Geographic Links and Predictable Returns

Zuben Jin and Frank Weikai Li*

Published in Journal of Business Finance and Accounting, 2024. DOI: 10.1111/jbfa.12782

Abstract: Using establishment-level data of 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. 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. The cross-firm information spillovers and return predictability are also stronger for geographic peers with economic linkages and with positive information. Our results are broadly consistent with sluggish price adjustment to nuanced geographic information.

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

1. Introduction

Economists have long recognized that location plays an important role in shaping economic growth through generating economies of scale in the 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 et al., 2015; Tuzel & Zhang, 2017), the speed of information transmission (Coval & Moskowitz, 2001; Malloy, 2005), the level of discount rate (Garcia & Norli, 2012), stock liquidity (Loughran & Schultz, 2005) and even financial misconduct (Parsons et al., 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 products are made (Bernile et al., 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

** 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. 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.

Funding information: Ministry of Education Academic Research, Fund Tier 1, Grant/Award Number: 20-C207-SMU-002

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 operating costs (such as labor costs and rents). In addition, natural disasters may occur in certain areas that disrupt the 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 neighboring firms (Jaffe et al., 1993) and consumption externalities among local residents (Glaeser et al., 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 linkage leads to comovement in firms' fundamentals, even for firms that operate in different industries and are 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 a 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 National Establishment Time-Series (NETS) database. This database provides addresses, as well as information on sales and employment, for each U.S. establishment owned by a public company over the period from 1989 to 2012. With these 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 the United States, $GEO_{ijt} = \frac{G_{it} * G'_{jt}}{\sqrt{(G_{it} * G'_{it}) * (G_{jt} * G'_{jt})}}$, where $G_{it} = (G_{it1}, G_{it2}, \ldots, G_{it3022})$ is a vector of firm i's proportional share of sales across 3022 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 captures the fundamental relationship between firms. We find that firm fundamentals (sales, costs and profits 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 had stores across major cities in the United States 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 state. 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-collar ones who buy organic food products and enjoy drinking premium coffee); hence, it is very likely that the operating performance and stock returns of the two firms comove with each other as both are exposed to the same local economic

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

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

conditions.[3] 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 industry 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 of year τ, we calculate the weighted average return of a portfolio of firms that are geographically linked to the focal firm, $GEORET_{it} = \sum_{j \neq i} GEO_{ij\tau-1} * RET_{jt} / \sum_{j \neq i} GEO_{ij\tau-1}$, where RET_{jt} is the return of firm j at month t and $GEO_{ij\tau-1}$ is the geographic linkage measure we construct using NETS data available at year $\tau - 1$.[4] 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 longs the focal firms whose geo-peers performed best in the prior month and shorts 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, including the geographic risk factor of Dissanayake (2021). 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 (Hoberg & Phillips, 2018; Moskowitz and Grinblatt, 1999), firms headquartered in the same regions (Parsons et al., 2020), firms that are linked along the supply chain (Cohen & Frazzini, 2008; Menzly & Ozbas, 2010), single- and multi-segment firms operating in the same industries (Cohen & Lou, 2012), and firms with similar technologies (Lee et al., 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 & Glaeser, 1997), it is likely that firms will have establishments largely overlapping with their industry peers geographically. Similarly, firms whose headquarters are 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 addition, we control for the focal firm's lagged tech-peer returns (Lee et al., 2019), focal firm's lagged pseudo-conglomerate returns (Cohen & Lou, 2012), focal firm's lagged supplier and customer industry returns (Menzly & Ozbas, 2010) and focal firm's product market peers' returns (Hoberg & Phillips, 2018). Lastly, a recent paper by Ali and Hirshleifer (2020) argues 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, same headquarter links, customer–supplier links, technology links, standalone-conglomerate firm links and shared analyst links.

After establishing the robustness of the lead–lag return effect among geographic peers, we conduct cross-sectional tests to examine factors impeding information diffusion across geo-peers. Our preferred explanation is that investors have limited attention and are slow to incorporate value-relevant information contained in the focal firm's geo-peers.

[3] In addition to fundamental comovement, we also predict a lead–lag return relation between Starbucks and Whole foods, which is built on an additional assumption that investors have limited attention and are slow in updating their expectation of a firm's value based on news of geographically linked firms. One reason behind investors' limited ability to infer information from related firms' news is that the structure of analyst business is organized at sector level, and there is insufficient degree of analyst common coverage on geographically linked firms (Parsons et al., 2020). In the case of Starbucks and Whole foods, we find that the number of analysts in both firms (as a fraction of analysts covering either firm) is only 7%. As a comparison, the number of analysts covering Starbucks and any other firms in the same Fama–French 48 industry is 12.7%.

[4] In our portfolio test, in order to ensure our results are distinct from the industry momentum effect (Moskowitz & Grinblatt, 1999) and same-headquarter lead–lag effect (Parsons et al., 2020), 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 constructing $GEORET_{it}$. The average percentages of geo-peers headquartered in the same state and from the same Fama–French 48 industry are about 12% and 6%, respectively.

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 have lower institutional ownership.[5] Second, the abnormal returns generated by our trading strategy raise the question of why the profits are not quickly arbitraged away by sophisticated 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. Third, using the Baker and Wurgler (2007) sentiment index, we find stronger return predictability during high-sentiment periods, suggesting that investors pay less attention to value-relevant fundamental information when they exhibit irrational exuberance.

We further explore the channel(s) through which information spillovers occur between geo-peers. We conjecture that economic linkages, such as supply chain relationships, industry, product market and technological linkages could facilitate information transfer across geographic peers. To test this conjecture, we separately construct *GEORET* based on whether geo-peers share some type of economic linkage with the focal firm. We then compare the differences in the predictive power of *GEORET* using Fama–MacBeth regressions. Consistent with our prediction, we find stronger return predictability of *GEORET* when the geo-peers share some type of economic linkage with the focal firm. We also find an asymmetry in the return predictability in the sense that positive news contained in geo-peers' past returns more strongly forecasts focal firm's future return. This is consistent with the in-built mechanism in conservative accounting for more timely recognition and reporting of bad news.

Although the return predictability effects we document are 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 explanations. First, we show the predictability of *GEORET* cannot be explained by firms' exposure to state-level macroeconomic conditions or aggregate opportunities as per Korniotis and Kumar (2013). Second, we examine the stock price reaction around earnings announcements. The idea is intuitive: earnings announcements help correct investor expectation errors about future cash flows; As a result, if the 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 the return predictability effect is driven by exposures to some unknown risks, strategy returns should accrue more evenly over subsequent trading days.[6] 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 nonannouncement day. This evidence is difficult to square with standard risk models.

Third, we find that geo-peers' returns significantly predict focal firms' subsequent earnings news (*SUEs*). *SUEs* are not return based, so this test is not confounded by imperfect controls for firm risks. In addition, this result, along with our finding that the return predictability of *GEORET* lasts for several months and does not reverse afterward, strongly suggests that the predictable return based on *GEORET* is driven by investor underreaction, not an overreaction or price-pressure effect. 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 Online Appendix provides a battery of other robustness tests. First, we document the robustness of the return predictability of *GEORET* to various perturbations such as removing microcap and low-priced 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 a significant geographic lead–lag effect. Third, we examine the sensitivity of our result to the staleness of the geographic linkage

[5] Also consistent with the idea that common analyst coverage expedites information flow between economically related firms (Parsons et al., 2020; Ali & Hirshleifer, 2020), we find weaker return predictability when the focal firm shares a large set of common analysts with its geo-peers, although this effect is not statistically significant.

[6] This test has been widely used in prior studies to separate mispricing from risk explanations (e.g., Bernard & Thomas, 1989; La Porta et al., 1997; Engelberg et al., 2018).

measure. Our results show that the effect declines slightly with a more "stale" geographic linkage measure but is still significant even when we use a 5-year-old geographic linkage measure. Fourth, the results are robust to various alternative thresholds used to define geo-peers. Fifth, our result persists if we construct geographic linkage using establishment employment data, which is less likely to be imputed than sales in NETS. Sixth, we find similarly strong geographic lead–lag effects using data from Exhibit 21, a section within or attached to 10-K fillings that provide the locations of firms' headquarters and material subsidiaries. Seventh, we show the predictive power of *GEORET* persists when we use the panel regression approach with firm and year-month fixed effects. Lastly, a placebo test shows insignificant return predictability based on the returns of matched geographically distant firms, suggesting our key result is not spurious.

