Social Ties, Comovements, and Predictable Returns*

Lin Peng Sheridan Titman Muhammed Yönaç Dexin Zhou

October 19, 2022

Abstract

We examine the relation between the social ties between firms' headquarters locations and comovements between their fundamentals and stock returns. Our evidence indicates that firms in the same industry with socially connected locations exhibit co-movement in fundamentals and stock returns that exceed those without socially connected locations. However, the stock returns reflect the location information with a lag. To exploit this lagged relationship, we form portfolios that buy (sell) stocks when their socially-weighted industry peer returns in the previous month is high (low). The value-weighted version of this portfolio generates a monthly alpha of 84 basis points. Social peer firm returns also predict firms' future earnings surprises, analysts' forecast errors, and earnings announcement returns. Further evidence indicates that the mispricing is stronger for low-visibility firms and for firms located outside of industry clusters, and that the effect is not subsumed by other sources of return predictability.

JEL Codes: G14, D85
Keywords: social networks, comovements, return predictability, machine learning

---

*The paper was previously circulated under the title "Social Ties and Predictable Returns." Lin Peng: Zicklin School of Business, Baruch College/CUNY, and the University of Cambridge, email: lin.peng@baruch.cuny.edu and lp604@cam.ac.uk. Sheridan Titman: McCombs School of Business, The University of Texas at Austin, e-mail: sheridan.titman@mccombs.utexas.edu. Muhammed Yönaç: University of Bristol Business School, email: muhammed.yonac@bristol.ac.uk. Dexin Zhou: Zicklin School of Business, Baruch College, email: dexin.zhou@baruch.cuny.edu. We thank Jawad Addoum (discussant), Nicholas Barberis, Vasco Carvalho, Tiago Cavalcanti, Sudheer Chava, Meredith Crowley, Jared DeLisle (discussant), Xi Dong, Elisa Faraglia, Diego Garcia (discussant), Xing Huang (discussant), Ted Joyce, Xin Li (discussant), Dong Lou, Sebastiano Manzan, Pierre Magontier (discussant), Ramana Nanda, Quoc Nguyen (discussant), Jun Wang, Junbo Wang (discussant), Wei Xiong, and conference and seminar participants at the 2021 China International Conference in Finance, the Urban Economics Association 2021 Meeting, the Financial Management Association 2021 Meeting, the Midwest Finance Association 2022 Meeting, 13th Florida State University SunTrust Beach Conference, the Eastern Finance Association 2022 Meeting, Spring 2022 NBER Behavioral Finance Working Group Meeting, Asian Finance Association 2022 Meeting, Baruch College, Durham University, Georgia Institute of Technology, Ozyegin University, Stockholm Business School, the University of Bristol, University of Birmingham, and the University of Cambridge for their helpful comments and suggestions. Lin Peng gratefully acknowledges financial support from the Krell research grant and the Keynes Fund. All remaining errors are our own.

1 Introduction

A growing body of literature highlights the interconnected nature of firms in the economy. For example, Acemoglu et al. (2012) and McNerney et al. (2022) find that production networks and input-output linkages facilitate the propagation of firm-level shocks, amplify economic growth, and contribute to aggregate fluctuations. This paper studies a source of cross-firm linkages that arise because of social ties between firms. Our analysis, which focuses on the spatial nature of social connections, builds on recent research that shows that social connections between regions are associated with important economic exchanges such as domestic and international trade flows, migration patterns, and knowledge spillovers (Bailey et al., 2018a, 2021; Breschi and Lenzi, 2016; Cohen et al., 2017). Given this research, it seems natural to expect that social connections between regions can also lead to strong ties between firms located in these regions, facilitating the flow of ideas, and leading the firms to adopt similar strategies as well as similar technologies. However, little is known about the extent to which such ties influence firm fundamentals and the extent to which capital market participants understand these ties and their relevance for firm valuation.

To explore the implications of these cross-firm interactions, we examine how social ties between regions influence the co-movements of the fundamentals and stock returns of firms headquartered in different locations. Our analysis is based on the idea that firms headquartered in locations that are socially linked will tend to be more similar, and as a result, will exhibit stronger co-movements. We proxy for firm social ties using Facebook's Social Connectedness Index (SCI) (see Bailey et al., 2018a), which measures the social connectedness of individuals both within and across U.S. counties.[1] Using this information, we construct specially tailored SCI-weighted industry portfolios for each firm in our sample. Relative to equally-weighted and value-weighted industry portfolios, these tailored portfolios place more weight on the focal firms' industry peers headquartered in counties with strong social ties to their home county. Our hypothesis is that industry peers located in socially connected counties are more similar, and as a result, a firm's changes in both its fundamentals and stock prices are more closely related to the SCI-weighted averages rather than the equally-weighted averages of their industry peers.

Our empirical analysis of contemporaneous co-movements are consistent with this hypothesis. We find that both the equally-weighted and SCI-weighted industry indexes explain changes in both firm fundamentals and stock returns, but the changes tend to be better explained by the SCI-weighted indexes. We then ask whether market prices fully reflect the importance of these so-

[1]As the world's largest online social networking service, Facebook's enormous scale and coverage (over 258 million active users in the United States as of 2020) and the relative representatives of its user body makes SCI a unique measure of the real-world geographic structure of U.S. social networks at population scale. See, Bailey et al. (2018a) and Kuchler et al. (2020) for further discussion on these points.

cial ties. We do this by calculating *SPFRET*, the SCI-weighted returns of each focal firm's industry peers, and examine the extent to which these returns explain the focal firm's returns. If market prices fail to fully reflect the importance of social connections, the SCI-weighted returns should lead the returns of the focal firms. Our evidence indicates that they do. Specifically, we show that a simple trading strategy that exploits these lead-lag relationships earn significant abnormal returns.

Further tests examine return predictability over longer horizons. In addition to investigating the return predictability of the SCI-weighted return over the past month, (*SPFRET*), we calculate the past one-year returns of each focal firm's SCI-weighted peer firms (*SPFMOM*, skipping the most recent month). We find that the SCI-weighted industry returns continue to predict a firm's future returns for up to 12 months. The one-year cumulative long-short portfolio alpha is 4.57% and 5.37% for equal weighted portfolios sorted by *SPFRET* and *SPFMOM*, respectively. The corresponding value weighted alphas, 2.12% and 3.60%, are somewhat smaller, but still reliably different than zero. These longer-term results provides further evidence that the predictability is driven by underreaction.

To gain insights into what market frictions drive the return predictability of *SPFRET*, we explore how firm characteristics influence the strength of the predictability. If the documented return predictability is due to investor inattention to value-relevant information from socially connected peer firm returns, we should expect stronger predictability for focal firms that are less visible to investors. Consistent with this conjecture, we find that return predictability results are stronger for smaller firms, firms with low institutional ownership, and firms with low analyst coverage. Similarly, our results show that information from social peer firms are more quickly reflected in the prices of firms located in industry clusters, consistent with the finding of Engelberg et al. (2018) that analysts tend to cover stocks inside industry clusters and firms located away from industry clusters are likely to receive less investor attention.

We also document that a firm's SCI-weighted industry portfolio predicts both its future earnings as well as its stock returns around earnings announcement dates. Specifically, we show that firms with the highest past cumulative social peer firm returns (*SPFMOM*) have, on average, standardized unexpected earnings (*SUE*) that are 34 percentage points higher in the next quarter than those with the lowest *SPFMOM* (or 24% of the standard deviation of *SUE*). Similarly, firms with the highest *SPFRET* exhibit 10.9 percentage points higher *SUE* in the following quarter. Our evidence indicates that this predictability persists for at least a year. Further analysis indicates that information embedded in SCI-weighted industry portfolios does not receive sufficient attention from sell-side analysts, who substantially underestimate the future earnings of firms with high social peer firm returns. Finally, we find that a firm's social peer firm returns positively predict its future earnings announcement returns. In summary, social peer returns contain important information regarding a firm's future earnings that is not fully taken into account by financial analysts and other market participants.

The evidence of return predictability documented in this paper is closely related to a large literature that examines economic linkages that generate lead-lag predictability. Most notably, Moskowitz and Grinblatt (1999) document evidence of industry momentum, which can be generated by the same lead-lag relationships revealed in our tests. We show that the lagged returns of equally-weighted industry portfolios do in fact predict the future returns of individual stocks that share the industry affiliation, but *SPFRET* generates significant predictability after controlling for the returns of both equally-weighted and value-weighted industry portfolios. More generally, we find that the predictability of *SPFRET* is not subsumed by any of the other predictors introduced in this literature.[2]

We analyze the contribution of our source of return predictability relative to those documented in the prior literature using three methodologies. First, we run a time series regression of the abnormal returns based on long-short portfolios sorted by *SPFRET* on the abnormal returns obtained using the alternative variables. We show that the abnormal returns based on other predictors explain roughly half of the *SPFRET*-based abnormal return. However, even after accounting for these economic linkages, the *SPFRET*-based strategy still generates an abnormal return of over 43 basis points per month. We next account for the alternative return predictors in the cross-section using Fama and MacBeth (1973) regressions. In addition to the variables listed above, we also include firm characteristics that are known to predict returns.[3] Even after accounting for this comprehensive list of controls, the firms with the highest *SPFRET* outperform those with the lowest *SPFRET* by 42 basis points in the next month. To further assess the incremental gain in return predictability, we use a machine learning methodology that accounts for the non-linear and the correlated nature of these predictors. Specifically, we follow Gu et al. (2020) and form a composite measure that aggregates the predictive information of all return predictors using the partial least square method. We show that *SPFRET* contributes substantially to the predictive power of the composite measure and is among the most important lead-lag return predictors. Economically, a long-short portfolio created as of 1994 using the composite predictor yields a cumulative return as of 2019 that is 86.4 percentage points higher than an alternative portfolio that does not utilize the

[2]In addition to industry momentum, we also include alternative industry momentum (e.g., Hoberg and Phillips, 2018), geographic lead-lag effects (e.g., Parsons et al., 2020), customer-supplier lead-lag effects (e.g., Cohen and Frazzini, 2008), lead-lag effects based on technological similarity (Lee et al., 2019), the complicated firm effect (Cohen and Lou, 2012), the lead-lag control based on shared analyst coverage (Ali and Hirshleifer, 2020), and based on labor flows across locations.

[3]The variables include lag one-month return, size, book-to-market ratio, momentum, illiquidity, idiosyncratic volatility, skewness, and co-skewness.

information in *SPFRET*.

Our paper is also closely related to the growing literature that examines how networks shape economic relationships (e.g., Acemoglu et al., 2012; Breschi and Lenzi, 2016; Cohen et al., 2017; Bailey et al., 2018b; McNerney et al., 2022; Bailey et al., 2021). Our paper contributes to this literature by showing that social ties between firm locations are associated with the comovement of firms' fundamentals and stock returns.

