Preliminary Draft

Near Is Dear: Remoteness, Soft Information, and Stock Returns

Chao Gao[*]

First version: June, 2022

Current version: October, 2022

Abstract

Geographic proximity plays an important role in dissemination of soft information. We document a causal link between geographic remoteness and lack of soft news in the market. Releases of public soft news surprise the market more and induce larger market reactions for remote firms. Consistent with incomplete information theories, we find that remotely headquartered stocks outperform proximate stocks by 8.59% annually on a risk-adjusted basis. The post-earnings announcement drift and return predictability of aggregate mutual fund trading are only observed among remote firms, highlighting the role of soft information in price discovery. Geographic dispersion of operations alleviates the information frictions associated with headquarter remoteness.

Keywords: proximity, remoteness, information frictions, soft information, cross-sectional stock returns, post earnings announcement drift, earnings momentum, geographic dispersion.

JEL Classification: D80, G12, G14, G39.

[*] Gao (chao.gao@anu.edu.au) is from the College of Business and Economics, Australian National University.

1. Introduction

The headquarter of a firm acts as an information interchange between the firm and the market. This is particularly true for soft information beyond quarterly earnings and any other mandatory disclosures. It is well documented that headquarter locations play an important role in investor preference, information flow, corporate behaviors, governance, and etc.[1] These previous findings are supportive of geographic locations being relevant to dissemination of soft information. While the proximity of the firm headquarter to certain groups of stakeholders of the firm is extensively studied, little is known about the informational role of the overall geographical remoteness of the firm and its asset pricing implications. We attempt to fill this gap by constructing general remoteness measures for conterminous U.S. firms using a large panel of headquarter location and population data.

The main questions we attempt to answer are 1) whether the geographic remoteness of the firm's headquarter impedes dissemination of soft information and, if so, 2) how it affects the pricing of its equity stock. A remote headquarter comes with benefits and costs. A common cost faced by all the remote firms should stem from the fact that they are farther away from their investors, financers, analysts, and other financial market participants. These parties both demand information from firms as well as supply information to assist firms' decision-making. However, common wisdom suggests that geographic remoteness increases information frictions between the firm and the financial market and impedes information flows both ways.

In this paper, we take the investors' perspective and focus on the information acquisition frictions faced by the firm investors. We attempt to directly and formally investigate the relation between headquarter remoteness and availability of soft information. Moreover, we hypothesize that

[1] Coval and Moskowitz (1999), Coval and Moskowitz (2001), Hau (2001), Feng and Seascholes (2004), Ivkovic and Weisbenner (2005), Loughran and Schultz (2005), Malloy (2005), Kang and Kim (2008), Chen, Gompers, Kovner, and Lerner (2010), Anand, Gatchev, Madureira, Pirinsky, and Underwood (2011), John, Knyazeva, and Knyazeva (2011), Ghoul, Guedhami, Ni, Pittman, and Saadi (2013), Giroud (2013), Korniotis and Kumar (2013), Alam, Chen, Ciccotello, and Ryan (2014), Hollander and Verriest (2016), Ellis, Madureira, and Underwood (2020), Da, Gurun, Li, and Warachka (2021), and Chen, Ma, Martin, and Michaely (2022).

the market requires higher compensation for holding firms with more remote headquarters due to the additional information frictions. This hypothesis rests on several previous theoretical works. Merton's (1987) theory of incomplete information suggests that investors who are not well diversified demand higher compensation from informationally segregated firms. Theoretical works by Easley, Hvidkjaer, and O'Hara (2002) and Easley and O'Hara (2004) suggest that investors require higher returns from firms with higher information asymmetry. All these theories easily apply to the pricing of geographic remoteness if remoteness impede the flow of soft information. Yet the association between remoteness and stock returns has not been empirically investigated.

To test these hypotheses, we rely on a set of novel geographic remoteness measures constructed from a large panel of U.S. firm headquarter locations and county-level population data. Several previous studies define a firm as remote if it is not located close to one of the financial centers or big cities.[2] These measures serve the purposes of the corresponding studies, however, they lack of certain features that we desire for testing the general informational and pricing implications of remoteness. Undoubtedly, people are the intermediary of information dissemination. And how accessible is the firm to these people is also critical. To serve the goal of this paper, we construct our remoteness measures based on both the firm's distance to the U.S. counties and the county-level populations. In our context, our measures have four major advantages compared to previous binary remoteness measures or point-to-point proximity measures. First, our remoteness measures are continuous and comparable across all firms. Second, being close to only one big city is intuitively different from being close to multiple big population centers. Thus, our measures are more comprehensive and account for the firm's overall relative location. Third, our measures explicitly incorporate population, which plays the essential role in information dissemination. Finally, several recent studies highlight the importance of reduction in travel time in mitigating the information frictions due to long physical distances.[3] The framework of our remoteness calculation allows us to take into account the air travel and construct remoteness measures that focus

[2] See, for example, Loughran and Schultz (2005), Chen, Gompers, Kovner, and Lerner (2010), John, Knyazeva, and Knyazeva (2011), Ghoul, Guedhami, Ni, Pittman, and Saadi (2013), and Chen, Ma, Martin, and Michaely (2022).

[3] See, for example, Giroud (2013), Ellis, Madureira, and Underwood (2020), Da, Gurun, Li, and Warachka (2021), and Chen, Ma, Martin, and Michaely (2022).

either on the best travel time or physical distance. This fact naturally facilitates an interesting comparison between the travel time based and the physical distance based remoteness in terms of their relevance. We discuss the measure construction in more details in the Section 2.

Does headquarter remoteness really impede the flow of soft information? The current literature does not provide a clear definition for soft information. In our study, soft information is defined as qualitative information beyond earnings news and news with a relatively clear valuation effect embedded in it, such as mergers and acquisitions, security transactions, and analyst ratings changes. We take three steps to provide a direct and formal answer to this question. First, we construct proxies for public soft news using a subset of RavenPack news. In a panel regressions setting, we show that the cross-sectional relation between remoteness and availability of public soft news is negative and highly significant. Second, we exploit exogenous shocks that temporarily reduce certain firms' accessibility and thus increase their remoteness without changing their news worthiness. Applying these shocks to a stringent difference-in-differences framework, we find that the treatment effect is associated with significantly less public soft news. In our third step, we investigate the market impact of the public soft news in the cross section. If it is more difficult to acquire soft information of remote firms, publically released soft information by news media should surprise the market more and thus induce a larger market reaction. Consistent with this notion, we find a significantly stronger association between the sentiment score of soft news and the market reaction for more remote firms.

We then turn to the stock price implications of the information frictions associated with geographic remoteness. The theories of incomplete information support a notion that remote firms earn higher returns than their proximate counterparts. We first examine the return predictability of the remoteness in the cross section and find that remote firms indeed tend to outperform their proximate counterparts. When we sort stocks based on their remoteness, the high decile (most remote firms) earns a monthly risk-adjusted return of 0.496%, highly statistically significant. The low decile (most proximate firms) earns a monthly risk-adjusted return of -0.192%, highly statistically significant. This is consistent with our "near is dear" hypothesis. In a Fama-MacBeth regression framework that controls other stock characteristics such as size, book-to-market ratio, past returns, volatility, and illiquidity, a one-standard deviation increase in remoteness leads to a 10-basis-point increase in return of the following month. This return predictability is strong and robust. A firm's remoteness is not expected to vary significantly over time. Consequently, we expect the return predictability of remoteness to be extended to longer windows. We test the return predictability of remoteness for the 1-year, 2-year, 3-year, 4-year, and 5-year ahead monthly returns after the portfolio formation. For these future months, the high-minus-low portfolio earns a monthly risk-adjusted return ranging from 0.532% to 0.756%, all highly statistically significant. The large magnitude of these returns echo the persistence of geographic remoteness as a firm characteristic. This persistence of return predictability also challenges other predictive variables as alternative explanations, most of which vary over time, and suggests a trading strategy that requires minimal portfolio rebalancing.

We further discuss the stock price implications of remoteness in the context of information frictions. Soft information helps evaluate firm performance, forecast fundamental information, and cross-validate disclosed fundamental information. Consequently, frictions associated with acquiring such information naturally distort stock price efficiency. We utilize several interesting settings to provide insights into how geographic remoteness influence stock returns through information frictions. First, we find that the famous post-earnings announcement drift (hereafter PEAD), indicating delayed information incorporation, is only observed among remote firms, suggesting that it takes longer to price the fundamental information in absence of soft information. Second, we show that actively managed mutual funds, who are diligent information acquirers, rebalance their remote holdings more frequently than other holdings. And consistent with an information advantage story, the aggregate trading by these mutual funds predicts future returns only among remote firms. Finally, the return pattern across remoteness portfolios disappear when the firms have geographically dispersed operations. That is, having dispersed information outlets effectively alleviates information frictions associated with headquarter remoteness.

Our study is closely related to two strands of literatures. First, there is a large body of literature that studies the role of the geographic locations of firm headquarters. Most of these studies relate firm geographic locations to information frictions or investor preferences, the two of which are often not mutually exclusive. Following this line of thoughts, it is well documented that geographic locations are related to a range of firm level variables, including trading activities and profits (Coval and Moskowitz (1999), Coval and Moskowitz (2001), Hau (2001), Feng and Seasholes (2004), Loughran and Schultz (2005), Ivkovic and Weisbenner (2005), Anand, Gatchev, Madureira, Pirinsky, and Underwood (2011), Ellis, Madureira, and Underwood (2020), and Da, Gurun, Li, and Warachka (2021)), implied cost of equity (El Ghoul, Guedhami, Ni, Pittman, and Saadi (2013)), borrowing cost (Hollander and Verrist (2016)), corporate governance (Kang and Kim (2008), Chen, Gompers, Kovner, and Lerner (2010), John, Knyazeva, and Knyazeva (2011), Giroud (2013), Alam, Chen, Ciccotello, and Ryan (2014)), correlation with local business cycles (Korniotis and Kumar (2013)), and information production (Malloy (2005) and Chen, Ma, Martin, and Michaely (2022)), among others. These studies indirectly support the role of geographic locations in information dissemination. Yet, whether geographic remoteness really impedes the flow of soft information and the corresponding equity pricing implications largely remains an open question.[4] Second, our paper is also directly related to the literature that investigates the pricing implications of information frictions using other proxies. Easley, Hvidkjaer, and O'Hara (2002) construct a measure of information asymmetry, PIN, and show that this measure is positively associated with future stock returns. Fang and Peress (2009) document a positive association between media coverage and future stock returns. Our paper attempt to add the missing link between these two literatures above by 1) formalizing the remoteness-information-frictions relation and 2) investigating the cross-sectional return predictability of geographic remoteness.

A part of our study also contributes to the large literature on the PEAD. The PEAD effect refers to the anomaly where a firm's stock price tends to drift in the direction of the previous earnings surprise. This anomaly is first formally documented by Bernard and Thomas (1989) and has attracted great attention ever since.[5] Almost all previous evidence points towards market underreaction as the explanation. The real question is what cause this underreaction. To our best knowledge, we are among the first to link information frictions, and soft information in particular, to the PEAD effect. A relatively

---

[4] An exception is Ghoul, et al. (2013), who investigate implied cost of equity, calculated using dividends and analyst earnings forecasts.

[5] See Richardson, Tuna, and Wysocki (2010) for a detailed review of this literature.

close literature studies investors' attention in explaining the PEAD effect (See, for example, DellaVigna and Pollet (2009), Hirshleifer, Lim, and Teoh (2009), and Hung, Li, and Wang (2015)).

The rest of the paper is organized as follows. In Section 2, we discuss the construction of remoteness measures. In Section 3, we introduce the data sources and our sample and summarize our remoteness measures. In Section 4, we establish the relation between geographic remoteness and soft information. In Section 5, we examine the stock return predictability of remoteness in the cross section. Section 6 further discusses the stock price implications of remoteness in the context of information frictions. Section 7 concludes the paper.

2. Remoteness of Firm Headquarters

In this section, we define the geographic remoteness measures used in our study. There is virtually no theoretical guidance on how remoteness should be measured. For the purpose of this study, the remoteness measures should highlight the geographic characteristics of the firm headquarters that are related to information dissemination. While the economic gravity models suggest that the amount of economic interactions between two entities are determined by both the distance in between and the characteristics of these two entities, the determination of the usable model is largely an empirical procedure that cannot be easily generalized to our remoteness definition.[6] That said, our remoteness measures still borrow the basic idea from the gravity model. Population is intuitively the intermediary of information dissemination. And more closely located population is more effective with this role. To avoid introducing noises with complicated methodologies, we rely on a very straight-forward calculation to incorporate 1) the firm headquarter's distance to surrounding county population centers and 2) the populations of the corresponding counties.