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 alternative explanations by controlling local economic conditions and examining nonreturn-based outcomes. 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 et al. (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.[7] Studies also document underreaction is more likely in settings where the nature of information is less salient (DellaVigna & Pollet, 2007; Giglio & Shue, 2014) or when investors are being distracted (DellaVigna & 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.

Our study is also related to a growing literature on the implication of investors' limited attention to information diffusion and market efficiency. Several theoretical works present a framework for understanding market price dynamics when a subset of investors has limited attention (e.g., Hirshleifer & Teoh, 2003; Peng & 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 (Hoberg & Phillips, 2018; Moskowitz & Grinblatt, 1999), customer–supplier links (Cohen & Frazzini, 2008; Menzly & Ozbas, 2010), technology links (Lee et al., 2019) and shared analyst links (Ali & 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 the 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 investments than on their distant investments. Similarly, Malloy (2005) shows that geographically proximate analysts are more accurate than other analysts. Loughran and Schultz (2005) document that firms headquartered in rural areas have poorer information environments 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 et al. (2020) document a lead–lag return

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

effect among firms headquartered in the same state. Korniotis and Kumar (2013) find that state-level macroeconomic factors (e.g., unemployment and housing collateral ratios) can predict returns of stocks headquartered in those states.[8]

One limitation of these studies is that they all use the firm's headquarters to identify geographic footprint. However, as shown by Bernile et al. (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.[9] Furthermore, a firm rarely changes its headquarter location. As a result, a firm's geographic peers are largely static, which cannot account for its geographic expansion over time. Our novel measure of geographic linkage improves upon these dimensions as we can identify the degree to which a specific firm is connected to its geo-peers and how this link changes over time.

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 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 postal records.[10] Recent studies employing NETS data include Neumark et al. (2011), Heider and Ljungqvist (2015) and Addoum et al. (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. Table IA.1 in the Online Appendix compares the matched NETS sample to the Compustat sample. The matched NETS sample represents about 54% of the Compustat universe in terms of the number of Cfirms. The representativeness of NETS is fairly consistent across different industries except for two industries. Compared to the Compustat firms, the most underrepresented industries are the finance & insurance and real estate, rental and leasing.[11]

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 New York Stock Exchange (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 6-month gap between the fiscal-year end month and the portfolio formation date. Specifically, we first match the NETS data in

[8] Conceptually, the return predictability in our paper arises from investor-limited ability to infer information from geographically related firms. This is different from the channel of return predictability emphasized by Korniotis and Kumar (2013), which focuses on localized systematic risks and local-trading-induced mispricing.

[9] 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.

[10] Barnatchez et al. (2017) conduct a thorough assessment of the NETS data and conclude that NETS data are useful and convenient for studying business activity in high detail.

[11] Since we exclude financial firms (Fama–French 48 industry code between 44 and 47) in our empirical tests, these two industries will not affect the results much.

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 a nonmissing stock price and SIC classification code from CRSP and nonnegative 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 the United States 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 $G_{it} = (G_{it1}, G_{it2}, \dots, G_{it3022})$ is a vector of firm i's proportional share of sales across 3022 U.S. counties over year t. GEO_{ijt} has the 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.[12] Furthermore, this measure is symmetric in firm ordering (i.e., $GEO_{ijt} = GEO_{jit}$) and not directly affected by the length of the G vectors.[13]

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 a 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 weights. *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. The 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 $t-1$. The book value equals the value of common stockholders' equity, plus deferred taxes and investment tax credits, and minus the book value of the 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 the short-term reversal effect. Following Cooper et al. (2008), asset growth (*AG*) is defined as the year-over-year growth rate of total assets. Following Novy-Marx (2013), gross profitability (*GP*) is defined as sales revenue minus the 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). Analyst forecast data are obtained from I/B/E/S.

[12] As an example, suppose there are three firms A, B and C, with establishment sales across three U.S. counties, as follows: $G_A = (0, 0, 1)$, $G_B = (0.6, 0.2, 0, 2)$ and $G_C = (1, 0, 0)$. 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 firms A and 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 is in the first county than in the third county.

[13] 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).

TABLE 1 Summary statistics.

Panel A: Descriptive statistics

	Mean	Std	Min	25PC	Median	75PC	Max
Number of firms	2320	347	1618	2082	2355	2556	2977
Percentage value of CRSP	0.57	0.07	0.46	0.50	0.58	0.63	0.70
Average number 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	

Note: 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 are 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 the Appendix. The sample consists of 668,117 firm-month observations spanning 1990–2013.

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 geo-peers 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 a large cross-sectional variation in geographic linkage among our sample firms. In Table IA.1, we report the average geographic linkage score (*GEO*) for each industry, based on two-digit North American Industry Classification System (NAICS). We find the average *GEO* ranges from 0.07 to 0.12, the highest for the mining and logging industry and the lowest for the educational services industry. 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}).[14] In subsequent analyses, we will control for these return predictors when examining the return predictability of $GEORET_{t-1}$.

4 | EMPIRICAL RESULTS

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 the fundamental relationship between geography-linked peer firms. Specifically, we regress growth in focal firms' annual sales, profits, cost of goods sold and Selling, General and Administrative Expenses (SG&A) expenses on the average growth measures of their geo-peers (e.g., *Geo sales growth*). The regression model is as follows:

$$Sales\ growth_{i,t} = \alpha + \beta_1 Geo\ sales\ growth_{i,t} + \gamma Control_{i,t} + Firm\ FE + Year\ FE + \varepsilon_{i,t}; \tag{3}$$

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 defined in Equation (1). *Geo cost growth*, *Geo SG&A growth* and *Geo profit growth* are constructed in a similar way. All regressions include firm and 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 of panel A shows that the coefficient on *Geo sales growth* is 0.135 ($t = 2.80$), 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 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 as 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 0.066 and on the margin of statistical significance. Columns 3 and 4 show that the same conclusions hold when operating performance is measured as firm profit growth.

---

[14] Although geo-peers partially overlap with industry peers and same-headquarter peers, they only represent a small fraction of geo-peers. The average percentages of geo-peers headquartered in the same state and from the same Fama–French 48 industry as the focal firm are 12% and 6%, respectively.

TABLE 2 Fundamental linkages.

Panel A: Comovement of sales and profit growth				
	(1)	(2)	(3)	(4)
	Sales growth$_t$	*Sales growth$_t$*	*Profit growth$_t$*	*Profit growth$_t$*
Geo sales growth$_t$	0.135**	0.066		
	(2.80)	(1.65)		
Same-state sales growth$_t$		0.027		
		(1.33)		
Industry sales growth$_t$		0.123***		
		(2.90)		
Analyst sales growth$_t$		0.317***		
		(4.05)		
Geo profit growth$_t$			0.659***	0.162**
			(5.37)	(2.15)
Same-state profit growth$_t$				0.044*
				(1.74)
Industry profit growth$_t$				0.334***
				(7.74)
Analyst profit growth$_t$				0.855***
				(9.91)
Size$_t$	0.026***	0.026***	0.0032	0.007**
	(8.18)	(7.92)	(1.10)	(2.33)
BM$_t$	−0.011***	−0.010***	−0.012***	−0.009***
	(−4.26)	(−3.67)	(−5.09)	(−3.71)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Cluster by	Firm, Year	Firm, Year	Firm, Year	Firm, Year
Adj. *R*-sq	0.105	0.128	0.047	0.076
N	36,069	29,655	39,932	29,888
Panel B: Comovement of cost of goods sold and SG&A growth				
	(1)	(2)	(3)	(4)
	Cost growth$_t$	*Cost growth$_t$*	SG&A growth$_t$	*SG&A growth$_t$*
Geo cost growth$_t$	0.009**	0.005*		
	(2.34)	(2.03)		
Same-state cost growth$_t$		0.004		
		(0.76)		
Industry cost growth$_t$		0.011***		
		(2.86)		
Analyst cost growth$_t$		0.043***		
		(6.11)		
Geo SG&A growth$_t$			0.004**	0.002*
			(2.48)	(1.76)

