6-2020

Geographic links and predictable returns

Frank Weikai LI
Singapore Management University, wkli@smu.edu.sg

Zuben JIN
Singapore Management University, zuben.jin.2017@pbs.smu.edu.sg

Geographic Links and Predictable Returns

Zuben Jin and Frank Weikai Li*

August 2020

Abstract

Using detailed information of establishments owned by U.S. public firms, we construct a novel measure of geographic linkage between firms. We show that the returns of geography-linked firms have strong predictive power for focal firm returns and fundamentals. A long-short strategy based on this effect yields monthly value-weighted alpha of approximately 60 basis points. This effect is distinct from other cross-firm return predictability and is not easily attributable to risk-based explanations. It is more pronounced for focal firms that receive lower investor attention, are more costly to arbitrage and during high sentiment periods. In addition, we find sell-side analysts similarly underreact, as their forecast revisions of geography-linked firms predict their future revisions of focal firms. Our results are broadly consistent with sluggish price adjustment to nuanced news affecting firms with geographically-overlapped establishmens.

JEL classification: G10, G12, G14, G24

Keywords: Geography, Limited attention, Cross-asset momentum, Market efficiency

*Li (corresponding author) is at the Lee Kong Chian School of Business, Singapore Management University. Address: 50 Stamford Road, Singapore 178899. E-mail: wkli@smu.edu.sg. Tel: +65-6828-9651. Fax: +65-6828-0427. Jin is at the Lee Kong Chian School of Business, Singapore Management University. E-mail: zuben.jin.2017@pbs.smu.edu.sg. The authors are grateful for the valuable comments received from Roger Loh, Jinfei Sheng, Baolian Wang, Jingda Yan, and Jialin Yu, as well as seminar participants at Singapore Management University and Tianjing University. We thank Ken French and Lu Zhang for making factor returns publicly available; Andrea Frazzini for making customer-supplier links publicly available; Gerard Hoberg and Gordon Phillips for making Text-based Network Industry Classifications publicly available; and Stephen Teng Sun for sharing us with the technological linkage data.

1 Introduction

Economists have long recognized that location plays an important role in shaping economic growth through generating economies of scale in production process and facilitating knowledge spillover among neighboring firms and workers (Marshall (1920)). A growing literature shows that geographic locations are also important for understanding firms' fundamental performance (Dougal, Parsons, and Titman (2015); Tuzel and Zhang (2017)), the speed of information transmission (Coval and Moskowitz (2001); Malloy (2005); Parsons, Sabbatucci, and Titman (2018)), the level of discount rate (Garcia and Norli (2012)), stock liquidity (Loughran and Schultz (2005)) and even financial misconduct (Parsons, Sulaeman, and Titman (2018)). However, existing studies mostly identify a firm's geographic location as its headquarter, while ignoring the fact that for many firms, the more economically relevant geographic unit should be its establishment location where sales are generated and goods are produced (Bernile, Kumar, and Sulaeman (2015)).

In this study, we examine the implications of firms' geographic linkage for the price discovery and information diffusion process. In particular, we hypothesize that a firm's fundamental and stock performance should comove with its geography-linked peer firms, which we identify based on firms' disaggregated establishment location information. This interdependence among firms that are geographically overlapped could arise for many reasons. For example, firms with establishments in the same areas are exposed to common local economic shocks, which will then affect demand for firms' products and input prices (such as labor costs and land prices). In addition, there are occasional natural disasters occuring in certain areas that may disrupt firms' production process (e.g, Hurricane Harvey in Texas and Louisiana in 2017). Firms also benefit from the local agglomeration effect due to knowledge diffusion between a city's workers (Moretti (2004)), technology spillover between neiboring firms (Jaffe, Trajtenberg, and Henderson (1993)), and consumption externalities among local residents (Glaeser, Kolko, and Saiz (2001)). These common shocks and spillover effects can naturally lead to fundamental and return comovement between firms that have geographically overlapped establishments.

Our empirical evidence verifies the conjecture that geographic links lead to comovement in firms' fundamentals, even for firms that operate in different industries and headquartered in different regions. More strikingly, we document significant return predictability across geography-linked firms. Specifically, we document a novel empirical relation wherein the stock returns of focal firms exhibit a predictable lag with respect to the recent returns of a portfolio of its geographic peers ("geo-peers"). Focal firms whose geo-peers earn higher (lower) returns will themselves earn higher (lower) returns in subsequent months. A trading strategy using a proxy based on lagged geo-peers' returns yields annual Carhart (1997) four-factor alpha of 6-7%. These results are robust to an extensive list of control variables and cannot be easily explained by risk-based explanations. Rather, our evidence appears most consistent with sluggish price adjustment to nuanced news affecting firms with geographically-overlapped establishments.

To study the comovement and lead-lag effect among geography-linked firms, we obtain establishment-level data from the NETS database. This database provides addresses, as well information on sales and employment, for each U.S. establishment owned by a public company over the period from 1989 to 2012. With this data, we construct a pairwise geographic linkage between firms using their establishment location information. Specifically, geographic linkage GEO_{ijt} is defined as the uncentered correlation of the distribution of sales between two firms i and j across all counties in US, $GEO_{ijt} = \frac{G_{it}*G'_{jt}}{\sqrt{(G_{it}*G'_{it})*(G_{jt}*G'_{jt})}}$, where $G_{it} = (G_{it1}, G_{it2}, ..., G_{it3022})$ is a vector of firm i's proportional share of sales across 3,022 U.S. counties over year t.[1] With this measure, we first verify a basic premise underlying our hypothesis, that geographic linkage constructed using establishment location capture fundamental relationship between firms. We find that firm fundamentals (sales and profit

[1]The geographic linkage measure is constructed in the same way as the product similarity used in Hoberg and Phillips (2016), text similarity used in Cohen, Malloy, and Nguyen (2020), and technological proximity measure used in Jaffe et al. (1986) and Lee, Sun, Wang, and Zhang (2019), among others.

growth) are strongly correlated with current fundamentals of geography-linked peer firms, even after controlling for the corresponding correlations using other linkage proxies including industry links, same-headquarter links, and shared analyst links.

Two companies can have geographically overlapped establishments, yet are not operating in the same industry and not headquartered in the same region. Consider the case of Starbucks Corporation which is a chain of coffeehouses headquartered in Seattle, Washington, and Whole Foods Market Inc., which is a supermarket chain headquartered in Austin, Texas. Both firms have stores across major cities in US. From 2010 to 2012, the average geographic linkage for these two firms is high: $GEO_{ijt} = \frac{G_{it} * G^{'}_{jt}}{\sqrt{(G_{it} * G^{'}_{it}) * (G_{jt} * G^{'}_{jt})}} = 0.68$. Yet these firms are not in the same industry (Standard Industrial Classification (SIC) code: 5812 vs. 5411) nor are they headquartered in the same region. Furthermore, they are not product market peers in the sense of Hoberg and Phillips (2016), as the text-based product similarity score for these firms is only 0.015.[2] However, these two firms generally target the same type of consumers (white-collars who buy organic food products and enjoy drinking premium coffee), hence it is very likely that sales and profits of the two firms comove with each other as both are exposed to the same local economic conditions. This example illustrates the potential importance of geographic linkage, as distinct from other economic linkages explored by prior studies. While it is natural for firms in the same industry to cluster in the same area, close geographic proximity can often transcend industrial boundaries.

Next, we implement a portfolio approach to study the return predictability among geography-linked firms. Specifically, for each focal firm i at month t, we calculate the weighted return of a portfolio of firms that share similar geographic locations as the focal firm, $GEORET_{it} = \sum_{j \neq i} GEO_{ijt} * RET_{jt} / \sum_{j \neq i} GEO_{ijt}$, where RET_{jt} is the return of firm j at month t and GEO_{ijt} is the geographic linkage measure we construct using information up to

[2]See Hoberg and Phillips (2016) for how product similarity scores are measured.

month t.[3] We then sort focal firms into deciles using returns earned by a portfolio of their geo-peers in the previous month. Our results show that the geo-peers' lagged returns can significantly predict focal firm returns. A portfolio that long the focal firms whose geo-peers performed best in the prior month and short the focal firms whose geo-peers performed worst in the prior month, yields a value-weighted Carhart (1997) four-factor alpha of 53 basis points per month (t=2.62). We further confirm these return prediction results are robust to using various factor models to adjust risk exposure. In addition, the return predictability persists in Fama-MacBeth regressions when we include standard controls such as firm size, book-to-market ratio, gross profitability, asset growth, short-term reversal, and medium-term price momentum.

Prior studies have documented several lead-lag return effects among economically-related firms, including firms operating in the same industries and product markets (Moskowitz and Grinblatt (1999);Hoberg and Phillips (2018)), firms headquartered in the same regions (Parsons, Sabbatucci, and Titman (2018)), firms that are linked along the supply chain (Cohen and Frazzini (2008); Menzly and Ozbas (2010)), single- and multi-segment firms operating in the same industries (Cohen and Lou (2012)), and firms with similar technologies (Lee, Sun, Wang, and Zhang (2019)). We conduct several tests to ensure that our novel return predictability among geography-linked firms is not a rediscovery of these existing interfirm linkages. First, given the well-known geographic agglomeration of firms in a single industry (Ellison and Glaeser (1997)), it is likely that firms will have establishments largely overlapping with their industry peers geographically. Similarly, firms whose headquarters located in the same region will likely have geographically overlapped business operations. To mitigate such concerns, we control for lagged industry return and lagged return of a portfolio of firms headquartered in the same state as the focal firm in Fama-MacBeth regressions. In

[3] In our portfolio test, in order to ensure our results are distinct from the industry momentum effect (Moskowitz and Grinblatt (1999)) and same-headquarter lead-lag effect (Parsons, Sabbatucci, and Titman (2018)), we exclude all firms from the same industry (based on Fama-French 48 industry classification) and headquartered in the same state as the focal firm when cosntructing $GEORET_{it}$.

addition, we control for the focal firm's lagged tech-peer returns (Lee, Sun, Wang, and Zhang (2019)), focal firm's lagged pseudo-conglomerate returns (Cohen and Lou (2012)), focal firm's lagged supplier and customer industry returns (Menzly and Ozbas (2010)), and focal firm's product market peers' returns (Hoberg and Phillips (2018)). Lastly, a recent paper by Ali and Hirshleifer (2020) show that all the existing cross-firm return predictability effects are a unified phenomenon captured by shared analyst coverage, that is, firms covered by the same set of analysts. We thus add the lagged returns of stocks that are connected to the focal stock through common analysts. The lead-lag return relationship among geo-peers is robust to the presence of all these controls. Taken together, these tests show that our measure of geographic linkage is distinct from existing interfirm links including industry links, product market links, headquarter links, customer-supplier links, techonology links, standalone-conglomerate firm links, and shared analysts links.

After establishing the robustness of lead-lag return effect among geography-linked firms, we conduct tests to examine the economic mechanisms underlying the return predictability results. Our preferred explanation is that investors have limited attention and are slow to incorporate value-relevant information contained in focal firm's geographic peers. If this is the case, we should observe stronger return predictability among firms that are more likely to be overlooked by investors. Consistent with this prediction, we find the return predictability is more pronounced for focal firms that are smaller and have lower institutional ownership. Also consistent with the idea that common analyst coverage expedite information flow between economically-related firms (Parsons, Sabbatucci, and Titman (2018); Ali and Hirshleifer (2020)), we find weaker return predictability when the focal firm share a large set of common analysts with its geo-peers. Second, the abnormal returns generated by our trading strategy raise the question of why the profits are not quickly arbitraged away by smart investors. Consistent with the idea that there are limits to arbitrage in real-world financial markets, we find stronger return predictability among firms that are more costly to trade, such as stocks with higher bid-ask spread, lower liquidity, and higher idiosyncratic volatility.

In addition, using the Baker and Wurgler (2007) sentiment index, we find stronger return predictability during high-sentiment periods. These cross-sectional and time-series tests help confirm that we are truly capturing a mechanism of delayed updating of focal firm prices to information important to their fundamental values.

Broadly speaking, there are two possible channels that can explain the comovement and lead-lag relation among geographically-linked firms. First, firms with geographically-overlapped establishments are naturally exposed to the same regional economic conditions (the "common exposure" channel). A second channel is that shocks originated from geographic peers *spillover* to focal firm due to complementarity in investment opportunities (the "spillover" channel). Using natural disasters as localized shocks, we provide evidence that the lead-lag relation we document (partially) results from shock spillover among geographic peers in addition to their common exposure to local economy. This is a novel channel in the context of cross-firm return predictability literature.

Although the return predictability effects we document is robust to adjustment using various asset pricing models, one may still be concerned that other unobserved risks could drive our results. We conduct several tests to further distinguish between mispricing and risk explanation. First, we examine the stock price reaction around subsequent earnings announcements. This test has been widely used in prior studies to separate mispricing from risk explanations (e.g., Bernard and Thomas (1989); La Porta, Lakonishok, Shleifer, and Vishny (1997); Engelberg, McLean, and Pontiff (2018)). The idea is intuitive: earnings announcements help correct investor expectation errors about future cash flows; As a result, if abnormal return is associated with investor biased beliefs about the firms' fundamentals, a disproportionate fraction of its returns should be realized around subsequent earnings announcements. In contrast, if return predictability effect is driven by exposures to some unknown risks, strategy returns should accrue more evenly over subsequent trading days. Our tests show that the return spread generated from geo-peers' return signal ($GEORET$) is 166% higher on a day during an earnings announcement window than on a non-announcement day. This evidence is difficult to square with standard risk models.[4]

Second, we also examine the predictability of $GEORET$ for focal firm's future standardized unexpected earnings (SUE). $SUEs$ are not return-based, so this test is not confounded by imperfect controls for firm risks. At the same time, earnings are fundamental drivers of firm value. If returns to the $GEORET$ hedged portfolio are driven by predictable changes in cash flows, rather than a compensation for risk, the $GEORET$ signal should also predict focal firms' future $SUEs$. Our results show that geo-peers' returns do strongly predict focal firms' subsequent $SUEs$. Consistent with a slow diffusion of earnings-related news, focal firms with high (low) $GEORET$ report higher (lower) future $SUEs$, even after controlling for each firm's own lagged $SUEs$. This result again suggests that the return predictability associated with $GEORET$ reflects incomplete price response to fundamental information, rather than compensation for risks. In addition, this result, along with our finding that the return predictability of $GEORET$ lasts for several months and does not reverse afterwards, strongly suggests that the predictable return based on $GEORET$ is driven by investor underreaction, not overreaction or liquidity effects. Lastly, we look at analyst forecasting behavior to provide direct evidence on the limited attention channel. We find that analysts are slow to carry information across geography-linked firms, as analyst forecast revisions of geo-peers significantly predict future forecast revision of focal firms.

