book to Market*, *Number of Analysts*, *Analyst Experience*, *Analyst Experience Covering a Firm* and *Forecast Age*. Standard errors are clustered by MSA.

			Error		
	(1)	(2)	(3)	(4)	(5)
Car Count Difference	0.0100***	0.0136***	0.0080*	0.0080***	
	(0.0030)	(0.0037)	(0.0044)	(0.0029)	
Car Count Difference × High Car Count					0.0103**
					(0.0047)
Car Count Difference × Low Car Count					0.0099***
					(0.0031)
Controls	Yes	Yes	Yes	Yes	Yes
Firm×Year×Quarter FE	Yes	Yes	Yes	Yes	Yes
MSA FE	No	No	Yes	Yes	Yes
Analyst×Firm FE	Yes	Yes	No	No	No
Analyst×Year×Quarter FE	No	Yes	No	No	No
Sample	Full	Full	No NYC	High Image	Full
R^2	0.849	0.896	0.855	0.817	0.816
Observations	18,005	16,591	6,120	10,369	18,189

Table 6

Variation in Local Information and the Spread of Individual Analyst Forecasts - Retail Sample

This table reports tests of local information dispersion on the range of individual analyst forecasts. The sample is 2,638 firm-quarters. The dependent variable is the range of individual analyst forecasts. The key independent variable is range of *Car Count Difference* across all analysts that cover firm j during quarter q. The specification include, but we do not report, coefficients for *Turnover*, *Book to Market*, and *Number of Analysts*. Standard errors are clustered by firm.

	Range of Individual Forecasts	
	(1)	(2)
Range of Car Counts	0.0228***	0.0118***
	(0.0042)	(0.0030)
Controls	Yes	Yes
Firm FE	No	Yes
Year×Quarter FE	Yes	Yes
R^2	0.128	0.377
Observations	2,638	2,635

Table 7

Analyst Geographic Dispersion, Consensus Forecasts, and Firm Liquidity
This table reports tests of geographic dispersion on consensus analyst forecast error. The sample in panel A is 6,053 firm-quarters, covering the firms in the Orbital Insight sample. Panel B includes 180,393 firm-quarters for all analysts we identify in the IBES sample. The dependent variable is the absolute consensus analyst forecast error in the first two columns and the bid-ask spread in the last two columns. *Bid-ask Spread* is the closing ask price minus the bid price scaled by the midpoint and multiplied by 100 for the three day window around the earnings announcement date. *MSA Dispersion Index* is a Herfindahl index of the concentration of analysts in MSAs, multiplied by negative one for ease of interpretation. *Number of MSAs* are the number of unique MSAs in which analysts covering a firm are located. The specifications include, but we do not report, controls for *Turnover*, *Book to Market*, and *Number of Analysts*. Standard errors are clustered by firm.

Panel A: Retail Sample

	Consensus Abs(Error)		Bid-ask Spread	
	(1)	(2)	(3)	(4)
MSA Dispersion Index	-0.0849***		-0.2806***	
	(0.0247)		(0.0460)	
Number of MSAs		-0.0125***		-0.0166**
		(0.0036)		(0.0072)
Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year×Quarter FE	Yes	Yes	Yes	Yes
Number of Analysts FE	No	Yes	No	Yes
R^2	0.398	0.403	0.731	0.729
Observations	6,053	6,053	5,967	5,967

Panel B: IBES Sample

	Consensus Abs(Error)		Bid-ask Spread	
	(1)	(2)	(3)	(4)
MSA Dispersion Index	-0.0887***		-0.2563***	
	(0.0059)		(0.0101)	
Number of MSAs		-0.0131***		-0.0174***
		(0.0014)		(0.0020)
Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year×Quarter FE	Yes	Yes	Yes	Yes
Number of Analysts FE	No	Yes	No	Yes
R^2	0.277	0.275	0.681	0.682
Observations	180,389	180,389	177,564	177,564

Table 8

Changes in Analyst Geographic Dispersion, Consensus Forecasts, and Liquidity - IBES Sample

This table examines the causal link between geographic dispersion and consensus analyst forecast error. The sample is firm-quarters where the firm lost analyst coverage due to an exogenous brokerage closure or merger. The dependent variable is the absolute consensus analyst forecast error in the first two columns and the bid-ask spread in the last two columns. *Loss of MSA* equals one if the firm is covered from fewer MSAs after an analyst departure. The specification includes, but we do not report, coefficients for *Turnover*, *Book to Market*, and *Number of Analysts*. Standard errors are clustered by firm.

	Consensus Abs(Error)		Bid-ask Spread	
	(1)	(2)	(3)	(4)
MSA Dispersion Index	-0.0150***		-0.1114***	
	(0.0043)		(0.0102)	
Loss of MSA		0.0179*		0.0341**
		(0.0107)		(0.0160)
Number of MSAs		-0.0102***		-0.0087*
		(0.0034)		(0.0047)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	No	Yes	No
Industry FE	No	Yes	No	Yes
Year×Quarter FE	Yes	Yes	Yes	Yes
R^2	0.542	0.237	0.791	0.586
Observations	179,689	6,550	176,865	6,481