Our first set of remoteness measures are based on the best travel times between firm headquarters and county population centers. A complication of this method stems from the fact that a traveller can take the ground transportation or travel by air from one location to another. And when

[6] See, for example, Anderson (1979), Anderson and Van Wincoop (2003), Anderson (2011), and Anderson (2016) for detailed discussions of the gravity model.

choosing the air travel, we still need to take into account the ground transportation to and from the airport. Therefore, to determine the best travel time between a firm headquarter and a county population center, we compare the driving time (if only the ground transportation is chosen) with the optimal itinerary by air. For the driving time calculation, although the U.S. interstates have speed limits well above 100km per hour, we set the travel speed at 100km per hour for all trips longer than 100km to account for the fact that we use straight-line distances, which underestimate driving distances.[7] For trips below 100km, the average speed is set at 60km per hour. For the optimal itinerary by air, we first calculate the total travel time by combining the airport-to-airport time and the time to and from the airports. We do so for all the flight/airport options available to a traveller and pick the optimal one.[8] Finally, we determine the best travel time by comparing the driving time with the optimal itinerary by air.

With the best travel time determined, we are able to construct the remoteness measures in two steps. In the first step, we calculate the accessibility (the opposite of remoteness) each year for each firm as the time-adjusted population within a certain radius from the firm headquarter:

$$AccessTime_{i,t} = \sum_j Pop_{j,t} * \frac{Time_{radius} - Time_{i,j,t}}{Time_{radius}}. \quad (1)$$

Here, $Pop_{j,t}$ is the total population of county j in year t. $Time_{radius}$ is the maximum time radius from the firm headquarter within which the counties are considered. $Time_{i,j,t}$ is the best travel time between the firm headquarter and the population center of county j. In the second step, we calculate the z-score of $AccessTime_{i,t}$, $AccessTime\_z_{i,t}$. Then the remoteness measure is given by:

$$RemoteTime_{i,t} = -AccessTime\_z_{i,t}. \quad (2)$$

Our second set of remoteness measures are based on physical distances between firm headquarters and county population centers. The calculation is similar to equations (1) and (2) above

---

[7] Although using a shortest-path distance along a road network is more accurate, the highway system in U.S. is very developed. This fact allows us to use the straight-line distance as a reasonable proxy.

[8] For example, a slightly farther airport may be the optimal option for the traveller as the closer airport may not have direct flights. We use an algorithm to determine the optimal itinerary by air that takes into account multiple origin and destination airports.

but only utilize the simple physical distance instead of the best travel time. The remoteness measure is given by the following two equations:

$$AccessDist_{i,t} = \sum_j Pop_{j,t} * \frac{Distance_{radius} - Distance_{i,j,t}}{Distance_{radius}} \tag{3}$$

and

$$RemoteDist_{i,t} = -AccessDist\_z_{i,t}. \tag{4}$$

Here, $Pop_{j,t}$ is the total population of county j in year t. $Distance_{radius}$ is the maximum distance radius from the firm headquarter within which the counties are considered. $Distance_{i,j,t}$ is the physical distance between the firm headquarter and the population center of county j. $AccessDist\_z_{i,t}$ is the z-score of $AccessDist_{i,t}$.

We construct four remoteness measures for our main empirical tests, two based on travel time and two based on physical distance. For the travel time based measures, we assign values 6 hours and 10 hours to $Time_{radius}$ in equation (1) to calculate Remote6h and Remote10h, respectively. Since there is no theoretical guidance on what $Time_{radius}$ should be, we believe the selection of this value ought to be the result from balancing the cost (noisiness) and benefit (inclusiveness). That is, although it is appealing to always include all the data, population located too far from the firm headquarter may not be relevant in the information dissemination and may introduce noise to our measure. We pick 6 hours for $Time_{radius}$ as it is approximately the longest direct flight in conterminous U.S.. It is also roughly the upper limit for a single-day travel. And we calculate an alternative measure using $Time_{radius}$ of 10 hours to be more inclusive. Similarly, we calculate physical distance based measures Remote600 and Remote1000 by assigning values 600km and 1000km to $Distance_{radius}$ in equation (3).

For both sets of remoteness measures, remoteness decreases in population in the surrounding area and increases in distances (travel time or physical distance) to the population centers. Additional details about the measure construction are discussed in the data section below.

3. Data and Sample

Our remoteness and soft information measures are the key elements of this study. We introduce the data used for variable constructions in Sections 3.1, 3.2, and 3.3. In Section 3.4, we discuss our final sample and filters. Section 3.5 presents summary statistics for understanding the remoteness measures.

3.1 Headquarter Locations and County Population Centers

Our remoteness measures are based on the distance between the firm headquarter location and the surrounding county population centers. We first obtain firm headquarter address data from the Forms 10-Q and 10-K for all publicly listed companies from SEC EDGAR system. To allow accurate calculation of point-to-point distances, we obtain the longitudes and latitudes for all U.S. zip codes from the U.S. Census Bureau Gazetteer files, which are used to locate each firm on the map. Each year, we define the firm's location using its last 10-Q or 10-K filing address. The county population center data in 2000, 2010, and 2020 are obtained from the U.S. Census Bureau Centers of Population files, which contain the longitude and latitude of each county population center. The center of population is defined as the balance point of the county when weights of identical size were placed on it so that each weight represented the location of one person.[9] Since these population centers are provided every 10 years, each year, we use the closest available data to define the county centers. For example, 2010 county population center data are used to define county centers for the ten years from 2005 to 2014.

3.2 Airport and Air Travel Data

In addition to the firm headquarter and county population center data, the travel-time based remoteness measures heavily rely on the airport and flight information. The airport data are obtained from the Federal Aviation Administration website, which include the longitude and latitude information of each airport in the U.S.. We manually fill in this information for a small number of historical airports that ceased to exist.

[9] A detailed explanation of the measurement can be found here: https://www2.census.gov/geo/pdfs/reference/cenpop2020/COP2020_documentation.pdf.

Flight information is acquired from the U.S. Bureau of Transportation Statistics. In order to identify the best itinerary between two locations, we utilize both the T-100 Domestic Market and the T-100 Domestic Segment data. Both datasets include the detailed information of the origin airport, the destination airport, the airline, and the number of passengers transported at monthly frequency. There are two main differences between these two datasets. First, the Market data define an itinerary as a single trip regardless the number of legs in-between the airports as long as the flight number remains the same. As a result, the passengers are "enplaned" and counted only once. In comparison, the Segment data contain information about each ramp-to-ramp flight and treat them as completely separate trips. Second, the Market data provide no travel time information. The Segment data, on the other hand, provide the ramp-to-ramp time between airports. Due to these differences, the Market data are useful in identifying itineraries, while the Segment data are useful in calculating travel times.[10]

To calculate the air travel time between two airports, we first eliminate the cargo flights using passenger number information. We then merge the Market data with the Segment data to identify the exact itineraries and number of legs:

A. Direct flights: We use the Segment data to identify the direct flights. An origin airport and destination airport pair can be served by multiple airlines. For a reasonable estimate of the travel time, we use the median value of all the ramp-to-ramp times for the trip in each year.
B. One-stop flights: We merge the Market data with the Segment data by the origin airport and the destination airport. The unmatched observations in the Market data represent itineraries with multiple legs. Therefore, we merge this remaining observations with the Segment data by the origin airport and the destination airport, separately, to obtain the Leg1 and Leg2 datasets, respectively. Finally, we match the former's destination airport with the latter's origin airport to identify the one-stop flights.

[10] A detailed explanation of the differences between the two datasets can be found here: https://transportation.libanswers.com/faq/166158

C. Multi-stop flights: We eliminate the identified direct flights and one-stop flights above from the Market data. Then we repeat the process in Step B to identify itineraries with multiple stops.

To calculate the final travel time, we make two adjustments to the ramp-to-ramp time. First, for each itinerary, we add a total one-hour transition time for the transitions at the origin and destination airports. Second, for non-direct flights, we add one hour for each overlay.

3.3 Soft Information

The current literature does not provide a clear definition for soft information. What complicate things even more is the fact that some soft information is only known to a limited group of investors, which causes an overlap with private information. To serve the purpose of our study, we quantify the extent to which there are publicly available soft news, which is defined as qualitative news beyond earnings news and news with a relatively clear valuation effect embedded in it, such as mergers and acquisitions, security transactions, and analyst ratings changes. Our measures of public soft news rely on RavenPack News Analytics database, which covers news articles from a broad range of news media since 2000.

We apply two layers of filters on the raw data from RavenPack. In the first step, we exclude news from less trustworthy news sources that have RavenPack's Source_Rank above 3. And we keep only news with relevance score of 100. In the second step, we identify soft news based on RavenPack's news event group variable, which categorizes each event contained in each news. Some common news event groups include earnings, analyst-ratings, and investor-relations among many other. To construct our soft news measure, we select the groups clearly identify soft news.[11] Detailed construction of our soft news measure is discussed in Section 4.

[11] Soft news are defined as news from groups assets, business-activities, civil-unrest, corporate-responsibility, crime, industrial-accidents, natural-disasters, legal, labor-issues, pollution, security, transportation, and war-conflict.

3.4 Sample and Filters

Our final sample includes firms with common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020 where the remoteness measures can be calculated. The sample period begins in 1996 due to the availability of the county population center data. In tests where we rely on RavenPack data, the sample begins in 2000. We limit our sample firms to those with headquarters in the conterminous U.S.. We obtain stock-level data from the Center for Research in Security Prices (CRSP) and exclude stocks with prices below $1 at the end of the previous month. The accounting information is from Compustat. Our final sample contains 12,775 firms and 1,191,456 firm-month observations.

3.5 Understanding Remoteness

In this section, we illustrate our remoteness measures in the time series and on the map and discuss some features of firms with remote headquarters. Our remoteness measures are transformed from the accessibility measures of equations (1) and (3), who themselves are distance-adjusted total populations in nature and thus easy to interpret. Therefore, we summarize these accessibility measures over our sample period to gain understanding of the remoteness measures and the differences between the measures based on travel time and physical distance. The results are reported in Table 1. The first observation is that Access6h is more than twice as large as Access600 across all the years. While both measures take into account population within roughly 6-hour travel time, when air travel is allowed, a much larger population becomes relevant. It is also worth noting that all four accessibility measures substantially increase over our sample period on average. This is not surprising as the changes of these measures are directly comparable to the population growth by construction. The interesting implications can be found in the magnitude of these changes. From 1995 to 2020, the total population of the U.S. grew by 23.73%. In comparison, the changes of Access6h and Access600 are 33.63% and 20.02%, respectively. The substantially larger change of the travel time based accessibility compared to the population growth demonstrates how the development of air travel has been advancing connectedness of different locations.

To demonstrate the relative remoteness of different locations, we illustrate the average remoteness across all firms in each U.S. county on the map for year 2020. In Figure 1, the darker blue indicates a higher value of average remoteness. Counties with no publicly listed firms in our sample are coded grey. Panel A and B use Remote6h and Remote600 as the remoteness measures, respectively. There are commonalities as well as some interesting differences characterizing these two measures. For the travel time based remoteness in Panel A, it is clear that Western counties are relatively more remote. And except for the most populated area between New York and Washington D.C., the relatively more proximate counties are usually close to large air hubs such as ATL of Atlanta, DFW of Dallas, and ORD of Chicago. In Panel B, the pattern of the distance based remoteness has the similar general pattern where the Western counties are relatively more remote. But slightly different from Panel A, the most proximate counties coincide with areas with the highest population density. And the Midwestern counties are more remote in this measure compared to when the travel time based measure is used.

Who are these remote firms? We calculate the average remoteness measures for the Fama-French 12 industries.[12] For each industry, we first average the remoteness measures across all the firms each year. Then we average across all the years. The results are reported in Table 2. The first column reports the industry average Remote6h. The top three industries with high average remoteness are Business Equipment, Healthcare, and Utilities. The most proximate three industries are Chemicals, Consumer Nondurables, and Manufacturing. Remote10h in the second column suggests a similar ranking. The third column reports the industry average Remote600. The top three industries with high average remoteness are Energy, Business Equipment, and Utilities. The most proximate three industries are Chemicals, Financial, and Consumer Nondurables. The ranking in the last column is similar to that of the third column. Overall, the travel time based and distance based remoteness measures broadly agree on the type of remote firms.

Finally, we summarize the firm characteristics across firms with different remoteness rankings. Market beta is estimated using daily returns of the last 3 months. Mcap is the firm's market capitalization. BM is the firm's book-to-market ratio where book value of equity is calculated as the Compustat book

[12] The definitions of the industries are obtained from the Kenneth R. French Data Library.

value of equity adding back deferred taxes and subtracting preferred equity. Lagret is the previous month return. Past 12 is the return of the previous twelve months with one month lag. Volatility is the standard deviation of the daily returns of the previous month. ILLIQ is the Amihud (2002) illiquidity measure. Every month, we assign firms into decile portfolios based on the remoteness measure, Remote6h, of the previous calendar year. For each decile portfolio, we first average the firm characteristics across all the firms each month. Then we average across all the months. We also calculate the differences between the high remoteness and low remoteness deciles. The results are quantitatively similar when sorting on other remoteness measures.