In addition to the tests reported in the main text of this study, our Internet Appendix provides a battery of other robustness tests. First, we document the robustness of the return predictability of $GEORET$ to various perturbations in such as removing micro-cap stocks and firms operating in few counties. Second, we report the robustness of return predictability by two subperiods: 1990-2001 and 2002-2013. In both subperiods, we find significant geographic lead-lag effect even after controlling for many other pricing anomalies. Third, we examine the sensitivity of our result to the staleness of the geographic linkage

[4] Although Patton and Verardo (2012) find stock betas increase on earnings announcement days, the increase in beta is symmetric for both positive and negative earnings surprises. As a result, time-varying beta cannot explain the large increase in the long-short portfolio's return spread on earnings announcement.

measure. Our results show that the effect declines slightly with more "stale" GEO data, but is still significant even when we use five-year-old geographic linkage measure. Fourth, the results are robust to alternative threshold used to define geographic peers. Finally, our result persists if we construct geographic linkage using establishment employment data, which is less likely to be imputed than sales in NETS data.

The remainder of this paper is organized as follows. Section 2 briefly surveys related literature and discusses the contribution of this study. Section 3 describes the data and presents summary statistics. Section 4 presents our main results on the lead-lag return relationship among geography-linked firms. Section 5 explores the underlying channels behind our results. Section 6 rules out risk-based explanations by conducting non-return-based tests and examining analyst forecast behavior. Section 7 concludes.

2 Related Literature and Contribution

Our paper contributes to several strands of existing literature. First, this study relates to a large literature that examines investor belief updating in response to new information. Tversky and Kahneman (1974), Daniel, Hirshleifer, and Subrahmanyam (1998), and Hong and Stein (1999), among others, suggest that investors may overweigh their own prior beliefs and underweight value-relevant public information, especially when the public information is less salient. A large set of empirical works lends support to this view.[5] Studies also document underreaction is more likely in settings where the nature of information is less salient (DellaVigna and Pollet (2007); Giglio and Shue (2014)) or when investors are being distracted (DellaVigna and Pollet (2009)). Our study is similar in spirit, but examines the slow diffusion of information contained in firms' geographic peers, an important diver of firm value that often transcends industry boundaries.

[5]For example, investors underreact to public announcements of corporate events including earnings announcements (Bernard and Thomas (1989)), and share repurchase and issuance (Ikenberry, Lakonishok, and Vermaelen (1995)) etc.

Our study is also related to a growing literature on the implicaton of investors' limited attention on information diffusion and market efficiency. Several theoretical works present a framework for understanding market price dynamics when a subset of investors have limited attention (e.g., Hirshleifer and Teoh (2003) and Peng and Xiong (2006)). The key message from these models is that slow information diffusion due to investors' limited attention can generate return predictability patterns that are difficult to explain with rational asset pricing models. These limited attention models have inspired a growing empirical literature. Particularly noteworthy are recent studies that document a lead-lag return effect between firms that have close economic links, such as industry links (Moskowitz and Grinblatt (1999); Hoberg and Phillips (2018)), customer-supplier links (Cohen and Frazzini (2008); Menzly and Ozbas (2010)), technology links (Lee, Sun, Wang, and Zhang (2019)) and shared analyst links (Ali and Hirshleifer (2020)). Our paper can be framed in terms of this literature, but we focus specifically on geographic links. We show that geographic linkage is distinct from other well-documented interfirm linkages.

Third, our study also contributes to the growing literature on the role of geography in information diffusion and price discovery process. For example, Coval and Moskowitz (2001) show that fund managers who are located close to firm headquarters earn higher returns on their local investment than distant investment. Similarly, Malloy (2005) show that geographically proximate analysts are more accurate than other analysts. Loughran and Schultz (2005) document firms headquartered in rural areas have poorer information environment and are traded less frequently compared to urban-based firms. Pirinsky and Wang (2006) document strong comovement in the stock returns of firms headquartered in the same geographic area. Parsons, Sabbatucci, and Titman (2018) document a lead-lag return effect among firms headquartered in the same state. Korniotis and Kumar (2013) find that state-level economic factors (e.g., unemployment and housing collateral ratios) can predict returns of stocks headquartered in those states. All these studies focus on firm's headquarter location. However, as shown by Bernile, Kumar, and Sulaeman (2015), the typical U.S.

public firm has economic interests in five states beyond its corporate headquarters location. A firm's headquarter may be in one state, while its plants and operations are located in other states, often far away from the headquarter. When the economic activities of a firm are geographically segmented, value-relevant information about the firm is also likely to be geographically dispersed.[6] The innovation of this study is to infer the arrival of geographic information for a firm from stock returns of other firms with economic activities in the same areas. Our approach thus identifies the geographic links between firms beyond their headquarter locations, and show the returns and fundamentals of focal firm can be predicted by its geographic peers.

3 Data and variables

3.1 Data

To capture firms' geographic footprints, we obtain establishment-level data from the NETS Publicly Listed Database produced by Walls & Associates using using Dun and Bradstreet (D&B) data. The NETS database provides annual employment and sales data for more than 63 million U.S. businesses and establishments (i.e., headquarters, subsidiaries, branches, and plants across the United States). This database maintains an essentially complete record of all establishments going back to 1989. Establishments are not legally required to report to D&B; however, D&B is a leading provider of business credit information and thus those establishments that wish to obtain lines of credit with suppliers or financial institutions have incentives to report to D&B. Additionally, D&B attempts to develop complete business lists by collecting information from independent sources, including phone calls, legal and bankruptcy filings, press reports, payment and collection activities, and government and

[6] A notable exception in the literature is Garcia and Norli (2012) that identifies U.S. states that are economically relevant for a company through textual analysis of annual reports.

postal records.[7] Recent studies employing NETS data include Neumark, Wall, and Zhang (2011), Heider and Ljungqvist (2015) and Addoum, Ng, and Ortiz-Bobea (2020), among others. We match each establishment with its parent company in Compustat by company name. The matching procedure includes both machine-matching and manual-matching.

We obtain monthly stock returns from the Center for Research in Security Prices (CRSP), and annual accounting data from Compustat. Our main sample consists of firms in the intersection of the NETS Publicly Listed data, CRSP, and Compustat. We include all common stocks (CRSP share codes 10 and 11) traded on the NYSE, Amex, and NASDAQ, and exclude financial firms (Fama-French 48 industry code between 44 and 47). To ensure that the relevant accounting information is publicly available to investors in the market, we impose at least a six-month gap between fiscal-year end month and the portfolio formation date. Specifically, we first match the NETS data in year t with Compustat accounting data for the most recent fiscal year (i.e., the fiscal year ended in calendar year t). We then match sample firms to CRSP stock returns from July of year $t+1$ to June of year $t+2$. We require firms to have non-missing stock price and SIC classification code from CRSP, and non-negative book equity data at the end of the previous fiscal year from Compustat. To reduce the impact of penny stocks, we exclude stocks that are priced below one dollar a share at the beginning of the holding period. We adjust the stock returns by delisting. If a delisting return is missing and the delisting is performance-related, we set the delisting return at -30% (Shumway (1997)).

We define our pairwise measure of geographic linkage, GEO_{ijt}, as the uncentered correlation of the distributions of sales across all counties in US between all pairs of firms i and j,

$$GEO_{ijt} = \frac{G_{it} * G'_{jt}}{\sqrt{(G_{it} * G'_{it}) * (G_{jt} * G'_{jt})}} \tag{1}$$

where $G_{it} = (G_{it1}, G_{it2}, ..., G_{it3022})$ is a vector of firm i's proportional share of sales across

[7] Barnatchez, Crane, and Decker (2017) conduct a through assessment of the NETS data and conclude that NETS data is useful and convenient for studying business activity in high detail.

3,022 U.S. counties over year t. GEO_{ijt} has fhe following properties: it is unity for firms whose geographic vectors are identical, and zero for firms whose vectors are orthogonal and it is bounded between zero and one for all other pairs. It is closer to unity the greater the degree of overlap of the two firms' establishment locations.[8] Further more, this measure is symmetric in firm ordering (i.e., $GEO_{ijt} = GEO_{jit}$) and not directly affected by the length of the G vectors.[9]

We then define geography-linked return ($GEORET$) as the weighted-average monthly return of geography-linked firms, with pairwise geographic linkage as weight. Formally, geography-linked return for firm i at month t is defined as:

$$GEORET_{it} = \sum_{j \neq i} GEO_{ij\tau-1} * RET_{jt} / \sum_{j \neq i} GEO_{ij\tau-1} \tag{2}$$

where RET_{jt} is the raw return of firm j at month t. Note that GEO naturally serves as a weighting function in calculating the portfolio return of geography-linked firms, such that firms more overlapped with the focal firm in geographic space receive higher weight. GEO is calculated at the end of each calendar year $\tau - 1$ based on NETS data in that year, and then mapped to the monthly stock return data from July of year τ to June of year $\tau + 1$.

We use standard control variables in our empirical analysis. *Size* is defined as the natural logarithm of market capitalization at the end of June in each year. Book-to-market ratio (BM) is the most recent fiscal year-end report of book value divided by the market capitalization at the end of calendar year $t - 1$. Book value equals the value of common

[8] As an example, suppose there are three firms A, B, and C, with establishment sales across three US counties, as follows: $G_A = (0, 0, 1)$, $G_B = (0.6, 0.2, 0, 2)$, $G_C = (1, 0, 0)$. In this example, G_{AB}=0.13, G_{AC}=0, and G_{BC}=0.90. Intuitively, firms A and C have no establishments in the same county and are thus assigned a geographic linkage measure of zero. These two firms would not be geo-peers for purposes of our analysis. Firm B has geographically overlapping establishments with both firm A and firm C. However, as shown above, firm B is more closely connected to firm C geographically (G_{BC} = 0.90), than it is to firm A (G_{AB} = 0.13). This is because a higher proportion of B's sales are in the 1st county than in the 3rd county.

[9] The length of the vector depends on the degree of geographic concentration of firms' economic activities. As a result, GEO will not capture the effect of geographic dispersion on stock returns as documented by Garcia and Norli (2012).

stockholders' equity, plus deferred taxes and investment tax credits, and minus the book value of preferred stock. Momentum (MOM) is defined as the cumulative holding-period return over the last 12 months skipping the most recent month. RET_{t-1} is the prior month's return to capture short-term reversal effect. Following Cooper, Gulen, and Schill (2008), asset growth (AG) is defined as year-over-year growth rate of total assets. Following Novy-Marx (2013), gross profitability (GP) is defined as sales revenue minus cost of goods sold scaled by assets. Institutional ownership data of stocks are available from the Thomson Reuters (formerly CDA/Spectrum) Institutional Holdings database (13F). Aanalyst forecast data are from I/B/E/S.

3.2 Summary Statistics

The final sample consists of 668,117 firm-month observations spanning July 1990 to December 2013. Panel A of Table 1 presents descriptive statistics for our sample firms. The average number of firms per month is 2,320. On average our sample firms cover around 57% of the CRSP common stock universe in terms of market capitalization. We note that the average number of geography-linked firms per focal firm is 795. The pairwise geographic linkage measure (GEO) has an average score of 0.09 with a standard deviation of 0.2, indicating large cross-sectional variation in geographic linkage among our sample firms. The remaining summary statistics are well known and do not require additional discussion.

In Panel B of Table 1, we present the pairwise correlation between our variables. Several correlation coefficients are noteworthy. Although $GEORET_{t-1}$ exhibits trivial correlations with a number of traditional return predictors (e.g., size, book-to-market, gross profitability, and asset growth), it is considerably more correlated with industry return ($INDRET_{t-1}$), return of a portfolio of firms headquartered in the same state ($HQRET_{t-1}$), and past one-month return (RET_{t-1}) (Pearson correlations are 0.095 for $INDRET_{t-1}$, 0.316 for $HQRET_{t-1}$, and 0.062 for RET_{t-1}). In subsequent analyses, we will control for these return predictors when examining the return predictability of $GEORET_{t-1}$.

4 Empirical results

We next turn to the main results of the paper. We first verify that geography-linked firms as identified by our measure are fundamentally related. We then show the lagged returns of geography-linked firms have strong predictability power for focal firm returns and this pattern is robust and distinct from existing cross-firm return predictability effects.

4.1 Fundamental comovement

We first verify our geographic linkage measure by examining whether our measure captures fundamental relationship bewteen geography-linked peer firms. Specifically, we regress focal firms' annual sales and profitability growth measures on the average growth measures of their geo-peers (*Geo sales growth*). We calculate the average growth variables of geo-peers using the same methodology as used in calculating $GEORET$. *Geo sales growth* is calculated as the weighted average sales growth of geo-peers using the weights in equation 3.1. All regressions include year fixed effects and size and book-to-market ratio as controls. To ensure that the growth variables for all firms are measured over the same horizon, we only include firms with December fiscal year ends.

Table 2 presents the results. Column 1 shows that the coefficient on *Geo sales growth* is 0.311 (t=3.25), indicating that there is a strong contemporaneous correlation between focal firm's and geo-peers' sales growth. In column 2, we add the average sales growth of other economically-linked peer firms. Specifically, industry sales growth is measured as the market capitalization-weighted average sales growth of all other firms in the same industry (based on Fama-French 48 industry classifications) as the focal firm. Same-state sales growth is measured as the average sales growth of all other firms headquartered in the same state as the focal firm. Analyst sales growth is calculated the weighted average sales growth of shared analyst-linked peers, using the weights defined in Ali and Hirshleifer (2020). The coefficient on *Geo sales growth* decreases to 0.199, but remains significant. Columns 3 and 4 show that the same conclusions hold when fundamental performance is measured as profitability growth instead of sales growth.

Overall, these results strongly suggest that our measure of geographic linkage captures fundamental relatedness between firms and that geographic linkage is distinct from other interfirm linkages identified in previous studies.

4.2 Portfolio tests

In this section, we show that stocks sorted based on their geography-linked peers' returns generate significant return spreads. We conduct the decile portfolio sorts as follows. At the beginning of each month, we sort stocks into deciles by the return earned by their geography-linked peers in the previous month ($GEORET_{t-1}$). To ensure our results are distinct from the industry momentum effect (Moskowitz and Grinblatt (1999)) and same-headquarter lead-lag effect (Parsons, Sabbatucci, and Titman (2018)), we exclude all firms from the same industry (based on Fama-French 48 industry classification) and headquartered in the same state as the focal firm when cosntructing $GEORET_{it}$ for the portfolio tests. These decile portfolios are then rebalanced at the beginning of each month to maintain either equal or value weights. We use the time series of monthly portfolio returns to compute the average excess return (and alphas) of the lowest decile (1) and the highest decile (10) portfolio over the entire sample. As we are most interested in the return spread between the two extreme deciles, we also report the return to a long–short portfolio, i.e., a zero-investment portfolio that longs the stocks in the highest $GEORET_{t-1}$ decile and shorts the stocks in the lowest decile (L/S). We compute these returns by subtracting either the risk-free return (excess returns) or by using a variety of factor models.