(Continues)

TABLE 2 (Continued)

Panel B: Comovement of cost of goods sold and SG&A growth				
	(1)	(2)	(3)	(4)
	Cost growth$_t$	*Cost growth$_t$*	SG&A growth$_t$	*SG&A growth$_t$*
Same-state SG&A growth$_t$				0.001
				(0.48)
Industry SG&A growth$_t$				0.003
				(1.68)
Analyst SG&A growth$_t$				0.019***
				(6.41)
Size$_t$	0.040***	0.042***	0.043***	0.044***
	(5.00)	(4.86)	(6.35)	(6.82)
BM$_t$	−0.279***	−0.231**	−0.094*	−0.038
	(−3.04)	(−2.39)	(−1.98)	(−0.64)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Cluster by	Firm, Year	Firm, Year	Firm, Year	Firm, Year
Adj. *R*-sq	0.083	0.106	0.117	0.133
N	36,132	29,711	32,213	26,294

Note: This table reports the panel regression results of fundamental linkages between the focal firm and its geography-linked peers. *Sales growth$_t$* is calculated as *Sales per share$_t$/Sales per share$_{t-1}$* − 1. *Cost growth$_t$* and *SG&A growth$_t$* are defined similarly. 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 the 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 as the weighted average sales growth of analyst-linked peers, using the weights defined in Ali and Hirshleifer (2020). The profit, cost of goods sold and SG&A growth of peer firms are 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 firm and year fixed effects and size and book-to-market ratio as control variables. *t*-Statistics based on standard errors clustered by firm and year are shown below as coefficient estimates.Coefficients marked with $*$, $**$ and $***$ are significant at 10%, 5% and 1%, respectively.

Firms with geographically overlapped establishments are not only exposed to the same demand shocks but also shocks to operating costs such as labor costs and rents. In panel B of Table 2, we examine whether there is a strong comovement among geo-peers on the cost side. We look at both costs of goods sold and SG&A expenses. Column 1 shows the coefficient on *Geo cost growth* is 0.009 ($t = 2.34$), indicating a strong contemporaneous correlation between focal firm's and geo-peers' cost of goods sold growth. Columns 3 and 4 show similar results when we look at SG&A expenses.

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 returns 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 & Grinblatt, 1999) and same-headquarter lead–lag effect (Parsons et al., 2020), we exclude all firms from the same industry (based on the Fama–French 48 industry classification) and headquartered in the same state as the focal firm when constructing $GEORET_{it}$ 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 each decile 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, that is, 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 six columns, we control for the portfolios' exposure to standard asset-pricing factors. The same L/S strategy delivers Capital Asset Pricing Model (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 price momentum (Carhart, 1997) does not significantly affect the strategy, as the four-factor alpha remains at 0.41% (0.53%) per month 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 et al. (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 factor loadings of the long-short portfolio on the Fama–French three factors, 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 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 Table 1 results that $GEORET_{t-1}$ is positively correlated with prior month return RET_{t-1}. As there is a well-documented short-term reversal effect in individual stock returns (Jegadeesh, 1990), the positive correlation between $GEORET_{i,t-1}$ and $RET_{i,t-1}$ actually weakens the *positive* return predictability of $GEORET_{i,t-1}$ if the short-term reversal effect is not accounted for. Consistent with this observation, Table IA.2 in the Online Appendix reports significant alphas to the L/S portfolio double sorted on $GEORET_{i,t-1}$ and $RET_{i,t-1}$.[15]

In Figure IA.1 in the Online Appendix, we show the cumulative returns of the *GEORET* strategy, where we take long positions in the 10% of firms with the highest lagged 1-month *GEORET* and short positions in the 10% of firms with the lowest lagged 1-month *GEORET*. The figure shows that the long-short geographic momentum strategy generates cumulative returns of about 300% from 1990 to 2013.

[15] Specifically, stocks are independently ranked and assigned into three and five groups at the beginning of every calendar month, based on their prior-month returns ($RET_{i,t-1}$) and geo-peers' return ($GEORET_{i,t-1}$), respectively. Within each quintile portfolio sorted on $GEORET_{i,t-1}$, we take the average return across the three portfolios sorted on $RET_{i,t-1}$. We then calculate returns and alphas to an L/S strategy that long stocks in the top quintile and short those in the bottom quintile based on $GEORET_{i,t-1}$. All stocks are equal- (value-) weighted within a given portfolio, and the portfolios are rebalanced every calendar month to maintain equal- (value-) weights.

TABLE 3 Geographic momentum strategy.

Panel A: Portfolio returns and alphas							
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)
2	0.55	−0.12	−0.03	−0.03	0.06	0.02	0.05
	(1.79)	(−0.87)	(−0.24)	(−0.24)	(0.44)	(0.33)	(0.16)
3	0.72	0.11	0.13	0.07	−0.04	0.01	−0.07
	(2.59)	(0.91)	(1.05)	(0.60)	(−0.34)	(−0.58)	(0.11)
4	0.85	0.23	0.24	0.23	0.20	0.23	0.20
	(3.02)	(1.93)	(2.05)	(1.99)	(1.48)	(1.52)	(1.66)
5	0.72	0.12	0.14	0.19	0.04	0.08	0.09
	(2.65)	(1.04)	(1.20)	(1.63)	(0.31)	(0.70)	(0.56)
6	0.64	0.00	0.02	0.08	0.00	0.08	0.04
	(2.22)	(−0.01)	(0.16)	(0.66)	(−0.02)	(0.31)	(0.58)
7	0.86	0.22	0.25	0.28	0.14	0.28	0.18
	(2.98)	(1.77)	(2.13)	(2.16)	(1.13)	(1.31)	(1.98)
8	0.44	−0.22	−0.19	−0.14	−0.17	−0.07	−0.13
	(1.45)	(−1.46)	(−1.33)	(−1.02)	(−1.14)	(−0.90)	(−0.52)
9	0.89	0.22	0.25	0.20	0.24	0.27	0.23
	(2.80)	(1.42)	(1.82)	(1.48)	(1.55)	(1.57)	(1.53)
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-weighted)	(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-weighted)	(2.72)	(0.79)	(2.00)	(−0.44)	(−0.60)	(−2.99)

(Continues)

TABLE 3 (Continued)

Note: 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 nonfinancial 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 the CAPM model; the Fama–French (1993) three-factor model; the four-factor model including Fama–French three-factor and Carhart's (1997) momentum factor; the Fama–French (2015) five-factor model; a seven-factor model (Fama–French five-factor plus the momentum and short-term reversal factor); and Q-factors model of Hou et al. (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 percentage; *t*-statistics are shown below the coefficient estimates, with 5% statistical significance indicated in bold.

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:

$$Ret_{i,t} = \beta_0 + \beta_1 GEORET_{i,t-1} + \gamma X_{i,t-1} + \epsilon_{i,t}, \tag{4}$$

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 4 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' returns 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 an approximately 50 basis points increase in focal firm return. The result from the 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.[16]

One of the stylized facts in urban economics is that firms from the same industry tend to cluster together geographically (Ellison & Glaeser, 1997). As a result, it is likely that firms will have establishments that largely overlap with their industry peers geographically. Similarly, firms whose headquarters are 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 the 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

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

TABLE 4 Fama–MacBeth regressions.

	(1) RET (%)	(2) RET (%)	(3) RET (%)	(4) RET (%)	(5) RET (%)	(6) RET (%)	(7) 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							14.300***
							(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	−0.043	−0.039	−0.069	0.001	−0.049	−0.074	−0.056
	(−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	−0.055	0.472
	(1.43)	(1.38)	(0.87)	(0.63)	(−0.24)	(−0.13)	(1.36)
Average *R*-sq	0.036	0.039	0.054	0.049	0.089	0.049	0.052
N	723,764	668,117	257,213	147,494	171,365	399,911	532,062

Note: 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 pseudoconglomerate 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 BEA Input-Output data (at the summary industry level)

(Continues)

TABLE 4 (Continued)

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 (2020). We also control 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 the Appendix. 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.

both positive and significant, consistent with the industry momentum effect documented by Moskowitz and Grinblatt (1999) and the same-headquarter lead–lag effect shown by Parsons et al. (2020).