We also contribute to the social finance literature, which provides evidence on the peer effects on retail investing and the effect of social interactions on the investment behavior of professional investors.[4] The literature on institutional investment shows that social ties between institutional managers can generate valuable information (e.g., Cohen et al., 2008; Pool et al., 2015; Hong and Xu, 2019). In contrast, the evidence for retail investors suggests that social effects are associated with the propagation of sentiment that influences both investment and housing decisions (e.g., Shiller, 2010; Hvide and Östberg, 2015; Bailey et al., 2018c) and contribute to behavioral biases (e.g., Heimer, 2016).[5] It is possible that the propagation of irrational sentiment can be a source of co-movement between stocks headquartered in socially tied regions, however, we would expect these effects to drive transitory price pressure that reverses in the long run. We do not find evidence of transitory price effects and long run reversals. Instead, our evidence suggests that the co-movement of firms associated with the social ties between their locations is driven by fundamentals rather than correlated market sentiment. In addition, we find that evidence of fundamental information that is not fully incorporated in the actions of professional market participants, such as financial analysts. Moreover, while the above studies find that social effects influence the decisions of both professional and retail investors, the studies do not document that these actions affect prices. Kuchler et al. (2021), however, does provide evidence of an effect on firm values and liquidity, but does not consider cross-firm comovements and lead-lag effects, which is the focus of our study.

The rest of the paper is organized as follows. In Section 2, we provide a detailed description of the data used in this study. In section 3, we investigate how social connectedness between locations relate to firms' comovement. In Section 4, we discuss the return predictability results. In Section 5, we analyze whether social peer firm returns capture earnings-related information about the focal firm and whether market participants fully incorporate this information. In section 6, we examine whether our results are explained by other well known predictive variables. Section 7

[4]See, for example, Ivković and Weisbenner (2007); Feng and Seasholes (2004); Brown et al. (2008); Hong et al. (2004, 2005); Cohen et al. (2008); Pool et al. (2015); Hvide and Östberg (2015); Heimer (2016); Hong and Xu (2019); Ouimet and Tate (2019).

[5]Related to this, Pool et al. (2012) finds evidence of home bias for mutual fund managers, who overweight stocks from their home state and hold excessively risky portfolios.

concludes.

2 Data and Variable Definitions

Our sample includes all common stocks traded on the NYSE, AMEX, and NASDAQ exchanges, covering the period from July 1963 through December 2019. The Center for Research in Security Prices (CRSP) provides the daily and monthly return and volume data. The accounting variables, including earnings, are obtained from the CRSP-Compustat merged database. Analyst earnings forecasts and institutional ownership data are from the Institutional Brokers' Estimate System (I/B/E/S) database and Thomson Reuters institutional (13F) holdings database, respectively. Data on the Fama-French five-factor and the Fama-French 48-industry classification are obtained from Kenneth French's data library. We eliminate stocks with a price per share less than $5 or more than $1,000. We require a minimum of 24 monthly observations for variables created using monthly data and 15 daily observations for those created using daily data. Unless otherwise stated, all variables are measured as of the end of the portfolio formation month (i.e., month t). The variables and the corresponding definitions are summarized in Table A1.

2.1 Key Variables

We follow Bailey et al. (2018a) and measure social connectedness between two U.S. counties using Facebook's *Social Connectedness Index* (*SCI*). The measure is the total number of Facebook friendship links between two U.S. counties (as of April 2016), divided by the product of the populations of the two counties. As the world's largest online social networking service, Facebook's scale and the relative representativeness of its user body make SCI a comprehensive measure of the geographic structure of the U.S. social networks.[6]

Figure 1 plots heat maps of social connectedness, measured with SCI, for Cook County, IL, in Panel A, and Bartholomew, IN, in Panel B. The focal counties are colored in red and darker colors indicate higher social connectedness to the focal counties. Although the two focal counties are located in two adjacent states, the maps show differences in their relative connectedness to other regions. Cook County, which includes the city of Chicago, is strongly connected to nearby counties as well as to counties in the southern states along the Mississippi River. The pattern has been documented by Bailey et al. (2018a) and Kuchler et al. (2021) and is attributed to the large-scale migration of African Americans from southern states to northern industrial cities during the

[6]Facebook had more than 2.1 billion monthly active users globally and 239 million active users in the United States and Canada as of 2017. This represents 68% of the adult population and 79% of online adults in the United States (Duggan et al., 2016). Facebook usage rates among U.S.-based online adults were relatively constant across various demographics and locations. Bailey et al. (2018a,b, 2019a,b,c, 2021); Kuchler et al. (2021, 2020), and Rehbein et al. (2020) provide evidence that friendships observed on Facebook are a good proxy for real-world connections.

Great Migration of 1916–1970.[7] Bartholomew County is strongly connected to nearby counties in the state of Indiana, counties in the neighboring state of Illinois, and counties in distant states such as Kentucky, Tennessee, North Dakota, Texas, and Florida. Hence, these plots show that there are substantial differences between the geographic proximity and the social connectedness across regions in the United States. These differences will help us identify the incremental information that social ties contain that is different from the effects of physical proximity.

[Insert Figure 1 here]

Our main variable of interest is the social peer firm return (*SPFRET*). For a given firm i and for a given month t, *SPFRET* is the average returns of i's same-industry peer firms that are located outside of the focal firm's headquarters states, weighted by the SCI between the headquarters locations of firm i and j. Formally, *SPFRET* is given by

$$SPFRET_{i,t} \equiv \frac{\sum_{j \in I_i} SCI_{i,j} RET_{j,t}}{\sum_{j \in I_i} SCI_{i,j}},$$

where I_i is the Fama-French 48-industry (FF48) to which firm i belongs, and S_i is the state where firm i's headquarters are located.[8] We exclude peer firms from the same state as the focal firm, thus alleviating the concerns that our results are driven by firms in close geographic proximity (e.g., Pirinsky and Wang, 2006; Parsons et al., 2020). For our long-run prediction analysis, we also consider *INDMOM*, the long-run version of *SPFMOM*, by compounding the variable from month $t-11$ to $t-1$.

We illustrate the construction of *SPFRET* in Figure 1 Panel C, in which the focal firm is Cummins Incorporated (ticker: CMI). CMI is headquartered in Bartholomew County. The firm belongs to the machinery industry (FF48 code Mach) and manufactures engines and filtration and power generation products. Figure 1 plots the heat maps of SCI for Bartholomew county, showing only the counties with the headquarters presence of firms in the machinery industry, whereas counties without the presence of such firms are presented in dark grey. Examples of CMI's industry peers include Caterpillar (ticker: CAT), Clarcor (ticker: CLC), and Stanley Black & Decker (ticker: SWK), located in Peoria, IL, Williamson, TN, and Hartford, CT, respectively.[9]

[7]Bailey et al. (2018a) show that social connections across regions are persistent and argue that SCI is a reasonable proxy for historical social connectedness between regions.

[8]We obtain firm headquarters location data from the "Augmented 10-X Header Data" for the period between 1994 and 2018, supplemented by data parsed from 10-Q and 10-K filings on SEC's Edgar database for 2019. For observations prior to 1994, for which firms' electronic filings are unavailable, we follow Parsons et al. (2020) and use the first available headquarters location in a firm's post-1994 SEC filings. As argued by Parsons et al. (2020), such measurement error would only create noise and bias against finding significant results. We thank Bill McDonald for providing the pre-1998 data through the Notre Dame Software Repository for Accounting and Finance (SRAF).

[9]Caterpillar designs, develops, engineers, manufactures, markets, and sells machinery, engines, financial products, and

When calculating *SPFRET*, the firms headquartered in counties with higher SCI with Bartholomew County receive higher weights. As a result, CAT and CLC have corresponding weights of 3.44% and 4.32%, respectively, due to the higher SCIs between their headquarters counties and Bartholomew County, whereas the weight on SWK is only 0.57%, due to the low SCI between Williamson and Bartholomew Counties. The SCI-based weighting of industry peers is distinctly different from the value-based weights of 13.66%, 0.83%, and 4.32% for CAT, CLC, and SWK, respectively. Hence, the *SPFRET* measure gives more weight to the returns of a firm's industry peers from socially connected regions, which captures a dimension of the linkages between firms and their industry peers that is different from those reflected in a value- or equally-weighted industry portfolio.

2.2 Firm Fundamentals

Following Parsons et al. (2020), we examine the comovement between focal firms and their peer firms' fundamental changes. We examine four firm fundamental variables, including ΔEPS, $\Delta Sales$, $\Delta Employees$ and $NewCapital$, all obtained using the annual fundamentals data set in Compustat. These variables are defined as follows: ΔEPS is the change in EPS scaled by lagged stock price (as in Kothari et al. (2006)), $\Delta Sales$ is the percentage growth in sales, $\Delta Employees$ is the percentage growth in the number of employees, and $NewCapital$ is the sum of net equity issuance plus net debt issuance scaled by lagged enterprise value [10].

We define a focal firm's social peer fundamentals as the SCI-weighted average fundamentals of all its social peers. In order to be counted as a social peer (for fundamentals calculation purposes), a peer firm must have the same fiscal year-end and be in the same FF48 industry as the focal firm. We further define a focal firm's industry fundamentals as the equal weighted average of the fundamentals of all firms that are in the same FF48 industry as the focal firm, and that have the same fiscal year-end as the focal firm. We exclude same-state industry peers from both social peer and industry fundamentals calculations in order to avoid confounding our results with geographic effects.

2.3 Controls

Lead-Lag Return Predictors We control for a list of variables that previous studies have shown to have cross-firm return predictability. The data are available for the entire sample period from

insurance to customers via a worldwide dealer network. It is the world's largest construction equipment manufacturer. CLARCOR, Inc. manufactures engine filtration and industrial/environmental filtration systems. Stanley Black & Decker, Inc. manufactures industrial tools and household hardware, and provides security products. The weights are measured as of April 2016.

[10] Following Loughran and Wellman (2011), enterprise value is defined as market value of equity (Compustat data item MKVALT) plus short-term debt (DLC) plus long-term debt (DLTT) plus preferred stock value (PSTK) minus cash and short-term investments (CHE).

July 1963 through December 2019 unless otherwise mentioned.

For a given stock and for a given month t, we define *INDRET* as the month-t equal-weighted average return of stocks from the same industry as the focal stock (see Moskowitz and Grinblatt, 1999). *TNICRET*, available starting July 1989, is the equal-weighted stock return of peer firms identified through 10-K product text (Hoberg and Phillips, 2018). *GEORET* is the month t equal-weighted average return of all stocks from the same economic area (EA) as the focal stock, excluding same-industry stocks (Parsons et al., 2020). *CFRET*, available since January 1982, is the month t weighted average return of stocks that share at least one analyst with the focal stock over the previous 12 months, where weights are the number of shared analysts between stocks.

CRET, available starting December 1976, is the equal-weighted average stock return of the main customers of the focal firm, where a six-month gap is required between the fiscal year-end of the supplier and stock returns (Cohen and Frazzini, 2008). *TECHRET*, available between July 1963 and July 2012, is the weighted average stock return of technology-linked peer firms, where the weights are the technological closeness between the peer firm and the focal firm, determined by the similarities between patent distributions across different technology categories (Lee et al., 2019).[11] *CONGRET* is the pseudo-conglomerate return, defined as the sales-weighted return of (value-weighted) single-segment firm portfolios, formed for each segment in which a conglomerate firm operates (Cohen and Lou, 2012); data is obtained from the Compustat segment files and starts from July 1977.