Table 3 Panel A provides strong evidence that geography-linked firms' returns predict focal firm returns. Specifically, we find that the equal-weighted long-short *GEORET* strategy (L/S), yields average monthly returns of 41 basis points ($t = 2.97$), or roughly 6% per year. Unlike most anomalies, the L/S strategy generates value-weighted returns that are even larger at 54 basis points per month ($t = 2.62$), or about 6.5% per year. In the next five columns, we control for the portfolios' exposure to standard asset-pricing factors. The same L/S strategy delivers CAPM alphas of 0.44% (0.54%) per month in equal- (value-) weighted portfolios. This strategy delivers Fama and French (1993) three-factor alphas of 0.44% (0.53%) per month in equal- (value-) weighted portfolios. Augmenting this model by adding the stock's own price momentum (Carhart (1997)) does not significantly affect the strategy, as the four-factor alpha remains at 0.41% (0.53%) per months in equal- (value-) weighted portfolios. We also adjust returns using the Fama and French (2015) five-factor model (5-Factor), and also conduct a test using the five-factor model plus the momentum factor and a short-term reversal factor (7-Factor). We find that the strategy's alpha only slightly changes after controlling for these factors, with the five-factor and seven-factor strategies earning abnormal monthly returns of 0.40% (0.57%) and 0.41% (0.60%), respectively, in equal- (value-) weighted portfolios. Finally, we report the portfolio alpha using the Q factors of Hou, Xue, and Zhang (2015) as the asset pricing model. The Q-factor alphas continue to be significant, with a value-weighted monthly alpha of 0.62% (t=2.64). These results show that focal firms with high (low) geo-peers' returns earn high (low) subsequent returns, after controlling for common risk factors.

In Panel B of Table 3, we report the portfolio alpha as well as the factor loadings on each of the Fama-French three factors and the Carhart (1997) momentum factor (*MOM*) and a short-term reversal factor (*ST_Rev*). The L/S portfolio has little exposure to most factors, as the loadings on factors are economically small and statistically insignificant. One important exception is the significant and negative loading on the short-term reversal factor (long prior month loser and short prior month winner), which is consistent with our observation that

$GEORET_{t-1}$ is positively correlated with prior month return RET_{t-1}. This finding can perhaps explain why risk-adjusted returns of the L/S strategy are not very different from excess return.

In Figure 1, we plot the value of a hypothetical dollar invested in each of three portfolios. The first, shown in red, shows the evolution of a dollar invested in the S&P 500 index. Dividends are assumed to be reinvested. Against this benchmark, we also plot the 10% of firms with the highest lagged 1-month $GEORET$ (blue), as well as the 10% of firms with the lowest lagged 1-month $GEORET$ (green). While the market portfolio grows by a factor of over 4 from 1990 to 2013, one dollar invested in the lowest decile barely exceeds \$3. On the other hand, a \$1 investment in the highest decile performs almost an order of magnitude better, growing to approximately \$20 by 2013.

4.3 Fama-MacBeth Regressions

In this section, we test the return predictability of $GEORET$ using the Fama and MacBeth (1973) regression methodology. One advantage of this methodology is that it allows us to examine the predictive power of $GEORET$ while controlling for other known predictors of cross-sectional stock returns. This is important because, as shown in Table 1, $GEORET$ is correlated with some of these predictors. We conduct the Fama-MacBeth regressions in the usual way. For each month, starting in July 1990 and ending with December 2013, we run the following cross-sectional regression:

$$Ret_{i,t} = \beta_0 + \beta_1 GEORET_{i,t-1} + \gamma X_{i,t-1} + \epsilon_{i,t} \tag{3}$$

where $Ret_{i,t}$ is the raw return of focal firm i in month t, $GEORET_{i,t-1}$ is the average return of the focal firm i's geo-peers in month $t-1$, and $X_{i,t-1}$ is a set of control variables known to predict returns, including the natural logarithm of the book-to-market ratio (BM), the natural logarithm of the market value of equity (Size), returns from the prior month (RET_{t-1}), returns from the prior 12-month period excluding month $t-1$ (MOM), gross profitability (GP), and asset growth (AG).

Table 3 reports the time-series averages of the coefficients of the independent variables, and the t-statistics are Newey-West adjusted (up to 12 lags) for heteroskedasticity and autocorrelation. Column 1 shows the coefficient on $GEORET_{t-1}$ is 8.317 with a t-statistics of 4.81, suggesting that geography-linked firms' return strongly predict next-month focal firm return even after controlling for well-known return predictors. Economically, a two-standard deviation increase in $GEORET_{t-1}$ leads to approximately 50 basis points increase in focal firm return. The result from Fama-MacBeth regression is consistent with time-series portfolio tests. The coefficients on control variables are also consistent with prior literature: asset growth and short-term reversal variables are significantly negatively correlated with future returns, while book-to-market ratio and gross profitability are significantly positively correlated with future returns.[10]

One of stylized facts in urban economics is that firms from the same industry tend to cluster together geographically (Ellison and Glaeser (1997)). As a result, it is likely that firms will have establishments largely overlap with their industry peers geographically. Similarly, firms whose headquarters located in the same areas will have geographically overlapped business operations by construction. To mitigate such concerns, in Column 2, we add the lagged value-weighted industry return ($INDRET$) and lagged value-weighted return of a portfolio of firms headquartered in the same state as the focal firm ($HQRET$) in regression. Compared to Column 1, the coefficient on $GEORET$ decreases to 5.957, but remains highly significant with a t-statistics of 4.01. The coefficients on $INDRET$ and $HQRET$ are both positive and significant, consistent with the industry momentum effect documented by Moskowitz and Grinblatt (1999) and the same-headquarter lead-lag effect

[10]The coefficient of MOM is positive but insignificant, potentially due to the 2009 momentum crash documented by Daniel and Moskowitz (2016).

shown by Parsons, Sabbatucci, and Titman (2018).

In Columns 3 to 7, we control for other interfirm linkages as documented by prior studies.[11] In Column 3, we add the focal firm's lagged technology-peer return ($TECHRET$) following Lee, Sun, Wang, and Zhang (2019), who document a lead-lag effect among firms overlapping in technology space. In Column 4, a portfolio of focal firm's pseudo-conglomerate returns ($PCRET$) is added based on Compustat Segment data following Cohen and Lou (2012), who show substantial return predictability from standalone firms to conglomerates. In Column 5, we add the lagged returns from a portfolio of the focal firm's supplier industry ($SUPPRET$) and customer industry ($CUSTRET$). These portfolios are constructed using Bureau of Economic Analysis (BEA) Input-Output data (at the summary industry level) following Menzly and Ozbas (2010). In Column 6, we add the lagged returns of focal firm's product market peers, which are identified based on textual analysis of firms' 10-K filings. Following Hoberg and Phillips (2018), we use the TNIC-3 network, which is calibrated to have a granularity to be comparable with SIC-3 code. In Column 7, we add the lagged return from a portfolio of firms that have shared analyst coverage with the focal firm ($CFRET$), following Ali and Hirshleifer (2020).

There are several noteworthy patterns. First, the coefficients on these variables are almost all siginificant and positive, consistent with prior literature. The only exception is that coefficients on customers' ($CUSTRET$) and supplers' returns ($SUPPRET$) are insignificantly positive, which could potentially due to difference in sample period. More importantly, we find the coefficient on $GEORET_{t-1}$ remains highly significant after controlling for these known interfirm linkages. In particular, we continue to find significant return predictability for $GEORET$ after controlling for interfirm links between stocks covered by common analysts, which as argued by Ali and Hirshleifer (2020), captures all the existing cross-firm return predictability effects. This finding may not be surprising

[11]Because the data availability on these additional linkage measures greatly reduce the sample size, we do not control for these variables in subsequent analyses.

as even skilled analysts may not closely track news about firms' geo-peers and quickly impound relevant information into focal firm's prices. We provide more evidence supporting underreaction on the part of analysts in subsequent sections.

While most of the previous cross-firm return predictability studies have focused on one-month lagged returns as predictors, some studies also examine longer horizon lags. Appendix Table A2 shows that returns of geographic peers over the past 6 and 12 months are still significant predictors of future focal firm return, while geo-peers' returns over past 24 months lose its predictive power. However, both the statistical significance and economic magnitude of the long-horizon effects are rather modest. This is consistent with prior studies that most of the cross-firm return predictability effects are strongest at the one-month horizon (Moskowitz and Grinblatt (1999); Ali and Hirshleifer (2020)). It suggests that although the market is not perfectly efficient, it reacts quickly enough to start incorporating value-relevant news into stock prices within a month.

We also examine the long-run return pattern of the lead-lag effect between geography-linked firms. If investors overreact to the news contained in lagged geo-peers' returns, we should observe some return reversal over longer holding periods. On the other hand, if the effect we document is primarily an underreaction to the news that affects focal firms' fundamental value, we should see no return reversal in the future. In Figure 2, we evaluate these two alternative hypotheses by plotting the cumulative return to the $GEORET$ hedged portfolio in the six months after portfolio formation. Consistent with slow diffusion of geographic information, we continue to observe a modest upward drift in portfolio returns through month six. In fact, we find no sign of a return reversal over the next 12 to 24 months. These findings are similar to the results from other cross-firm return predictability studies (Cohen and Frazzini (2008); Cohen and Lou (2012); Lee, Sun, Wang, and Zhang (2019)). Overall, the evidence seems to be most consistent with delayed response of focal firm prices to fundamental information contained in returns of geo-peers, and not an overreaction phenomenon.

4.4 Robustness tests

In this section, we conduct a battery of robustness tests on geography-linked return predictability, and report the results in Appendix Table A3.

4.4.1 Excluding micro-cap stocks

First, to alleviate the concern that our results are driven by micro-cap stocks, we exclude stocks with price less than $5 or market capitalization below the 10th NYSE percentile. Columns 1 and 2 of Table A3 shows that the coefficients of $GEORET_{t-1}$ are still positive and highly significant in both settings, suggesting that our result is not driven by micro-cap stocks. Given that small firms are more likely to operate in a single area, another way to remove micro-cap stocks is to restrict our sample to focal firms with establishments in at least two counties. Column 3 shows the predictive power of $GEORET_{t-1}$ is robust to this sample selection criteria.

4.4.2 Geography-linked return predictability across time

In Columns 4 and 5 of Table A3, we examine whether the return predictability of geography-linked firms varies over time. We divide our full sample period into two subperiods: 1990-2001 and 2002–2013. We then repeat our baseline Fama-MacBeth regression for each subperiod. Our results hold up well in both periods, after controlling for various return predictors. The coefficients of $GEORET_{t-1}$ are similar in two subperiods, being 6.406 (t=2.75) during 1990-2001 and 5.426 (t=3.24) during 2002-2013. This remarkable persistence in the coefficient of $GEORET_{t-1}$ is in sharp constrast with that of some other return predictors, which declines substantially in the recent period. Consistent with prior studies, we find the effect of industry momentum reduce by more than half and the own price momentum effect becomes insignificant over the 2002-2013 period (Parsons, Sabbatucci, and Titman (2018)). What is more noteworthy from our perspective is that the return predictability of geography-linked firms is robust in both subperiods.

4.4.3 Persistence of the geographic linkage measure

We also examine the sensitivity of our main result to the age of the geographic linkage measure. Untabulated analysis shows the correlations between $GEO_{i,j,t}$ and its corresponding one-year lagged measures is 0.95, suggesting that firms' geographic footprints are relatively persistent over time. Columns 6-8 of table A3 shows $GEORET_{t-1}$ constructed using lagged values of GEO also predict focal firm returns. While predictability decreases with the number of lagged years, even five-year-old geographic linkage measures work quite well. One implication is that investors do not need extremely timely information on firms' establishments location information to implement this strategy. Even relatively "stale" geographic information have some predictive power for focal firm returns.

4.4.4 Using alternative geographic peers

In our main tests, a geographic peer is defined as a firm with any geographic overlap with the focal firm (i.e. any firm whose GEO value is greater than zero). To evaluate the sensitivity of our results to this cut-off value, we conduct a test where the peer sample is limited to just the top 50 geo-peers of the focal firm. Finally, we also construct alternative geographic linkage measure using establishment employment data, as the number of employees at establishments are less likely to be imputed than sales in NETS data. Column 9 and 10 of Table A3 show that the predictive power of $GEORET$ is still robust using these alternative proxies of geographic peers.

5 Mechanism

The results so far suggest that the lead-lag effects between geography-linked firms we document may be driven by slow dissemination of geographic news. In this section, we further explore the cross-sectional heterogeneity of our main results to various firm and stock characteristics associated with: (a) the extent to which investors might be attentive to such news, and (b) the costs that investors face if they attempt to profit from the mispricing. In addition, we examine whether the return predictability of $GEORET$ varies with aggregate investor sentiment.

5.1 Limited attention

If investors are fully rational and have unlimited capacity to analyze all value-relevant information, the news contained in geo-peers' returns should be reflected in focal firm's prices in a timely fashion. However, a large set of theorectical and empirical studies show that due to limited attention, investors tend to underweight public information, especially when the information is less visible (DellaVigna and Pollet (2009); Giglio and Shue (2014)) or more complicated to analyze (Cohen and Lou (2012)). If this is the case, the return predictability of $GEORET$ should be stronger among firms that receive less investor attention. Prior literature proposes several measures of investor attention including firm size, institutional ownership, and analyst common coverage.[12] We posit that smaller firms, and firms that have lower institutional ownership, and have fewer common analysts with their geographic peers, receive less attention from investors and, therefore, will exhibit a more sluggish stock price reaction to the information contained in $GEORET$.

To test this prediction, we define a size-based dummy variable that equals one if a focal firm is above the sample median in terms of the log value of market capitalization at the end

[12] See, for example, Bali, Peng, Shen, and Tang (2014), Parsons, Sabbatucci, and Titman (2018), and Ali and Hirshleifer (2020).

of the previous fiscal year, and zero otherwise. Similarly, we define a dummy variable that equals one if the institutional ownership (IO) at the end of the previous year is above the sample median. Finally, we define a dummy $CANALYST$ equals one if the average number of analysts covering the focal firm and its geo-peers at the previous year-end is above sample median, and zero otherwise. The results of these tests are reported in columns 1 to 3 of Table 5. Consistent with the prediction of limited attention channel, the coefficient estimates on the three interaction terms between the investor attention dummies and $GEORET_{t-1}$ are all negative, and in the case of firm size and institutional ownership, the interaction term is statistically significant. This result lends support to our hypothesis that the return predictability of $GEORET$ is driven by investors' inattention to the geographic linkage information.

5.2 Costs of arbitrage

In addition to attention proxies, we consider how the return predictability varies across our sample with different degrees of arbitrage costs. The evidence indicates that sophisticated investors, such as arbitrageurs, also fail to incorporate the information embedded in $GEORET$ and bring stock prices to full-information value. We thus expect that our results to be more pronounced among firms subject to greater limits to arbitrage. To test this conjecture, we use three measures to proxy for the cost of arbitrage: idiosyncratic volatility ($IDVOL$), bid-ask spread ($Spread$), and Amihud illiquidity ($Illiquidity$). Wurgler and Zhuravskaya (2002) and Pontiff (2006) argue that arbitrageurs' demand for a stock is inversely related to its arbitrage risk, which is reflected in its idiosyncratic volatility.[13] In addition, prior research suggests that information diffusion into price is slower when trading costs are higher and stocks are less liquid (Bali, Peng, Shen, and Tang (2014)). Therefore, we expect the return predictability of $GEORET$ will be more pronounced for less liquid

[13] Evidence supporting idiosyncratic return volatility as one of the most significant limits to arbitrage is documented in Stambaugh, Yu, and Yuan (2015), for instance.

stocks with higher bid-ask spread.