In columns 3–7, we control for other interfirm linkages as documented by prior studies.[17] In column 3, we add the focal firm's lagged technology-peer return (*TECHRET*) following Lee et al. (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 the 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 the 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 significant and positive, consistent with prior literature. The only exception is that coefficients on customers' (*CUSTRET*) and suppliers' returns (*SUPPRET*) are insignificantly positive, which could potentially be due to differences in the 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 the interfirm link 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 as even skilled analysts may not closely track news about firms' geo-peers and quickly impound relevant information into the 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 1-month lagged returns as predictors, some studies also examine longer horizon lags. Table IA.3 in the Online Appendix 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 the past 24 months lose their 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 (Ali and Hirshleifer, 2020; Moskowitz and Grinblatt, 1999). 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 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 IA.2 of the Online Appendix, we evaluate

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

these two alternative hypotheses by plotting the cumulative return to the *GEORET* hedged portfolio in the 6 months after portfolio formation. Consistent with the slow diffusion of geographic information, we continue to observe a modest upward drift in portfolio returns through month six. Untabulated analysis shows no sign of a return reversal over the next 12–24 months. Overall, the evidence seems to be most consistent with the delayed response of focal firm prices to fundamental information contained in returns of geo-peers and not an overreaction phenomenon.

4.4 | Robustness checks and placebo tests

In this section, we conduct a battery of robustness checks on geography-linked return predictability. We also conduct a placebo test using pseudo geographic peers. We report these results in Tables IA.4– IA.6 in the Online Appendix.

4.4.1 | Excluding microcap stocks

First, to alleviate the concern that our results are driven by microcap stocks, we exclude stocks with a price less than $5 or market capitalization below the 10th NYSE percentile. Columns 1 and 2 of Table IA.4 show that the coefficients of $GEORET_{t-1}$ are still positive and highly significant in both settings, suggesting that our result is not driven by microcap stocks. Given that small firms are more likely to operate in a single area, another way to remove microcap 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 | Excluding large geographic peers

Hou (2007) finds that within the same industry, common information gets incorporated into the prices of larger stocks quickly than into the prices of smaller stocks. It is possible that the return predictability we document is due to large firms leading small firms in incorporating common information. To rule out this possibility, we only include geo-peers with market capitalization smaller than the focal firm when constructing *GEORET*. Column 4 shows that the predictive power of $GEORET_{t-1}$ is robust to this construction, suggesting that our result is not driven by small firms reacting more sluggishly to common information than large firms.

4.4.3 | Geography-linked return predictability across time

In columns 5 and 6 of Table IA.4, 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 contrast with that of some other return predictors, which have declined substantially in the recent period. Consistent with Parsons et al. (2020), we find the effect of industry momentum was reduced by more than half and the own price momentum effect became insignificant over the 2002–2013 period. What is more noteworthy from our perspective is that the return predictability of geography-linked firms is robust in both subperiods.

4.4.4 | Persistence of the geographic linkage measure

We also examine the sensitivity of our main result to the age of the geographic linkage measure. The untabulated analysis shows the correlation between $GEO_{i,j,t}$, and its corresponding 1-year lagged measures is 0.95, suggesting that firms' geographic footprints are relatively persistent over time. Columns 7–9 of Table IA.4 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 5-year-old geographic linkage measures work quite well. One implication is that investors do not need extremely timely information on firms' establishment location information to implement this strategy. Even relatively "stale" geographic information has some predictive power for focal firm returns.

4.4.5 | Alternative measures of 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 cutoff value, we conduct a test where the peer sample is limited to the top 50 peers with the strongest geographic linkage with the focal firm. We also construct an alternative geographic linkage measure using establishment-level employment data, as the number of employees at establishments is less likely to be imputed than sales in NETS data. Columns 10 and 11 of Table IA.4 show that the predictive power of *GEORET* is still robust using these alternative measures of geographic peers.

4.4.6 | Panel regression approach

In our baseline tests, we use the Fama–MacBeth regression approach to examine the return predictability of *GEORET*, which is commonly used in asset pricing studies. The advantage of Fama–MacBeth regression is that this approach avoids using forward-looking return data when estimating the coefficients of return predictors. To ensure the robustness of our results, we conduct panel regressions with firm and year-month fixed effects and cluster standard errors at firm and year-month levels. Column 12 shows that the coefficient of *GEORET* is still positive ($t = 2.71$) with economic magnitudes similar to the baseline result, suggesting that the return predictability of *GEORET* is robust using alternative estimation methods.

4.4.7 | Using Exhibit 21 data to construct a geographic linkage

In this subsection, we rerun the return predictability test using an alternative data set that contains firms' geographic footprints. Specifically, we gather the locations of firms' headquarters and material subsidiaries from Exhibit 21, a section within or attached to 10-K fillings.[18] Dyreng et al. (2013) gather and compile these Exhibit 21 data using a text search program, which identifies the states in which firms' headquarters or domestic material subsidiaries are located for each company in each year.[19] With these data, we construct *GEORET*(*state*) in the same way as *GEORET* but based on the distributions of firms' material subsidiaries across all states in the United States. We then rerun Fama–MacBeth regressions as in Table 4, but use *GEORET*(*state*) as the main return predictor. The sample period runs from 1993 to 2014.

[18] A subsidiary that accounts for more than a certain percentage (usually 5% or 10%) of the consolidated assets or revenues of the parent firm and its subsidiaries is considered a material subsidiary.

[19] We thank Scott Dyreng for sharing the data on his personal website: https://sites.google.com/site/scottdyreng/Home/data-and-code/EX21-Dataset.

Table IA.5 reports the results. The coefficient on *GEORET*(*state*) is 17.93 ($t = 3.62$) in column 1, suggesting that geography-linked peer firms' returns strongly predict the next-month return of the focal firm after controlling for well-known return predictors.[20] In column 2, we add the lagged 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*(*state*) decreases to 14.25 but remains highly significant. In columns 3–7, we add other interfirm linkages as in Table 4. The return predictability of *GEORET*(*state*) remains statistically significant across all the models. Overall, the consistent evidence we obtain with alternative data of firms' geographic footprints suggests that our findings of geographic lead–lag return relationship are robust to measurement errors in the NETS database.

4.4.8 | Placebo tests

Finally, we conduct a placebo test by constructing the *GEORET_placebo* variable as the average returns of geographically distant firms. Specifically, each year we sort the geo-peers of each focal firm into deciles based on their geographic linkages with the focal firm. Then for each geo-peer of the focal firm, we select a firm in the same industry, closest in firm size but with a geographic linkage value in the bottom decile. We then construct the *GEORET_placebo* as the weighted average returns of these matched geographically distant firms and rerun the Fama-MacBeth regression using the *GEORET_placebo* as the return predictor. Table IA.6 shows that the coefficients on the *GEORET_placebo* are statistically insignificant and economically small (0.293 vs. 5.96). The insignificant results from the placebo test indicate that the strong return predictability of *GEORET* is not spurious.

5 | MECHANISMS

The results so far suggest that the lead–lag effects between geography-linked firms we document may be driven by the 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, (b) the costs that investors face if they attempt to profit from the mispricing, (c) aggregate investor sentiment and (d) the nature of the information. In addition, we explore the channel(s) through which information spillovers occur between geo-peers and the types of information that slowly diffuse across geo-peers.

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 the focal firm's prices in a timely fashion. However, a large set of theoretical and empirical studies shows that due to limited attention, investors tend to underweight public information, especially when the information is less visible (DellaVigna & Pollet, 2009; Giglio & Shue, 2014) or more complicated to analyze (Cohen & 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 a stock's institutional ownership and shared analyst coverage of related firms.[21] We posit that firms with lower institutional ownership and 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*.

[20] Note that while the coefficient estimate of *GEORET*(*state*) in Fama–MacBeth regressions is larger than that of *GEORET* as reported in Table 4 of the paper, the standard deviation of *GEORET*(*state*) is much smaller. As a result, the return predictability of *GEORET*(*state*) is economically similar to that of *GEORET*.

[21] See, for example, Bali et al. (2014) and Ali and Hirshleifer (2020).