We also consider a potential lead-lag effects driven by cross-firm labor flows. The job-to-job (J2J) flow statistics, available quarterly from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) database, provide estimates of worker flows across jobs and across different geographic labor markets by worker and firm characteristics (see Hyatt et al., 2014; Hahn et al., 2017). We track the within-industry labor flows between firms' headquarters locations for each industry and define *J2JRET* as the labor flow intensity-weighted industry peer returns.[12]

For our long-run prediction analysis, we also consider the long-run versions of the above

[11] The economic area-ZIP Code link file is obtained from Riccardo Sabbatucci's website. The textual network industry relatedness classification (TNIC) for individual firms are obtained from the online data library of Gerard Hoberg and Gordon Phillips. Customer-supplier firm links are obtained through the Linking Suite by WRDS. We thank Usman Ali and Stephen Teng Sun for providing their data on technological overlap among firms.

[12] For a given industry, we define the directional labor migration probability from state i to state j as the labor flow from i to j (relative to the number of employees in j) divided by the sum of i's relative labor outflows. We then define the (two-way) labor flow intensity between states i and j, $J2J$, as the average of the labor migration probabilities from i to j and from j to i. We average the quarterly values to annual frequencies to reduce seasonality effects. $J2JRET$ is then the $J2J$-weighted industry peer firm returns (excluding same-state peers) where monthly returns between July of year t and June of $t+1$ are matched to the $J2J$ of year $t-1$. The LEHD data are available since year 2000, and the $J2J$ measures are persistent, with an annual AR(1) coefficient of 0.93. Given the persistence, we backfill the observations prior to 2000 with the corresponding value as of year 2000. To the extent that we use this as a control variable, the backfilling would only bias against finding results for our main variable of interest, *SPFRET*.

variables (i.e., *INDMOM, GEOMOM, CFMOM, CMOM, TECHMOM, CONGMOM, TNICMOM*, and *J2JMOM*), by compounding the variable from month $t-11$ to $t-1$. Following Ali and Hirshleifer (2020), *CFMOM* is calculated as the coverage-weighted average of cumulative returns of analyst-coverage-connected stocks from month $t-11$ to $t-1$.

Other Controls We also control for a stock's own characteristics that the literature has shown to predict returns. The variables are measured as of the end of month t unless otherwise stated and are described below.

RET is the monthly stock return, adjusted for delisting to avoid survivorship bias (Shumway, 1997). We estimate firm size (*SIZE*) and book-to-market ratio (*BMKT*) following Fama and French (1992). *MOM* is obtained by compounding *RET* from month $t-11$ to $t-1$ (Jegadeesh and Titman, 1993). *IVOL* is the monthly idiosyncratic volatility, computed as the standard deviation of the daily residuals obtained by regressing the stock return on the Fama and French (1992) three factors over the previous month (Ang et al., 2006). *ILLIQ* is Amihud's illiquidity (Amihud, 2002), defined as the average daily ratio of the absolute stock return to the dollar trading volume within the previous month. *MAX* is the maximum daily stock return realized over the previous month (Bali et al., 2011). *SKEW* is the sample skewness of the daily stock returns from the previous month (Bali and Hovakimian, 2009). *COSKEW* is the stock's monthly coskewness, following (Harvey and Siddique, 2000).

2.4 Summary Statistics

Table 1 presents the time-series averages of cross-sectional statistics (Panel A) and correlations (Panel B) of the main variables in our analyses. All variables except RET_{t+1} are measured at the end of month t. Summary statistics shown in Panel A are consistent with earlier studies such as Ali and Hirshleifer (2020).

We standardize independent variables in our regressions using the methodology of Gu et al. (2020). Specifically, each month, we cross-sectionally rank all independent variables from lowest to highest and map the ranks into the $[0,1]$ interval through scaling them by the number of observations. This non-parametric transformation allows the analysis to focus on the ordering of the variables as opposed to the magnitudes, which makes the estimates less sensitive to outliers(e.g., Kelly et al., 2019; Freyberger et al., 2020).

In Panel B, we report correlation between variables, with the lead-lag variables normalized to between 0 and 1. *SPFRET* and *INDRET* are both positively correlated with contemporaneous returns (*RET*), with *SPFRET* having a higher correlation than *INDRET* (0.166 and 0.143 respectively). We find that *SPFRET* is highly correlated with *CFRET* of Ali and Hirshleifer (2020), with a coefficient of 0.49. This may be because the firms that share analyst coverage are likely to be in the same industry. Furthermore, the average cross-sectional correlation between *SPFRET* and *INDRET* is considerable, at 0.70. This is likely due to the construction of *SPFRET*, which relies on the SCI-weighted, same-industry firm returns. Given the finding that industry momentum strongly predicts a stock's future returns (e.g., Moskowitz and Grinblatt, 1999; Hoberg and Phillips, 2018), we first conduct preliminary analysis in the next subsection to gain insight into the extent to which social ties between firms are associated with the return predictability of industry momentum.

[Insert Table **1** here]

3 Social Ties and Firm Comovements

A growing body of literature shows that social ties between regions foster economic interactions (e.g., Cohen et al., 2017; Bailey et al., 2018b), suggesting that social ties can potentially serve as a proxy for the strength of economic links between firms located in these regions. In this section, we examine if strong social connections are associated with higher comovement in firm fundamentals and the firms' stock returns.

Following Parsons et al. (2020), we conduct the panel regression:

$$X_{i,j,t} = \beta_1 X_{SPF,i,j,t} + \beta_2 X_{EW,j,t} + \alpha_t + \epsilon_{i,j,t}. \tag{1}$$

The dependent variable $X_{i,j,t}$ is a fundamental measure for focal firm i, in industry j, and at time t. We consider four fundamental measures, including ΔEPS, $\Delta Sales$, $\Delta Employees$, and $NewCapital$. These fundamental variables are measured at the annual frequency. We also consider firm returns, measured at the monthly frequency.[13] The key independent variable X_{SCI} is the change in the corresponding fundamental measure of industry peers, weighted by their SCI to the focal firm's headquarters. We exclude firms from the focal firm's headquarters states to reduce the influence of geographically co-located firms studied in (Pirinsky and Wang, 2006; Parsons et al., 2020). To ensure that our results are not solely driven by fundamental comovement due to industry affiliation, we control for the average change in the fundamental for industry peers X_{EW}.[14] We additionally control for time fixed effects.

We start our analyses with univariate analyses. In Table **2** Panel A, we focus on the relationship between between SCI-weighted portfolio and the focal firm. Column 1 presents the result

[13] In columns 1-4, for each fiscal year end-focal firm pair we look for peer firms that have the same fiscal year-end as the focal firm to make sure that the fundamentals are calculated over the same time period. In column 5, this constraint is not imposed. This difference results in a further drop in observations for fundamental variables.

[14] We exclude the focal form when forming the equal-weighted industry peer.

based on ΔEPS. We find that focal firms' EPS growth significantly comove with that of social peer firms. Specifically, focal firms with highest social peer firm EPS growth exhibit 245 basis points higher ΔEPS than those firms with low social peer firm ΔEPS. Similarly, as reported in column 2, we find that even after controlling for industry sales growth ($\Delta Sales$), firms with the highest social peer $\Delta Sales$ are 31 percentage points higher than those firms with the lowest social peer firm sales growth. We also find significant comovement between focal firms and their social peers in both employee number and newly raised capital. Given strong comovements between focal firms' and social peer firms' fundamentals, we expect that focal firms' returns to comove strongly with their social peer firms. To investigate this hypothesis, we conduct the regression 1 with monthly returns as the variable of interest. We report this result in column 5. We find that firms with the highest social peer firm returns outperform those with the lowest social peer firm returns by 7.91%. This shows that firms headquartered in locations with close social ties to the focal firm county are more informative about the focal firm than average industry returns. In Panel B, we report univariate relationship between equal weighted industry portfolio and the focal firm. While we find a strong positive comovement between the equal weighted industry peers and the focal firm, the economic magnitudes of these coefficients are smaller than those reported in Panel A.

In Panel C, we further conduct multivariate regression analyses by including both SCI-weighted and the equal weighted portfolios. This analyses further highlight that SCI-weighted portfolio tends to comove more with the focal firm compared with equal weighted portfolio. Out of the four fundamental performance variables, SCI-weighted peer consistently exhibit strong and positive coefficients, while equal weighted portfolio only positively and significantly relate to focal firms' sales growths. Similarly, we also find that SCI-weighted returns exhibit strong return comovement with focal firms, but equal weighted firms do not exhibit significant comovement with that of focal firms.

We also conduct both univariate and multivariate analyses with value-weighted industry portfolio instead of equal-weighted industry portfolio and our results are not only robust but even stronger (see Table A2). These results suggest that SCI-weighted portfolio are more tightly connected with focal firms than equally- or value-weighted industry peers.

4 Social Ties and Return Predictions

As shown in Section 3, focal firms' returns strongly comove with its social peer firm returns. If investors do not fully incorporate the information from social peer firms, their returns can potentially predict focal firms' future return performances. In this section, we formally examine whether returns of social peer firms can explain and predict the returns of focal firms.

We begin our analysis with portfolio analyses, showing that a long-short portfolio sorted by

SPFRET produces significant abnormal returns. We then conduct spanning tests to investigate the extent to which the abnormal returns can be explained by existing economic mechanisms.

4.1 Portfolio Analysis

We first perform portfolio analysis and investigate the returns to a long-short portfolio sorted by *SPFRET*. Specifically, for each month, we sort our sample firms into deciles based on *SPFRET*. We then hold the portfolios for a month and calculate returns for each portfolio and the return differential between the portfolios with the highest and the lowest *SPFRET*.

Table 3 reports the results of the univariate portfolio analysis. The first row in each panel shows the raw returns for equal-weighted portfolios. To ensure that our results are not driven by risk factors, we also report the Fama-French five-factor alphas (FF5) for equal-weighted portfolios in the second row. The third and fourth rows in each panel report the raw returns and FF5 alphas for value-weighted portfolios, respectively. We find that SPFRET-sorted long-short portfolio produces economically large and highly significant excess returns. For example, the FF5 alpha is 157 basis points per month (t = 7.06) for equal-weighted portfolio and 84 basis points (t = 5.22) for the value-weighted portfolio.

[Insert Table 3 here]

Figure 2 presents the average monthly long-short portfolio alpha over time. Panel A shows that a positive alpha exists for most of the years in our sample. Given that *SPFRET* is highly correlated with industry momentum, we examine the incremental return predictability of *SPFRET* by examining the abnormal returns of portfolios sorted by $SPFRET_{\perp}$ (after controlling for industry momentum). Panel B shows that the $SPFRET_{\perp}$-based portfolios also consistently produce positive returns. This suggests that *SPFRET* captures additional information that is distinct from the average industry momentum effects.

Panel A also shows that the abnormal returns for the post-2000 period tend to be somewhat smaller. One reason for this is that industry momentum is substantially weaker and becomes insignificant post 2000 (Ali and Hirshleifer, 2020). In contrast, Panel B shows that, after filtering out the industry returns, the $SPFRET_{\perp}$-based abnormal returns perform reasonably well in more recent periods, again confirming the relevance of the incremental information contained in the social peer firm returns.

[Insert Figure 2 here]

Next, we conduct two-way sorted portfolio analysis to further isolate the effect of *SPFRET* from the effects of industry momentum. At the end of each month, stocks are first sorted into quintiles based on *INDRET*. Then, within each *INDRET* quintile, stocks are further sorted into quintiles based on *SPFRET*. We calculate the returns for each of the 25 portfolios in the month that follows, along with the returns of five high-minus-low portfolios based on *SPFRET* for each *INDRET* quintile.