To test this prediction, we calculate idiosyncratic volatility ($IDVOL$) as the standard deviation of the residuals from a regression of daily excess stock returns on Fama and French (1993) factors within a month (at least ten daily returns required) following Ang, Hodrick, Xing, and Zhang (2006). Following Amihud (2002), $ILLIQUIDITY$ is the average daily ratio of absolute stock return to the dollar trading volume within each month. Following Corwin and Schultz (2012), we calculate the bid-ask spread ($SPREAD$) from daily high and low prices.[14] For all three variables, we create a dummy variable equals one if the corresponding proxy is above sample median in a month, and zero otherwise.

The results are reported in Columns 4 to 6 of Table 5. Column 4 shows that the coefficient estimate on the interaction term between the idiosyncratic volatility dummy and $GEORET_{t-1}$ is positive and statistically significant, 5.318 (t=2.61). Column 5 and 6 shows that the interaction term between an indicator of high bid-ask spread and higher Amihud illiquidity and lagged geo-peers' return ($GEORET_{t-1}$) is also positive and statistically significant. These findings lend support to our prediction that the return predictability effect is stronger for more difficult-to-arbitrage stocks.

5.3 Investor Sentiment

Recent studies show that stock market mispricings are typically more pronounced when the overall sentiment is high (Stambaugh, Yu, and Yuan (2012); Antoniou, Doukas, and Subrahmanyam (2016)), potentially due to amplification of investors' behavioral biases during high-sentiment periods. In our setting, this suggests that investors may pay less

[14] The Corwin and Schultz (2012) spread estimate is based on two reasonable assumptions. First, daily high-prices are almost always buyer-initiated trades and daily low-prices are almost always seller-initiated trades. The ratio of high and low prices for a day therefore reflects both the fundamental volatility of the asset and its bid-ask spread. Second, the component of the high-to-low price ratio that is due to volatility increases proportionately with the length of the trading interval while the component due to bid-ask spreads do not. Corwin and Schultz (2012) show via simulations that, under realistic conditions, the correlation between their spread estimates and true spreads is about 0.9 and their estimates are substantially more precise than other spread estimators.

attention to the performance of focal firm's geographic peers, which are value-relevant but less salient fundamental information. In addition, any level of mispricing would be more difficult to be arbitraged away due to increased noise trader risks and short-sale constraints (De Long, Shleifer, Summers, and Waldmann (1990)). As a result, we should expect the lead-lag return effect among geography-linked firms to be stronger during high sentiment periods. To test this idea, we use the Baker and Wurgler (2007) sentiment index ($SENTIMENT$) to proxy for aggregate investor sentiment. We create a dummy variable equals one if $SENTIMENT$ is above sample median, and zero otherwise. Column 7 of Table 5 shows that the coefficient estimate on the interaction term between $SENTIMENT$ dummy and $GEORET_{t-1}$ is indeed positive and significant. This finding provides further evidence that return predictability of $GEORET$ is likely a result of mispricing due to investors' underreaction to geographic information, especially during high-sentiment periods.

5.4 Common exposure to regional economy or spillover effect?

Broadly speaking, there are two possible channels that can explain the comovement and lead-lag relation among geographically-linked firms. First, firms with geographically-overlapped establishments are naturally exposed to the same regional economic fundamentals ("common exposure" channel). A second channel is that shocks originated from geographic peers "spillover" to focal firm due to complementarity in investment opportunities or technology spillover between neiboring firms ("spillover" channel). A key challenge in empirical tests is to differentiate between these two channels.

We conduct two tests to examine whether common exposure to regional economy can fully explain the return predictability of $GEORET$. Our first test is a subsample test that groups firms into manufacturing and non-manufacturing firms. The idea is that because manufacturing firms rely on national or even global demand for their products, common exposure to local economic fundamentals is unlikely to explain the lead-lag return effects among manufacturing firms that are geographically linked. Appendix Table A4 reports the Fama-MacBeth regression results for manufacturing and non-manufacturing firms separately. We find the coefficients on $GEORET_{t-1}$ are significant and have similar economic magnitude for both manufacturing and non-manufacturing firms, suggesting that our result is not fully explained by the "common exposure" channel. Second, we directly capture firms' exposure to regional economic condition by constructing a firm-specific predicted regional economic activity proxy ($PREA$) following Smajlbegovic (2019). Specifically, $PREA$ is the sales-weighted average of economic activity growth rate across all states that the firm operates:

$$PREA_{it} = \sum_{s=1}^{50} SALE\_SHARE_{i,s,\tau-1} * \frac{\Delta \widehat{SCI_{s,t+6}}}{SCI_{s,t}} \tag{4}$$

where $\frac{\Delta \widehat{SCI_{s,t+6}}}{SCI_{s,t}}$ is the predicted growth rate of the State Coincident Index of state s in month t for the next 6 months and $SALE\_SHARE_{i,s,\tau-1}$ is firm's fraction of sales in state s in last year. $PREA$ can be interpreted as the average forecast of the economic activity growth rate over all firm-relevant U.S. states. The orthogonalized proxy $PREA^{\perp}$ is the sum of a constant and the residuals of cross-sectional regressions of $PREA$ on return sensitivities to national economic activity and the Fama and French (1993) risk factors.

If the return predictability of $GEORET$ is derived solely from a firm's exposure to the economic conditions of all regions where it operates, we should find the effect of $GEORET$ to be absorbed by $PREA^{\perp}$ when predicting future returns. We report the corresponding Fama-MacBeth regression result in column (1) of Table 6. Consistent with Smajlbegovic (2019), the predicted regional economic activity meausre $PREA^{\perp}$ significantly and positively predict future stock return, indicating a slow diffusion of local macroeconomic information into stock prices. More importantly, the coefficient of $GEORET$ is still significantly positive, suggesting the cross-firm return predictability is not fully explained by common exposure to local economy.

To test any spillover effect among geographic peers, we examine whether a focal firm is negatively affected if its geographic peers are exposed to some negative shocks, even if the focal firm itself have no establishments in the areas affected by the shock. Our test is motivated by Giroud and Mueller (2019), who argue that when a firm's establishment experiences a negative shock to its cash flow, the firm will optimally "spreads" the shock across multiple establishments in an effort to equate their marginal productivity. As a result, local shocks not only lead to sales declines at local business units but also at business units in distant regions. Giroud and Mueller (2019) show that local shocks indeed propagate across US regions through firms' internal networks of establishments. To operationalize the idea, we construct a measure GEO_HAZARD as the weighted average of dummies indicating geo-peers headquartered in states affected by any natural disaster in a month, using geographic linkage (GEO) as weights. The advantage of using natural disaster is that natural disasters are exogenous events that occurred throughout our sample period, and their impacts are highly localized.[15] In untabulated analysis, we find that GEO_HAZARD leads to lower sales growth over the next year for focal firms that themselves are not exposed to any disasters. Column 2 of Table 6 reports the Fama-MacBeth regression results. The coefficient of GEO_HAZARD is -1.227 with a t-statistics of -2.71. This novel finding provides support to the "spillover" channel that local shocks to geographic peers transmit to the focal firm through geographic linkage. Overall, we provide evidence that the lead-lag relation between geographically-connected firms can result from a spillover effect in addition to their common exposure to local economy, which has not been explored by previous studies including Parsons, Sabbatucci, and Titman (2018).

[15] We obtain the natural hazards (including flooding, hurricane/Tropical Storm, Severe Storm/Thunderstorm, Tornado, Wildfire) records at county-year level from SHELDUS (Spatial Hazard Events and Losses Database for the United States).

6 Risk explanations

In Section 4, we find that the return predictability of $GEORET$ cannot be explained by well-known risk factors, such as the Fama-French five-factors and the momentum factor. Nevertheless, it is still possible that other unknown risks could drive our results. This would be the case, for example, if geo-peers' returns can somehow proxy for regional macroeconomic risks, which would then lead to changes in focal firms' discount rates. We conduct several tests in this section to examine this possibility.

6.1 Returns around earnings announcements

First, we examine stock price reactions around subsequent earnings announcements. This approach is widely used in the literature (see, for example, Bernard and Thomas (1989); La Porta, Lakonishok, Shleifer, and Vishny (1997); Engelberg, McLean, and Pontiff (2018)). The idea is intuitive: if an anomaly is associated with mispricing, then it will be stronger in the earnings announcement window, as the announcement of these earnings helps to correct investor expectation errors about firms' future cash flows. In contrast, if abnormal return is driven by exposure to unobserved risks, then the subsequent returns should accrue more evenly over subsequent periods. To conduct this test, we conduct panel regression analysis following the methdology of Engelberg, McLean, and Pontiff (2018). Our unit of observation is firm-day rather than firm-month in this test. Specifically, we regress the daily return of a stock ($DRET$) on the last month geo-peers' return ($GEORET$), an earnings announcement window dummy ($EDAY$), and the interaction between the two variables. We also control for day fixed effects and a set of control variables, including the lagged values for each of the past ten days for stock returns, stock returns squared, and trading volume.

We present our results in Table 7. The earnings announcement window is defined as either the one-day window (Columns 1 to 2) or a three-day window (Columns 3 to 4), centered on the earnings announcement date. The significantly positive coefficient on $GEORET$ suggest that returns of geographic peers can predict focal firm's return on non-earnings announcement days. Also consistent with the earnings announcement premium literature (Frazzini and Lamont (2007)), the coefficient on earnings announcement date dummy is positive and highly significant. More importantly, we find the coefficient on the interaction term is positive and significant under all specifications. Consistent with the mispricing explanation, returns to the $GEORET$ strategy are much larger when earnings news are released. For example, in Column 1, the coefficient on $GEORET$ is 0.347 (t=2.82), while the interaction coefficient on $GEORET*EDAY1$ is 0.578 (t=2.04). The coefficients indicate that for a $GEORET$ value of 0.06 (two standard deviation change), expected returns are higher by 2.08 basis points on non-earnings announcement days, and by an additional 3.47 basis points on earnings announcement days. In other words, the return spread generated by the $GEORET$ hedged strategy is 166% higher during an earnings announcement window than that on non-announcement days. These results are extremely difficult to square with standard risk-based explanations.

6.2 Evidence from non-return-based outcomes

6.2.1 Forecasting earnings surprises

Our preferred explanation for the return predictability results is that investors have limited attention and are slow to incorporate information contained in returns of focal firm's geo-peers, and evidence so far support this mispricing explanation. However, disagreements about whether return predictability reflects risk versus mispricing are often difficult to resolve using only realized returns and risk proxies. This is because return predictability can be attributed to risk, even if the source of risk is not directly observable or measurable. As an alternative approach, we also examine whether $GEORET$ has predictive power for focal firm's non-return-based outcomes. Our first non-return-based metric is focal firms' standardized unexpected earnings (SUE). $SUEs$ capture unanticipated changes in firm's earnings and are not return-based, so this test would not be confounded by imperfect risk controls. At the same time, unexpected earnings are fundamental drivers of firm value, so results on earnings predictability could further confirm that the return predictability is due to changes in unexpected firm cash flows, rather than compensation for some unobservable risk.

To that end, we test whether the stock return of the geography-linked firms predicts future unexpected earnings of the focal firm. The dependent variable is standardized unexpected earnings (SUE), defined as the difference between the actual quarterly earnings per share (EPS) and analyst consensus forecast of quarterly EPS scaled by stock prices in the month before quarterly earnings announcement. The main explanatory variable of interest is lagged $GEORET$, computed using the past three month returns of focal firm's geo-peers. Control variables include the focal firm's own lagged $SUEs$, up to four quarters.

Table 8 contains regression results under various model specifications. Column 1 presents a simple regression of SUE on lagged $GEORET$, with firm and year-quarter fixed effects. The estimated coefficient on $GEORET_{t-1}$ is 0.0019 ($t = 2.12$). In Columns 2 and 3, we add the focal firms' own lagged SUEs as control variables, while Column 3 includes industry and year-quarter fixed effects. The results show that $GEORET$ continues to positively predict future $SUEs$. These results further confirm that the short-window announcement returns we documented in Section 6.1 are driven by $GEORET$'s ability to anticipate the directional changes in focal firm's future earnings.

6.2.2 Forecasting short interests

Because returns of geo-peers contain value-relevant information about focal firm, sophisticated investors may exploit such information in their trading decisions. Short sellers appear to fit the profile of informed traders in the equity market. A large literature shows that short sellers' positions can predict future stock returns and short sellers are particularly skilled at analyzing public information (see, among others, Boehmer, Huszar, and Jordan (2010) and Engelberg, Reed, and Ringgenberg (2012)). If short sellers exploit information contained in $GEORET$, they should trade in the direction indicated by $GEORET$.

To examine whether $GEORET$ predict short sellers' trades, we run Fama–MacBeth regressions of change in short interest ratio on lagged $GEORET$. Specifically, the dependent variable is the change in short interest ratio from the previous month, where short interest ratio is defined as number of shares shorted over total number of shares outstanding.[16] The main explanatory variable of interest is lagged $GEORET$. Controls include firm size, book-to-market ratio, gross profitability, asset growth, the stock's own lagged monthly return. If short sellers make use of information contained in stock performance of geo-peers, the coefficient on $GEORET_{t-1}$ should be negative.

Table 9 reports the time series averages of the cross-sectional regression coefficients. Column 1 shows the coefficient on $GEORET_{t-1}$ is -0.19 (t=-3.0). In column 2, we add the lagged industry return ($INDRET_{t-1}$) and lagged return of a portfolio of firms headquartered in the same state as the focal firm ($HQRET_{t-1}$) in regression. Compared to Column 1, the coefficient on $GEORET$ barely changes and remains highly significant with a t-statistics of -2.84. These results suggest that short sellers increase their short positions on focal firm when the recent stock performance of its geography-linked peers are poor. To sum up, the tests based on earnings surprises and short interests (both are non-return-based metrics) lend further support to the mispricing explanation.

6.3 Evidence from analyst information updating

In this section, we examine analyst forecasting behavior to provide direct evidence on the limited attention channel. This setting is particularly useful because analyst earnings forecast

[16]We get the month-end short interests data from Compustat.

revisions directly measure investors' belief updating process. If analysts are slow to carry information across geography-linked firms due to limited processing capacity, we should observe past forecast revisions of geographic peers predict future forecast revisions of focal firms. To test this hypothesis, we conduct a test similar to the return predictability test except that we use analyst forecast revisions of annual EPS instead of stock returns.

Table 10 presents the results. All of the regressions include lagged forecast revision, past 1-month and past 12-month (skipping the most recent month) return, log of market capitalization, and log of book-to-market ratio as control variables. The dependent variables, FRP and FRB, are the one-month-ahead revision in consensus annual EPS forecast of the focal firm scaled by lagged stock price (Columns 1 and 2) and book value of equity per share (Columns 3 and 4), respectively. Our variable of interest is $GEOFRP_{t-1}$ ($GEOFRB_{t-1}$), defined as the average forecast revisions of the focal firm's geo-peers in the previous month, using the geographic linkage measure (GEO) constructed in Section 3 as weights. Consistent with our hypothesis, column 1 shows that the coefficient on $GEOFRP_{t-1}$ is 0.0528 (t=4.20), suggesting that the average forecast revision of geography-linked firms is a strong predictor of future revisions of focal firm.