To test this prediction, 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. Second, we define a dummy *CANALYST* that equals one if the average number of analysts covering the focal firm and its geo-peers at the previous year-end is above the sample median and zero otherwise. The results of these tests are reported in columns 1 and 2 of Table 5. Consistent with the prediction of a limited attention channel, the coefficient estimates on the interaction terms between the investor attention measures and $GEORET_{t-1}$ are all negative, and in the case of 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 information contained in the returns of geo-peers.

5.2 | Costs of arbitrage

In addition to investor attention, 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 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.[22] In addition, prior research suggests that information diffusion into the price is slower when trading costs are higher and stocks are less liquid (Bali et al., 2014). Therefore, we expect the return predictability of *GEORET* will be more pronounced for less liquid 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 et al. (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.[23] For all three variables, we create a dummy variable that equals one if the corresponding proxy is above the sample median in a month and zero otherwise.

The results are reported in columns 3–5 of Table 5. Column 3 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). Columns 4 and 5 show that the interaction terms between an indicator of higher bid-ask spread and higher Amihud illiquidity and lagged geo-peers' return ($GEORET_{t-1}$) are also positive and statistically significant. These findings lend support to our prediction that the return predictability effect is stronger for stocks that are more costly to arbitrage.

5.3 | Investor sentiment

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

[22] Evidence supporting idiosyncratic return volatility as one of the most significant limits to arbitrage is documented in Stambaugh et al. (2015), for instance.

[23] 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.

TABLE 5 Cross-sectional heterogeneity.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	RET (%)	*RET* (%)	*RET* (%)	*RET* (%)	*RET* (%)	*RET* (%)	*RET* (%)
GEORET	7.879***	5.882***	2.621**	3.044**	3.339*	4.838***	2.782**
	(4.40)	(3.62)	(1.98)	(2.11)	(1.71)	(4.00)	(2.23)
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*	1.633**
						(1.87)	(2.40)
GEORET× (Investor Attention>Median)							−3.482**
							(−2.03)
GEORET× (Arbitrage Costs>Median)							3.929**
							(2.15)
Controls	Y	Y	Y	Y	Y	Y	Y
Average R-sq	0.041	0.047	0.043	0.043	0.040	0.042	0.050
N	661,189	546,712	668,116	667,847	668,117	668,117	538,365

Note: This table reports the results of cross-sectional heterogeneity tests to evaluate the sensitivity of geographic momentum to proxies for investor attention, arbitrage costs and investor sentiment. The tests are Fama–MacBeth return forecasting regressions where the dependent variable *RET* is the focal firm's monthly stock return (in percentage). *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 (2020). *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 10 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. In column 7, we construct a composite measure *Investor Attention* as the average of the standardized *IO* and *CANALYST* to capture investors' attention on the stock. We construct a composite measure *Arbitrage Costs* as the average of standardized *IDVOL*, *ILLIQUIDITY* and *SPREAD* to capture arbitrage costs of the stock. 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 that is equal to one if *SENTIMENT* is above sample media and zero otherwise. The usual firm-level controls are also included. All variables are defined in the Appendix. The regression specification is the same as in Table 4. Coefficients marked with $*$, $**$ and $***$ are significant at 10%, 5% and 1%, respectively.

the 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 arbitrage away due to increased noise trader risks and short-sale constraints (De Long et al., 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 that equals one if *SENTIMENT* is above the sample median and zero otherwise. Column 6 of Table 5 shows that the coefficient estimate on the interaction term between the *SENTIMENT* dummy and $GEORET_{t-1}$ is indeed positive and significant. This finding provides further evidence that the return predictability of *GEORET* is likely a result of mispricing due to investors' underreaction to geographic information, especially during high-sentiment periods.

Furthermore, we include all three types of measures in the same model to examine whether their effects on the return predictability of *GEORET* are independent of each other. Because variables capturing the same effect could be highly correlated, we construct two composite measures, one proxy for investor attention and another proxy for arbitrage costs. Specifically, the composite measure of *Investor Attention* is the average value of standardized *IO* and *CANALYST* for a stock. Similarly, we construct a composite measure of *Arbitrage Costs* as the average value of standardized *IDVOL*, *ILLIQUIDITY* and *SPREAD*. The key variables of interest in this test are the interaction terms between *GEORET* and three dummies, indicating each measure is above the sample median. Column 7 of Table 5 shows that the coefficients of all three interaction terms have the predicted signs and are statistically significant. This suggests that limited attention, arbitrage costs and investor sentiment are important independent factors that amplify the return predictability of *GEORET*.

5.4 | The role of economic linkage

In this subsection, we explore the channel through which information spillover occurs between geo-peers. Prior studies suggest that firms sharing economic linkages are often clustered geographically (Jaffe et al., 1993). For example, geographic clustering for firms in the supply chain relationships can be expected when an industrial organization is efficient or the local labor market or fiscal incentives attract specific types of firms. As another example, innovation activities are often spatially concentrated as agglomeration allows local firms to share ideas and talents more efficiently. We thus expect a stronger information spillover and cross-firm return predictability among geo-peers with economic linkages.

To test this conjecture, we separately construct *GEORET* based on whether the focal firm and its geo-peers share some type of economic linkage or not. We consider four types of well-documented economic linkages between firms, including technological links, supply chain relationships, industry and product market linkages. Specifically, *GEORET_techlink_low* (*GEORET_techlink_high*) is the weighted average returns of geo-peers whose technological linkage with the focal firm is below (above) median in year $t-1$. *GEORET_with_sclink* (*GEORET_without_sclink*) is the weighted average returns of geo-peers with (without) customer–supplier linkage with the focal firm in year $t-1$.[24] *GEORET_with_INDlink* (*GEORET_without_INDlink*) is the weighted average returns of geo-peers that are in the same (different) Fama–French 48 industry group with the focal firm in year $t-1$. *GEORET_with_productlink* (*GEORET_without_productlink*) is the weighted average returns of geo-peers with (without) product market linkage with the focal firm in year $t-1$, using the text-based product market measure of Hoberg and Phillips (2016)). Then we run Fama–MacBeth regressions including both *GEORET* variables to compare the differences in the predictive power of *GEORET* in forecasting focal firm's next-month return.

Table 6 reports the results. To compare the predictive power of the *GEORET* measures, we focus on *t*-values.[25] Based on the *t*-statistics of the *GEORET* coefficients, we find the return predictability is stronger for

[24] The Customer–supplier linkage is constructed using BEA Input-Output data following Menzly and Ozbas (2010).

[25] The reason we focus on *t*-values is that the average coefficient estimates in the Fama–MacBeth regression can be interpreted as monthly returns on long-short trading strategies that trade on that part of the variation in each regressor that is orthogonal to every other regressor. The *t*-values associated with the Fama and MacBeth coefficients are, therefore, proportional to the Sharpe ratios of the long-short strategies.

TABLE 6 Decomposing GEORET based on geo-peers with and without economic linkage with the focal firm.

	(1) *RET* (%)	(2) *RET* (%)	(3) *RET* (%)	(4) *RET* (%)
GEORET_techlink_low	1.137			
	(1.54)			
GEORET_techlink_high	1.816**			
	(2.49)			
GEORET_with_sclink		2.441***		
		(4.33)		
GEORET_without_sclink		6.523**		
		(2.02)		
GEORET_with_INDlink			2.610***	
			(5.26)	
GEORET_without_INDlink			3.652***	
			(2.88)	
GEORET_with_productlink				2.925***
				(4.56)
GEORET_without_productlink				3.055*
				(1.68)
INDRET	7.760***	11.250***	11.250***	7.619***
	(3.32)	(5.22)	(5.22)	(2.71)
HQRET	7.372***	5.942***	5.942***	4.015***
	(3.85)	(5.53)	(5.53)	(3.39)
Controls	Y	Y	Y	Y
Average *R*-sq	0.052	0.040	0.040	0.046
N	253,339	641,740	641,740	389,944

Note: This table shows the return predictability of *GEORET* by constructing *GEORET* based on the economic linkage between the focal firm and its geographic peers. The dependent variable *RET* is the focal firm's monthly return (in percentage). In column 1, we construct *GEORET_techlink_low* (*GEORET_techlink_high*) as the weighted average returns of geo-peers whose technology link with the focal firm is below (above) median. In column 2, we construct *GEORET_with_sclink* (*GEORET_without_sclink*) as the weighted-average returns of geo-peers with (without) the customer-supplier linkage with the focal firm in year $t-1$. We obtain the customer-supplier linkage using BEA Input-Output data following Menzly and Ozbas (2010). In column 3, we construct *GEORET_with_INDlink* (*GEORET_without_INDlink*) as the weighted-average returns of geo-peers that belong to the same (different) Fama–French 48 industry as the focal firm in year $t-1$. In column 4, we construct *GEORET_with_productlink* (*GEORET_without_productlink*) as the weighted-average returns of geo-peers with (without) product market linkage with the focal firm in year $t-1$. The usual firm-level controls are also included. All variables are described in the Appendix. The regression specification is the same as in Table 4.Coefficients marked with $*$, $**$ and $***$ are significant at 10%, 5% and 1%, respectively.