We report the results of a bivariate portfolio sort with equal-weighted and value-weighted Fama and French (2015) abnormal returns in Table **4**, Panels A and B, respectively. The final rows in both panels report the average alphas for the *SPFRET* quintiles. The high-minus-low average alpha in Panel A is 56 basis points per month with a t-statistic of 7.38 whereas the economic magnitude under value-weight is 39 basis points with a t-statistic of 4.84, as reported in Panel B. Hence, the bivariate sort results further confirm that *SPFRET* has economically large return predictability that is above and beyond the effect of the traditional industry momentum of Moskowitz and Grinblatt (1999).

Another way to control for industry momentum is to sort *SPFRET* within each industry. We report these results in Table A3. Consistent with the double-sort results presented in Table **4**, *SPFRET*-based strategy still generates significant abnormal returns.

4.2 Heterogeneity in Return Predictability

Our analysis has so far established that social ties between industry peer firm locations capture important linkages between firms. Even so, these linkages would not necessarily result in return predictability if the linkages are well understood by investors and timely reflected in stock prices. Therefore in this subsection, we conduct additional analysis to provide insight into the underlying market frictions that may contribute to the failure to incorporate such information. We do so by exploring differences in a firm's information environment and the firm's headquarters location relative to industry cluster areas.

Information Environment One explanation for the return predictability that we document is slow information diffusion (e.g., Hong and Stein, 1999), due to investors' limited cognitive resources (e.g., Hirshleifer and Teoh, 2003; Peng and Xiong, 2006). The limited attention mechanism further predicts that the predictability should be stronger for firms with poor information environments. Thus, we examine how our results vary across firms with different information environments using three common proxies in the literature: firm size, analyst coverage, and institutional ownership.

We conduct a double-sort analyses by *SPFRET* and the three information environment proxies. We report the value-weighted results for the short and the long legs of the strategy, as well as the long-short returns in Table **5** Panel A. We find that the return predictability is mainly driven by firms with low market capitalization. Long-short returns are 181 basis points per month for small firms, which is over three times as large as that for large firms. Similarly, we find that *SPFRET*-based strategy generates more positive returns for firms with low institutional ownership and analyst coverage.

Taken together, these results support the slow information diffusion hypothesis and suggest that a firm's visibility to its investors influences the speed at which information about the firm's peer returns are incorporated into the firm's stock prices.

[Insert Table 5 here]

Firms in Industry Cluster Areas The next hypothesis is motivated by Engelberg et al. (2018) that firms located inside industry clusters are watched more closely by analysts and other market participants. As a result, social peer firm information is more timely impounded into these firms. Therefore the hypothesis predicts that the return predictability of *SPFRET* would be stronger for firms located outside of industry clusters.

We first define a county as an industry cluster county if the county is ranked among the top 20% or 10% in the market capitalization for a given Fama-French 48 industry. We then define an indicator variable, *OUTCLS*, as one if a firm's headquarters is located in one of the industry cluster counties and zero otherwise.

We double sort the firms based on their *SPFRET* and *OUTCLS* and report the portfolio return results in Panel B. The first three columns define industry cluster counties using the top 20% of counties, and the next two columns define top 10% of counties as industry cluster counties. In both specifications, we find that *SPFRET* generates stronger long-short returns for firms that located outside of industry clusters compared with firms headquartered inside industry clusters. The relation suggest that stock prices of firms located outside of industry centers are slow to incorporate important information in social peer firm returns.

5 Predictability in the Long Run

In this section, we test whether the returns of social peer firms can predict stock returns in longer horizons. We also examine the types of the fundamental information about focal firms that social peer firms' returns contain. In addition, we investigate the extent to which investors and analysts fail to incorporate such information into their expectations.

5.1 Predicting Long-Run Returns

As shown in subsection 4.2, our evidence is consistent with investors' underreaction to fundamental information (e.g., Hirshleifer and Teoh, 2003; Peng and Xiong, 2006). However, the same set of results can also be consistent with investors overreacting to information from socially connected firms (e.g., De Bondt and Thaler, 1985; Daniel et al., 1998). Therefore, we analyze the return predictability over longer horizons to help us differentiate these two hypotheses. If the predictability is a result of investors' underreaction, we should not observe a reversal in the long run, whereas the overreaction-based mechanism would imply long-run reversals.

For long-horizon analysis, we consider both *SPFRET* and the cumulative version of social peer firm returns, *SPFMOM*, following previous studies (e.g., Moskowitz and Grinblatt, 1999; Parsons et al., 2020; Ali and Hirshleifer, 2020). *SPFMOM* is also likely to reflect longer-run fundamental information more accurately than the monthly *SPFRET* as the latter tends to be noisier. Table **6** reports the long-horizon return predictability results using an calendar time portfolio approach following Jegadeesh and Titman (1993). Panel A presents the five-factor alpha of the equal-weighted calendar time portfolios. Both *SPFRET* and *SPFMOM* generate high return predictability in the month after portfolio formation (157 basis points based on *SPFRET* and 69 basis points based on *SPFMOM* per month). The long-short strategy continues to generate positive returns in months 2-3 after the portfolio formation month, generating a return differential of 24 and 61 basis points, respectively. The effect of *SPFRET* remains significant for months 4–6 and 7–12, and both variables lose their significance in generating excess returns after the first twelve months.[15] The cumulative returns for the long-short portfolio sorted by *SPFRET* is 4.332%. As indicated by the regression coefficient in Table 2 Panel A, column (5), the long-short strategy based on *SPFRET* corresponds to a contemporaneous return of 7.52%. That is, over 35% of the information in *SPFRET* is associated with delayed reaction.[16] Panel B presents the Fama and French (2015) five factor alpha for value-weighted portfolios. The patterns are similar to those presented in Panel A. In particular, we find that both *SPFRET* and *SPFMOM* can strongly predict returns in the short run and there is no evidence for long-run reversals.

Overall, the findings show that the return predictability of social peer firm returns lasts for up to one year and the lack of reversal provides further support to the underreaction-based explanations.

[Insert Table **6** here]

5.2 Predicting Future Fundamental Performances

We next investigate the nature of the information that social peer firm returns capture and consider whether such information help predict a firm's future fundamental performances. Specifically,

[15] We conduct the these analyses using the orthogonalized versions of *SPFRET* and *SPFMOM*, after controlling for all other variables in Panel A of Table A6. The results are similar to those obtained by using raw *SPFRET* and *SPFMOM*.

[16] The delayed reaction is calculated as $4.332\%/(7.52\% + 4.332\%)$.

we will focus on three earnings-related variables, including standardized unexpected earnings, forecast errors, and earnings announcement returns.

Empirically, we estimate panel regressions in the following form:

$$Fundamental_{i,t+s} = \alpha + \beta_1 SPFRET_{i,t} + \beta_2 SPFMOM_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+1}, \tag{2}$$

where *Fundamental* represents a fundamental variable. *SPFRET* and *SPFMOM* are social peer firm returns and momentum, respectively.[17] X is a vector of control variables, including time and firm fixed effects. We report these results In Table 7. In Panel A, we do not include additional fixed effects besides fixed effects. In Panel B, we additionally control for long-term and short-term industry momentum *INDMOM* and *INDRET*. Standard errors are clustered by month and firm

5.2.1 Predicting Earnings Growth

We first examine standardized unexpected earnings (SUE) in quarters 1–4 following the portfolio formation month, which captures firms' growths in earnings.[18] We find that *SPFMOM* and *SPFRET* exhibit strong power in predicting future SUEs. Specifically, an increase from the lowest *SPFMOM* to the highest leads to a 34 percentage points increase in the following quarter's SUE, which equals 24% of the average cross-sectional standard deviation of SUE. In panel B, we additionally control for industry return and industry momentum. Our results remain robust after adding these controls.

Overall, this result indicates that social peer firm returns are positively associated with future earnings growth, confirming that social peer firm returns contain valuable information about firms' fundamentals.

[Insert Table 7 here]

5.2.2 Do Investors Incorporate Information on Social Ties?

Next, we investigate the extent to which market participants incorporate information regarding the performance of social peer firms.

---

[17] We conduct robustness checks of results with $SPFRET_{\perp}$ and $SPFMOM_{\perp}$, the versions of *SPFRET* and *SPFMOM* that are orthogonalized to the other predictive variables. Table A6 presents the results and shows that our findings are robust.

[18] The standardized unexpected earnings (constructed following Bernard and Thomas (1990) and Mendenhall (1991)). We calculate the unexpected earnings (UE) of stock i in calender quarter q of an earnings announcement as $UE_{i,q} = EPS_{i,q} - EPS_{i,q-4}$, where $EPS_{i,q}$ and $EPS_{i,q-4}$ are the stock's basic earnings per share (EPS) excluding extraordinary items in quarters q and $q-4$ respectively. EPS is adjusted for stock splits and reverse splits through division by the Compustat item AJEXQ, the quarterly cumulative adjustment factor. Standardized unexpected earnings in quarter q ($SUE_{i,q}$) is defined as $UE_{i,q}$ scaled by its standard deviation over the past eight quarters, with a minimum of four UE observations available.

We first focus on a very important information provider in the financial market, the financial analyst. In a frictionless market in which analysts are fully aware of information about socially-connected industry peers, such information should have been incorporated into their earnings forecasts. On the other hand, if social peer firm return information has not been fully incorporated by financial analysts, then social peer firm returns would be able to predict future analyst forecast errors.

Similarly, if marginal investors of a stock are fully aware of the peers' information and correctly price the focal stock's prices, social peer firm returns would not be able to forecast future earnings announcement returns. But if social peer firm information has not been fully incorporated by prices, then the peer returns would be able to forecast future earnings announcement returns. Therefore, we next examine whether social peer firm returns can predict future analyst forecast errors and earnings announcement returns.

Analyst Forecast Errors Financial analysts obtain and communicate information about the firms they follow and serve as crucial information intermediaries between firms and investors. Therefore, focusing on the information set of this class of sophisticated players in the financial market helps us to gain insight into the extent to which peer firm information is incorporated by market professionals.

Specifically, we analyze whether analysts have incorporated the peer firm information into their earnings forecasts by examining their forecast errors. Following DellaVigna and Pollet (2009), we consider the analyst forecast error in quarters $t+1$ to $t+4$ as key dependent variables.[19] We show that both *SPFMOM* and *SPFRET* can predict forecast errors over the following year and this remains robust after including industry momentum controls (as reported in Panel B). These results suggest that even professional market participants such as sell-side analysts are sluggish in incorporating the information of social peer firms into their earnings forecasts, with a substantial delay of up to one year.

Announcement Returns An alternative method to assess whether relevant fundamental information of social peer firms has been incorporated by investors is to examine future earnings announcement returns. If investors fail to consider such information in a timely manner, such underreaction would be corrected upon future earnings announcements. Hence, we expect that social

[19]Forecast errors are defined as the difference between the announced earnings for that quarter and analyst consensus forecast (i.e., the median forecast), scaled by the stock price five trading days before the earnings announcement date. We adjust actuals, forecasts, and prices for stock splits and reverse splits through scaling them by the item CFACSHR—the cumulative adjustment factor from the daily CRSP file. We only use the forecasts for the next quarter and within 90 days of the earnings announcement date. If an analyst makes multiple forecasts within this time window, we use the latest forecast.

peer firms' returns would positively predict focal firms' earnings announcement returns in the future.