In column 2, we add average forecast revisions of other economically related firms. Specifically, $INDFRP_{t-1}$ is the market capitalization-weighted average forecast revisions of all other firms in the same Fama-French 48 industry as the focal firm. $STATEFRP_{t-1}$ is the average forecast revisions of all other firms headquartered in the same state as the focal firm. $ANALYSTFRP_{t-1}$ is calculated as the weighted average forecast revisions of shared analyst-linked peers, using the weights defined in Ali and Hirshleifer (2020). The coefficient on $GEOFRP_{t-1}$ decreases to 0.0313 but remains highly significant (t=2.73). The coefficients on $INDFRP_{t-1}$, $STATEFRP_{t-1}$ and $ANALYSTFRP_{t-1}$ are also significantly positive, consistent with the results in Ali and Hirshleifer (2020). In columns 3 and 4, we show the same pattern holds using $GEOFRB_{t-1}$ (forecast revision scaled by book value of equity per share) as the measure. These results suggest that the return lead-lag effects that we show may at least partially be driven by analyst sluggish information updating. This is consistent with studies documenting inefficient forecast behaviors by analysts (Bouchaud, Krueger, Landier, and Thesmar (2019)). In addition, we find that the coefficients on lagged forecast revisions (FRP_{t-1} and FRB_{t-1}) are highly significant, consistent with prior studies that past forecast revisions of a stock are strong predictors of subsequent forecast revisions of the same stock. Given that analysts underreact to news about the same firm, it is very plausible that they might also underreact to information from other firms that are merely geographically linked to the focal firm.

7 Conclusion

Using detailed information of establishments owned by U.S. public firms from 1990 to 2012, we construct a novel measure of geographic linkage between firms that are from different industries and headquartered in different regions. We show that the returns of geography-linked firms have strong predictive power for focal firm returns and fundamentals. A long-short strategy based on this effect yields annual value-weighted alpha of approximately 6.5%. This effect is distinct from other cross-firm return predictability and is not easily attributable to risk-based explanations. It is more pronounced for focal firms that receive lower investor attention and are more costly to arbitrage. In addition, we find sell-side analysts similarly underreact, as their forecast revisions of geography-linked firms predict their future revisions of focal firms. Our results are broadly consistent with sluggish price adjustment to more nuanced geographic information.

References

Addoum, J. M., D. T. Ng, and A. Ortiz-Bobea, 2020, "Temperature shocks and establishment sales," *The Review of Financial Studies*, 33(3), 1331–1366.

Ali, U., and D. Hirshleifer, 2020, "Shared analyst coverage: Unifying momentum spillover effects," *Journal of Financial Economics*, 136(3), 649–675.

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

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

Antoniou, C., J. A. Doukas, and A. Subrahmanyam, 2016, "Investor sentiment, beta, and the cost of equity capital," *Management Science*, 62(2), 347–367.

Baker, M., and J. Wurgler, 2007, "Investor sentiment in the stock market," *Journal of economic perspectives*, 21(2), 129–152.

Bali, T. G., L. Peng, Y. Shen, and Y. Tang, 2014, "Liquidity shocks and stock market reactions," *The Review of Financial Studies*, 27(5), 1434–1485.

Barnatchez, K., L. D. Crane, and R. Decker, 2017, "An assessment of the national establishment time series (nets) database," .

Bernard, V. L., and J. K. Thomas, 1989, "Post-earnings-announcement drift: delayed price response or risk premium?," *Journal of Accounting research*, pp. 1–36.

Bernile, G., A. Kumar, and J. Sulaeman, 2015, "Home away from home: Geography of information and local investors," *The Review of Financial Studies*, 28(7), 2009–2049.

Boehmer, E., Z. R. Huszar, and B. D. Jordan, 2010, "The good news in short interest," *Journal of Financial Economics*, 96(1), 80–97.

Bouchaud, J.-P., P. Krueger, A. Landier, and D. Thesmar, 2019, "Sticky expectations and the profitability anomaly," *The Journal of Finance*, 74(2), 639–674.

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

Cohen, L., and A. Frazzini, 2008, "Economic links and predictable returns," *The Journal of Finance*, 63(4), 1977–2011.

Cohen, L., and D. Lou, 2012, "Complicated firms," *Journal of financial economics*, 104(2), 383–400.

Cohen, L., C. Malloy, and Q. Nguyen, 2020, "Lazy prices," *The Journal of Finance*, 75(3), 1371–1415.

Cooper, M. J., H. Gulen, and M. J. Schill, 2008, "Asset growth and the cross-section of stock returns," *the Journal of Finance*, 63(4), 1609–1651.

Corwin, S. A., and P. Schultz, 2012, "A simple way to estimate bid-ask spreads from daily high and low prices," *The Journal of Finance*, 67(2), 719–760.

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

Daniel, K., D. Hirshleifer, and A. Subrahmanyam, 1998, "Investor psychology and security market under-and overreactions," *the Journal of Finance*, 53(6), 1839–1885.

Daniel, K., and T. J. Moskowitz, 2016, "Momentum crashes," *Journal of Financial Economics*, 122(2), 221–247.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, 1990, "Noise trader risk in financial markets," *Journal of political Economy*, 98(4), 703–738.

DellaVigna, S., and J. M. Pollet, 2007, "Demographics and industry returns," *American Economic Review*, 97(5), 1667–1702.

——— , 2009, "Investor inattention and Friday earnings announcements," *The Journal of Finance*, 64(2), 709–749.

Dougal, C., C. A. Parsons, and S. Titman, 2015, "Urban vibrancy and corporate growth," *The Journal of Finance*, 70(1), 163–210.

Ellison, G., and E. L. Glaeser, 1997, "Geographic concentration in US manufacturing industries: a dartboard approach," *Journal of political economy*, 105(5), 889–927.

Engelberg, J., R. D. McLean, and J. Pontiff, 2018, "Anomalies and news," *The Journal of Finance*, 73(5), 1971–2001.

Engelberg, J. E., A. V. Reed, and M. C. Ringgenberg, 2012, "How are shorts informed?: Short sellers, news, and information processing," *Journal of Financial Economics*, 105(2), 260–278.

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

——— , 2015, "A five-factor asset pricing model," *Journal of financial economics*, 116(1), 1–22.

Fama, E. F., and J. D. MacBeth, 1973, "Risk, return, and equilibrium: Empirical tests," *The Journal of Political Economy*, pp. 607–636.

Frazzini, A., and O. A. Lamont, 2007, "The earnings announcement premium and trading volume," *NBER working paper*, (w13090).

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

Giglio, S., and K. Shue, 2014, "No news is news: do markets underreact to nothing?," *The Review of Financial Studies*, 27(12), 3389–3440.

Giroud, X., and H. M. Mueller, 2019, "Firms' Internal Networks and Local Economic Shocks," *American Economic Review*, 109(10), 3617–49.

Glaeser, E. L., J. Kolko, and A. Saiz, 2001, "Consumer city," *Journal of economic geography*, 1(1), 27–50.

Heider, F., and A. Ljungqvist, 2015, "As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes," *Journal of financial economics*, 118(3), 684–712.

Hirshleifer, D., and S. H. Teoh, 2003, "Limited attention, information disclosure, and financial reporting," *Journal of accounting and economics*, 36(1-3), 337–386.

Hoberg, G., and G. Phillips, 2016, "Text-based network industries and endogenous product differentiation," *Journal of Political Economy*, 124(5), 1423–1465.

Hoberg, G., and G. M. Phillips, 2018, "Text-based industry momentum," *Journal of Financial and Quantitative Analysis*, 53(6), 2355–2388.

Hong, H., and J. C. Stein, 1999, "A unified theory of underreaction, momentum trading, and overreaction in asset markets," *The Journal of finance*, 54(6), 2143–2184.

Hou, K., C. Xue, and L. Zhang, 2015, "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, 28(3), 650–705.

Ikenberry, D., J. Lakonishok, and T. Vermaelen, 1995, "Market underreaction to open market share repurchases," *Journal of financial economics*, 39(2-3), 181–208.

Jaffe, A. B., et al., 1986, "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," *American Economic Review*, 76(5), 984–1001.

Jaffe, A. B., M. Trajtenberg, and R. Henderson, 1993, "Geographic localization of knowledge spillovers as evidenced by patent citations," *the Quarterly journal of Economics*, 108(3), 577–598.

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

La Porta, R., J. Lakonishok, A. Shleifer, and R. Vishny, 1997, "Good News for Value Stocks: Further Evidence on Market Efficiency," *Journal of Finance*, pp. 859–874.

Lee, C. M., S. T. Sun, R. Wang, and R. Zhang, 2019, "Technological links and predictable returns," *Journal of Financial Economics*, 132(3), 76–96.

Loughran, T., and P. Schultz, 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.

Marshall, A., 1920, "Principles of Economics, 8-th edition," *Marshall7621Principles of Economics*, p. 621.

Menzly, L., and O. Ozbas, 2010, "Market segmentation and cross-predictability of returns," *The Journal of Finance*, 65(4), 1555–1580.

Moretti, E., 2004, "Human capital externalities in cities," in *Handbook of regional and urban economics*. Elsevier, vol. 4, pp. 2243–2291.

Moskowitz, T. J., and M. Grinblatt, 1999, "Do industries explain momentum?," *The Journal of finance*, 54(4), 1249–1290.

Neumark, D., B. Wall, and J. Zhang, 2011, "Do small businesses create more jobs? New evidence for the United States from the National Establishment Time Series," *The Review of Economics and Statistics*, 93(1), 16–29.

Novy-Marx, R., 2013, "The other side of value: The gross profitability premium," *Journal of Financial Economics*, 108(1), 1–28.

Parsons, C. A., R. Sabbatucci, and S. Titman, 2018, "Geographic lead-lag effects," *Available at SSRN 2780139*.

Parsons, C. A., J. Sulaeman, and S. Titman, 2018, "The geography of financial misconduct," *The Journal of Finance*, 73(5), 2087–2137.

Patton, A. J., and M. Verardo, 2012, "Does beta move with news? Firm-specific information flows and learning about profitability," *The Review of Financial Studies*, 25(9), 2789–2839.

Peng, L., and W. Xiong, 2006, "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, 80(3), 563–602.

Pirinsky, C., and Q. Wang, 2006, "Does corporate headquarters location matter for stock returns?," *The Journal of Finance*, 61(4), 1991–2015.

Pontiff, J., 2006, "Costly arbitrage and the myth of idiosyncratic risk," *Journal of Accounting and Economics*, 42(1), 35–52.

Shumway, T., 1997, "The delisting bias in CRSP data," *The Journal of Finance*, 52(1), 327–340.

Smajlbegovic, E., 2019, "Regional economic activity and stock returns," *Journal of Financial and Quantitative Analysis*, 54(3), 1051–1082.

Stambaugh, R. F., J. Yu, and Y. Yuan, 2012, “The short of it: Investor sentiment and anomalies,” *Journal of Financial Economics*, 104(2), 288–302.

——— , 2015, “Arbitrage asymmetry and the idiosyncratic volatility puzzle,” *The Journal of Finance*, 70(5), 1903–1948.

Tuzel, S., and M. B. Zhang, 2017, “Local risk, local factors, and asset prices,” *The Journal of Finance*, 72(1), 325–370.

Tversky, A., and D. Kahneman, 1974, “Judgment under uncertainty: Heuristics and biases,” *science*, 185(4157), 1124–1131.

Wurgler, J., and E. Zhuravskaya, 2002, “Does arbitrage flatten demand curves for stocks?,” *The Journal of Business*, 75(4), 583–608.

Table 1 Summary Statistics

This table presents summary statistics for the key variables used in the cross-sectional regressions. The sample includes all NYSE/Amex/Nasdaq-listed securities with share codes 10 or 11 that are contained in the CRSP/Compustat merged data file. Financial firms (Fama-French 48 industry code between 44 and 47) and stocks with prices less than $1 at portfolio formation are excluded. All variables except for future stock returns are winsorized within each cross-section at 1% and 99% levels. All statistics is computed cross-sectionally (for each calendar month) and then averaged across all months. % Value of CRSP is the total market capitalization of our sample firms as a percentage of the total market capitalization of the CRSP universe, computed each month and averaged across all months. Panel A reports the sample coverage statistics and descriptive statistics for the key variables. Panel B reports pairwise correlations, with 5% statistical significance indicated in bold. All variable definitions are in Appendix Table A1. The sample consists of 668,117 firm-month observations spanning 1990 to 2013.

Panel A: Descriptive statistics

	Mean	Std	Min	25PC	Median	75PC	Max
# of Firms	2320	347	1618	2082	2355	2556	2977
% Value of CRSP	0.57	0.07	0.46	0.50	0.58	0.63	0.70
Average # of geo-peers per focal firm	795	651	1.70	307	594	1098	3022
GEO	0.09	0.20	0.00	0.00	0.02	0.07	1.00
GEORET	0.01	0.03	-0.06	0.00	0.01	0.03	0.10
RET	0.01	0.15	-0.67	-0.07	0.00	0.08	1.72
INDRET	0.02	0.03	-0.04	0.00	0.02	0.03	0.08
HQRET	0.01	0.03	-0.06	0.00	0.01	0.03	0.09
RET(t-1)	0.01	0.14	-0.33	-0.07	0.00	0.08	0.54
SIZE	12.40	1.96	8.43	10.98	12.30	13.68	17.44
BM	0.68	0.59	0.04	0.29	0.52	0.87	3.36
GP	0.39	0.29	-0.55	0.22	0.36	0.53	1.32
AG	0.25	0.73	-0.46	-0.02	0.08	0.24	6.08
MOM	0.16	0.58	-0.71	-0.19	0.05	0.36	2.78

Panel B: Pearson (Spearman) correlations below (above) the diagonal

		1	2	3	4	5	6	7	8	9	10
GEORET(t-1)	1		**0.014**	**0.078**	**0.324**	**0.058**	**0.006**	0.005	**0.006**	-0.005	**0.016**
RET	2	**0.014**		**0.018**	**0.012**	**-0.023**	**0.040**	**0.011**	**0.024**	-0.005	**0.045**
INDRET(t-1)	3	**0.095**	**0.021**		**0.079**	**0.115**	**-0.011**	0.000	**0.020**	-0.002	0.003
HQRET(t-1)	4	**0.316**	**0.012**	**0.104**		**0.158**	-0.002	0.009	0.002	**-0.010**	**0.015**
RET(t-1)	5	**0.062**	**-0.018**	**0.118**	**0.174**		**0.017**	**0.020**	**0.022**	**-0.016**	**0.016**
SIZE	6	-0.003	-0.009	**-0.012**	-0.004	**-0.030**		**-0.370**	**-0.026**	**0.204**	**0.109**
BM	7	0.005	**0.020**	0.003	**0.009**	**0.035**	**-0.387**		**-0.173**	**-0.269**	**-0.142**
GP	8	0.001	**0.016**	**0.017**	0.001	**0.014**	**-0.016**	**-0.105**		**0.015**	**0.069**
AG	9	-0.004	**-0.022**	-0.004	**-0.008**	**-0.027**	**0.074**	**-0.159**	**-0.096**		**0.021**
MOM	10	**0.014**	**0.024**	0.006	**0.013**	0.003	**0.030**	**-0.115**	**0.058**	0.008	