GEORET constructed using geo-peers sharing economic linkages with the focal firm. For example, the *t*-stats of *GEORET_without_sclink* is 2.02, while the *t*-stats of *GEORET_with_sclink* is 4.33. The same pattern holds for technological, industry and product market linkages. These results support the channel that economic linkages facilitate information spillovers among geo-peers.

5.5 | Variation based on the information type

5.5.1 | Positive versus negative information

The practices of conservative accounting suggest there is more timely recognition and reporting of bad news (Basu, 1997). As a result, there could be an asymmetry in the lead–lag return relation between geographic peers. To test this idea, we construct two *GEORET* variables based on whether the news implied by the return of geo-peers is good or bad. Specifically, we construct *GEORET_Pos* and *GEORET_Neg* as the weighted-average returns of geo-peers with positive and negative returns in month $t-1$, respectively. Then we run the Fama–MacBeth regression including both measures of *GEORET_Pos* and *GEORET_Neg*. Column 1 of Table IA.7 in the Online Appendix shows that while *GEORET_Pos* significantly predicts focal firm's next-month return, the coefficient on *GEORET_Neg* is insignificant and economically small (3.09 vs. 0.54). This asymmetric return predictability is consistent with the in-built mechanisms in conservative accounting for more timely recognition and reporting of bad news.

5.5.2 | Accounting versus nonaccounting information

Our results so far show both the fundamental comovement and cross-stock return predictability across geography-linked firms, suggesting that the slow diffusion of accounting information across geo-peers could drive the lead–lag return relationship. Alternatively, the return predictability could also arise from the slow diffusion of nonaccounting information (such as changes in state regulations/policies). To explore the nature of the information driving the return predictability, we decompose *GEORET* into two parts, *GEORET_Acct* and *GEORET_Other*, based on geo-peer firms with and without quarterly earnings announcements in month $t-1$, respectively. The rationale is that as accounting information is usually disclosed in firms' earnings announcements, *GEORET_Acct* should capture the accounting information of geo-peers. We then rerun Fama–MacBeth regressions of focal firm's return in month t on two predictors, *GEORET_Acct* and *GEORET_Other*.

Column 2 of Table IA.7 reports the results. The coefficient on *GEORET_Acct* is 0.117 with a t-statistics of 4.77, while the coefficient of *GEORET_Other* is 0.110 with a t-statistics of 2.57. Based on t-values in the Fama–MacBeth regression, we conclude that the return predictability of *GEORET_Acct* is stronger than that of *GEORET_Other*, although both are significant. The evidence suggests that the lead–lag return relation among geographically overlapped firms is driven by the slow diffusion of both accounting information disclosed in firms' earnings announcements and nonaccounting fundamental information, but the channel of slow diffusion of accounting-related information is more important.

6 | ALTERNATIVE 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 some local risks could drive our results. For example, if geo-peers' returns can somehow proxy for local macroeconomic risks, which would then lead to changes in focal firms' discount rates in a geographically segmented market. We conduct several tests to examine this possibility.

6.1 | Exposure to state-level economic conditions

We conduct several tests to examine whether exposure to the state's economic conditions can fully explain the return predictability of *GEORET*. First, we directly capture firms' exposure to regional economic conditions 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 in which the firm operates:

$$PREA_{it} = \sum_{s=1}^{50} SALE\_SHARE_{i,s,\tau-1} * \frac{\Delta\widehat{SCI_{s,t+6}}}{SCI_{s,t}}, \tag{5}$$

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 the 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 states where it operates, we should find the effect of *GEORET* weakens significantly after controlling $PREA^{\perp}$. Column 1 of Table 7 panel A shows that consistent with Smajlbegovic (2019), the predicted regional economic activity variable $PREA^{\perp}$ significantly and positively predicts future stock return, indicating a slow diffusion of local macroeconomic information into stock prices. However, the return predictability of *GEORET* is not affected by $PREA^{\perp}$.

Second, we add several state-level macroeconomic variables as per Korniotis and Kumar (2013) in the Fama–MacBeth regressions. Specifically, in column 2 of panel A, we add the state labor income growth and the unemployment rate at the quarterly frequency. Following Korniotis and Kumar (2013), state *Income Growth* is calculated as the log differences between state labor income in a given quarter and its value in the same quarter of last year. The relative state *Unemployment Rate* is measured as the ratio of the current state unemployment rate to the moving average of the state employment rates over the previous 16 quarters.[26] In column 3, we further add the *Housing Collateral Ratio*, which is calculated as the log ratio of housing equity to labor income.[27] We match monthly stock returns with these state macroeconomic variables available in the most recent quarter. The results show that the return predictability of *GEORET* is not affected by including these state-level macroeconomic variables.

It is possible that these state-level macroeconomic variables may not fully capture state-level aggregate opportunities. Therefore, in our third test, we use panel regressions with state×year-month (or county×year-month) fixed effects. The inclusion of state×year-month fixed effects should absorb any unobserved time-varying state-level opportunities.[28] Panel B of Table 7 shows that the coefficient of *GEORET* is still positive and highly significant. Overall, all three tests suggest that the return predictability of *GEORET* is unlikely driven by firms' exposure to time-varying state economic conditions or aggregate opportunities.

6.2 | Exposure to the geographic risk factor

Dissanayake (2021) documents a geographic risk premium, measured as the difference in expected returns between geographically dispersed industries and agglomerated industries. The risk premium arises because firms in geographically agglomerated industries are better naturally hedged against aggregate shocks, and the investor is willing to pay a higher price, hence lower expected return for stocks in geographically agglomerated industries.

In this subsection, we conduct portfolio tests and Fama–MacBeth regressions to show that our finding is not driven by the geographic risk factor documented by Dissanayake (2021). In panel A of Table IA.8, we adjust the returns to the geographic momentum strategy by the geographic risk factor (GDMA), in addition to the Fama–French three factors. The GDMA factor is a long-short portfolio that goes long on geographically dispersed industries and short

[26] We obtain the labor income data from the BEA and unemployment data from the Bureau of Labor Statistics.

[27] We obtain the housing equity data from the Census Bureau, and the sample period starts from 2000.

[28] The state and county are based on the firm's headquarter location. Standard errors are double clustered at firm and year-month levels.

TABLE 7 Controlling state-level economic conditions.