The results are reported in Table 3. All the remoteness measures monotonically increase from low to high Remote6h deciles, suggesting strong correlations among these measures. Market beta measures the market risk exposure of the firm. The average beta is 0.958 and 0.952 for the low and high remoteness portfolios, respectively. The high-minus-low difference is statistically insignificant. And market beta shows no clear pattern across remoteness deciles. Firm size is not only related to systematic risk profile of the firm, but also used to proxy for various firm characteristics including information frictions. The average firm size shows a U-shaped pattern across the remoteness deciles with the low point at decile 8. The high and low remoteness deciles contain relatively larger firms with average size of $5.3 billion and $6.3 billion, respectively. The average book-to-market ratio shows an inverted U-shaped pattern across the remoteness deciles with the high point at decile 7. The high and low deciles have average BM of 0.716 and 0.791, respectively. Turning to the past returns, remote firms seem to have better past performance than proximate firms. This is more evident for the past 12 month returns with one month lag. Remote firms also have more volatile daily price movements suggested by Vol. Finally, the top decile has higher average illiquidity than the bottom decile. The average illiquidity is much higher among deciles 2 to 7.

4. Remoteness and Soft Information

The previous literature provides suggestive evidence that remotely located firms suffer from information frictions, especially for soft information. In this section, we attempt to formalize this idea.

And we take three steps to provide convincing evidence of the geographic effect on soft information. In Section 4.1, we construct proxies for public soft news and examine the cross-sectional relation between public soft news and remoteness. In Section 4.2, we exploit exogenous shocks to certain firms' accessibility and establish the causal link between remoteness and public soft news using a difference-in-differences setting. In Section 4.3, we investigate the market impact of the public soft news in the cross section.

4.1 Remoteness and Soft News in Panel Regressions

Is headquarter remoteness really associated less soft information? To examine this relation, we first construct public soft news measures using news data from RavenPack. We aggregate soft news based on news event group described in Section 3.3 every month into a softness measure SoftRatio, which is the proportion of soft news in all the news with non-missing group identifier in the month. This variable measures the softness of the overall public news of the firm. We then estimate panel regressions of soft news measures on remoteness as specified below:

$$SoftRatio_{i,m} = b_0 + b_1 Remoteness_{i,y-1} + b_2 Controls_{i,m-1} + \gamma_j + \eta_m + \varepsilon_{i,m}. \quad (5)$$

Here, $SoftRatio_{\mathrm{i,m}}$ is news softness of firm i for month m. The variable $Remoteness_{i,\mathrm{y}-1}$ is one of our remoteness measure for firm i from the previous year. For the lagged control variables $Controls_{i,m-1}$, we include the log of one plus the total number of news in the previous month, the log of the market capitalization from the previous month, the log of the book-to-market ratio from the previous month, the previous month return, the past 12-month return (skipping the most recent month), the volatility from the previous-month daily returns, and the Amihud (2002) illiquidity measure. We include the industry (3-digit SIC) fixed effect to control for industry specific characteristics that could drive a correlation between a firm's remoteness and availability soft news. We also include a year-month fixed effect to control for time trends. If headquarter remoteness impedes flow of soft information and results in fewer public soft news, we expect the coefficient b_1 to be significantly negative.

The estimated coefficients are reported in Table 4 using four different remoteness measures, as shown in column (1) to (4). In the first column, we use travel-time based Remote6h as the remoteness measure. The coefficient is -0.221 and statistically significant at 1% level. Given that our remoteness measures are standardized, this coefficient suggests that a one standard deviation increase in remoteness reduces SoftRatio by -0.221. To put this value in perspective, SoftRatio has a median of 0 and 75 percentile of 1.124. Moving to control variables, the firm's news softness is positively related to the number of news coverage LagNews. Firm size carries a negative coefficient as it is highly correlated with the number of news coverage. And dropping the LagNews turns this coefficient positive and highly significant. News softness is also positively associated with book-to-market ratio and volatility and negatively associated with past performance. In column (2) to (4), we use Remote10h, Remote600, and Remote1000 as the remoteness measure, respectively. The coefficients on remoteness are quantitatively similar and all statistically significant at 1% level.[13]

4.2 An Identification Strategy: Flight Delay, Diversion, and Cancellations

To ensure that the strongly negative coefficients documented above are not due to latent variables that drive both firms' headquarter choices and soft information, we exploit exogenous shocks to firms' accessibility that do not have direct impact on news worthiness of the firm for identification. Namely, we utilize cases where an abnormally large portion of flights are significantly delayed, diverted, or cancelled in a month for the airport that is most commonly used by potential visitors of the firm. These cases create exogenous variations in the cross-section as well as in the time-series and thus facilitate highly stringent difference-in-differences tests to help us establish causality.

We obtain the flight On-Time Performance dataset from the U.S. Bureau of Transportation Statistics. The dataset contains identifiers that indicate if the scheduled flight was on-time, delayed, diverted, or cancelled. We define a flight as significantly delayed if the delay is over 120 minutes. Each month, we count the delays, diversions, and cancellations for each airport as the origin airport and then as the destination airport. A delay or a cancellation can be caused by either of the airports. However, a

[13] Results based on an alternative soft news measure, log of one plus the number of soft news, are qualitatively similar to the results in Table 4.

diversion can only be caused by the destination airport. Thus, we sum up the counts of delays and cancellations where the airport is either the origin or the destination. And then we add to it the count of diversions where the airport is the destination. We scale this number by the total number of scheduled flights to and from this airport to calculate the monthly DelayRate for each airport. We plot the average DelayRate by calendar month in Figure 2.[14] Apparently, DelayRate significantly in winter months due to blizzards in summer months due to thunderstorms. The three months associated with the highest DelayRates are January, February, and June.

As some airports have severer delay than others in general, we gauge the severity of delay for each airport using its own historical data. Specifically, abnormally high delay, AbnormalDelay, is defined for each firm's most commonly used airport by comparing the current month DelayRate with the average DelayRate of the previous 12 months. AbnormalDelay is equal to 1 if the current month DelayRate is at least 50% higher than the historical average and 0 otherwise. We then examine how soft information reacts to these accessibility shocks using a difference-in-differences regression specification:

$$SoftRatio_{i,m} = \; c_0 + c_1 AbnormalDelay_{i,m} + c_2 Controls_{i,m-1}$$

$$+\gamma_i + \eta_m + \varepsilon_{i,m}. \tag{6}$$

Here, $SoftRatio_{i,m}$ is news softness of firm i for month m defined in the previous section. $AbnormalDelay_{i,m}$ equal to 1 assigns firm i into treatment group and control group otherwise. For the lagged control variables are defined the same as in Table 4. To facilitate a stringent difference-in-differences test, we include both the firm fixed effect and the year-month fixed effect. As $AbnormalDelay_{i,m}$ varies across firms and overtime, these fixed effects facilitate a comparison between treatment and control firms (the first difference) and how the effect changes the firm moves in

[14] Here we use data from 2000 to 2019. We exclude 2020 as there were significant cancellations of scheduled flights between March and May of 2020 caused by COVID outbreak.

and out of the treatment group (the second difference). If headquarter remoteness has a negative causal effect on the flow of soft information, we expect the coefficient c_1 to be significantly negative.

The estimated coefficients are reported in Table 5. In column (1) where we include only the treatment indicator, the coefficient on this indicator is -0.260 with a highly significant t-statistic of -3.48. After including all control variables, the coefficient on the treatment indicator remains negative and statistically significant at -0.238. This result confirms the causal effect of headquarter remoteness/accessibility on availability of soft news.

4.3 Market Reactions to Soft News

Although a causal link between remoteness and soft news is established, we take a further step to characterize the lack of soft news for remotely located firms. For a firm that is transparent to a large base of investors, the market relies less on news media to collect soft information. On the contrary, if a firm is remotely located and thus lack of publicly available soft information, any publicly released soft news would help update investor believes and thus induce larger market reactions.

To test this hypothesis, we apply two additional filters to soft news identified in Section 3.3. First, in order for the news to reach a broad investor base, we follow the literature and retain only soft news covered by *Dow Jones Financial Wire*, *Barron's*, and *The Wall Street Journal*. Second, to ensure that the news is fresh to the market, we exclude news with preceding similar events within 90 days. We calculate cumulative abnormal returns (CARs hereafter) for these soft news as the compounded market-adjusted returns in a short three-day window (CAR[-1,1]) and a long twenty-one-day window (CAR[-10,10]). Then we estimate the following specification:

$$CAR_{i,k} = \ d_0 + d_1 CSS \times Remoteness_{i,y-1} + d_2 CSS_{i,k} + d_3 Remoteness_{i,y-1}$$

$$+ d_2 Controls_{i,m-1} + \gamma_j + \eta_m + \varepsilon_{i,m}. \qquad (7)$$

Here, CAR_i is the market reaction to soft news k of firm i in month m. $CSS_{i,k}$ is the Composite Sentiment Score from RavenPack that gauges the textual sentiment of the news. $Remoteness_{i,\mathrm{y}-1}$ and the lagged control variables are defined the same as in Table 4. We include the industry (3-digit SIC) fixed effect and the year-month fixed effect. We expect d_2 to carry a positive coefficient due to the news sentiment effect. However, if soft information for remote firms is scarce and thus the releases of soft news about these firms surprise the market more, we expect the coefficient d_1 on the interaction term to be significantly positive.

We report the regression results using CAR[-1,1] in Table 6 Panel A. This short-window market reaction by design captures the actual market reactions to the soft news releases. In column (1) where we use Remote6h to gauge the remoteness, remote firms do not seem to have systematically more positive or negative soft news. As expected, the sentiment score carries a positive and statistically significant coefficient. Moving to the interaction term, we find that the association between the sentiment score and the market reaction is significantly stronger for more remote firms. Based on the coefficient of 4.556 on CSS, the coefficient on the interaction term suggests one standard deviation increase in remoteness lead to a roughly 50% increase in the market reaction, holding CSS constant. The results are similar in column (2) to (4) where we use Remote10h, Remote600, and Remote1000 as the remoteness measure, respectively. Comparing the magnitude of the coefficients suggests a relatively weaker effect from the distance-based remoteness measures. In Table 6 Panel B, we change the dependent variable to the market reaction from a longer window CAR[-10,10] for robustness. And this longer window also helps us check whether the difference in market reactions between remote and proximate firms is due to fast learning before news releases for proximate firms. If the news content was learned by the market within 10 days before the news release, we expect this longer window to capture the corresponding return and diminish the market reaction difference between remote and proximate firms. As shown in this panel, the interaction term still carries a large positive coefficient that is highly significant in column (1) and (2) where we use travel time based remoteness measures. This result suggests that the proximate firms might have been incorporating these soft information long before this return window. In column (3) and (4), this coefficient is less significant for distance based remoteness measures.

5. Remoteness and Cross Section of Stock Returns

The previous section establishes the causal link between headquarter remoteness and availability of soft information. The theories of incomplete information suggest that investors require a return premium from firms with more severe information frictions. We hypothesize that these theories apply to soft information, and as a result, remote firms should be priced more conservatively and associated with higher future stock returns. In this section, we present our analysis of remoteness and the cross section of stock returns. We present results from our portfolio analysis in Section 5.1. Section 5.2 documents the Fama-MacBeth regression analysis. Section 5.3 and 5.4 examine the return persistence and patterns across other firm characteristics.

5.1 Performance of Remoteness Sorted Portfolios

To begin examining the remoteness-return relation, we first conduct a conventional portfolio analysis. This approach has several benefits. First, a portfolio approach replicates the returns that would be realized if remoteness was used as the basis for a trading strategy. Second, this methodology makes no assumptions regarding the linearity of the relationship between remoteness and subsequent returns. Every month, we sort stocks into decile portfolios based on remoteness measures of the previous calendar year. For each portfolio, we first calculate the average return in the cross section using value weights. We then calculate the time-series average returns and estimate the regression coefficients relative to the state-of-art Fama and French (2015) five-factor model.

In Table 7, we report the results for the decile portfolios as well as the spread portfolio formed by longing the high decile and shorting the low decile. We use Remote6h as the remoteness measure in Table 7 Panel A. The results confirm a remoteness premium. For average portfolio returns, the low decile earns 0.700% per month. The high decile earns 1.384% per month. The spread portfolio earns 0.684% per month, statistically significant at 1% level. The return pattern is not perfectly monotonic as the lowest average return of 0.592% is earned by decile 6.

If the cross-sectional return variation fully explained by Fama-French five factors, then remote firms should not earn significantly a higher alpha than the proximate firms. However, the results show that the lowest two deciles earn negative monthly alphas of -0.192% and -0.181%, with t-statistics of -2.45 and -1.82 respectively. And the highest two deciles earn positive monthly alphas of 0.496% and 0.416%, both statistically significant at 1% level. All other deciles in the middle earn statistically insignificant alphas. The spread portfolio earns a monthly alpha of 0.689%, statistically significant at 1%. This magnitude is comparable to the raw return spread, suggesting that little return premium is explained by the risk factors. This result confirms the premium from the univariate result and shows that the return difference is from both the long and the short legs of the strategy.[15]

The coefficients on the risk factors are also interesting. The spread portfolio positively loads on the market factor (MKT), suggesting a positive exposure to the market risk. The coefficients on the book-to-market factor (HML) and the investment factor (CMA) are significantly negative, suggesting that the spread strategy has positive exposure to growth firms and firms with aggressive investment. This strategy does not seem to have exposure to size factor or profitability factor.