To examine this hypothesis, we conduct a panel regression in the following form with the dependent variable *CAR*, which represents the market-adjusted cumulative abnormal returns computed for three-day windows around earnings announcements in the next four quarters.

We find that both *SPFRET* and *SPFMOM* are strongly related to earnings announcement returns in the following quarter. In fact, *SPFMOM* can predict earnings announcement CAR in two quarters. In terms of the economic magnitude, companies with the highest *SPFMOM* or *SPFRET* have 14 basis points higher returns in the following quarter's CAR compared with those with the lowest *SPFMOM* (*SPFRET*). As shown in Panel B, while controlling for industry momentum variables reduce the economic significance of the predictive coefficients in the first two quarters, it leads to more prolonged return predictability for *SPFMOM*.

Taken together, our results suggest that the strong return predictability of social peer returns on focal firm returns that we document in Section 3 are due to the fundamental linkages between the firms that have not been fully recognized by market participants.

6 Social Peer Firm Returns and Other Lead-lag Predictors

So far, we show that *SPFRET* significantly predicts future returns. However, past research also documents many other lead-lag variables that can predict firm returns. Thus, in this section, we examine the robustness of our predictability results after controlling for lead-lag relationships and other well known cross-sectional predictors.

6.1 Spanning Tests

In this subsection, we conduct a portfolio spanning test to confirm that the return predictability is not driven by other well-documented lead-lag relationships documented in the existing literature.

Specifically, we follow Ali and Hirshleifer (2020) and estimate the following time-series regression:

$$RET_{SPFRET,t} = \alpha + \beta F_t + \epsilon_t, \tag{3}$$

where RET_{SPFRET} is the value-weighted long-short portfolio returns constructed in Section 4.1. F represents a vector of long-short portfolios. We consider the Fama-French five-factor (Fama and French, 2015) returns and the returns of portfolios based on firms' economic linkages as identified by previous studies. To ensure that the coefficients are comparable across different columns, we require that return information exists for all the portfolios considered in our test.

Table 8 reports the results for a sample period in which all portfolios under our consideration have valid return data to facilitate the comparison across different specifications. Column 1 is the baseline specification with a constant and the Fama-French five-factors as the explanatory variables. The column shows that the coefficient on the constant, which captures the Fama-French five-factor alpha of the *SPFRET*-based portfolio strategy, is 92 basis points per month and is significant at the 1% level. The result is consistent with the portfolio analysis that we show in Table **3**.

Next, we gradually include additional long-short portfolios one at a time to examine the extent that the alpha of a *SPFRET* portfolio can be explained by other lead-lag effects. We first consider the industry momentum strategy based on the FF 48 industry (i.e., *INDRET*) in column 2. Industry momentum is highly relevant as *SPFRET* is constructed using firms from the same industry. Indeed, we find that the *SPFRET* strategy loads highly positively on the industry momentum portfolio. However, after controlling for *INDRET*, the alpha of *SPFRET* is 77 basis points per month and remains significant at the 1% level.

We next include the geographic momentum documented in Parsons et al. (2020). Column 3 shows that the coefficient of *GEORET* is insignificant, suggesting that geographic momentum is not a significant contributor to the alpha of *SPFRET*. This may be because we exclude industry peers that are headquartered in the focal firm's headquarters county when constructing the *SPFRET* variable.

We also consider a new channel, which is the labor flows between firms' headquarters locations, as a possible contributor to our findings. Column 4 shows that the labor flow-based industry peer firm returns, *J2JRET*, explains an economically meaningful portion of the returns to the *SPFRET*-based portfolio. However, the *SPFRET* alpha remains robust, at 60 basis points.

Cohen and Lou (2012) document that a conglomerate firm's returns can be predicted by the return of its industry segments. While it is unclear how this effect can directly explain our results, we nonetheless consider it as an additional control. Again, as reported in column 5, we find that our portfolio loads on *CONGRET* positively and significantly, but our alpha remains highly significant.

In column 6, we additionally include the lead-lag effect due to the customer-supplier relationship (Cohen and Frazzini, 2008). It is possible that firms located in socially connected areas are more likely to form customer-supplier relationships. The *SPFRET* alpha remains highly significant (54 basis points per month).

In column 7, we further include lead-lag effect due to product market similarities (see Hoberg and Phillips, 2018, for more detail). We find that the *SPFRET* portfolio loads significantly on both industry return portfolios. However, the alpha of the *SPFRET* portfolio remains highly significant, at 51 basis points.

We also add the lead-lag effect due to technological similarity between firms (Lee et al., 2019), as Bailey et al. (2018a) find evidence consistent with social connectedness facilitating technological spillover across regions. As reported in column 8, although the coefficient on the *TECHRET* portfolio is positive, the variable does not fully subsume the alpha of the *SPFRET* portfolio.

Finally, we control for portfolio returns based on *CFRET*, which captures returns of lagged returns of firms with shared analyst coverage (e.g., Ali and Hirshleifer, 2020). This strategy is not intended to capture a specific economic relationship, but it is shown that it accurately summarize economic relevance between two firms and thus subsumes many existing lead-lag relationship. We include the return based on *CFRET* in column 9. While *SPFRET*-based portfolio significantly loads on the *CFRET*-based portfolio returns, we find that *SPFRET*-based portfolio still generate a 43.5 basis points alpha in the following month.

Overall, this result indicates that the most relevant lead-lag effects to the *SPFRET* portfolio are industry momentum, labor flow-based industry peer firm returns, and the shared analyst lead-lag. After such stringent control, our alpha remains both economically as well as statistically significant.

[Insert Table **8** here]

Figure 3 further illustrates the relative contributions of the economic linkage-based variables, using coefficient estimates from the last column of Table **8** and the average returns of the corresponding long-short portfolios. The figure shows that, among the largest contributors of the *SPFRET* portfolio excess returns, shared analyst returns explain close to 15 basis points and *J2JRET* explains roughly 14 basis points. Industry momentum explains roughly 8 basis points. The contribution from the other factors are relatively small, all below 5 basis points. More important, even after accounting for the economic linkages documented in the existing literature, the *SPFRET* portfolio still generates a substantial incremental alpha, which is not only higher than the contribution from each of the individual long-short portfolios but is also greater than the contribution from the sum of the contribution of the risk factors and other lead-lag relationships.

[Insert Figure 3 here]

In sum, the spanning tests show that social connectedness between firm locations corresponds to important economic ties across firms such as industry or product similarity, labor flow, customer-supplier links, and technology similarity. But our social tie–based peer firm measure also captures a substantial amount of cross-firm linkages that existing measures do not account for. Our results therefore suggest that the linkages across firms can be complex and nuanced and that these linkages have not been fully understood by the market participants or incorporated into prices.

6.2 Fama-MacBeth Regression Analysis

Our portfolio sorting analyses and spanning tests show that a long-short portfolio based on social peer firm returns generates strongly abnormal returns that are not subsumed by the abnormal returns obtained by sorting on alternative variables. In this subsection, we conduct further analysis using Fama-MacBeth regression, which allows us to examine whether social peer firm returns predict future stock returns at the stock level while accounting for a comprehensive set of controls.

We estimate the following regression:

$$RET_{i,t+1} = \alpha + \beta SPFRET_{i,t} + \gamma X_{i,t} + \epsilon_{i,t+1}, \tag{4}$$

where $RET_{i,t+1}$ is the one-month-ahead firm return and *SPFRET* is the social peer firm returns. X is a vector of control variables that includes short-term industry momentum (*INDRET*), the geographic lead-lag effect (*GEORET*), the shared analyst coverage lead-lag effect (*CFRET*), the customer-supplier lead-lag effect (*CRET*), the technological spillover effect (*TECHRET*), the complicated firm effect (*CONGRET*), the labor market lead-lag effect (*J2JRET*), and the text-based industry momentum (*TNICRET*). We also include a battery of well-known cross-sectional predictive variables, including *RET*, *SIZE*, *BMKT*, *BM*, *MOM*, *IVOL*, *ILLIQ*, *MAX*, *SKEW*, and *COSKEW*. We standardize the independent variables following Gu et al. (2020) and Kelly et al. (2019).

Table 9 reports the results of Fama-MacBeth regressions. In column 1, we only include *SPFRET* and a vector of return predictability controls. We find that *SPFRET* is positive and significant at the 1% level. Moving from the lowest to the highest *SPFRET* leads to a 1.4% higher average return in the following month. Next, we gradually include additional lead-lag controls in columns 2 through 9.

In columns 2, we include the industry momentum effect documented in Moskowitz and Grinblatt (1999). As expected, the industry momentum variables are highly significant and their inclusion attenuates the effect of *SPFRET* to some extent, due to the strong positive correlation between these variables. Nevertheless, the coefficient of *SPFRET* remains large and highly significant, at 0.787, with a *t*-statistic of 7.669.

Column 3 additionally controls for the geographic lead-lag effects and shows that the coefficient of *GEORET* is positive and highly significant, consistent with Parsons et al. (2020). More importantly, the coefficient of *SPFRET* remains very similar to the corresponding value in column 2, suggesting that the geographic lead-lag effects does not contribute significantly to the effect of *SPFRET*.

In column 4, we add a control for the labor flow–based return predictor that we propose and shows that *J2JRET* is positive and significant. Consistent with the spanning test, we find that

J2JRET slightly weakens the predictive power of *SPFRET*. However, the coefficient of *SPFRET* remains highly significant both economically and statistically.

Similarly, in columns 5 through 8, we further control for the "complicated firm" effect (Cohen and Lou, 2012) by including *CRET*, the customer-supplier effect (Cohen and Frazzini, 2008), the product market similarity effect (Hoberg and Phillips, 2018) (*TNICRET*), and the effect of technological linkage (Lee et al., 2019) (*TECHRET*), respectively. We show that these variables have positive coefficients, consistent with the previous findings, but these controls do not significantly affect the economic magnitude and the statistical significance of *SPFRET*.

Finally, we control for *CFRET*, which is a proxy for the analyst lead-lag effect. This variable is known to explain many other lead-lag effects in the literature (Ali and Hirshleifer, 2020). Column 9 shows that the return predictability of *SPFRET* remains strong and economically meaningful even after including this control. To further ensure the validity of these key empirical analyses, we present a comprehensive set of robustness tests in subsection 6.4.

In sum, our results show that *SPFRET* continues to significantly predict next month's return despite including a comprehensive set of predictive variables associated with specific economic linkages across firms. The economic magnitude of *SPFRET* is substantial. Firms with the highest *SPFRET* outperform firms with the lowest *SPFRET* by 41.7 basis points in the following month. This set of results is consistent with the analyses presented in Table **8**, supporting the notion that industry peer firms located in socially proximate areas contain important information about the focal firm that is not fully incorporated into the focal firm's stock prices.

The comprehensive set of economic linkages that we consider partially explains the predictability of *SPFRET*, suggesting that social ties between firms' headquarters locations capture some of these known economic linkages. More importantly, our results also show that these economic linkages do not fully explain the predictability of *SPFRET*. Hence, our evidence suggests that the cross-firm linkages are multifaceted and that our social tie–based peer returns help provide incremental information regarding some of these nuanced linkages.