Panel A: FM regressions with state-level macroeconomic variables			
	(1)	(2)	(3)
	RET (%)	*RET* (%)	*RET* (%)
GEORET	5.828***	5.422***	4.079***
	(3.97)	(3.78)	(2.85)
$PREA^{\perp}$	0.153***		
	(2.77)		
Income Growth		3.916	1.986
		(1.26)	(0.78)
Unemployment Rate		−0.019	−0.055
		(−0.06)	(−0.16)
Housing Collateral Ratio			−0.732
			(−0.62)
INDRET	11.900***	11.820***	6.658**
	(5.31)	(5.45)	(2.60)
HQRET	5.659***	4.973***	3.515**
	(5.94)	(4.87)	(2.46)
RET_{t-1}	−3.290***	−3.178***	−2.256***
	(−6.82)	(−6.17)	(−3.69)
SIZE	−0.040	−0.039	−0.077
	(−0.76)	(−0.76)	(−1.24)
BM	0.476***	0.444***	0.425**
	(2.98)	(2.78)	(2.17)
GP	0.669***	0.648***	0.682**
	(2.95)	(2.87)	(2.24)
AG	−0.426***	−0.424***	−0.464***
	(−6.15)	(−6.03)	(−4.46)
MOM	0.452	0.454	−0.213
	(1.34)	(1.36)	(−0.40)
Average *R*-sq	0.043	0.039	0.037
N	662,877	668,117	346,926

Panel B: Panels regressions with firm and state/county by year-month fixed effects		
	(1)	(2)
	RET (%)	*RET* (%)
GEORET	7.625***	7.500***
	(2.63)	(2.84)
INDRET	10.420***	10.480***
	(3.42)	(3.75)
HQRET	3.674	−13.980*
	(1.42)	(−1.89)
RET_{t-1}	−5.179***	−5.084***
	(−4.76)	(−4.33)

(Continues)

TABLE 7 (Continued)

Panel B: Panels regressions with firm and state/county by year-month fixed effects		
	(1)	(2)
	RET (%)	*RET* (%)
SIZE	−2.030***	−2.167***
	(−14.99)	(−17.91)
BM	0.320**	0.283*
	(2.08)	(1.78)
GP	1.701***	1.743***
	(5.34)	(4.74)
AG	−0.043	−0.017
	(−0.50)	(−0.23)
MOM	−0.042	−0.113
	(−0.15)	(−0.46)
State × Year-month FE	Y	N
County × Year-month FE	N	Y
Firm FE	Y	Y
Cluster by	Firm, Year-month	Firm, Year-month
Adj. *R*-sq	0.136	0.151
N	663,796	549,990

Note: This table reports the result for return predictability by controlling local macroeconomic conditions. 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 panel A, we run Fama–MacBeth regressions by adding state-level macroeconomic variables. Specifically, in column 1, we control for the predicted regional economic activity ($PREA^{\perp}$). Following Smajlbegovic (2019), the predicted regional economic activity variable, *PREA*, is constructed from a linear combination of predicted state economic activity growth rates weighted by the fraction of sales in each state a firm 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. In column 2, we add the state-level labor income growth and the unemployment rate at quarterly frequency. Following Korniotis and Kumar (2013), state-level *Income Growth* is defined as the log differences between state income in a given quarter and state income in the same quarter of the last year. The relative state *Unemployment Rate* is defined as the ratio of the current state unemployment rate to the moving average of the state employment rates over the previous 16 quarters. We obtain the labor income data from the BEA and unemployment data from the Bureau of Labor Statistics. In column 3, we further add the Housing Collateral Ratio following Korniotis and Kumar (2013). This measure is defined as the log ratio of housing equity to labor income. We obtained the housing data from the Census Bureau, and the sample period starts from 2000. We match the returns with macroeconomic variables from the most recent quarter and rerun the Fama–MacBeth regressions. In panel B, we run panel regressions by adding state-by-year-month fixed effects in column 1 and county-by-year-month fixed effects in column 2, respectively. We also add firm fixed effects. The state and county are defined based on the firm's headquarters location. Standard errors are clustered at firm and year-month levels. The usual firm-level controls are also included. Other variables are defined in the Appendix. Coefficients marked with $*$, $**$ and $***$ are significant at 10%, 5% and 1%, respectively.

geographically agglomerated industries.[29] The alphas of our geographic momentum strategy are still significant and have similar economic magnitudes as those in Table 3. Panel B reports the factor loadings of our L/S portfolio on the GDMA factor. We see that the loadings on the GDMA factor are all statistically insignificant, which explains the limited effect of controlling the GDMA factor. In panel C, we construct the firm-level geographic risk exposure

[29] We are grateful to Ruchith Dissanayake for sharing the GDMA factor data with us.

(*GEORISK*) as a stock's return sensitivity to the GDMA factor.[30] We then rerun the Fama–MacBeth regression by adding *GEORISK* in the regressions. The significant return predictability of *GEORET* still holds. Overall, the significant return predictability of *GEORET* documented in our study is not explained by firms' exposure to geographic risk factors as per Dissanayake (2021).

6.3 | Exposure to unobserved risks

Our preferred explanation for the return predictability results is that investors have limited attention and are slow to incorporate information contained in returns of the focal firm's geo-peers, and evidence so far supports 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. In this subsection, we conduct further tests to rule out the possibility that the profitability of our geographic momentum strategy is explained by exposure to unobserved risks.

6.3.1 | Returns around earnings announcements

First, we examine stock price reactions around subsequent earnings announcements. The idea is intuitive: if an anomaly is associated with mispricing, 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 the abnormal return is driven by exposure to unobserved risks, the subsequent returns should accrue more evenly over subsequent periods. To conduct this test, we conduct a panel regression analysis following the methodology of Engelberg et al. (2018). Our unit of observation is firm-day rather than firm-month in this test. Specifically, we regress the daily return of stock (*DRET*) on the last month's 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 10 days for stock returns, stock returns squared and trading volume.

We present our results in Table 8. The earnings announcement window is defined as either a 1-day window (columns 1 and 2) or a 3-day window (columns 3 and 4), centered on the earnings announcement date. The significantly positive coefficient on *GEORET* suggests that returns of geographic peers can predict the focal firm's return on non-earnings announcement days. Also consistent with the earnings announcement premium literature (Frazzini & Lamont, 2007), the coefficient on the 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 is released. For example, in column 1, the coefficient on *GEORET* is 0.347 ($t = 2.82$), while the interaction coefficient on *GEORET* $\times$ *EDAY* is 0.578 ($t = 2.04$). The coefficients indicate that for a *GEORET* value of 0.06 (two standard deviation changes), 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 geographic momentum 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.[31]

[30] We use a rolling 60-month window to estimate the firm-level geographic risk exposure.

[31] 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 the earnings announcement.

TABLE 8 Returns on earnings announcement days.

	(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***	0.082***	0.119***
	(13.50)	(15.66)	(7.37)	(10.55)
Lagged controls	N	Y	N	Y
Day FE	Y	Y	Y	
Adj. *R*-sq	0.048	0.069	0.048	0.069
N	17,953,058	17,875,817	17,953,058	17,875,817

Note: 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 (columns 1 and 2), or 3-day window (columns 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 10 days for stock returns, stock returns squared and trading volume. Key variables are described in the Appendix. Standard errors are clustered on time. T-statistics are in parentheses.Coefficients marked with $*$, $**$ and $***$ are significant at 10%, 5% and 1%, respectively.

6.3.2 | Evidence from nonreturn-based outcomes

As an alternative approach, we examine whether *GEORET* has predictive power for the focal firm's standardized unexpected earnings (*SUE*). *SUEs* capture unanticipated changes in the 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 construct the dependent variable as standardized unexpected earnings (*SUE*), which is defined as the difference between the actual quarterly earnings per share (EPS) and the analyst consensus forecast of quarterly EPS scaled by stock prices in the month before the quarterly earnings announcement. The main explanatory variable of interest is lagged *GEORET*, computed using the past 3 month returns of the focal firm's geo-peers. Control variables include the focal firm's own lagged *SUEs*, up to four quarters.

Table 9 contains regression results under various model specifications. Column 1 presents a simple regression of *SUE* on lagged *GEORET*, with the firm and year-quarter fixed effects. The estimated coefficient on $GEORET_{t-1}$ is 0.002 (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 to predict future *SUEs*. These results further confirm that the short-window announcement returns we documented in Subsection 6.3.1 are driven by *GEORET*'s ability to anticipate the directional changes in the focal firm's future earnings.

TABLE 9 Predicting earnings surprises.