Table 2 Fundamental Linkages

This table reports the panel regression results of fundamental linkages between focal firm and its geography-linked peers. Sales growth(t) is calculated as Sales per share(t) /Sales per share(t-1) −1. Profit growth is calculated as (Profit(t) – Profit(t-1))/average (Assets(t), Assets(t-1)), where Profit is measured as operating income before depreciation (Compustat data item OIBDP). Geo sales growth is the weighted average sales growth of focal firm's geography-linked peers, using the geographic linkage measure defined in section 3. Industry sales growth is measured as the market capitalization-weighted average sales growth of all other firms in the same Fama-French 48 industry as the focal firm. Same-state sales growth is measured as the equal-weighted average sales growth of all other firms headquartered in the same state as the focal firm. Analyst sales growth is calculated the weighted average sales growth of analyst-linked peers, using the weights defined in Ali and Hirshleifer (2019). The profit growth of peer firms is defined similarly. The sample is limited to firms with December fiscal year ends. All variables are measured at the end of each calendar year and are winsorized at the 1% and 99% levels. All regressions include year fixed effect and size and book-to-market ratio as control variables. *T*-statistics based on standard errors clustered by year are shown below coefficient estimates. The sample period is from 1990 to 2013.

	sales_growth(t)	sales_growth(t)	profit_growth(t)	profit_growth(t)
Geo sales growth(t)	0.311***	0.199*		
	(3.25)	(1.90)		
Same-state sales growth(t)		0.112***		
		(3.00)		
Industry sales growth(t)		0.186***		
		(4.23)		
Analyst sales growth(t)		0.802***		
		(6.50)		
Geo profit growth(t)			0.526***	0.124**
			(6.26)	(2.26)
Same-state profit growth(t)				0.0550***
				(2.83)
Industry profit growth(t)				0.308***
				(8.28)
Analyst profit growth(t)				0.887***
				(7.95)
Size(t)	-0.0292***	-0.0156***	0.00371**	0.00208
	(-7.37)	(-6.05)	(2.72)	(1.62)
BM(t)	-0.0351***	-0.0186*	-0.00625***	-0.00839***
	(-3.91)	(-2.06)	(-3.54)	(-4.21)
Constant	0.528***	0.253***	-0.0353*	-0.0207
	(9.76)	(6.78)	(-1.98)	(-1.21)
Year FE	YES	YES	YES	YES
adj. R-sq	0.015	0.038	0.020	0.061
N	36553	30070	40185	30277

Table 3 Geographic momentum strategy, abnormal returns 1990–2013.

This table reports abnormal returns and factor loadings for a geographic momentum strategy. Firms are ranked and assigned into decile portfolios at the beginning of every calendar month, based on the prior-month return to a portfolio of their geography-linked peer firms (*GEORET*). We exclude geographic peers from the same industry (based on Fama-French 48 industry groups) and headquartered in the same state as the focal firm when constructing *GEORET*. All stocks are equal- (value-) weighted within a given portfolio, and the portfolios are rebalanced every calendar month to maintain equal- (value-) weights. All non-financial stocks with stock price greater than $1 at portfolio formation are included. Excess return is the raw return of the portfolio over the risk-free rate. Alpha is the intercept from a regression of monthly excess return on factor returns. Factor returns are from the Kenneth French Data Library, and factor models include: CAPM model; the Fama-French (1993) three-factor model; a four-factor model including Fama-French three-factor and Carhart's (1997) momentum factor, Fama-French (2015) five-factor model, seven-factor model (Fama-French five-factor plus the momentum and short-term reversal factor) and Q-factors model of Hou, Xue and Zhang (2015). L/S is the alpha of a zero-cost portfolio that holds the top 10% stocks ranked by *GEORET* and sells short the bottom 10%. Panel B reports the alpha and the risk factor loadings, where the benchmark is a five-factor model (Fama-French three-factor plus the momentum and short-term reversal factor). Returns and alphas are in monthly percent, *t*-statistics are shown below the coefficient estimates, with 5% statistical significance indicated in bold.

Panel A: Portfolio returns

Decile	Excess returns (%)	CAPM alpha (%)	3-Factor alpha (%)	4-Factor alpha (%)	5-Factor alpha (%)	7-Factor alpha (%)	Q-Factor alpha (%)
1	0.40	-0.28	-0.21	-0.28	-0.17	-0.24	-0.15
(low)	(1.24)	(-1.77)	(-1.45)	(-1.92)	(-1.07)	(-1.50)	(-0.86)
10	0.94	0.25	0.32	0.26	0.40	0.36	0.47
(high)	(2.79)	(1.35)	(1.81)	(1.41)	(2.14)	(2.07)	(2.26)
L/S	**0.41**	**0.44**	**0.44**	**0.41**	**0.40**	**0.41**	**0.38**
(Equal-weights)	(2.97)	(3.19)	(3.24)	(3.08)	(2.73)	(3.09)	(2.59)
L/S	**0.54**	**0.54**	**0.53**	**0.53**	**0.57**	**0.60**	**0.62**
(value-weights)	(2.62)	(2.50)	(2.45)	(2.40)	(2.37)	(2.65)	(2.64)

Panel B: Risk factor loadings

	Alpha (%)	MKT	SMB	HML	MOM	ST_Rev
1	-0.30	1.03	0.01	-0.21	0.09	0.002
(low)	(-2.04)	(24.52)	(0.20)	(-3.18)	(2.28)	(2.40)
10	0.27	1.08	0.22	-0.25	0.06	-0.001
(high)	(1.55)	(18.19)	(2.74)	(-2.64)	(1.27)	(-0.94)
L/S	**0.44**	0.02	0.06	0.01	0.01	-0.003
(Equal-weights)	(3.46)	(0.64)	(1.13)	(0.14)	(0.22)	(-5.14)
L/S	**0.57**	-0.04	0.21	-0.04	-0.03	-0.003
(value-weights)	(2.72)	(0.79)	(2.00)	(-0.44)	(-0.60)	(-2.99)

Table 4 Fama-MacBeth regressions, 1990-2013

This table reports the result for Fama-MacBeth return forecasting regressions. The sample period is from 1990 to 2013. The dependent variable is the focal firm's monthly return (in percentage) *RET* and the key explanatory variable of interest is lagged geography-linked firms' return (*GEORET*). In Column 2, we add focal firm's lagged value-weighted industry return (*INDRET*) and the lagged value-weighted return of a portfolio of firms headquartered in the same state as the focal firm *(HQRET)*. In Column 3, we add the focal firm's lagged tech-peer return *(TECHRET)* constructed following Lee et al. (2019). In Column 4, a portfolio of focal firm's pseudo-conglomerate returns (*PCRET*) is added based on Compustat Segment data following Cohen and Lou (2012). In Column 5, we add the lagged returns from a portfolio of the focal firm's supplier (*SUPPRET*) and customer (*CUSTRET*) industries. These portfolios are constructed using Bureau of Economic Analysis (BEA) Input-Output data (at the summary industry level) following Menzly and Ozbas (2010). In Column 6, we add the lagged returns of focal firms' product market peers (*TNICRET*) following Hoberg and Phillips (2018). In Column 7, we add the lagged returns of stocks are connected through shared analyst coverage *(CFRET)* following Ali and Hirshleifer (2019). We also control for firm size (*SIZE*), book-to-market ratio (*BM*), gross profitability (*GP*), asset growth (*AG*), the firm's own lagged monthly return (*RET*(*t*-1)), and medium-term price momentum (*MOM*). Other variables are defined in Appendix Table A1. The sample excludes financial firms (Fama-French 48 industry code between 44 and 47) and stocks with a price less than $1 at portfolio formation. Cross-sectional regressions are run every calendar month, and the standard errors are Newey-West adjusted (up to 12 lags) for heteroskedasticity and autocorrelation. Fama-MacBeth t-statistics are reported below the coefficient estimates. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	RET (%)	RET (%)	RET (%)	RET (%)	RET (%)	RET (%)	RET (%)
GEORET	8.317***	5.957***	5.489***	4.198**	9.463**	6.281***	3.086**
	(4.81)	(4.01)	(2.60)	(2.26)	(2.58)	(4.09)	(2.35)
INDRET		12.08***	5.818***	7.569***	-2.130	8.432***	5.140***
		(5.48)	(2.67)	(4.37)	(-0.25)	(3.45)	(2.78)
HQRET		5.849***	4.755***	6.381***	6.273**	4.635***	3.050***
		(5.64)	(3.11)	(4.19)	(2.05)	(3.64)	(2.80)
TECHRET			8.583***				
			(3.95)				
PCRET				5.800***			
				(3.61)			
SUPPRET					3.442		
					(0.42)		
CUSTRET					4.725		
					(0.54)		
TNICRET						0.891**	
						(2.10)	
CFRET							14.30***
							(7.13)
RET(t-1)	-2.760***	-3.173***	-4.485***	-4.704***	-4.451***	-2.630***	-4.007***
	(-5.50)	(-6.16)	(-7.23)	(-7.96)	(-3.84)	(-3.96)	(-7.33)
SIZE	-0.0425	-0.0392	-0.0687	0.000813	-0.0485	-0.0735	-0.0560
	(-0.81)	(-0.75)	(-1.01)	(0.02)	(-0.53)	(-1.07)	(-1.10)
BM	0.421**	0.444***	0.438*	0.622***	0.779***	0.317	0.367**
	(2.50)	(2.74)	(1.90)	(3.71)	(2.98)	(1.54)	(2.53)
GP	0.681***	0.649***	0.771**	0.833***	1.287**	0.578*	0.611***
	(2.89)	(2.85)	(2.38)	(3.47)	(2.53)	(1.78)	(2.73)
AG	-0.429***	-0.420***	-0.430***	-0.383**	-0.482	-0.544***	-0.422***
	(-6.23)	(-6.00)	(-4.39)	(-2.45)	(-0.88)	(-5.39)	(-6.00)
MOM	0.465	0.460	0.274	0.262	-0.108	-0.0546	0.472
	(1.43)	(1.38)	(0.87)	(0.63)	(-0.24)	(-0.13)	(1.36)
Constant	0.951	0.711	1.070	-0.189	0.705	0.936	1.000
	(0.99)	(0.75)	(0.87)	(-0.21)	(0.46)	(0.73)	(1.02)
Average R-sq	0.0357	0.039	0.0536	0.0493	0.0894	0.0489	0.0518
N	723764	668117	257213	147494	171365	399911	532062

Table 5 Arbitrage costs and limited attention

This table reports the results of a set of cross-sectional (time-series) analyses to evaluate the sensitivity of geographic momentum to proxies for limited attention, arbitrage costs and investor sentiment. The tests are Fama-MacBeth return forecasting regressions where the dependent variable *RET* is the monthly focal firm stock return (in percentage). The explanatory variables are the lagged geography-linked firms' return (*GEORET*), lagged industry return (*INDRET*), lagged value-weighted return of a portfolio of firms headquartered in the same state as the focal firm *(HQRET)*, plus a number of interaction terms. *SIZE* is the natural log of market capitalization at the end of the previous fiscal year. *IO* is the percentage of institutional ownership at the end of the previous fiscal-year end. *CANALYST* is the average number of analysts covering the focal firm and geography-linked firms at the previous year end following Ali and Hirshleifer (2019). *IDVOL* is the standard deviation of the residuals from a regression of daily stock excess returns in the pre-30 days on the Fama and French (1993) factors (at least ten daily returns required). *ILLIQUIDITY* and *SPREAD* are the Amihud illiquidity and bid-ask spread of the firm at the end of the previous month, respectively. *SENTIMENT* is the Baker and Wurgler (2007) sentiment index based on first principal component of five sentiment proxies. All variables are defined in Appendix Table A1. All the interaction terms except for the *SENTIMENT* are based on indicator variables that take the value of one if the underlying variable is above the cross-sectional median, and zero otherwise. For investor sentiment, we create a dummy variable equals one if SENTIMENT is above sample media, and zero otherwise. The usual firm-level controls are also included. The sample excludes financial firms (Fama-French 48 industry code between 44 and 47) and stocks with a price less than $1 at portfolio formation. Cross-sectional regressions are run every calendar month, and the standard errors are Newey-West adjusted (up to 12 lags) for heteroskedasticity and autocorrelation. Fama-MacBeth t-statistics are reported below the coefficient estimates. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from 1990 to 2013.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	RET (%)	RET (%)	RET (%)	RET (%)	RET (%)	RET (%)	RET (%)
GEORET	8.319***	7.879***	5.882***	2.621**	3.044**	3.339*	4.838***
	(4.36)	(4.40)	(3.62)	(1.98)	(2.11)	(1.71)	(4.00)
INDRET	12.18***	11.80***	11.15***	12.11***	12.04***	11.94***	12.080***
	(5.51)	(5.50)	(5.09)	(5.55)	(5.51)	(5.33)	(5.48)
HQRET	5.991***	5.894***	4.983***	5.498***	5.480***	5.468***	5.849***
	(5.76)	(5.72)	(4.28)	(5.50)	(5.42)	(5.01)	(5.64)
GEORET*SIZE>Median	-5.463**						
	(-2.48)						
GEORET*IO>Median		-4.146**					
		(-2.02)					
GEORET*CANALYST>Median			-1.316				
			(-0.70)				
GEORET*IDVOL>Median				5.318***			
				(2.61)			
GEORET*ILLIQUIDITY>Median					4.595**		
					(2.54)		
GEORET*SPREAD>Median						5.753**	
						(2.41)	
GEORET*SENTIMENT>Median							1.119*
							(1.87)
RET(t-1)	-3.172***	-3.220***	-3.368***	-3.278***	-3.242***	-3.172***	-3.173***
	(-6.15)	(-6.24)	(-6.18)	(-6.43)	(-6.34)	(-6.15)	(-6.16)
SIZE	-0.0556	-0.0516	-0.0980	-0.0660	-0.0393	-0.0556	-0.039
	(-1.14)	(-1.02)	(-1.51)	(-1.58)	(-0.88)	(-1.14)	(-0.75)
BM	0.446***	0.410***	0.302*	0.413***	0.437***	0.446***	0.444***
	(2.76)	(2.64)	(1.83)	(2.70)	(2.84)	(2.76)	(2.74)
GP	0.653***	0.638***	0.631***	0.635***	0.648***	0.653***	0.649***
	(2.89)	(2.83)	(2.62)	(2.90)	(2.98)	(2.89)	(2.85)
AG	-0.419***	-0.419***	-0.429***	-0.408***	-0.411***	-0.419***	-0.420***
	(-6.07)	(-6.14)	(-6.10)	(-5.79)	(-6.03)	(-6.07)	(-6.00)
MOM	0.463	0.464	0.549	0.500	0.483	0.463	0.460
	(1.39)	(1.41)	(1.58)	(1.56)	(1.47)	(1.39)	(1.38)
CONSTANT	0.768	0.867	1.477	1.176	0.782	0.768	0.879
	(0.84)	(0.92)	(1.32)	(1.53)	(0.95)	(0.84)	(1.21)
Average R-sq	0.04	0.041	0.047	0.043	0.043	0.04	0.0424
N	668117	661189	546712	668116	667847	668117	668117

Table 6 Testing the Common Exposure Channel and the Spillover Channel

This table reports the Fama-MacBeth regressions of returns on *GEORET*, predicted regional economic activity ($PREA^{\perp}$) and natural hazards experienced by focal firm's geographic peers (*GEO_HAZARD*). The sample period is from 1990 to 2013. The dependent variable is the focal firm's monthly return (in percentage) *RET*. Following Smajlbegovic (2019), the predicted regional economic activity proxy, PREA, is constructed from a linear combination of predicted state economic activity growth rates weighted by firm's fraction of sales in all states it operates. The orthogonalized proxy $PREA^{\perp}$ is the sum of a constant and the residuals of cross-sectional regressions of PREA on return sensitivities to national economic activity and the Fama and French (1993) risk factors. *GEO_HAZARD* is the weighted average of dummies indicating geo-peers that are headquartered in states affected by any natural disaster at month t-1. In column (2), we restrict the sample to firms that do not have establishments in the areas affected by natural disasters at month t and month t-1. The natural disaster data is from SHELDUS (Spatial Hazard Events and Losses Database for the United States). We also include focal firm's lagged value-weighted industry return (INDRET) and the lagged value-weighted return of a portfolio of firms headquartered in the same state as the focal firm (HQRET), focal firm's firm size (*SIZE*), book-to-market ratio (*BM*), gross profitability (*GP*), asset growth (*AG*), the firm's own lagged monthly return (*RET*(*t*-1)), and medium-term price momentum (*MOM*). All variables are described in Appendix Table A1. All variables are winsorized at 1% and 99% in the cross-section. The sample excludes financial firms (Fama-French 48 industry code between 44 and 47) and stocks with a price less than $1 at portfolio formation. Cross-sectional regressions are run every calendar month, and the standard errors are Newey-West adjusted (up to 12 lags) for heteroskedasticity and autocorrelation. *T*-statistics are in parentheses. Coefficients marked with $*$, $**$, and $***$ are significant at 10%, 5%, and 1%, respectively.