Panel B of Table 7 documents average portfolio returns and alphas using other remoteness measures for portfolio formation. The results are consistent with Panel A overall. All of the spread portfolios earn positive and statistically significant returns and alphas. When using a more inclusive travel time based remoteness Remote10h, the spread portfolio earns an average return of 0.605% and an alpha of 0.700%, both statistically significant. When using physical distance based measures, Remote600 and Remote1000, the spread portfolio returns are slightly smaller in magnitude and less statistically significant compared to when the travel time based measures, Remote6h and Remote10h, are used. The return pattern is also less monotonic when sorted on distance based remoteness measures.

5.2 Fama-MacBeth Regressions

The portfolio approach is a straight-forward and intuitive way of establishing the remoteness-return relation and demonstrate the trading profit. In this section, we control for other relevant firm-

[15] With respect to the CAPM, Fama and French (1993) three-factor, and Carhart (1997) four factor models, high remoteness deciles earn alphas of 0.402%, 0.421%, and 0.416%, all statistically significant. The low remoteness deciles earn alphas of -0.057%, -0.069%, and -0.072%, respectively, all statistically insignificant. The spread portfolios earn alphas of 0.459%, 0.490%, and 0.488%, respectively. The t-statistics are 1.92, 2.59, and 2.46, respectively. The results are reported in the Internet Appendix.

level characteristics at the same time by adopting the two-stage estimation of Fama and MacBeth (1973). For the first stage of Fama-MacBeth regression, for each month m, we estimate a cross-sectional specification:

$$Return_{i,m} = f_{0m} + f_{1m}Remoteness_{i,y-1} + f_{2m}Controls_{i,m-1} + \varepsilon_{i,m}. \qquad (8)$$

Here, $Return_{i,\mathrm{m}}$ is the cumulative pre-announcement abnormal return of stock i for month m. The variable $Remoteness_{i,\mathrm{y}-1}$ is one of our remoteness measure for stock i from the previous year. For the lagged control variables $Controls_{i,m-1}$, we include the log of the market capitalization from the previous month, the log of the book-to-market ratio from the previous month, the previous month return, the past 12-month return (skipping the most recent month), the volatility from the previous-month daily returns, and the Amihud (2002) illiquidity measure.

The first stage estimation yields the time-series of coefficients $\{f_{0m}, f_{1m}, f_{2m}\}$. In the second stage, we average these time-series of coefficients to obtain estimates for f_0, f_1, and f_2. The time-series t-statistics were calculated using Newey-West standard errors. If the market requires higher returns from remote firms, we expect the coefficient f_1 to be significantly positive.

We report the time-series averages of the coefficients estimated from the above regression in Table 8 using four different remoteness measures, as shown in column (1) to (4). In column (1), we use the travel time based remoteness measure, Remote6h. The coefficient on the uncertainty proxy is 0.096 with a t-statistic of 2.93. As our remoteness measures are standardized, this coefficient implies that a one standard deviation increase of $Remote6h$ increases the monthly return by 0.096%. The positive and significant coefficient supports our hypothesis that the market requires higher returns from more remote firms.

In column (2) to (4), we use alternative remoteness measure Remote10h, Remote600, and Remote1000, respectively. The more inclusive travel time based measure Remote10h carries a coefficient of 0.091, statistically significant at 5% level. In comparison, the coefficients on the physical distance based measures, Remote600, and Remote1000, are 0.047 and 0.067, respectively. The corresponding t-statistics are 1.57 and 2.10, respectively. That is, the travel time based measures seem to have stronger association with returns than the physical distance based measures as shown in both the magnitude and statistical significance. This is consistent with our portfolio results.

Overall, both the portfolio and regression results support the existence of a remoteness premium. The return predictability is distinct from a number of variables associated with subsequent returns. And interestingly, the travel time based remoteness measures seem to be more relevant to the returns. The additional fact that both Remote6h and Remote1000 have stronger associations with returns compared to Remote600 seems to advocate the importance of including physically distant but chronologically proximate populations in remoteness calculation. From this point forward, we use Remote6h as the main remoteness measure.

5.3 Return Persistence

Changes to regional population and accessibility to air travel are intuitively sedate over time. Consequently, our remoteness measures, especially the relative rankings, should be highly persistent. Then, can we extend the remoteness-return relation in the previous section to returns farther into the future? The answer to this question is important as most of other predictive variables fluctuate substantially over time, resulting in short-lived return predictability. Thus, high persistence in the remoteness-return relation would separate the remoteness effect from majority of alternative explanations. This question should also be of interest to investors who intend to exploit remoteness in a trading strategy. A highly persistent return predictability implies little need for portfolio rebalancing and thus lower transaction costs.

To investigate this question, we adopt a similar portfolio approach as in Table 7, which uses remoteness measure of the previous calendar year. For the persistence check, we lag the sorting variables by additional 1 to 5 years to examine its return predictability for monthly returns 1 to 5 years after the original portfolio formation. We expect the remote firms to persistently outperform the proximate firms in the future.

Table 9 reports the Fama-French five-factor alphas of remoteness sorted portfolios. The results echo the persistent nature of remoteness. For the portfolios sorted on 1- to 5-year lagged $Remote6h$, the low deciles earn negative alphas of -0.186%, -0.193%, -0.241%, -0.182%, and -0.109%, respectively, all statistically significant at 5% level except for the 5-year lag. The corresponding high deciles earn positive alphas of 0.358%, 0.562%, 0.364%, 0.376%, and 0.423%, respectively, all statistically significant at 5% level. The spread portfolios earn alphas ranging from 0.532% to 0.756%, all statistically significant at 1% level. More importantly, the return predictability for the 5-year ahead return is still highly significant and large in magnitude. Such strong performance persistence hints at a trading strategy requiring minimal portfolio rebalancing.

5.4 Sequentially Sorted Portfolios

In the previous sections, we document robust evidence that remote firms persistently earn higher returns than proximate firms. Is this remoteness-return association more evident among stocks with certain characteristics? To answer this question, we form sequentially sorted portfolios. Every month, we first sort firms into quintiles based on a previous-month firm characteristic and then based on Remote6h. Size, market beta, BM, Lagret, Past12, and ILLIQ are all defined in Table 3. Turnover is the total trading volume scaled by the number of shares outstanding. IVOL is the idiosyncratic volatility of Ang, Hodrick, Xing, and Zhang (2006). For each portfolio, we first calculate the average return in the cross section using value weights. We then calculate the time-series average returns and estimate the regression coefficients relative to the Fama-French five-factor model. For brevity, we limit our reporting to the high-minus-low remoteness spread portfolios within each characteristic quintile.

Table 10 reports the alphas of these spread portfolios. One might be concerned that our previous results are driven by smaller and less visible firms. However, the spread alphas show an interesting pattern across the size quintiles. From the small to large quintiles, the spread alpha increases monotonically from 0.077% to 0.658%. The t-statistic also monotonically increases from 0.46 to 3.74.[16] Apparently, the remoteness effect is significantly weaker among smaller firms. A likely explanation is that information about larger firms relates more to industry- or market-level shocks. Thus, information frictions impeding flows of the market-level information post larger undiversifiable uncertainty on these

---

[16] This pattern explains why remoteness is much less correlated with future returns in the portfolio results when portfolios are formed using equal weights.

firms. The market beta quintiles provide additional insights along this line of thought. The largest alpha of 0.763% is observed in the high market beta quintile. This result echoes the results for size quintiles and advocates a larger remoteness effect among informationally important firms.

Moving to other characteristics, for BM quintiles, the remoteness spread is positive and significant in the low BM quintile and quintile 3, but statistically insignificant in the other quintiles. It implies remoteness is more relevant for firms with high growth opportunities. For past return quintiles, the remoteness spread is positive and significant in all the quintiles except for the low Past12 quintile. There is also no clear pattern across the idiosyncratic volatility quintiles. The remoteness spread earns an alpha of 0.709% with a t-statistic of 3.44 in the low IVOL quintile, suggesting that the remoteness effect does not rely on potential idiosyncratic shocks. Finally, the remoteness effect cannot be explained by stock liquidity. The spread alphas are significant in most of the turnover and ILLIQ quintiles. And interestingly, remoteness spread portfolio earns the largest alphas in the high turnover quintile and the low ILLIQ quintile. This is likely related to the size effect discussed above.

6. Further Discussions

In this section, we further discuss the stock price implications of remoteness in the context of information frictions. Soft information helps evaluate firm performance, forecast fundamental information, and cross-validate disclosed fundamental information. Consequently, frictions associated with acquiring such information naturally distort stock price efficiency. We utilize several interesting settings to provide insights into how geographic remoteness influence stock returns through soft information frictions. In Section 6.1, we examine the famous post-earnings announcement drift in the context of remoteness. In Section 6.3, we study the trading behaviors of diligent information acquirers across the remoteness groups. In Section 6.4, we investigate the role of geographically dispersed operations in alleviating the information frictions associated with remote firms.

6.1 Post-Earnings Announcement Drift

While soft information (by our definition) may contain fundamental information per se, it could also help cross-validate and effectively incorporate other public information in the stock prices. For example, the Chief Financial Officer of the firm frequently having dinners with bankers does not necessarily convey bad information. But it becomes a different story if the firm makes a negative earnings announcement at the same time. Combining both pieces of information, the market could infer financial distress and bad prospect in a longer horizon. Therefore, frictions impeding the flow of soft information lead to delayed price reactions to hard information.

A natural test of this hypothesis is the well-known PEAD effect. The PEAD effect refers to the anomaly that a firm's stock price tends to drift in the direction of the previous earnings surprise, which is hard public information. The previous evidence overwhelmingly points towards a market underreaction story. In this section, we jointly test two hypotheses: 1) remote firms suffer from lack of soft information and thus price inefficiency; and 2) the PEAD effect is more evident among remote firms due to the lack of soft information. To test these hypotheses, we utilize sequentially sorted calendar-time portfolios suggested by Fama (1998). To account for the stock performance in the 6 months after the earnings announcement, we line up the event months for all the earnings announcements by calendar month. Every calendar month, we first sort firms into quintiles based on Remote6h then based on an earnings surprise measure.[17] We utilize two earnings surprise measures. CAR is the market reaction based earnings surprise, which is the cumulative abnormal return in the three-day window around the earnings announcement. We first compute the abnormal daily returns by subtracting the CRSP value-weighted daily market returns from the stock daily returns, then obtain the cumulative abnormal returns by compounding the daily abnormal returns. SUE is the analyst forecast based earnings surprise, which is calculated as the median value of analyst forecasts minus the actual earnings per share value deflated by the last quarter-end stock price. We hold each stock for 6 months after the earnings announcement. The portfolios are rebalanced monthly to include new earnings announcements. For each portfolio, we first calculate the average return in the cross section using value weights. We then estimate the regression coefficients relative to the Fama-French five-factor model.

[17] The results are quantitatively similar when using independently sorted portfolios.

We expect the outperformance of positive surprise firms compared to negative surprise firms to be stronger among more remote firms.

Table 11 reports the Fama-French five factor alphas of the calendar-time portfolios as well as spread portfolios of longing high (most positive) surprise quintiles and shorting low (most negative) surprise quintiles for each remoteness quintile. Table 11 Panel A uses CAR as the earnings surprise measure. From low remoteness to high remoteness, the low surprise quintile earns a monthly alpha of -0.442%, -0.323%, -0.369%, -0.229%, and 0.036%, respectively. The corresponding t-statistics are -2.21, -2.26, -2.49, -1.45, and 0.18, respectively. In comparison, from low remoteness to high remoteness, the high surprise quintile earns a monthly alpha of -0.179%, -0.137%, -0.368%, 0.188%, and 0.692%, respectively. The alpha is only statistically significant in the high remoteness quintile. The earnings surprise spread portfolio earns an alpha of 0.263%, 0.187%, 0.001%, 0.417%, and 0.657%, respectively, from low remoteness to high remoteness. The alphas are only statistically significant in the two highest remoteness quintiles. In Table 11 Panel B, we use SUE as the earnings surprise measure and document a similar pattern. The earnings surprise spread portfolio only generates a significant alpha of 0.766% in the highest remoteness quintile. That is, it takes longer to price the fundamental information for remote firms in absence of soft information.

6.2 Trading of Diligent Information Acquirers: Actively Managed Mutual Funds

A recent study by Ellis, Madureira, and Underwood (2020) documents that the introduction of direct flights between a mutual fund and a firm increases the mutual fund investment in the firm and the investment performance. This finding advocates diligent acquisition of soft information by mutual funds to assist their investment decisions. While acquiring information from remote firms is associated with higher costs, it is also more rewarding as the information is less likely to have been fully incorporated in the current price. Thus, we expect the mutual fund trading to predict future stock performance of the remotely located firm.