	(1)	(2)	(3)
	SUE_t	SUE_t	SUE_t
$GEORET_{t-1}$	0.002**	0.002*	0.003*
	(2.12)	(2.02)	(1.98)
SUE_{t-1}		0.067*	0.134***
		(1.97)	(3.25)
SUE_{t-2}		0.031	0.081***
		(1.45)	(3.44)
SUE_{t-3}		−0.007	0.036**
		(−0.54)	(2.73)
SUE_{t-4}		0.013	0.054
		(0.45)	(1.51)
Firm FE	Y	Y	N
Industry FE	N	N	Y
Year-quarter FE	Y	Y	Y
Adj. *R*-sq	0.065	0.069	0.043
N	163,169	90,493	90,000

Note: 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 the analyst consensus forecast of quarterly EPS scaled by stock prices in the month before a quarterly earnings announcement. *GEORET* is calculated based on the past 3-month returns of geography-linked peers of the focal firm. We include firm fixed effect and year-quarter fixed effect in columns 1 and 2. In column 3, we include industry fixed effect and year-quarter fixed effects. We add one-quarter to four-quarter lags of the firm's own *SUE*s as control variables. Other variables are described in the Appendix. 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.

6.3.3 | Evidence from analyst forecast revisions

Lastly, we examine analyst forecasting behavior to provide direct evidence on the limited attention channel. This setting is particularly useful because analyst earnings forecast 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 to 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, the log of market capitalization and the log of book-to-market ratio as control variables. The dependent variables, *FRP* and *FRB*, are the 1-month-ahead revision in the consensus annual EPS forecast of the focal firm scaled by lagged stock price (columns 1 and 2) and the 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 Equation (1) as weights. Consistent with our hypothesis, column 1 shows that the coefficient on $GEOFRP_{t-1}$ is 0.053 ($t = 4.20$), suggesting that the average forecast revision of geography-linked firms is a strong predictor of future revisions of the focal firm.

TABLE 10 Lead-lag effects in analyst forecast revisions.

	(1) FRP_t	(2) FRP_t	(3) FRB_t	(4) FRB_t
$GEOFRP_{t-1}$	0.053***	0.031***		
	(4.20)	(2.73)		
$STATEFRP_{t-1}$		0.016**		
		(2.18)		
$INDFRP_{t-1}$		0.134***		
		(7.31)		
$ANALYSTFRP_{t-1}$		0.092***		
		(8.84)		
$GEOFRB_{t-1}$			0.037***	0.025*
			(2.83)	(1.94)
$STATEFRB_{t-1}$				0.008
				(1.39)
$INDFRB_{t-1}$				0.045***
				(5.10)
$ANALYSTFRB_{t-1}$				0.023***
				(4.38)
FRP_{t-1}	0.047***	0.041***		
	(9.36)	(8.50)		
FRB_{t-1}			0.050***	0.048***
			(7.59)	(7.40)
RET_{t-1}	0.009***	0.009***	0.020***	0.020***
	(14.49)	(14.51)	(19.60)	(19.65)
$RETt_{-13,\,t-2}$	0.001***	0.001***	0.003***	0.003***
	(8.52)	(8.51)	(9.34)	(9.35)
SIZE	0.0003***	0.0003***	0.001***	0.001***
	(15.24)	(15.22)	(14.73)	(15.51)
BM	−0.0005***	−0.0005***	0.002***	0.002***
	(−5.59)	(−5.98)	(6.87)	(6.95)
Average *R*-sq	0.038	0.043	0.034	0.036
N	456,785	454,719	443,155	441,177

Note: 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 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. All other variables are described in the Appendix. The regression specification is the same as in Table 4.Coefficients marked with $*$, $**$ and $***$ are significant at 10%, 5% and 1%, respectively.

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 0.031 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 the 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 analysts' sluggish information updating. This is consistent with studies documenting inefficient forecast behaviors by analysts (Bouchaud et al., 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 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 1989 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 an 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, are more costly to arbitrage and during high-sentiment periods. The cross-firm information spillovers and return predictability are also stronger for geographic peers sharing economic linkages and with positive information. 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 a sluggish price adjustment to nuanced geographic information.

ACKNOWLEDGEMENT

We thank Bin Han, Byoung-Hyoun Hwang, Claire Yurong Hong, Mark Kamstra (discussant), Da Ke, Roger Loh, Wenlan Qian, Johan Sulaeman, Stephen Teng Sun, Jinfei Sheng, Baolian Wang, Chen Wang (discussant), Jingda Yan, Jialin Yu, Dexin Zhou, Qifei Zhu as well as seminar participants at 2022 China International Conference in Finance, 2021 Northern Finance Association Annual Conference, 2020 Singapore Scholars Symposium, 2021 FMCG conference, PanAgora Asset Management, and Tianjing University for helpful comments. 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. We are grateful for the 2020 Crowell Memorial Award for Best Paper (second prize) in Quantitative Investments by PanAgora Asset Management. Li acknowledges that this study was partially funded at the Singapore Management University through a research grant (20-C207-SMU-002) from the Ministry of Education Academic Research Fund Tier 1.

ORCID

Frank Weikai Li https://orcid.org/0000-0002-7906-6680

REFERENCES

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

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

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

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–152.

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

Barnatchez, K., Crane, L. D., & Decker, R. (2017). *An assessment of the National Establishment Time Series (NETS) database* (Finance and Economics Discussion Series 2017-110). Board of Governors of the Federal Reserve System.

Basu, S. (1997). The conservatism principle and the asymmetric timeliness of earnings1. *Journal of Accounting and Economics*, *24*(1), 3–37.

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

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

Bouchaud, J.-P., Krueger, P., Landier, A., & Thesmar, D. (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., & Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance*, *63*(4), 1977–2011.

Cohen, L., & Lou, D. (2012). Complicated firms. *Journal of Financial Economics*, *104*(2), 383–400.

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

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

Corwin, S. A., & Schultz, P. (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., & Moskowitz, T. J. (2001). The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, *109*(4), 811–841.

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

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

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

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

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

Dissanayake, R. (2021). Geographic distribution of firms and expected stock returns. *Journal of Economic Dynamics and Control*, *133*, 104267.

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

Dyreng, S. D., Lindsey, B. P., & Thornock, J. R. (2013). Exploring the role Delaware plays as a domestic tax haven. *Journal of Financial Economics*, *108*(3), 751–772.

Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy*, *105*(5), 889–927.

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

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

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

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, *81*, 607–636.

Frazzini, A., & Lamont, O. A. (2007). The earnings announcement premium and trading volume (NBER working paper, w13090). National Bureau of Economic Research.

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

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

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, *1*(1), 27–50.

Heider, F., & Ljungqvist, A. (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., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1-3), 337–386.

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

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

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

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *The Review of Financial Studies*, *20*(4), 1113–1138.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, *28*(3), 650–705.

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

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

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance*, *45*(3), 881–898.

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

La Porta, R., Lakonishok, J., Shleifer, A., & Vishny, R. (1997). Good news for value stocks: Further evidence on market efficiency. *Journal of Finance*, 859–874.

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

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

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

Marshall, A. (1920). *Principles of economics* (8th ed., p. 621). Macmillan, London.

Menzly, L., & Ozbas, O. (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* (Vol. 4, pp. 2243–2291). Elsevier.

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

Neumark, D., Wall, B., & Zhang, J. (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., Sabbatucci, R., & Titman, S. (2020). Geographic lead-lag effects. *The Review of Financial Studies*, *33*(10), 4721–4770.

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

Patton, A. J., & Verardo, M. (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., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80*(3), 563–602.

Pirinsky, C., & Wang, Q. (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., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288–302.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, *70*(5), 1903–1948.
Tuzel, S., & Zhang, M. B. (2017). Local risk, local factors, and asset prices. *The Journal of Finance*, *72*(1), 325–370.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.
Wurgler, J., & Zhuravskaya, E. (2002). Does arbitrage flatten demand curves for stocks? *The Journal of Business*, *75*(4), 583–608.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

APPENDIX

Variables	Definition
GEO	Geographic linkage measure GEO_{ijt} is defined as the uncentered correlation of the distribution of establishment sales between two firms *i* and *j* across all counties in the United States. Establishment-level sales data is from the National Establishment Time Series (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 the year-over-year growth rate of total assets, following Cooper et al. (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_{t-1}	Lagged monthly raw return, or short-term return reversal variable, defined as focal firm's stock return in month $t-1$.

(Continues)

Variables	Definition
SUE	Standardized unexpected earnings (SUE) is defined as the difference between the actual quarterly earnings per share (EPS) and the analyst consensus forecast of quarterly EPS scaled by stock prices in the month before the 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 10 daily returns required) following Ang et al. (2006).