[Insert Table **9** here]

6.3 Relative Importance in the Composite Return Predictor

So far we have shown that our main variable, *SPFRET*, can significantly predict one-month-ahead focal stock returns in a way that is incremental to the 18 other predictors used in the literature. In this section, we assess the extent to which our variable can help enhance the joint return predictability of all the existing predictors using a machine learning approach. That is, we follow Kelly et al. (2019) and Gu et al. (2020) and evaluate the incremental contribution of *SPFRET* relative to a composite predictor that aggregates all 19 individual predictors (included in equation

(4). This test is particularly important since *SPFRET* is significantly correlated with many other lead-lag variables, which can lead to exaggerated statistical significance. The partial least square method can account for these problems (Abdi, 2010).

Specifically, starting with July 1994, each month, we train a partial least square regression (PLS) model using all the data available up to that point.[20] We generate a composite predictor that combines all 19 variables. Then, for each training set, we obtain the in-sample R^2s from the full PLS model and from a restricted model omitting one of the nine key cross-firm lead-lag return predictors (namely, *SPFRET*, *INDRET*, *GEORET*, *J2JRET*, *CONGRET*, *CRET*, *TNICRET*, *TECHRET*, and *CFRET*). For each omitted variable, we calculate the R^2 difference between the full model and the restricted model.[21] We then define the relative importance of the variable as the corresponding R^2 difference divided by the sum of the R^2 differences associated with each of the five predictors.

Figure 4 presents the monthly average relative importance of the nine cross-firm lead-lag return predictors for the full sample period of 1994–2019. It shows that, among the predictors, *SPFRET* contributes substantially to the predictive power of the composite predictor, with a relative importance of 18% for the full sample period.[22] In particular, *SPFRET* is more important than *INDRET*, with the relative importance of *SPFRET* exceeding that of *INDRET* by 7.9%. In sum, our analysis shows that *SPFRET* substantially improves the predictive power of the PLS-based composite return predictor, beyond the variables analyzed in the prior literature.

[Insert Figure 4 here]

To illustrate the economic magnitude of *SPFRET*'s contribution to the composite predictor, we follow Gu et al. (2020) and compare the performance of portfolios based on the composite predictor that includes information from all 19 predictive variables (i.e., full model predictor) and an alternative composite predictor, "No SPFRET", that excludes *SPFRET*. Each month, we sort stocks into deciles based on the two predictors respectively and obtain the one-month-ahead value-weighted returns of the corresponding portfolios. We then compute the cumulative log returns over the period of August 1994 through December 2019.

Figure 5 presents the cumulative log returns of portfolios based on the full model predictor (in orange) and the "No SPFRET" predictor (in blue). Panel A plots the returns to the long-short portfolios and shows that portfolio based on the composite predictor substantially outperforms

[20] *TNICRET* data starts in July 1989, and the length of the initial training period is five years. We employ a five fold cross-validation procedure to determine the number of PLS components for each training set.

[21] We follow Kelly et al. (2019) and Gu et al. (2020) and use the coefficient estimates obtained from the complete model to compute R^2 for the restricted model.

[22] The nine key predictors contribute to 42% of the R^2 differences in the complete model, whereas the remaining ten own-stock characteristics together contribute 58%.

the alternative "No SPFRET" portfolio, by 86.4 percentage points at the end of 2019. Panel B plots the returns of the long- and the short-legs separately and shows that the full model composite predictor outperforms the "No SPFRET" predictor for both the long and the short legs. Hence, the result suggests that *SPFRET* contains additional information that helps enhance the predictive power of the existing cross-sectional predictors in an economically meaningful way.

[Insert Figure 5 here]

6.4 Robustness

We have shown that *SPFRET* strongly predicts the one-month-ahead stock returns. In this subsection, we present a battery of robustness checks, based on the full model of the Fama-MacBeth specification as in Table **9**. Table A4 presents the results, with column 1 the same as column 9 of Table **9** to facilitate the comparison.

Panel Regressions with Time Fixed Effects In column 2, we use a panel regression to examine the predictive relationship instead of the Fama-MacBeth regression specification. We also double-cluster the standard errors by month and by stock. Pástor et al. (2017) argue that an OLS regression specification with time fixed effects weighs each cross-section differently whereas the Fama-MacBeth method uses equal weightings. Thus, our panel regression analysis ensures that our results are robust under this alternative weighting scheme. The coefficient of *SPFRET* is highly significant, and the magnitude is also large, at 0.65, which is even higher than the coefficient obtained from the Fama-MacBeth specification.

Controlling for Value-Weighted Industry Returns In our main analysis, we control for the equal-weighted industry returns (*INDRET*) because the variable is highly correlated with our main variable, *SPFRET*, with a coefficient of 63.8%. The correlation between *SPFRET* and the value-weighted *INDRET*, on the other hand, is 52.8%. Hence the equal-weighted *INDRET* serves as a more stringent control and allows the report of more conservative results. Column 3 presents the results when value-weighted *INDRET* are used instead. The coefficient on *SPFRET* is now substantially larger, both in magnitude and in statistical significance, compared to column 1, and *INDRET* becomes insignificant. This comparison confirms that our main findings are robust to the construction of *INDRET*.

Alternative SCI Measures One concern regarding the use of *SPFRET* is that the measure is based on Facebook's SCI as of 2016. We believe that SCI serves as a useful proxy that reflects stable historical social ties between regions. As shown in Bailey et al. (2018a), the Facebook SCI measure can be mapped to labor migration patterns dating back to the 1930s, suggesting that the SCI measure closely corresponds to historical social connections between regions. Nevertheless, there is a legitimate concern that traders may not have been able to use this measure in real time prior to 2016 and thus may not have been able to incorporate the specific signal. There is also a concern that SCI between counties are determined endogenously. For example, the presence of highly economically related firms could lead to higher social connectedness between firms' headquarters locations.

To address these concerns, we assess whether traders could use an alternative measure of social ties between two locations. As shown in Bailey et al. (2018a), the Facebook SCI is highly negatively correlated with geographic distance. Hence we use geographic proximity as an alternative proxy for social ties between regions and define $SPFRET_{DIST}$ as the inverse distance–weighted industry peer firm returns. Thus, the alternative *SPFRET* measure relies on the historically available information and is less likely to be subject to the reverse causality concern. Column 4 shows that $SPFRET_{DIST}$ significantly predicts future returns, with a comparable coefficient of 0.411. This suggests that one can use geographic proximity to proxy for social ties and generate significant abnormal profits in real time portfolios.[23]

Analysis at The ZIP-Code Level In column 5, we show that our results are robust to using a ZIP code–based SCI measure, which defines geographic regions more granularly. The coefficient of *SPFRET* remains positive and highly significant. Firms with the highest *SPFRET* outperform those with the lowest *SPFRET* by 40 basis points in the following month. This magnitude is consistent with the coefficient reported in column 1.

Alternative Industry Classification In column 6, we show that our result is robust to an alternative industry classification that is based on TNIC (Hoberg and Phillips, 2018). We find that the economic magnitude of the *SPFRET* coefficient becomes even higher compared with column 1.

Alternative Standardization Our main regression analysis follows Gu et al. (2020) and uses ranked independent variables that are scaled to the $[0, 1]$ interval to alleviate the influences of potential outliers. In column 7 of A4, we show that our results are robust under an alternative standardization procedure in which independent variables are demeaned and then divided by the variable's standard deviation. The coefficient on *SPFRET* is 0.109 and is statistically significant.

[23] The average cross-sectional correlation between *SPFRET* and $SPFRET_{DIST}$ is 45%. Despite this relatively high correlation, when we include both *SPFRET* and $SPFRET_{DIST}$ in the same regression, *SPFRET* remains highly significant at 0.385. Thus, consistent with Kuchler et al. (2021), our result shows that SCI captures the component of social connectedness between two locations that is distinctly different from geographic proximity.

Note that the coefficient in column 8 is not directly comparable with those from the other columns because of the scale differences. We therefore compare the economic magnitude of coefficients by looking at the return differences predicted by a *SPFRET* increase of 10% to 90%. The resulting return difference is 36 basis points based on the coefficient estimates in column 7, qualitatively similar to the corresponding value of 44 basis points for column 1.

Orthogonalized *SPFRET* We have so far directly controlled for 18 other return predictors in the regression analysis. To further address the concern that the coefficient of *SPFRET* may be influenced by its correlations with these other variables, we conduct an additional test using orthogonalized *SPFRET* as the main independent variable. We define $SPFRET_{\perp}$ as the regression residual of $SPFRET$ on all the other 18 independent variables (including the basic controls) and report the result based on this variable in column 8. The economic magnitude of the coefficient drops to some extent, consistent with our prior results that part of the effect of *SPFRET* is attributable to labor flow and other economic linkages documented in the literature. However, our results remain highly statistically significant and economically meaningful, at 0.265, confirming that the incremental information in *SPFRET* remains relevant.

Accounting for Overlaps in Firms' Economic Presence In column 9, we investigate whether the predictability of *SPFRET* can be explained by firms' economic presences in socially connected locations. We follow Garcia and Norli (2012) and Bernile et al. (2015) to extract the frequency of states in firms' 10-K filings. We exclude peer firms that are headquartered in a state in which the focal firm has an economic presence. This exercise is highly conservative, as it eliminates 28% of peer firms. We find that even under this very restrictive specification, *SPFRET* still delivers considerable predictive power.

Accounting for Homophily Another potential explanation for our result is that socially connected counties tend to have similar socioeconomic characteristics. Thus, firms located in connected counties are more likely to face correlated economic shocks. Thus, we control for the lagged portfolio returns that are weighted based on the similarity between the focal county and peer firms' counties with respect to four county characteristics: population density, education, political inclination, and income.[24] In column 10, we report the Fama-MacBeth regression that includes this control and find that *SPFRET* still exhibit significant return predictability. In contrast,

[24]For each county, we form a four dimensional vector based on these four dimensions. Each dimension is normalized to a value between 0 and 1 following Gu et al. (2020). For a pair of county, we calculate the Euclidean distance based on their county characteristic vectors. When calculating *SIMRET*, we weight out-of-state industry peer firm returns based on the inverse Euclidean distance.

SIMRET does not exhibit a significant predictive coefficient. Thus, this result suggests that our results are not driven by county homophily.

In-State v.s. Out-of-State Peers We also delve deeper into the construction of *SPFRET* and evaluate the predictive power of two alternative variables. Our main measure *SPFRET* excludes returns of same-state peers, alleviating the concerns that our results are driven by geographic proximity (e.g., Parsons et al., 2020). To directly account for the returns of same-state industry peers, we define $SPFRET_{STATE}$ as the SCI-weighted returns of same-state industry peers as the focal firm. Table A5, column 1, presents the results. We find that the coefficient of *SPFRET* remains highly significant and comparable to the corresponding coefficients in Table **9**. The coefficient on $SPFRET_{STATE}$ is also positive and significant, although somewhat smaller than *SPFRET*. This result further confirms that our results are not driven solely by the return predictability of geographically proximate industry firms.

Returns of Firms From Other Industries Similarly, to examine whether returns from socially connected firms from other industries exhibit predictive power, we define *NPFRET* as the SCI-weighted returns of firms from other industries. Table A5, column 2 presents the results and shows that *SPFRET* remains robust, whereas *NPFRET* is insignificant. This suggest that the power of *SPFRET* mostly comes from industry-based fundamental linkages between firms.

7 Conclusion

As Amazon's recent search for a second headquarter illustrates, firms take their location choices very seriously. At least for firms in technology industries, vibrant environments that provide access to an educated workforce are a primary consideration. As we show in this paper, the social connections between locations can also affect a firm's fundamentals and its stock returns.