To test this hypothesis, we include all the actively managed U.S equity mutual funds in the intersection of Center for Research in Securities Prices (CRSP) mutual fund database and Thomson Reuters S12 mutual fund holdings file. We keep only domestic actively managed equity mutual funds that hold at least 10 stocks and have at least \$5 million assets. As mutual fund holdings are disclosed on a quarterly basis, we assume that any holding changes during the quarter occur at the end of the quarter and study the stock returns of the holdings in the following quarter. By doing so, we avoid any positive associations between mutual fund trading and stock performance caused by window dressing. However, we acknowledge that the within-quarter behaviors of the mutual funds, as documented by Kacperczyk, Sialm, and Zheng (2008), are inevitably overlooked due to the low data frequency.

First, do actively managed mutual funds take advantage of their diligent information acquisition? Do they more frequently trade on the remote firms where they intuitively have more information advantage? We answer these questions by calculating the aggregate mutual fund turnover for different remoteness groups. Each quarter, we calculate the aggregate mutual fund holding of each remoteness group and aggregate mutual fund buying and selling, all in dollar value. Then, each quarter q for each remoteness group j, mutual fund turnover of is defined as

$$Mutual\ \ Fund\ \ Turnover_{j,q} = \frac{Max(Dollar\ \ Buy_{j,q},\quad Dollar\ \ Sell_{j,q})}{Total\ \ Dollar\ \ Holding_{j,q-1}}. \tag{9}$$

The upper panel of Table 12 reports the time series average of aggregate mutual fund turnover in firms with different remoteness. From low remoteness to high remoteness quintiles, mutual fund turnover increases from 0.128 to 0.147 where the difference between these two numbers are 0.019 with a t-statistic of 4.39. It suggests that, as diligent acquirers of soft information, actively managed mutual funds more actively trade on remote firms.

We then turn to the return predictability of net mutual fund trading. Every calendar month in quarter q, we sort firms into quintiles based on Remote6h and further assign them into Net Buy and Net Sell groups based on aggregate mutual fund trading in quarter $q-1$. We expect mutual fund trading to be more informative of future stock return among remote firms. Again, due to the low frequency on trading data, any potential trading-return relation would indicate a long-lived information advantage due to mutual funds' diligent information acquisition.

The bottom panel of Table 12 reports the Fama-French five factor alphas of the trading-based calendar-time portfolios for each remoteness quintile. Net Buy portfolio only earns a statistically significant monthly alpha of 0.635% in the highest remoteness quintile. For Net Sell portfolios, all of alphas are statistically insignificant. The spread portfolios suggest that, when comparing the Net Buy and Net Sell portfolios, the former only significantly outperforms the latter in the highest remoteness quintile by 0.373% with a t-statistic of 1.86. Overall, these results are consistent with the actively managed mutual funds' diligent information acquisition playing an important role in investing in remote firms.

6.3 Does Geographic Dispersion Alleviate the Information Frictions?

The headquarter of a firm is the most important but not the only hub of its soft information. For firms with geographically dispersed operations, each local operation inherits at least part of the overall soft information of the firm and serves as the local representative of these information. Thus, firms with geographically dispersed operations can be effectively reached by larger population. Consistent with this notion, Garcia and Norli (2012) find that investors require a compensation for holding less geographically dispersed firms.

To investigate whether geographic dispersion alleviates information frictions, we closely follow Garcia and Norli (2012) to construct a proxy for geographic dispersion. For all our sample firms, we use a computerized algorithm to read all the annual reports (10-K fillings with SEC) and count the U.S. state names mentioned in the sections "Item 1: Business," "Item 2: Properties," "Item 6: Consolidated Financial Data," and " Item 7: Management's Discussion and Analysis." Firms that do not mention any U.S. state names are excluded from this analysis. We utilize the most recent state count by the end of June each year to construct geographic dispersion portfolios for the next 12 months. If geographic dispersion indeed alleviates information frictions and facilitates dissemination of soft information, we expect the return predictability of remoteness weakens among geographically dispersed firms. Symmatrically, we expect geographic dispersion to play a more critical role in stock pricing among firms with more remote headquarters.

We report the portfolios independently double-sorted on remoteness and geographic dispersion in Table 13. The upper panel reports the time series average of geographic remoteness across different remoteness quintiles. Unsurprisingly, proximately located firms have more geographically dispersed operations than remote firms as they are more likely to be national firms. On average, firms in the low remoteness quintile have about 3 more state counts than those in the high remoteness quintile. The bottom panel reports the Fama-French five-factor alphas of the portfolios. As expected, the remote firms significantly outperform proximate firms by 0.779%, 1.138%, and 0.641% in the bottom three dispersion quintiles. And the outperformance disappears in the two highest dispersion quintiles. This is consistent with our hypothesis that geographic dispersion alleviates information frictions. Along the remoteness dimension, truly local firms only earn significantly higher returns than dispersed firms in the two highest remoteness quintiles, suggesting that geographic dispersion plays a more critical role in investor recognition and thus stock pricing among firms with remote headquarters.

7. Conclusion

In this paper, we construct remoteness measures that characterize the geographic features of the firm including the surrounding populations and distances from these population centers. We use these measures and exogenous shocks to firms' accessibility to establish causal impact of remoteness on the flow of soft information. We also hypothesize that the market requires a positive premium for holding stocks with more remote headquarters due to the information frictions. We find that remote stocks significantly outperform proximate stocks. This outperformance is from both the long and short legs and highly persistent due to the nature of remoteness. Several of our results also hightlight the important role of flight travel in mitigating information frictions. Finally, we further discuss the stock price implications of remoteness in the context of information frictions in a few interesting settings. Our results add the missing piece between the literature on the role of firm headquarter geographic locations and the literature on the asset pricing implications of information frictions.

References

Alam, Z. S., Chen, M. A., Ciccotello, C. S., & Ryan, H. E. (2014). Does the location of directors matter? Information acquisition and board decisions. *Journal of Financial and Quantitative Analysis*, 49(1), 131-164.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31-56.

Anand, A., Gatchev, V. A., Madureira, L., Pirinsky, C. A., & Underwood, S. (2011). Geographic proximity and price discovery: Evidence from NASDAQ. *Journal of Financial Markets*, 14(2), 193-226.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, 69(1), 106-116.

Anderson, J. E. (2011). The Gravity Model. *Annual Review of Economics*, 2011, 3:133–60.

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1), 170-192.

Anderson, J. E. (2016). The gravity model of economic interaction. *Boston College and NBER*, 3, 391-7.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1), 259-299.

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: delayed price response or risk premium? *Journal of Accounting Research*, 27, 1-36.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57-82.

Chen, D., Ma, Y., Martin, X., & Michaely, R. (2022). On the fast track: Information acquisition costs and information production. *Journal of Financial Economics*, 143(2), 794-823.

Chen, H., Gompers, P., Kovner, A., & Lerner, J. (2010). Buy local? The geography of venture capital. *Journal of Urban Economics*, 67(1), 90-102.

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6), 2045-2073.

Coval, J. D., & Moskowitz, T. J. (2001). The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, 109(4), 811-841.

Da, Z., Gurun, U. G., Li, B., & Warachka, M. (2021). Investment in a smaller world: The implications of air travel for investors and firms. *Management Science*, 67(1), 417-435.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *The Journal of Finance*, 64(2), 709-749.

Easley, D., Hvidkjaer, S., & O'hara, M. (2002). Is information risk a determinant of asset returns? *The Journal of Finance*, 57(5), 2185-2221.

Easley, D., & O'hara, M. (2004). Information and the cost of capital. *The Journal of Finance*, 59(4), 1553-1583.

El Ghoul, S., Guedhami, O., Ni, Y., Pittman, J., & Saadi, S. (2013). Does information asymmetry matter to equity pricing? Evidence from firms' geographic location. *Contemporary Accounting Research*, 30(1), 140-181.

Ellis, J., Madureira, L., & Underwood, S. (2020). The causal effects of proximity on investment: Evidence from flight introductions. *Journal of Financial and Quantitative Analysis*, 55(6), 1978-2004.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3), 283-306.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5), 2023-2052.

Feng, L., & Seasholes, M. S. (2004). Correlated trading and location. *The Journal of Finance,* 59(5), 2117-2144.

Gao, M., & Huang, J. (2020). Informing the market: The effect of modern information technologies on information production. *The Review of Financial Studies*, 33(4), 1367-1411.

Garcia, D., & Norli, Ø. (2012). Geographic dispersion and stock returns. *Journal of Financial Economics*, 106(3), 547-565.

Giroud, X. (2013). Proximity and investment: Evidence from plant-level data. *The Quarterly Journal of Economics*, 128(2), 861-915.

Guest, N. M. (2021). The information role of the media in earnings news. *Journal of Accounting Research*, *59*(3), 1021-1076.

Hau, H. (2001). Location matters: An examination of trading profits. *The Journal of Finance*, 56(5), 1959-1983.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, 64(5), 2289-2325.

Hollander, S., & Verriest, A. (2016). Bridging the gap: the design of bank loan contracts and distance. *Journal of Financial Economics*, 119(2), 399-419.

Hou, K., & Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *The Review of Financial Studies*, 18(3), 981-1020.

Hung, M., Li, X., & Wang, S. (2015). Post-earnings-announcement drift in global markets: Evidence from an information shock. *The Review of Financial Studies*, 28(4), 1242-1283.

Ivković, Z., & Weisbenner, S. (2005). Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*, 60(1), 267-306.

John, K., Knyazeva, A., & Knyazeva, D. (2011). Does geography matter? Firm location and corporate payout policy. *Journal of Financial Economics*, 101(3), 533-551.

Kacperczyk, M., Sialm, C., & Zheng, L. (2008). Unobserved actions of mutual funds. *The Review of Financial Studies*, *21*(6), 2379-2416.

Kang, J. K., & Kim, J. M. (2008). The geography of block acquisitions. *The Journal of Finance*, 63(6), 2817-2858.

Korniotis, G. M., & Kumar, A. (2013). State-level business cycles and local return predictability. *The Journal of Finance*, *68*(3), 1037-1096.

Li, F. W., & Sun, C. (2022). Information acquisition and expected returns: Evidence from EDGAR search traffic. *Journal of Economic Dynamics and Control*, 104384.

Loughran, T., & Schultz, P. (2005). Liquidity: Urban versus rural firms. *Journal of Financial Economics*, 78(2), 341-374.

Malloy, C. J. (2005). The geography of equity analysis. *The Journal of Finance*, 60(2), 719-755.

Merton, R.C., 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42, 483–511.

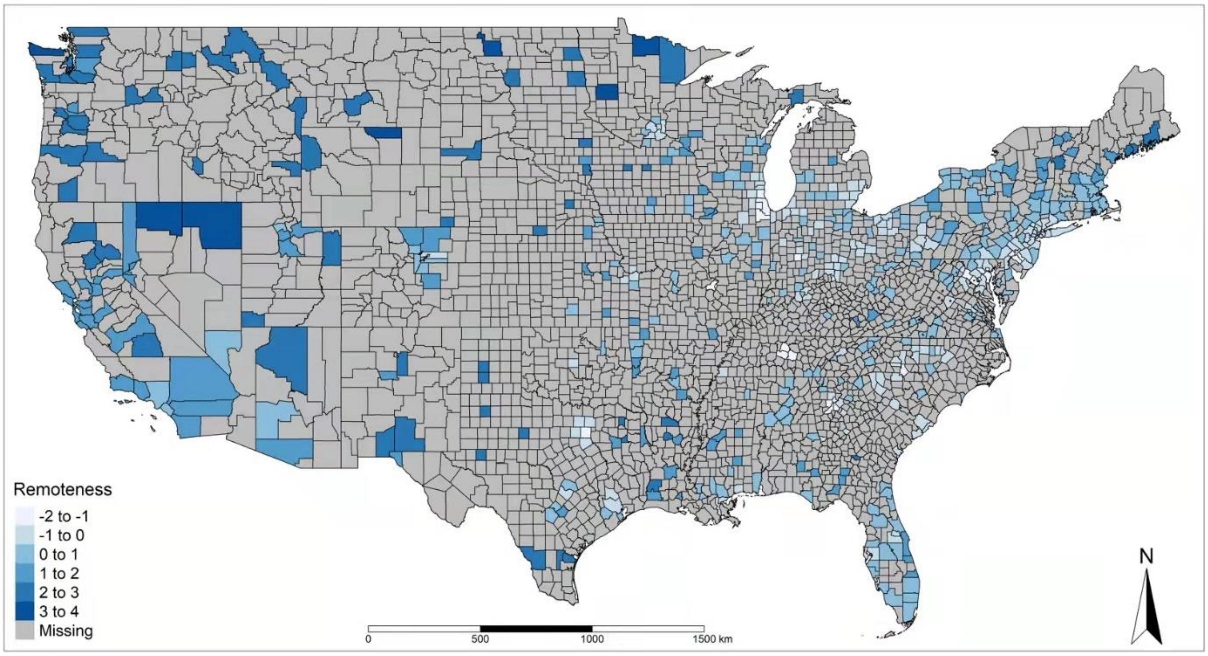
Richardson, S., Tuna, I., & Wysocki, P. (2010). Accounting anomalies and fundamental analysis: A review of recent research advances. *Journal of Accounting and Economics*, 50(2-3), 410-454.

Ryans, J. (2017). Using the EDGAR log file data set. *Working Paper*.