	(1)	(2)
	RET(%)	RET(%)
GEORET	5.828***	
	(3.97)	
GEO_HAZARD		-1.227***
		(-2.71)
INDRET	11.900***	17.490***
	(5.31)	(5.63)
HQRET	5.659***	5.427*
	(5.94)	(1.82)
$PREA^{\perp}$	0.153***	
	(2.77)	
RET(t-1)	-3.290***	-3.764***
	(-6.82)	(-6.16)
SIZE	-0.040	-0.068
	(-0.76)	(-0.91)
BM	0.476***	0.606***
	(2.98)	(2.80)
GP	0.669***	0.069
	(2.95)	(0.17)
AG	-0.426***	-0.382***
	(-6.15)	(-4.14)
MOM	0.452	0.476
	(1.34)	(1.57)
Constant	0.537	1.758
	(0.58)	(1.38)
Average R-sq	0.039	0.044
N	662,877	107,111

Table 7 Returns on Earnings Announcement Days

This table reports regressions of announcement window daily returns *DRET* (in percentage) on the geography-linked firms' return (*GEORET*), earnings announcement date dummy variable (*EDAY*), and the interaction term between earnings announcement date dummy and *GEORET*. Geography-linked firms' return (*GEORET*) of a focal firm is calculated as the average monthly return of geographic peers weighted by pairwise geographic linkage measure defined in section 3. *EDAY* is a dummy variable which equals one if the daily observation is during an earnings announcement window, and zero otherwise. An earnings announcement window is defined as the one-day (Column 1 and 2) or three-day window (Column 3 and 4) centered on an earnings announcement date. Following Engelberg et al. (2018), we obtain earnings announcement dates from the Compustat quarterly database, examine the firm's trading volume scaled by market trading volume for the day before, the day of, and the day after the reported earnings announcement date, and define the day with the highest volume as the earnings announcement day. We control for day-fixed effect and other lagged control variables including lagged values for each of the past ten days for stock returns, stock returns squared, and trading volume. Key variables are described in Appendix Table A1. Standard errors are clustered on time. T-statistics are in parentheses, coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from 1990 to 2013.

	(1)	(2)	(3)	(4)
	One-day window		Three-day window	
	DRET (%)	DRET (%)	DRET (%)	DRET (%)
GEORET	0.347***	0.443***	0.339***	0.434***
	(2.82)	(3.44)	(2.76)	(3.38)
GEORET * EDAY	0.578**	0.623**	0.423**	0.458**
	(2.04)	(2.23)	(2.09)	(2.26)
EDAY	0.227***	0.264***	0.0823***	0.119***
	(13.50)	(15.66)	(7.37)	(10.55)
Lagged controls	No	Yes	No	Yes
Day FE	Yes	Yes	Yes	Yes
adj. R-sq	0.048	0.069	0.048	0.069
N	17953058	17875817	17953058	17875817

Table 8 Future Earnings Surprise

This table reports forecasting regressions of next-quarter's standardized unexpected earnings (*SUE*) on *GEORET*. *SUE* is defined as the difference between the actual quarterly earnings per share (EPS) and analyst consensus forecast of quarterly EPS scaled by stock prices in the month before quarterly earnings announcement. *GEORET* is calculated based on past three-month returns of geography-linked peers of the focal firm. We include firm fixed effect, year-quarter fixed effect in columns 1 and 2. In column 3, we include industry fixed effect and year-quarter fixed effect. We add one-quarter to four-quarter lags of the firm's own SUEs as control variables. Key variables are described in Appendix Table A1. All variables are winsorized at 1% and 99% in the cross-section. In parentheses below the coefficient estimates, t-statistics are reported using standard errors clustered in firm and time dimensions. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from 1990 to 2013.

	(1)	(2)	(3)
	SUE(t)	SUE(t)	SUE(t)
GEORET(t-1)	0.00191**	0.00239*	0.00257*
	(2.12)	(2.02)	(1.98)
SUE(t-1)		0.0667*	0.134***
		(1.97)	(3.25)
SUE(t-2)		0.0306	0.0812***
		(1.45)	(3.44)
SUE(t-3)		-0.00743	0.0362**
		(-0.54)	(2.73)
SUE(t-4)		0.0129	0.0543
		(0.45)	(1.51)
Constant	-0.000798***	-0.000386***	-0.000394***
	(-28.04)	(-8.90)	(-5.85)
Firm FE	YES	YES	NO
Industry FE	NO	NO	YES
Year-quarter FE	YES	YES	YES
adj. R-sq	0.065	0.069	0.043
N	163169	90493	90000

Table 9 Change in Short Interest Ratios

This table reports the Fama-MacBeth regressions of change in short interest ratios (*SR_change*) on *GEORET*. The dependent variable is the focal firm's monthly change in short interest ratios (*SR_change*) and the key explanatory variable of interest is one-month lagged geography-linked firms' return (*GEORET*). We also include focal firm's firm size (*SIZE*), book-to-market ratio (*BM*), gross profitability (*GP*), asset growth (*AG*), the firm's own lagged monthly return (*RET*(*t*-1)), and medium-term price momentum (*MOM*). Key variables are described in Appendix Table A1. All variables are winsorized at 1% and 99% in the cross-section. The sample excludes financial firms (Fama-French 48 industry code between 44 and 47) and stocks with a price less than $1 at portfolio formation. Cross-sectional regressions are run every calendar month, and the standard errors are Newey-West adjusted (up to 12 lags) for heteroskedasticity and autocorrelation. T-statistics are in parentheses. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from 1990 to 2013.

	(1) SR_change (%) (t)	(2) SR_change (%) (t)
GEORET(t-1)	-0.190***	-0.196***
	(-3.00)	(-2.84)
INDRET(t-1)		0.010
		(0.20)
HQRET(t-1)		0.142***
		(3.14)
RET(t-1)	0.176***	0.174***
	(3.89)	(3.70)
SIZE	0.002	0.002
	(1.31)	(1.24)
BM	-0.010***	-0.010***
	(-3.89)	(-4.43)
GP	-0.017***	-0.018***
	(-3.12)	(-3.23)
AG	0.021***	0.022***
	(3.66)	(3.66)
MOM	0.052***	0.052***
	(5.35)	(5.43)
Constant	-0.010	-0.009
	(-0.44)	(-0.37)
Average R-sq	0.013	0.014
N	630,026	581,907

Table 10 Lead-lag Effects in Analyst Forecast Revisions

This table reports the results of Fama-MacBeth regressions in which the dependent variable is the analyst forecast revision. *FRP* and *FRB* are the monthly change in analyst consensus forecast of annual EPS scaled by lagged stock price and book value of equity per share, respectively. *GEOFRP(t-1)* is the weighted average analyst forecast revisions of a focal firm's geography-linked peers in the previous month, using the geographic linkage measure defined in section 3 as weights. *INDFRP(t-1)* is measured as the market capitalization-weighted average forecast revisions of all other firms in the same Fama-French 48 industry as the focal firm. *STATEFRP(t-1)* is measured as the equal-weighted average forecast revisions of all other firms headquartered in the same state as the focal firm. *ANALYSTFRP(t-1)* is calculated as the weighted average forecast revisions of analyst-linked peers, using the weights defined in Ali and Hirshleifer (2019). *GEOFRB*, *STATEFRB*, *INDFRB*, *and ANALYSTFRB* are constructed in a similar way based on *FRB*. Control variables include the 1-month lagged forecast revisions, past 1-month return, past 12-month return (excluding the most recent month), log of market capitalization, and log of book-to-market ratio. Key variables are described in Appendix Table A1. The sample excludes financial firms (Fama-French 48 industry group between 44 and 47) and stocks with a price less than $1 at portfolio formation. Fama-MacBeth t-statistics are reported below the coefficient estimates. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively. The sample period is from 1990 to 2013.

	FRP(t)	FRP(t)	FRB(t)	FRB(t)
GEOFRP(t-1)	0.0528***	0.0313***		
	(4.20)	(2.73)		
STATEFRP(t-1)		0.0155**		
		(2.18)		
INDFRP(t-1)		0.134***		
		(7.31)		
ANALYSTFRP(t-1)		0.0920***		
		(8.84)		
GEOFRB(t-1)			0.0370***	0.0248*
			(2.83)	(1.94)
STATEFRB(t-1)				0.00777
				(1.39)
INDFRB(t-1)				0.0448***
				(5.10)
ANALYSTFRB(t-1)				0.0231***
				(4.38)
FRP(t-1)	0.0472***	0.0407***		
	(9.36)	(8.50)		
FRB(t-1)			0.0503***	0.0484***
			(7.59)	(7.40)
RET(t-1)	0.00905***	0.00905***	0.0204***	0.0204***
	(14.49)	(14.51)	(19.60)	(19.65)
RET(t-13, t-2)	0.00113***	0.00110***	0.00264***	0.00261***
	(8.52)	(8.51)	(9.34)	(9.35)
SIZE	0.000273***	0.000271***	0.000662***	0.000668***
	(15.24)	(15.22)	(14.73)	(15.51)
BM	-0.000488***	-0.000475***	0.00200***	0.00201***
	(-5.59)	(-5.98)	(6.87)	(6.95)
Constant	-0.00573***	-0.00551***	-0.0156***	-0.0154***
	(-16.21)	(-15.72)	(-17.10)	(-17.33)
Average R-sq	0.0383	0.0428	0.0341	0.0363
N	456785	454719	443155	441177

Figure 1. Cumulative performance of the trading strategy.
This graph shows the time series evolution of a $1 investment in each of three portfolios. The red line is a market (S&P500) portfolio, where dividends are reinvested in the market. The blue (green) line represents a long-only strategy that value-weights the top (bottom) 10% of firms ranked by the return of its geography-linked peers (*GEORET*) at the end of the previous month. The portfolios are monthly rebalanced and sample period is from 1990 to 2013.

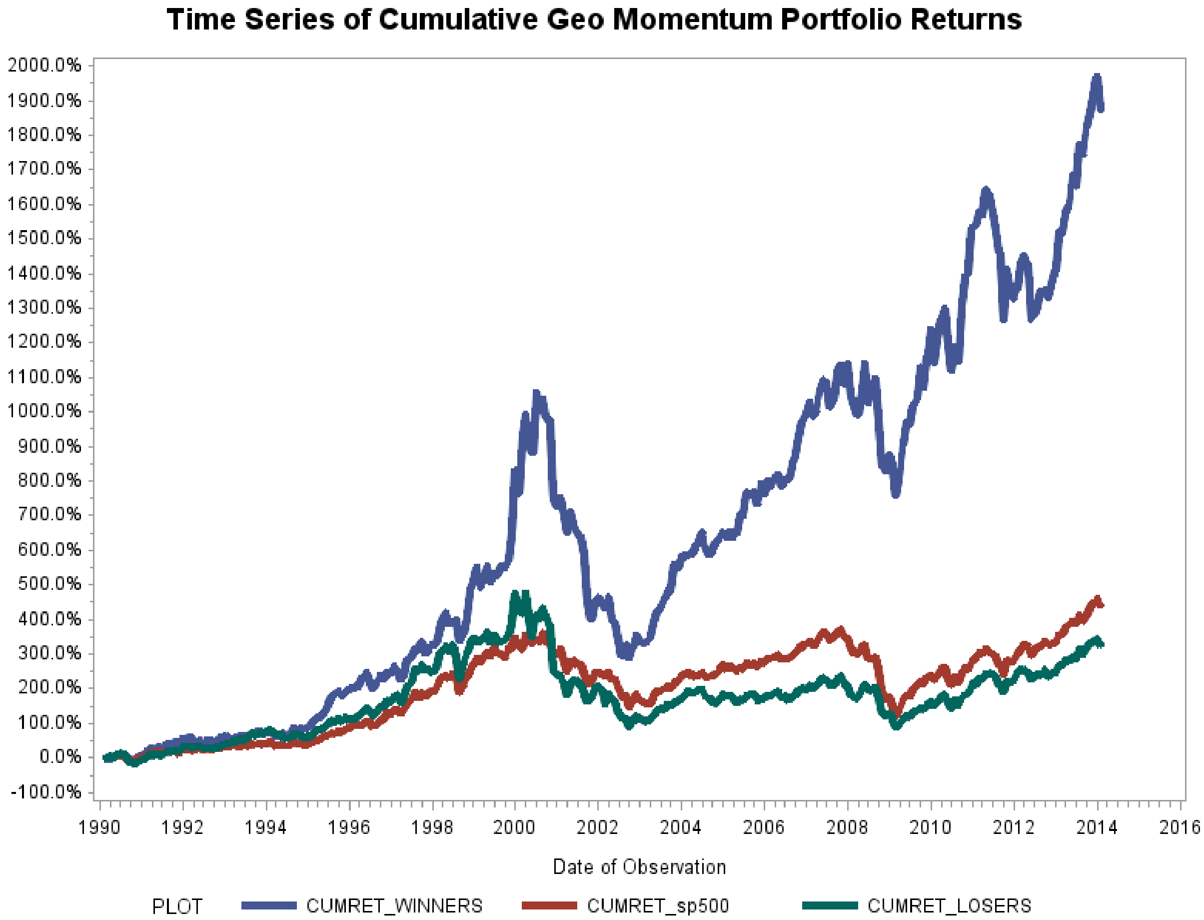
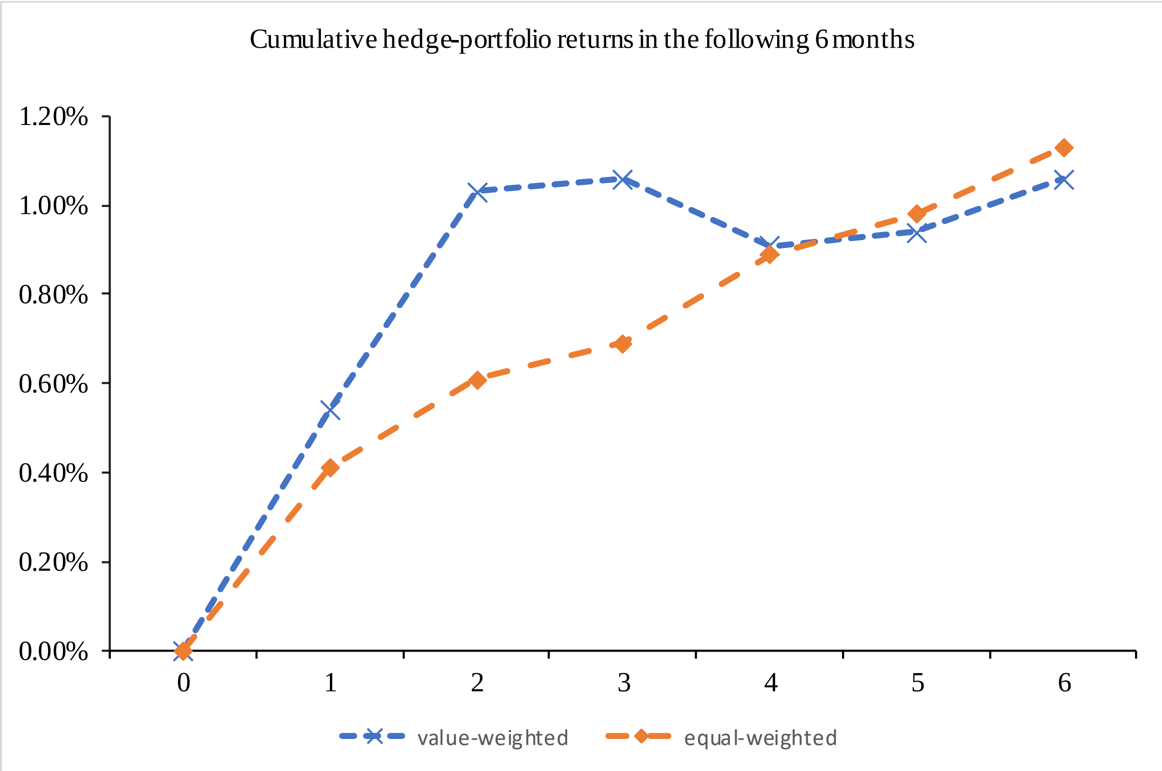