Our evidence indicates that firms that belong to the same industry and are headquartered in socially connected locations exhibit strong co-movements in their stock returns as well as their fundamentals, measured by earnings growth, sales growth, change in the number of employees, and new capital raised. However, a measure of the returns of a firm's same-industry social peers, (*SPFRET*), strongly predicts its future returns, suggesting that information about a firm's social peers is not immediately reflected in its stock price. A long-short portfolio based on *SPFRET* generates a highly significant Fama and French (2015) monthly alpha of 157 basis points (equal-weighted) and 84 basis points (value-weighted).

We show that only about half of the long-short portfolio alpha can be explained by economic connections that have been documented in previous studies (such as industry momentum, technology similarity, customer-supplier linkages, and labor flows across regions), leaving unexplained a substantial abnormal return of 43 basis points per month. The result remains robust when we examine two-way sorted portfolios by *SPFRET* and industry returns as well as with Fama and MacBeth (1973) regression analysis, in which we control for 18 other lead-lag return predictors or stock characteristics established in previous studies. *SPFRET* is among the most important contributors to the power of a PLS-based composite predictor that aggregates the effects of all 18 individual predictors.

Consistent with investor inattention generating sluggish price adjustments, our results are stronger for firms with low visibility (measured by market capitalization, institutional ownership, or low analyst coverage), as well as for firms located outside of industry cluster counties. Furthermore, the predictability of *SPFRET* lasts for up to one year and does not reverse in the long run. In addition, social peer firm returns help predict focal firms' future earnings, analyst forecast errors, and future earnings announcement returns.

Our findings raise a number of issues that suggest future avenues of continuing research. In particular, the evidence highlights how the sphere of a city's influence may go beyond its borders. In future research, we will examine the extent to which firms in locations with greater social ties influence the returns of their industry peers more than those with locations with weaker social ties. The determinants of social ties and how they affect innovation also warrant future research. For example, we might expect firms in locations with greater social ties to be more innovative because they have access to more ideas, but if individuals from these locations are more influential, we might expect them to have more social ties. So identifying cause and effect is likely to be an important challenge.

References

Abdi, Hervé, 2010, Partial least squares regression and projection on latent structure regression (pls regression), *Wiley interdisciplinary reviews: computational statistics* 2, 97–106.

Acemoglu, Daron, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, 2012, The network origins of aggregate fluctuations, *Econometrica* 80, 1977–2016.

Ali, Usman, and David Hirshleifer, 2020, Shared analyst coverage: Unifying momentum spillover effects, *Journal of Financial Economics* 136, 649–675.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *The Journal of Finance* 61, 259–299.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, 2018a, Social connectedness: Measurements, determinants, and effects, *Journal of Economic Perspectives* 32, 259–80.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, 2018b, The economic effects of social networks: Evidence from the housing market, *Journal of Political Economy* 126, 2224–2276.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, 2018c, The economic effects of social networks: Evidence from the housing market, *Journal of Political Economy* 126, 2224–2276.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel, 2019a, House price beliefs and mortgage leverage choice, *The Review of Economic Studies* 86, 2403–2452.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel, 2019b, Social connectedness in urban areas, Technical Report w26029, NBER Working Paper.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel, 2021, International trade and social connectedness, *Journal of International Economics* 129, 103418.

Bailey, Michael, Drew M Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, 2019c, Peer effects in product adoption, Technical report, National Bureau of Economic Research.

Bali, Turan G., Nusret Cakici, and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *The Journal of Financial Economics* 99, 427–446.

Bali, Turan G., and Armen Hovakimian, 2009, Volatility spreads and expected returns, *Management Science* 55, 1797–1812.

Bernard, Victor L, and Jacob K Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305–340.

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

Breschi, Stefano, and Camilla Lenzi, 2016, Co-invention networks and inventive productivity in us cities, *Journal of Urban Economics* 92, 66–75.

Brown, Jeffrey R., Zoran Ivković, Paul A. Smith, and Scott Weisbenner, 2008, Neighbors matter: Causal community effects and stock market participation, *The Journal of Finance* 63, 1509–1531.

Cohen, Lauren, and Andrea Frazzini, 2008, Economic links and predictable returns, *The Journal of Finance* 63, 1977–2011.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy, 2008, The small world of investing: Board connections and mutual fund returns, *Journal of Political Economy* 116, 951–979.

Cohen, Lauren, Umit G Gurun, and Christopher Malloy, 2017, Resident networks and corporate connections: Evidence from World War II internment camps, *The Journal of Finance* 72, 207–248.

Cohen, Lauren, and Dong Lou, 2012, Complicated firms, *Journal of Financial Economics* 104, 383–400.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under-and overreactions, *The Journal of Finance* 53, 1839–1885.

De Bondt, Werner FM, and Richard Thaler, 1985, Does the stock market overreact?, *The Journal of Finance* 40, 793–805.

DellaVigna, Stefano, and Joshua Pollet, 2009, Investor inattention and Friday earnings announcements, *Journal of Finance* 64, 709–749.

Duggan, Maeve, Shannon Greenwood, and Andrew Perrin, 2016, Social media update 2016. Pew Research Center.

Engelberg, Joseph, Arzu Ozoguz, and Sean Wang, 2018, Know thy neighbor: Industry clusters, information spillovers, and market efficiency, *Journal of Financial and Quantitative Analysis* 1937–1961.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Feng, Lei, and Mark S. Seasholes, 2004, Correlated trading and location, *The Journal of Finance* 59, 2117–2144.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *The Review of Financial Studies* 33, 2326–2377.

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

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *The Review of Financial Studies* 33, 2223–2273.

Hahn, Joyce K, Henry R Hyatt, Hubert P Janicki, and Stephen R Tibbets, 2017, Job-to-job flows and earnings growth, *American Economic Review* 107, 358–63.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.

Heimer, Rawley Z, 2016, Peer pressure: Social interaction and the disposition effect, *The Review of Financial Studies* 29, 3177–3209.

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.

Hoberg, Gerard, and Gordon M Phillips, 2018, Text-based industry momentum, *Journal of Financial and Quantitative Analysis* 53, 2355–2388.

Hong, Harrison, Jeffrey D Kubik, and Jeremy C Stein, 2004, Social interaction and stock-market participation, *The journal of finance* 59, 137–163.

Hong, Harrison, Jeffrey D Kubik, and Jeremy C Stein, 2005, Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, *The Journal of Finance* 60, 2801–2824.

Hong, Harrison, and Jeremy C Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *The Journal of Finance* 54, 2143–2184.

Hong, Harrison, and Jiangmin Xu, 2019, Inferring latent social networks from stock holdings, *Journal of Financial Economics* 131, 323–344.

Hvide, Hans K., and Östberg, 2015, Social interaction at work, *Journal of Financial Economics* 117, 628–652.

Hyatt, Henry R, Erika McEntarfer, Kevin L McKinney, Stephen Tibbets, and Doug Walton, 2014, Job-to-job (j2j) flows: New labor market statistics from linked employer-employee data, *US Census Bureau Center for Economic Studies Paper No. CES-WP-14-34* .

Ivković, Zoran, and Scott Weisbenner, 2007, Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices, *The Review of Financial Studies* 20, 1327–1357.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* 48, 65–91.

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Kothari, S.P., Jonathan Lewellen, and Jerold B. Warner, 2006, Stock returns, aggregate earnings surprises, and behavioral finance, *Journal of Financial Economics* 79, 537–568.

Kuchler, Theresa, Yan Li, Lin Peng, Johannes Stroebel, and Dexin Zhou, 2021, Social proximity to capital: Implications for investors and firms, *The Review of Financial Studies* .

Kuchler, Theresa, Dominic Russel, and Johannes Stroebel, 2020, The geographic spread of Covid-19 correlates with structure of social networks as measured by Facebook, Technical report, National Bureau of Economic Research.

Lee, Charles MC, Stephen Teng Sun, Rongfei Wang, and Ran Zhang, 2019, Technological links and predictable returns, *Journal of Financial Economics* 132, 76–96.

Loughran, Tim, and Jay W. Wellman, 2011, New evidence on the relation between the enterprise multiple and average stock returns, *Journal of Financial and Quantitative Finance* 46, 1629–1650.

McNerney, James, Charles Savoie, Francesco Caravelli, Vasco M. Carvalho, and J. Doyne Farmer, 2022, How production networks amplify economic growth, *Proceedings of the National Academy of Sciences* 119, e2106031118.

Mendenhall, Richard R, 1991, Evidence on the possible underweighting of earnings-related information, *Journal of Accounting Research* 29, 170–179.

Moskowitz, Tobias J, and Mark Grinblatt, 1999, Do industries explain momentum?, *The Journal of Finance* 54, 1249–1290.

Newey, Whitney K, and Kenneth D West, 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631–653.

Ouimet, Page, and Geoffrey Tate, 2019, Learning from coworkers: Peer effects on individual investment decisions, *The Journal of Finance* 75, 133–172.

Parsons, Christopher A, Riccardo Sabbatucci, and Sheridan Titman, 2020, Geographic lead-lag effects, *The Review of Financial Studies* 33, 4721–4770.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, 2017, Do funds make more when they trade more?, *The Journal of Finance* 72, 1483–1528.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.

Pirinsky, Christo, and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, *The Journal of Finance* 61, 1991–2015.

Pool, Veronika K., Noah Stoffman, and Scott E. Yonker, 2012, No Place Like Home: Familiarity in Mutual Fund Manager Portfolio Choice, *The Review of Financial Studies* 25, 2563–2599.

Pool, Veronika K, Noah Stoffman, and Scott E Yonker, 2015, The people in your neighborhood: Social interactions and mutual fund portfolios, *The Journal of Finance* 70, 2679–2732.

Rehbein, Oliver, Simon Rother, et al., 2020, Distance in bank lending: The role of social networks, Technical report, University of Bonn and University of Mannheim, Germany.