Figure 1. Remoteness on the map

This figure shows the average remoteness across all firms in each U.S. county on the map for year 2020. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. The darker blue indicates a higher value of average remoteness. Counties with no publicly listed firms in our sample are coded grey. Panel A and B use Remote6h and Remote600 as the remoteness measures, respectively.

Panel A. Average Remote6h by county



Panel B. Average Remote600 by county

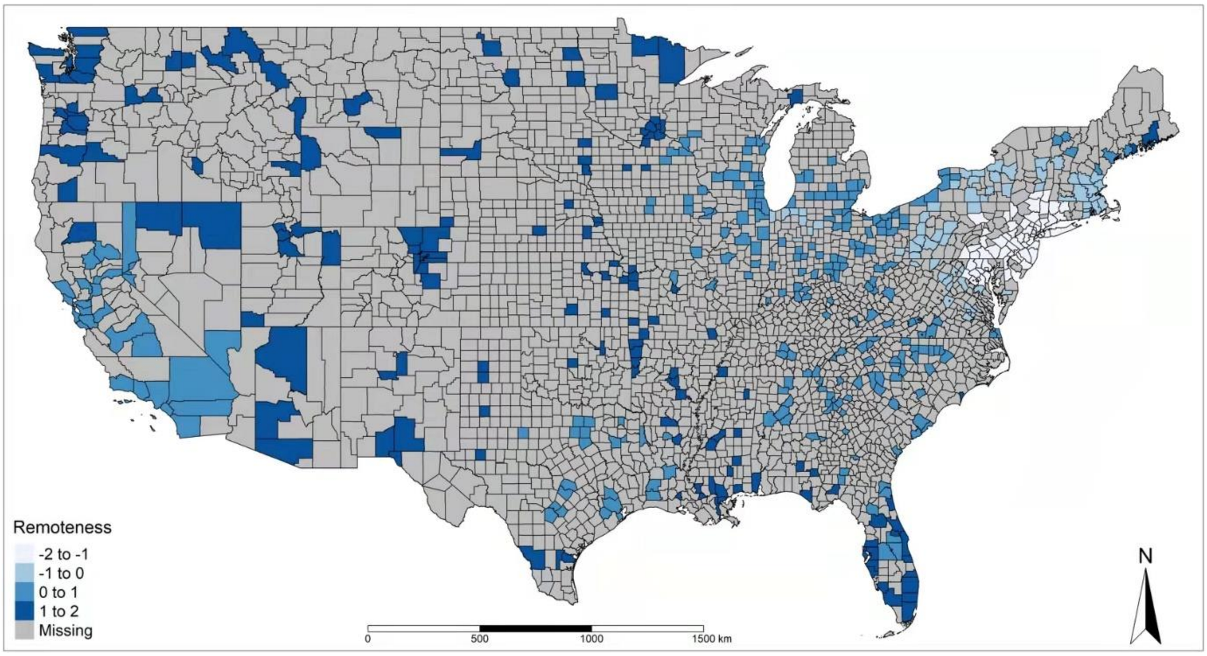


Figure 2. Flight delay rate

This figure plots the average DelayRate by calendar month. We sum up the counts of significant delays and cancellations where the airport is either the origin or the destination. And then we add to it the count of diversions where the airport is the destination. We scale this number by the total number of scheduled flights to and from this airport to calculate the monthly DelayRate for each airport. We define a flight as significantly delayed if the delay is over 120 minutes. For each calendar month, we first average across all the firms each year and then average across all the years.

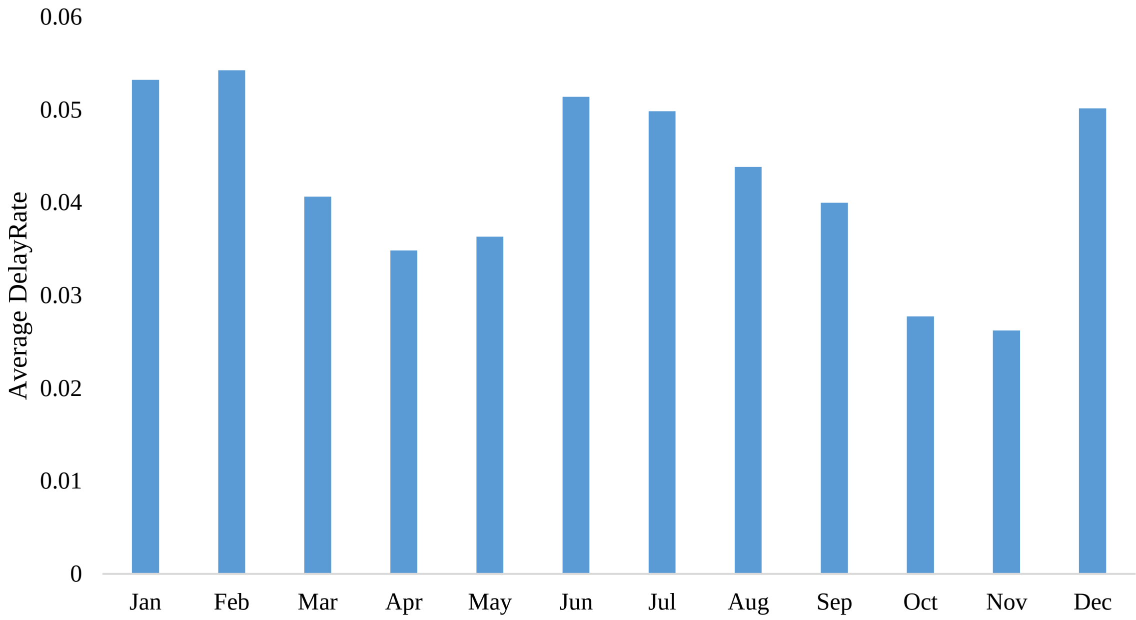


Table 1. Accessibility over time

This table reports the average accessibility across all sample firms by calendar year. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. Access6h and Access10h are defined in equation (1). Access600 and Access1000 are defined in equation (3).

Year	Access6h	Access10h	Access600	Access1000
1995	56.746	125.904	20.456	36.776
1996	56.063	126.635	20.569	36.869
1997	56.619	128.208	20.821	37.283
1998	57.512	130.090	20.987	37.649
1999	57.719	131.503	21.270	38.084
2000	58.162	132.255	21.483	38.386
2001	58.214	133.713	21.504	38.395
2002	60.742	135.307	21.623	38.587
2003	63.602	140.010	21.659	38.770
2004	64.474	141.099	21.634	38.725
2005	64.939	142.849	22.338	40.009
2006	65.306	143.488	22.616	40.511
2007	66.800	146.182	22.922	41.005
2008	67.315	146.993	23.011	41.172
2009	68.335	148.345	23.028	41.447
2010	69.853	150.365	23.146	41.902
2011	70.058	151.316	23.267	42.090
2012	71.357	152.919	23.376	42.291
2013	71.305	153.740	23.488	42.475
2014	72.216	155.282	23.600	42.616
2015	72.537	156.673	23.675	42.577
2016	74.029	158.379	23.867	42.858
2017	75.229	160.341	24.006	43.103
2018	74.812	160.487	24.080	43.078
2019	74.900	161.472	24.161	43.185
2020	75.829	161.441	24.552	43.726

Table 2. Remoteness by industry

This table reports the average remoteness measures by Fama-French 12 industries. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. Remoteness measures are defined in equation (2) and (4). For each industry, we first average the remoteness measures across all the firms each year. Then we average across all the years.

	Remote6h	Remote10h	Remote600	Remote1000
Consumer Nondurables	-0.311	-0.307	-0.150	-0.182
Consumer Durables	-0.176	-0.234	-0.088	-0.282
Manufacturing	-0.200	-0.212	-0.033	-0.161
Energy	-0.051	-0.310	0.637	0.698
Chemicals	-0.401	-0.386	-0.258	-0.289
Business Equipment	0.183	0.106	0.086	0.165
Telecom	-0.292	-0.339	-0.189	-0.103
Utilities	-0.039	0.008	0.060	-0.019
Shops	-0.221	-0.252	0.054	0.020
Healthcare	0.086	0.047	-0.122	-0.008
Financial	-0.090	-0.040	-0.192	-0.296
Other	-0.084	-0.160	0.060	0.087

Table 3. Firm characteristics by remoteness

This table reports the firm characteristics of firms with different remoteness rankings. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. Remoteness measures are defined in equation (2) and (4). Market beta is estimated based on CAPM using daily returns of the last 3 months. Mcap is the firm's market capitalization. BM is the firm's book-to-market ratio where book value of equity is calculated as the Compustat book value of equity adding back deferred taxes and subtracting preferred equity. Lagret is the previous month return. Past 12 is the return of the previous twelve months with one month lag. Volatility is the standard deviation of the daily returns of the previous month. ILLIQ is the Amihud (2002) illiquidity measure. Every month, we assign firms into decile portfolios based on the remoteness measure, Remote6h, of the previous calendar year. For each decile portfolio, we first average the firm characteristics across all the firms each month. Then we average across all the months. In the last column, we report the difference between the high remoteness and low remoteness deciles. *, **, or *** indicates significance level of 1%, 5%, or 10%, respectively.

	By Remote6h										
	Low	2	3	4	5	6	7	8	9	High	H - L
Remote6h	-1.447	-1.101	-0.892	-0.615	-0.337	-0.044	0.281	0.751	1.101	1.592	3.038***
Remote10h	-1.254	-0.883	-0.782	-0.542	-0.377	-0.093	0.105	0.403	0.802	1.492	2.746***
Remote600	-0.307	-1.028	-0.518	-0.524	-0.130	-0.079	0.356	0.455	0.587	0.967	1.274***
Remote1000	-0.699	-0.922	-0.678	-0.520	-0.173	-0.105	0.343	0.595	0.782	1.114	1.813***
Market beta	0.958	0.980	0.946	0.947	0.931	0.963	0.931	0.901	1.200	0.952	-0.006
Mcap	6339	6833	4730	3906	3220	3200	2410	1824	5277	5303	-1036***
BM	0.791	0.921	0.797	0.759	0.781	0.804	0.839	0.824	0.585	0.716	-0.075***
Lagret	1.484	1.418	1.384	1.325	1.383	1.506	1.459	1.488	1.990	1.651	0.167
Past12	14.490	16.088	15.330	14.339	15.756	14.693	17.396	16.637	20.210	17.318	2.828***
Vol	0.448	0.482	0.480	0.491	0.497	0.503	0.498	0.515	0.580	0.513	0.065***
ILLIQ	0.011	0.025	0.022	0.024	0.024	0.017	0.022	0.028	0.012	0.017	0.006***

Table 4. Panel regressions of news softness on remoteness

This table reports panel regressions of news softness on remoteness measures. The sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 2000 to 2020. SoftRate is defined as the proportion of soft news in all the news with non-missing RavenPack group identifier in the month in percentage. The remoteness measures are defined in equation (2) and (4). LagNews is log of the one plus the total number of news in the previous month. LogSize is the log of the market capitalization from the previous month. BM is the log of the book-to-market ratio from the previous month. Lagret the previous month return. Past12 is the past 12-month return (skipping the most recent month). Vol is the volatility from the previous-month daily returns. ILLIQ is the Amihud (2002) illiquidity measure. We include the industry (3-digit SIC) fixed effect and year-month fixed effect in all columns. t-statistics calculated based on standard errors clustered at the firm level are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

	(1)	(2)	(3)	(4)
VARIABLES	SoftRate	SoftRate	SoftRate	SoftRate
Remote6h	-0.221***			
	[-3.83]			
Remote10h		-0.210***		
		[-3.39]		
Remote600			-0.229***	
			[-4.14]	
Remote1000				-0.160***
				[-2.92]
LagNews	1.386***	1.385***	1.383***	1.384***
	[43.80]	[43.78]	[43.79]	[43.77]
LogSize	-0.112***	-0.111***	-0.105***	-0.104***
	[-3.15]	[-3.12]	[-2.97]	[-2.94]
BM	0.245***	0.246***	0.247***	0.246***
	[6.13]	[6.15]	[6.17]	[6.15]
Lagret	-0.034***	-0.034***	-0.034***	-0.034***
	[-18.88]	[-18.87]	[-18.92]	[-18.92]
Past 12	-0.007***	-0.007***	-0.007***	-0.007***
	[-10.49]	[-10.50]	[-10.53]	[-10.52]
Std	2.062***	2.058***	2.066***	2.069***
	[18.19]	[18.15]	[18.24]	[18.25]
ILLIQ	-0.178	-0.174	-0.187	-0.173
	[-0.73]	[-0.72]	[-0.77]	[-0.71]
Constant	5.327***	5.285***	5.219***	5.201***
	[7.46]	[7.41]	[7.32]	[7.28]
Observations	722,652	722,652	722,652	722,652
R-squared	0.033	0.033	0.033	0.033
Industry FE	Yes	Yes	Yes	Yes
YearMonth FE	Yes	Yes	Yes	Yes