Figure 2. Long-Short Portfolio Performance Persistence
This figure plots the cumulative returns of the long-short portfolio in the six months after portfolio formation. At the beginning of every calendar month, all firms are ranked in ascending order on the basis of the return of a portfolio of its geography-linked peers (*GEORET*) at the end of the previous month. The ranked stocks are assigned to one of ten decile portfolios. All stocks are value- (equal-) weighted within each portfolio, and the portfolios are rebalanced every calendar month to maintain value- (equal-) weights. The long-short portfolio is a zero-cost portfolio that buys the top decile and sells short the bottom decile. The graph depicts the cumulative returns to both an equal-weighted (dashed line) and a value-weighted (dotted line) long-short portfolio. The sample excludes financial firms and stocks with a price of less than $1 at portfolio formation. The sample period is from 1990 to 2013.



Appendix Table A1: Variable Definitions

Variables	Definition
GEO	Geographic linkage measure GEO_{ijt} is defined the uncentered correlation of the distribution of establishment sales between two firms i and j across all counties in US. Establishment-level sales data is from NETS publicly listed database.
GEORET	Geography-linked return is defined as the weighted average return of a focal firm's geography-linked firms, using the geographic linkage GEO as weights.
RET	Stock monthly raw return adjusted for delisting bias following Shumway (1997).
INDRET	Industry return, defined as value-weighted average return of Fama-French 48 industries.
HQRET	Value-weighted return of a portfolio of firms headquartered in the same state as the focal firm.
SIZE	The natural logarithm of market capitalization at the end of June in each year.
BM	Book-to-market ratio is the most recent fiscal year-end report of book value divided by the market capitalization at the end of calendar year t-1. Book value equals the value of common stockholders' equity, plus deferred taxes and investment tax credits, and minus the book value of preferred stock.
GP	Gross profitability is defined as sales revenue minus cost of goods sold scaled by assets, following Novy-Marx (2013).
AG	Asset growth is defined as year-over-year growth rate of total asset, following Cooper, Gulen, and Schill (2008).
MOM	Medium-term price momentum variable, defined as focal firm's stock return for the last 12 months excluding the most recent month.
RET (-1)	Lagged monthly raw return, or short-term return reversal variable, defined as focal firm's stock return in month t-1.
SUE	Standardized unexpected earnings (SUE) is defined as the difference between the actual quarterly earnings per share (EPS) and analyst consensus forecast of quarterly EPS scaled by stock prices in the month before quarterly earnings announcement.
FRP (FRB)	One-month-ahead revision in consensus annual EPS forecast on the focal firm scaled by lagged stock price (book value of equity per share).
IO	The percentage of institutional ownership at the end of the previous fiscal-year end.
CANALYST	Average number of analysts covering the focal firm and geography-linked peers at the previous year-end.
SPREAD	Bid-ask spread is calculated based on daily high and low prices following Corwin and Schultz (2012).
ILLIQUIDITY	Following Amihud (2002), Illiquidity is defined as the average daily ratio of absolute stock return to the dollar trading volume within a month.
IDVOL	Idiosyncratic volatility is defined as the standard deviation of the residuals from a regression of daily excess stock returns on Fama and French (1993) three factors within a month (at least ten daily returns required) following Ang, Hodrick, Xing, and Zhang (2006).

Appendix Table A2 Long-horizon lags

This table reports Fama-MacBeth return forecasting regressions. The sample period is from 1990 to 2013. The dependent variable is the focal firm's monthly return (in percentage) *RET*. The key explanatory variables of interest are lagged geography-linked firms' return over past 6 months in column 1 (*GEORET (t-6, t-1)*), over past 12 months in column 2 (*GEORET(t-12, t-1)*), and over past 24 months in column 3 (*GEORET(t-24,t-1)*). Other variables are defined in Appendix Table A1. The sample excludes financial firms (Fama-French 48 industry code between 44 and 47) and stocks with a price less than $1 at portfolio formation. Cross-sectional regressions are run every calendar month, and the standard errors are Newey-West adjusted (up to 12 lags) for heteroskedasticity and autocorrelation. Fama-MacBeth t-statistics are reported below the coefficient estimates. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)
GEORET (t-6, t-1)	1.678**		
	(2.25)		
GEORET (t-12, t-1)		0.918**	
		(2.40)	
GEORET (t-24, t-1)			0.381
			(1.28)
INDRET	12.02***	11.48***	10.73***
	(5.43)	(5.33)	(5.02)
HQRET	6.249***	6.114***	5.901***
	(5.17)	(5.33)	(4.94)
RET(t-1)	-3.213***	-3.200***	-3.313***
	(-6.20)	(-6.10)	(-6.19)
SIZE	-0.0446	-0.0593	-0.0556
	(-0.83)	(-1.09)	(-1.01)
BM	0.469***	0.461***	0.430**
	(2.86)	(2.77)	(2.49)
GP	0.616***	0.606***	0.532**
	(2.70)	(2.61)	(2.18)
AG	-0.435***	-0.431***	-0.558***
	(-6.20)	(-5.90)	(-6.30)
MOM	0.412	0.365	0.266
	(1.23)	(1.07)	(0.78)
CONSTANT	0.727	0.945	1.001
	(0.79)	(0.99)	(0.98)
Average R-sq	0.0398	0.0397	0.0399
N	658799	647766	588004

Appendix Table A3 Robustness Tests.

This table reports various robustness tests for Fama-MacBeth return forecasting regressions. The sample period is from 1990 to 2013. The dependent variable is the focal firm's monthly return (in percentage) *RET* and the key explanatory variable of interest is lagged geography-linked firms' return (*GEORET*). In Column 1 and 2, we exclude stocks with price less than $5 or market capitalization below the 10th NYSE percentile, respectively. In column 3, we restrict our sample to focal firms with establishments in at least two counties. Columns 4 and 5 report the results for two subperiods: 1990-2001 and 2002–2013. Columns 6-8 shows the results by using GEORET constructed using geographic linkage measures lagged by 1, 3 and 5 years, respectively. In column 9, we construct the GEORET using the top 50 geo-peers of the focal firm. In Column 10, we construct geographic linkage measure using number of employees at firm establishments. Other variables are defined in Appendix Table A1. The sample excludes financial firms (Fama-French 48 industry code between 44 and 47) and stocks with a price less than $1 at portfolio formation. Cross-sectional regressions are run every calendar month, and the standard errors are Newey-West adjusted (up to 12 lags) for heteroskedasticity and autocorrelation. Fama-MacBeth t-statistics are reported below the coefficient estimates. Coefficients marked with *, **, and *** are significant at 10%, 5%, and 1%, respectively.

	Stock price greater than $5	Mktcap above 10th NYSE percentile	Establishments in >=2 counties	1990-2001	2002-2013	GEO lagged by 1 year	GEO lagged by 3 year	GEO lagged by 5 years	Top 50 geo-peer firms	Construct GEO using employment
GEORET	3.741***	4.989***	6.574***	6.406***	5.426***	5.870***	4.706***	4.638***	2.721***	6.112***
	(2.78)	(2.98)	(3.85)	(2.75)	(3.24)	(3.91)	(3.59)	(3.97)	(3.86)	(4.13)
INDRET	10.26***	8.992***	10.88***	17.24***	5.987**	11.79***	10.67***	8.789***	12.26***	12.21***
	(4.95)	(4.00)	(4.67)	(7.05)	(2.11)	(5.22)	(5.26)	(3.98)	(5.53)	(5.52)
HQRET	4.033***	3.558***	5.625***	6.585***	4.978***	6.043***	5.587***	5.450***	6.418***	5.758***
	(3.71)	(2.93)	(5.19)	(4.72)	(3.25)	(5.88)	(4.90)	(4.33)	(5.54)	(5.69)
RET(t-1)	-2.364***	-2.393***	-3.680***	-4.408***	-1.714***	-3.136***	-3.055***	-2.567***	-3.167***	-3.181***
	(-4.74)	(-4.07)	(-6.83)	(-6.68)	(-2.84)	(-6.10)	(-5.92)	(-4.96)	(-6.15)	(-6.21)
SIZE	-0.0101	-0.0410	-0.0985**	-0.0282	-0.0521	-0.0444	-0.0145	-0.00686	-0.0396	-0.0376
	(-0.26)	(-0.78)	(-2.14)	(-0.37)	(-0.77)	(-0.86)	(-0.29)	(-0.14)	(-0.76)	(-0.72)
BM	0.376*	0.502**	0.435***	0.488**	0.393*	0.457***	0.476***	0.368**	0.442***	0.447***
	(1.84)	(2.29)	(2.81)	(2.06)	(1.83)	(2.82)	(2.91)	(2.30)	(2.72)	(2.76)
GP	0.703***	0.562**	0.686***	0.912**	0.337	0.611***	0.597***	0.592**	0.647***	0.639***
	(2.87)	(2.02)	(3.08)	(2.45)	(1.63)	(2.71)	(2.69)	(2.54)	(2.86)	(2.81)
AG	-0.272***	-0.324***	-0.450***	-0.370***	-0.480***	-0.392***	-0.373***	-0.386***	-0.422***	-0.424***
	(-3.72)	(-4.85)	(-4.60)	(-4.60)	(-4.00)	(-5.16)	(-4.38)	(-3.72)	(-6.02)	(-6.00)
MOM	0.565*	0.479	0.222	1.175***	-0.385	0.399	0.308	0.255	0.461	0.457
	(1.76)	(1.34)	(0.56)	(7.04)	(-0.64)	(1.22)	(0.95)	(0.80)	(1.38)	(1.38)
CONSTANT	0.339	0.870	1.550*	0.274	1.228	0.777	0.335	0.330	0.773	0.702
	(0.43)	(0.87)	(1.82)	(0.21)	(0.86)	(0.83)	(0.36)	(0.35)	(0.81)	(0.74)
Average R-sq	0.0476	0.0591	0.0431	0.0448	0.0322	0.0396	0.0418	0.0422	0.0386	0.039
N	511251	397499	483354	374609	293508	668997	659754	630981	668117	669820

Appendix Table A4 Subsample tests for manufacturing and non-manufacturing firms

This table reports the Fama-MacBeth regressions of returns on *GEORET* for the subsample of manufacturing and non-manufacturing firms. Firms with 2-digit NAICS code in the range of 31-33 are classified as manufacturing firms while others are classified as non-manufacturing firms. The sample period is from 1990 to 2013. The dependent variable is the focal firm's monthly return (in percentage) *RET*. We also include focal firm's lagged value-weighted industry return (INDRET) and the lagged value-weighted return of a portfolio of firms headquartered in the same state as the focal firm (HQRET), focal firm's firm size (*SIZE*), book-to-market ratio (*BM*), gross profitability (*GP*), asset growth (*AG*), the firm's own lagged monthly return (*RET*(*t*-1)), and medium-term price momentum (*MOM*). All variables are described in Appendix Table A1. All variables are winsorized at 1% and 99% in the cross-section. The sample excludes financial firms (Fama-French 48 industry code between 44 and 47) and stocks with a price less than $1 at portfolio formation. Cross-sectional regressions are run every calendar month, and the standard errors are Newey-West adjusted (up to 12 lags) for heteroskedasticity and autocorrelation. *T*-statistics are in parentheses. Coefficients marked with ∗, ∗∗, and ∗∗∗ are significant at 10%, 5%, and 1%, respectively.

	Manufacturing	Non-manufacturing
	RET (%)	RET (%)
GEORET	5.385**	6.254***
	(2.57)	(3.78)
INDRET	9.546***	15.48***
	(3.99)	(4.97)
HQRET	4.881***	5.853***
	(3.33)	(5.04)
RET(t-1)	-3.993***	-2.465***
	(-7.15)	(-4.42)
SIZE	-0.0228	-0.0455
	(-0.41)	(-0.85)
BM	0.625***	0.315**
	(3.21)	(2.20)
GP	0.588*	0.750***
	(1.78)	(3.16)
AG	-0.427***	-0.390***
	(-4.93)	(-4.29)
MOM	0.368	0.538
	(1.10)	(1.56)
CONSTANT	0.657	0.626
	(0.65)	(0.65)
Average R-sq	0.0444	0.0438
N	346086	322031