Table 5. Difference-in-differences regressions

This table reports panel regressions of news softness on the AbnormalDelay indicator. The sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 2000 to 2020. We sum up the counts of significant delays and cancellations where the airport is either the origin or the destination. And then we add to it the count of diversions where the airport is the destination. We scale this number by the total number of scheduled flights to and from this airport to calculate the monthly DelayRate for each airport. We define a flight as significantly delayed if the delay is over 120 minutes. For each firm's most commonly used airport, AbnormalDelay is equal to 1 if the current month DelayRate is at least 50% higher than the historical average and 0 otherwise. SoftRate is defined as the proportion of soft news in all the news with non-missing RavenPack group identifier in the month in percentage. LagNews is log of the one plus the total number of news in the previous month. LogSize is the log of the market capitalization from the previous month. BM is the log of the book-to-market ratio from the previous month. Lagret the previous month return. Past12 is the past 12-month return (skipping the most recent month). Vol is the volatility from the previous-month daily returns. ILLIQ is the Amihud (2002) illiquidity measure. We include the firm fixed effect and year-month fixed effect in all columns. t-statistics calculated based on standard errors clustered at the firm level are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

	(1)	(2)
VARIABLES	SoftRate	SoftRate
AbnormalDelay	-0.260***	-0.238***
	[-3.48]	[-3.19]
LagNews		1.190***
		[41.96]
LogSize		-0.702***
		[-11.22]
BM		0.368***
		[8.41]
Lagret		-0.023***
		[-13.40]
Past 12		-0.004***
		[-10.21]
Std		1.053***
		[10.85]
ILLIQ		-0.153
		[-0.94]
Constant	7.248***	18.388***
	[566.43]	[14.67]
Observations	716,850	710,531
R-squared	0.073	0.079
Firm FE	Yes	Yes
YearMonth FE	Yes	Yes

Table 6. Market reactions to soft news releases

This table reports panel regressions of news softness on remoteness measures. The sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 2000 to 2020. We calculate the market reactions as the compounded market-adjusted returns in a short three-day window (CAR[-1,1]) for Panel A and in a long twenty-one-day window (CAR[-10,10]) for Panel B. CSS is the composite sentiment score from RavenPack and gauges the textual sentiment of the news. The remoteness measures are defined in equation (2) and (4). LagNews is log of the one plus the total number of news in the previous month. We also include the following control variables. LogSize is the log of the market capitalization from the previous month. BM is the log of the book-to-market ratio from the previous month. Lagret the previous month return. Past12 is the past 12-month return (skipping the most recent month). Vol is the volatility from the previous-month daily returns. ILLIQ is the Amihud (2002) illiquidity measure. We include the industry (3-digit SIC) fixed effect and year-month fixed effect in all columns. t-statistics calculated based on standard errors clustered at the firm level are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

Panel A. Short-window market reaction

	(1)	(2)	(3)	(4)
VARIABLES	CAR[-1,1]	CAR[-1,1]	CAR[-1,1]	CAR[-1,1]
Remote =	Remote6h	Remote10h	Remote600	Remote1000
CSS×Remote	2.324***	2.666***	1.182***	1.603***
	[4.75]	[5.05]	[2.91]	[3.44]
CSS	4.556***	4.596***	4.378***	4.356***
	[9.57]	[9.63]	[9.55]	[9.54]
Remote	-0.054	-0.041	-0.072*	-0.068
	[-0.94]	[-0.70]	[-1.73]	[-1.49]
Constant	0.645	0.626	0.663	0.641
	[0.92]	[0.88]	[0.93]	[0.89]
Observations	122,529	122,529	122,529	122,529
R-squared	0.011	0.011	0.011	0.011
Firm Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
YearMonth FE	Yes	Yes	Yes	Yes

Panel B. Long-window market reaction

VARIABLES	(1) CAR[-10,10]	(2) CAR[-10,10]	(3) CAR[-10,10]	(4) CAR[-10,10]
Remote =	Remote6h	Remote10h	Remote600	Remote1000
CSS*Remote	2.673***	3.100***	1.082	1.608*
	[2.85]	[2.87]	[1.38]	[1.85]
CSS	7.356***	7.415***	7.119***	7.106***
	[7.92]	[7.87]	[7.83]	[7.91]
Remote	0.107	0.101	0.036	0.081
	[1.24]	[1.08]	[0.47]	[1.03]
Constant	1.387	1.414	1.463	1.459
	[1.18]	[1.20]	[1.23]	[1.23]
Observations	122,529	122,529	122,529	122,529
R-squared	0.043	0.043	0.043	0.043
Firm Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
YearMonth FE	Yes	Yes	Yes	Yes

Table 7. Portfolio returns by remoteness

This table reports the monthly returns of the portfolios sorted on remoteness. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. We report both the raw returns as well as the time-series alphas relative to the Fama and French (2015) five-factor model. We sort stocks into decile portfolios every month based on remoteness of the previous calendar year. In Panel A, the sorting variable is Remote6h. In Panel B, we use alternative sorting variables Remote10h, Remote600, and Remote1000. Remoteness measures are defined in equation (2) and (4). For each portfolio, we first calculate the average return in the cross section using value weights. We then report the time-series average returns and the regression coefficients relative to the five-factor model. The returns are in percentage. t-statistics are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

Panel A. Sorted by Remote6h

	Low	2	3	4	5	6	7	8	9	High	H - L
Return	0.700***	0.689***	0.862***	0.778***	0.764***	0.592	0.830**	0.867***	1.154***	1.384***	0.684***
	[2.97]	[2.45]	[3.55]	[2.89]	[2.83]	[1.57]	[2.54]	[2.93]	[2.90]	[4.08]	[2.79]
Alpha	-0.192**	-0.181*	-0.059	-0.177	-0.137	-0.237	-0.052	0.042	0.416***	0.496***	0.689***
	[-2.45]	[-1.82]	[-0.67]	[-1.43]	[-1.25]	[-1.08]	[-0.39]	[0.40]	[2.65]	[3.34]	[3.54]
MKT	0.892***	1.004***	0.930***	0.937***	0.964***	0.925***	1.002***	0.948***	1.110***	1.056***	0.164***
	[35.84]	[39.90]	[45.18]	[30.29]	[31.31]	[17.67]	[29.42]	[31.86]	[27.43]	[27.10]	[2.93]
SMB	-0.019	-0.202***	-0.047	-0.025	0.027	0.086	0.308***	0.216***	-0.003	0.043	0.063
	[-0.44]	[-4.18]	[-1.24]	[-0.42]	[0.62]	[0.71]	[4.57]	[5.26]	[-0.05]	[0.65]	[0.65]
HML	0.195***	0.239***	0.124***	0.057	0.085	0.249***	-0.118*	-0.027	-0.247***	-0.265***	-0.460***
	[4.60]	[5.91]	[3.19]	[0.98]	[1.46]	[2.87]	[-1.66]	[-0.61]	[-3.31]	[-4.27]	[-5.07]
CMA	0.184***	0.009	0.233***	0.178***	-0.040	-0.290**	-0.109	-0.113	-0.279***	-0.428***	-0.612***
	[3.27]	[0.18]	[3.82]	[2.66]	[-0.50]	[-2.51]	[-1.05]	[-1.31]	[-2.46]	[-4.02]	[-4.35]
RMW	0.155***	0.024	0.151***	0.260***	0.116*	0.073	-0.140	-0.149**	-0.580***	0.070	-0.085
	[3.65]	[0.59]	[3.16]	[6.17]	[1.86]	[1.14]	[-1.36]	[-2.29]	[-6.58]	[0.85]	[-0.80]

Panel B. Sorted by other remoteness measures

		Low	2	3	4	5	6	7	8	9	High	H - L
By Remote10h	Return	0.745***	0.685**	0.761***	0.770***	0.673*	0.736***	0.733**	1.021***	1.136***	1.350***	0.605**
		[3.43]	[2.42]	[2.70]	[2.84]	[1.89]	[2.73]	[2.46]	[3.48]	[2.89]	[3.83]	[2.26]
	Alpha	-0.181**	-0.189*	-0.107	-0.125	-0.202	-0.191*	-0.138	0.193*	0.353**	0.520***	0.700***
		[-2.37]	[-1.66]	[-1.43]	[-1.12]	[-0.98]	[-1.91]	[-1.33]	[1.91]	[2.41]	[3.31]	[3.54]
By Remote600	Return	0.682**	0.827***	0.739***	0.893***	0.951***	0.568	0.589	1.263***	0.794***	1.157***	0.475**
		[2.40]	[3.04]	[2.87]	[3.76]	[3.30]	[1.61]	[1.55]	[4.21]	[3.27]	[3.66]	[2.32]
	Alpha	-0.147	-0.122	-0.128	-0.012	0.035	-0.264	-0.267	0.340***	-0.086	0.366***	0.513***
		[-1.32]	[-1.28]	[-1.60]	[-0.13]	[0.45]	[-1.25]	[-1.43]	[2.75]	[-0.95]	[2.75]	[2.97]
By Remote1000	Return	0.688**	0.747***	0.804***	0.791***	0.885***	0.736***	0.509	1.172***	1.031***	1.224***	0.535**
		[2.38]	[2.89]	[3.17]	[2.98]	[3.14]	[3.01]	[1.42]	[3.08]	[3.60]	[3.56]	[2.19]
	Alpha	-0.278**	-0.129	-0.095	-0.099	0.032	-0.153	-0.325	0.309**	0.177	0.401**	0.680***
		[-2.50]	[-1.47]	[-1.02]	[-1.20]	[0.28]	[-1.52]	[-1.50]	[2.01]	[1.49]	[2.47]	[3.13]

Table 8. Fama-MacBeth regressions

This table reports Fama-MacBeth regressions of monthly return on remoteness measures. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. The remoteness measures are defined in equation (2) and (4). LogSize is the log of the market capitalization from the previous month. BM is the log of the book-to-market ratio from the previous month. Lagret the previous month return. Past12 is the past 12-month return (skipping the most recent month). Vol is the volatility from the previous-month daily returns. ILLIQ is the Amihud (2002) illiquidity measure. The returns are in percentage. t-statistics calculated based on Newey-West standard errors are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

	(1)	(2)	(3)	(4)
VARIABLES	Monthly Return	Monthly Return	Monthly Return	Monthly Return
Remote6h	0.096***			
	[2.93]			
Remote10h		0.091**		
		[2.48]		
Remote600			0.047	
			[1.57]	
Remote1000				0.067**
				[2.10]
LogSize	-0.059	-0.058	-0.070*	-0.069*
	[-1.46]	[-1.43]	[-1.76]	[-1.75]
BM	0.064	0.063	0.059	0.061
	[1.37]	[1.36]	[1.26]	[1.30]
Lagret	-0.011**	-0.011**	-0.011**	-0.010*
	[-2.00]	[-2.06]	[-1.97]	[-1.83]
Past12	0.002	0.002	0.002	0.002
	[0.93]	[0.93]	[1.00]	[0.97]
Vol	-1.175***	-1.165***	-1.185***	-1.196***
	[-4.00]	[-3.98]	[-4.00]	[-4.05]
ILLIQ	2.025***	2.035***	1.883***	1.937***
	[2.90]	[2.92]	[2.63]	[2.73]
Constant	2.500***	2.492***	2.700***	2.690***
	[3.01]	[2.97]	[3.34]	[3.32]
Adj. Rsq	0.042	0.042	0.042	0.042

Table 9. Persistence of return predictability

This table reports the monthly alphas of the portfolios sorted on remoteness measure, Remote6h, with different lags. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. Remote6h is defined in equation (2). We sort stocks into decile portfolios every month based on Remote6h lagged by 1 to 5 whole years to examine its return predictability for up to 5 years in the future. For each portfolio, we first calculate the average return in the cross section using value weights. We then estimate the portfolio alphas relative to the five-factor model. The alphas are in percentage. t-statistics are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

Sort on	Low	2	3	4	5	6	7	8	9	High	H - L
1-year lagged	-0.186**	-0.057	-0.150*	-0.036	-0.312	-0.019	-0.023	0.137	0.358**	0.434***	0.620***
	[-2.36]	[-0.70]	[-1.70]	[-0.38]	[-1.44]	[-0.16]	[-0.17]	[1.26]	[2.17]	[2.87]	[3.23]
2-year lagged	-0.193**	-0.096	-0.148	-0.061	-0.250	-0.289**	0.097	0.115	0.348**	0.562***	0.756***
	[-2.25]	[-1.08]	[-1.21]	[-0.63]	[-1.08]	[-2.40]	[0.65]	[1.06]	[2.03]	[3.38]	[3.52]
3-year lagged	-0.241**	-0.083	-0.035	-0.057	-0.269	-0.173	0.092	0.261*	0.384**	0.364**	0.605***
	[-2.54]	[-1.01]	[-0.35]	[-0.58]	[-1.22]	[-1.36]	[0.65]	[1.93]	[2.22]	[2.14]	[2.70]
4-year lagged	-0.182**	-0.180**	-0.026	0.066	-0.415	0.053	0.016	0.364**	0.325*	0.376**	0.558***
	[-1.99]	[-2.18]	[-0.25]	[0.59]	[-1.75]	[0.44]	[0.13]	[2.51]	[1.92]	[2.22]	[2.58]
5-year lagged	-0.109	-0.147*	-0.041	-0.008	-0.289	0.004	-0.054	0.216**	0.155	0.423***	0.532***
	[-1.29]	[-1.81]	[-0.43]	[-0.09]	[-1.10]	[0.04]	[-0.51]	[2.02]	[0.96]	[2.66]	[2.70]

Table 10. Sequentially-sorted portfolios

This table reports the monthly alphas of the high-minus-low remoteness spread portfolios based on Remote6h across different firm characteristics. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. Remote6h is defined in equation (2). Each month, we sequentially double-sort portfolios on a firm characteristic and Remote6h in that order. For each portfolio, we first calculate the average return in the cross section using value weights. We then estimate the portfolio alphas relative to the Fama-French five-factor model. For brevity, we limit our reporting to the high-minus-low remoteness spread portfolios within each characteristic quintile. The alphas are in percentage. t-statistics are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

Sorted on	Low	2	3	4	High
Size	0.077	0.241*	0.365***	0.394***	0.658***
	[0.46]	[1.78]	[2.70]	[3.45]	[3.74]
Market beta	0.598**	0.631***	0.388*	0.468**	0.763***
	[2.29]	[2.81]	[1.90]	[1.99]	[2.72]
BM	0.851***	0.166	0.547***	-0.049	0.099
	[3.82]	[0.88]	[2.68]	[-0.23]	[0.31]
Lagret	0.589*	0.721***	0.416*	0.516**	0.587**
	[1.76]	[3.27]	[1.84]	[2.25]	[2.00]
Past12	0.395	0.571**	0.477**	0.383**	0.513**
	[1.24]	[2.17]	[2.41]	[2.11]	[1.97]
IVOL	0.709***	0.391*	0.538**	0.925***	0.865*
	[3.44]	[1.85]	[1.97]	[2.59]	[1.94]
Turnover	0.313*	0.596***	0.097	0.611***	0.806***
	[1.67]	[2.79]	[0.42]	[3.12]	[3.04]
ILLIQ	0.699***	0.423***	0.269**	0.263	0.506***
	[3.85]	[3.64]	[2.09]	[1.50]	[2.70]

Table 11. Post-earnings announcement drift

This table reports the PEAD effect across remoteness quintiles using a calendar-time portfolio approach. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. Remote6h is defined in equation (2). To account for the stock performance in the 6 months after the earnings announcement, we line up the event months for all the earnings announcements by calendar month. Every calendar month, we sort firms into quintiles based on Remote6h and earnings surprise in that order. For earning surprise, we use the cumulative abnormal return in the 3-day window around the earnings announcement and analyst forecast based earnings surprise in Panel A and B, respectively. For each portfolio, we first calculate the average return in the cross section using value weights. We then estimate the portfolio alphas relative to the Fama-French five-factor model. The alphas are in percentage. t-statistics are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

Panel A. Market reaction based earning surprise

		By Remote6h				
		Low	2	3	4	High
By CAR	Low	-0.442**	-0.323**	-0.369**	-0.229	0.036
		[-2.21]	[-2.26]	[-2.49]	[-1.45]	[0.18]
	2	-0.314***	-0.069	-0.321	0.148	0.258
		[-2.61]	[-0.54]	[-1.45]	[0.95]	[1.44]
	3	-0.071	-0.191**	0.082	-0.110	0.464***
		[-0.69]	[-2.19]	[0.79]	[-0.87]	[2.77]
	4	-0.107	-0.158	-0.018	0.037	0.367**
		[-1.34]	[-1.10]	[-0.16]	[0.31]	[2.22]
	High	-0.179	-0.137	-0.368	0.188	0.692***
		[-1.45]	[-1.10]	[-1.57]	[1.12]	[3.45]
	H - L	0.263	0.187	0.001	0.417*	0.657***
		[1.20]	[1.11]	[1.11]	[1.95]	[2.93]

Panel B. Analyst forecast based earnings surprise

		By Remote6h				
		Low	2	3	4	High
By SUE	Low	-0.314*	-0.190	0.239	-0.072	-0.026
		[-1.88]	[-1.00]	[0.99]	[-0.42]	[-0.11]
	2	-0.140	-0.254**	-0.220*	-0.079	0.301*
		[-1.39]	[-2.06]	[-1.80]	[-0.53]	[1.83]
	3	-0.097	0.028	-0.139	0.041	0.511***
		[-1.10]	[0.28]	[-1.34]	[0.32]	[3.01]
	4	0.008	-0.134	-0.026	0.128	0.510***
		[0.09]	[-1.37]	[-0.20]	[0.77]	[2.67]
	High	-0.041	-0.072	-0.075	0.205	0.740***
		[-0.33]	[-0.51]	[-0.42]	[1.14]	[3.31]
	H - L	0.273	0.118	-0.314	0.277	0.766***
		[1.33]	[0.54]	[-1.24]	[1.34]	[2.91]

Table 12. Trading of diligent information acquirers: actively managed mutual funds

This table reports the trading and the return predictability of trading by actively managed mutual funds among remoteness groups using a calendar-time portfolio approach. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. Remote6h is defined in equation (2). The upper panel reports the time series average of aggregate mutual fund turnover in firms with different remoteness. Each quarter, we calculate the aggregate mutual fund holding of each remoteness group and aggregate mutual fund buying and selling, all in dollar value. Then mutual fund turnover of is defined in equation (8). The bottom panel reports the alphas of mutual fund trading portfolios. Net Buy (Sell) indicates net buying (selling) by actively managed mutual funds in the calendar quarter q-1. Every calendar month in quarter q, we sort firms into quintiles based on Remote6h and further assign them into Net Buy or Net Sell group based on aggregate mutual fund trading in quarter q-1. We also report the performance difference between the two mutual fund trading portfolios for each remoteness quintile. For each portfolio, we first calculate the average return in the cross section using value weights. We then estimate the portfolio alphas relative to the Fama-French five-factor model. The alphas are in percentage. t-statistics are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

	By Remote6h					
	Low	2	3	4	High	H - L
Mutual Fund Turnover	0.128	0.132	0.136	0.147	0.147	0.019***
						[4.39]
	FF5 alpha					
Net Buy	-0.148	0.003	-0.069	0.047	0.635***	0.783***
	[-1.32]	[0.02]	[-0.58]	[0.37]	[4.21]	[4.27]
Net Sell	-0.136	-0.142	-0.016	0.011	0.262	0.398**
	[-1.53]	[-1.31]	[-0.15]	[0.08]	[1.60]	[2.06]
Buy - Sell	-0.012	0.145	-0.053	0.036	0.373*	0.384*
	[-0.12]	[0.77]	[-0.39]	[0.19]	[1.86]	[1.82]

Table 13. Remoteness and geographic dispersion of operations

This table reports the portfolios independently sorted on Remote6h and Garcia and Norli (2012) geographic dispersion measure. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. Remote6h is defined in equation (2). Geographic dispersion is the state name counts from SEC Form 10-K following Garcia and Norli (2012). The upper panel reports the time series average of geographic remoteness across different remoteness quintiles. The bottom panel reports the alphas of the portfolios. Every month, we sort firms independently into quintiles based on Remote6h and geographic dispersion. For each portfolio, we first calculate the average return in the cross section using value weights. We then estimate the portfolio alphas relative to the Fama-French five-factor model. The alphas are in percentage. t-statistics are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

		By Remote6h					
		Low	2	3	4	High	H - L
Geographic Dispersion		10.922	10.200	9.260	9.210	7.657	-3.265***
							[71.83]
		FF5 alpha					
By Geographic Dispersion	Low	-0.048	0.001	-0.010	0.305	0.731***	0.779**
		[-0.33]	[0.01]	[-0.05]	[1.35]	[3.12]	[2.58]
	2	0.120	-0.032	-0.073	-0.163	1.258***	1.138***
		[0.40]	[-0.14]	[-0.36]	[-0.66]	[4.35]	[2.61]
	3	-0.168	0.066	-0.208	0.232	0.473*	0.641**
		[-1.08]	[0.43]	[-1.11]	[1.22]	[1.94]	[2.02]
	4	-0.175	0.055	-0.202	0.041	0.126	0.301
		[-1.20]	[0.39]	[-1.18]	[0.25]	[0.53]	[1.05]
	High	-0.300**	-0.205*	-0.268	-0.258*	-0.276	0.024
		[-2.26]	[-1.87]	[-1.09]	[-1.81]	[-1.44]	[0.18]
	H - L	-0.252	-0.207	-0.257	-0.563**	-1.007***	-0.755**
		[-0.96]	[-0.99]	[-0.32]	[-2.11]	[-3.04]	[-2.01]

Internet Appendix

Table IA.1. Portfolio results based on alternative models

This table reports the monthly returns of the portfolios sorted on remoteness. Our sample covers common stocks listed on the NYSE, Nasdaq, and Amex from 1996 to 2020. We report the time-series alphas relative to the CAPM, Fama and French (1993) three-factor, and Carhart (1997) four-factor models in Panel A, B, and C, respectively. We sort stocks into decile portfolios every month based on Remote6h of the previous calendar year. Remote6h is defined in equation (2). For each portfolio, we first calculate the average return in the cross section using value weights. We then report the time-series average returns and the regression coefficients relative to the models. The returns are in percentage. t-statistics are in the brackets below. ***, **, or * indicates significance level of 1%, 5%, or 10%, respectively.

Panel A. CAPM

	By Remote6h										
	Low	2	3	4	5	6	7	8	9	High	H - L
Alpha	-0.057	-0.164	0.083	0.003	-0.086	-0.265	-0.143	-0.054	0.041	0.402**	0.459*
	[-0.54]	[-1.35]	[0.80]	[0.02]	[-0.77]	[-1.06]	[-1.00]	[-0.48]	[0.20]	[2.44]	[1.92]
MKT	0.811***	0.947***	0.842***	0.838***	0.941***	0.951***	1.114***	1.041***	1.309***	1.125***	0.314***
	[27.86]	[34.02]	[30.36]	[24.13]	[31.00]	[16.78]	[34.56]	[33.88]	[23.53]	[26.78]	[4.89]

Panel B. Fama-French three-factor model

	By Remote6h										
	Low	2	3	4	5	6	7	8	9	High	H - L
Alpha	-0.069	-0.166	0.075	-0.003	-0.090	-0.275	-0.150	-0.061	0.058	0.421***	0.490***
	[-0.87]	[-1.58]	[0.87]	[-0.02]	[-0.83]	[-1.12]	[-1.19]	[-0.60]	[0.34]	[2.88]	[2.59]
MKT	0.841***	0.999***	0.873	0.869***	0.950***	0.952***	1.040***	0.988***	1.245***	1.102***	0.261***
	[34.99]	[47.73]	[39.46]	[29.69]	[31.82]	[23.25]	[31.87]	[34.01]	[26.29]	[28.34]	[4.82]
SMB	-0.064	-0.210***	-0.088*	-0.106*	-0.014	0.047	0.351***	0.261***	0.183**	0.000	0.064
	[-1.26]	[-4.66]	[-1.89]	[-1.78]	[-0.35]	[0.42]	[3.80]	[6.32]	[2.13]	[0.00]	[0.62]
HML	0.319***	0.251***	0.264***	0.213***	0.109**	0.163***	-0.207***	-0.120***	-0.550***	-0.404***	-0.723***
	[9.77]	[9.36]	[6.69]	[4.73]	[2.34]	[3.06]	[-2.91]	[-2.85]	[-7.76]	[-6.45]	[-9.07]

Panel C. Carhart four-factor model

	By Remote6h									
	Low	2	3	4	5	6	7	8	9	High	H - L
Alpha	-0.072	-0.172	0.098	-0.013	-0.129	-0.241	-0.199	-0.042	0.101	0.416***	0.488**
	[-0.91]	[-1.57]	[1.10]	[-0.09]	[-1.19]	[-0.92]	[-1.58]	[-0.42]	[0.61]	[2.70]	[2.46]
MKT	0.843***	1.002***	0.858***	0.875***	0.974***	0.931***	1.071***	0.977***	1.219***	1.105***	0.262***
	[35.40]	[48.79]	[35.39]	[27.05]	[31.50]	[23.74]	[32.02]	[35.09]	[25.62]	[25.07]	[4.52]
SMB	-0.064	-0.211***	-0.085*	-0.107*	-0.019	0.052	0.345***	0.264***	0.188**	-0.001	0.064
	[-1.27]	[-4.60]	[-1.90]	[-1.77]	[-0.47]	[0.45]	[4.02]	[6.55]	[2.12]	[-0.01]	[0.62]
HML	0.321***	0.255***	0.248***	0.220***	0.136***	0.139**	-0.173**	-0.133***	-0.580***	-0.401***	-0.722***
	[9.89]	[8.61]	[6.54]	[4.40]	[2.86]	[2.21]	[-2.49]	[-2.91]	[-7.76]	[-5.79]	[-8.38]
MOM	0.006	0.009	-0.037	0.016	0.062**	-0.055	0.079*	-0.029	-0.068	0.008	0.002
	[0.32]	[0.46]	[-1.46]	[0.58]	[2.28]	[-1.18]	[1.76]	[-1.19]	[-1.46]	[0.14]	[0.03]