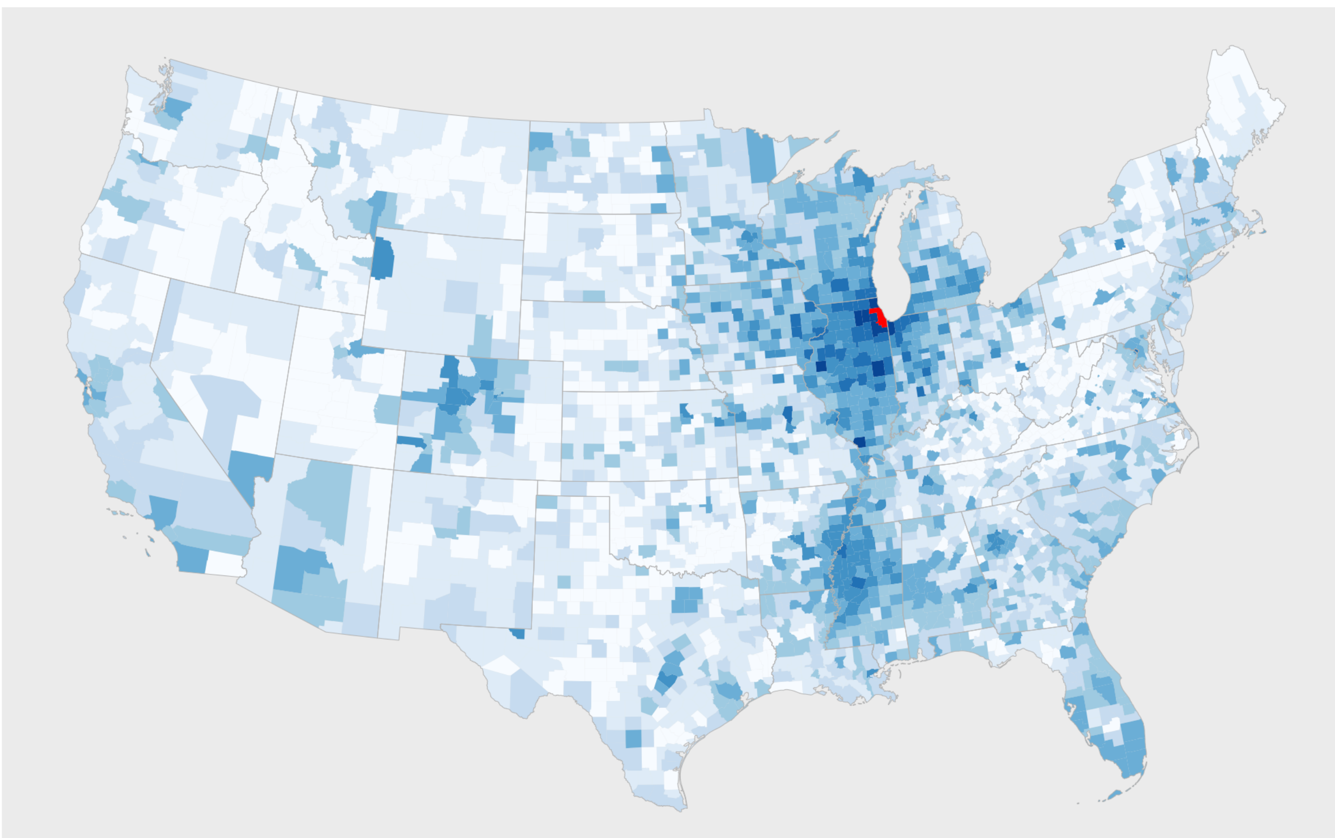
Shiller, Robert J., 2010, *Irrational Exuberance* (Princeton University Press, Princeton, N.J.).

Shumway, Tyler, 1997, The delisting bias in crsp data, *Journal of Finance* 52, 327–340.

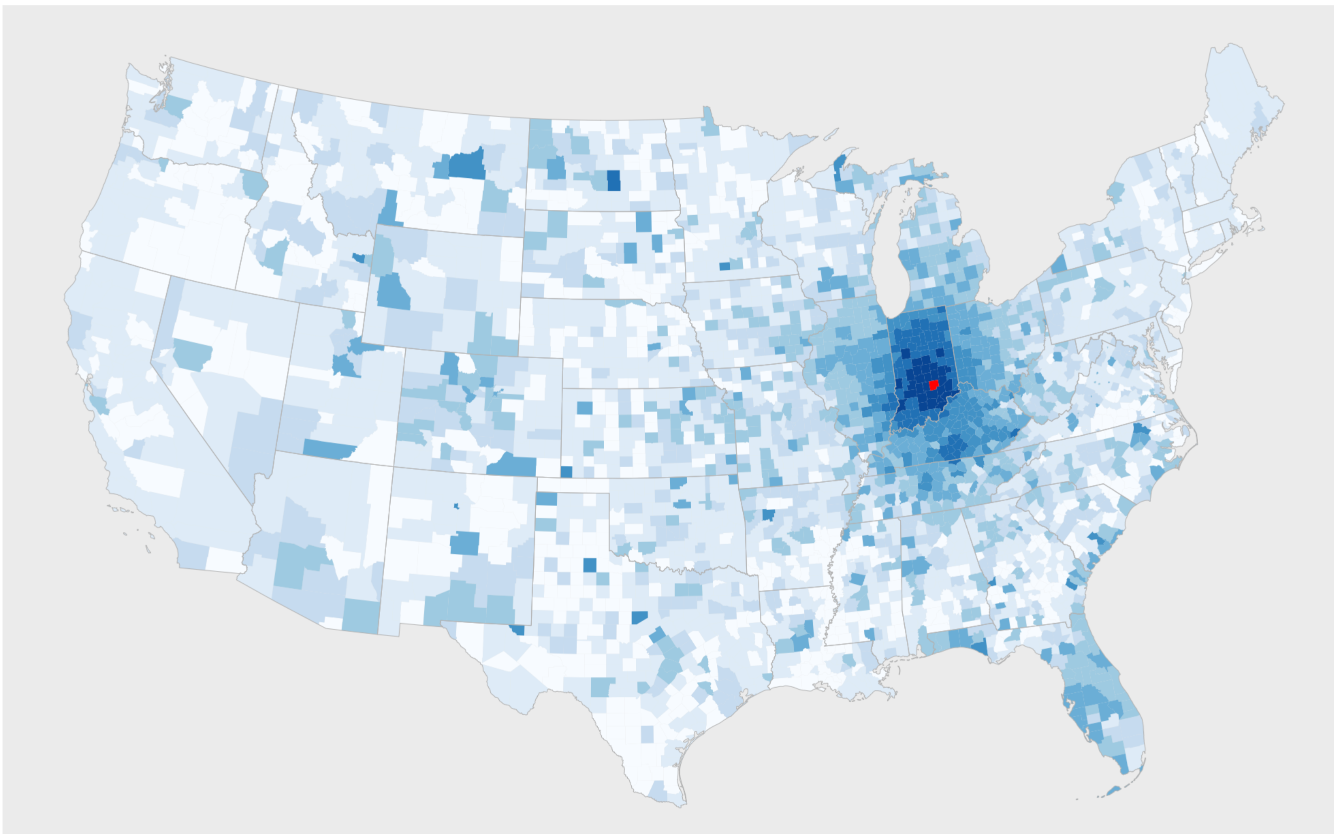
Figure 1: Examples of Social Connectedness

This figure shows county-level heat maps of the social connectedness to Cook County, IL, in Panel A, and Bartholomew County, IN, in Panel B. The focal counties are in red and darker colors indicate higher social connectedness to the focal counties. Panel C The figure presents CMI's headquarters county, Bartholomew County, IN, and the locations of CMI's industry peers, with dark blue indicating higher social connectedness to Bartholomew County. Examples of the industry peer firms include Caterpillar (ticker: CAT), Clarcor (ticker: CLC), and Stanley Black & Decker (ticker: SWK), located in Peoria, IL, Williamson, TN, and Hartford, CT, respectively. Counties without any peer firms' presence are presented in dark grey.

Panel A: Social Connectedness to Cook County, IL



Panel B: Social Connectedness to Bartholomew County, IN



Panel C: Social Connectedness to Bartholomew County, IN, and the Presence of "Machinery" Firms

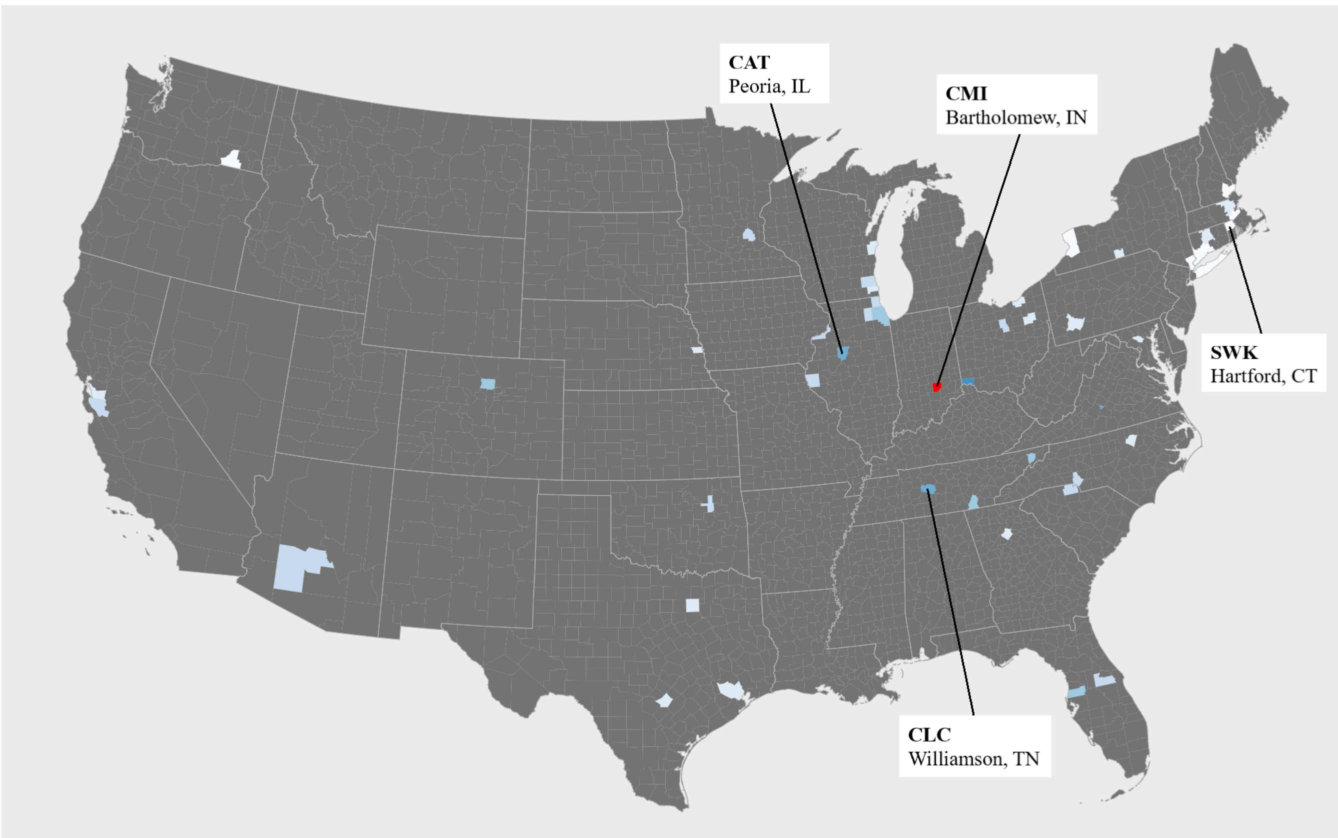


Figure 2: Average Monthly Alpha of Long Short Portfolio Over Time

The figure depicts the time series of monthly average Fama-French five-factor (FF5) alphas for long-short (LS) portfolios based on *SPFRET* (Panel A) and $SPFRET_{\perp}$ (Panel B), computed for each year in our sample. *SPFRET* is the SCI-weighted average return of stocks from the same Fama-French 48-industry as the focal stock but from a different state. $SPFRET_{\perp}$ is the abnormal return generated from a panel regression of *SPFRET* on the equal-weighted industry return, using all available stock-month observations prior to the portfolio formation month (i.e., month t) and month fixed effects. At the end of each month from July 1963 through November 2019, all common stocks are sorted into deciles based on either *SPFRET* (the left panel) or $SPFRET_{\perp}$ (the right panel). Then, next month's value-weighted return of the decile-1 portfolio is subtracted from that of the decile-10 portfolio to calculate the return of the LS portfolio for the following month. Monthly LS abnormal returns are then calculated using the FF5 model. Each bar in the figure shows the average monthly abnormal return for the year that it represents. All returns are in percentages.

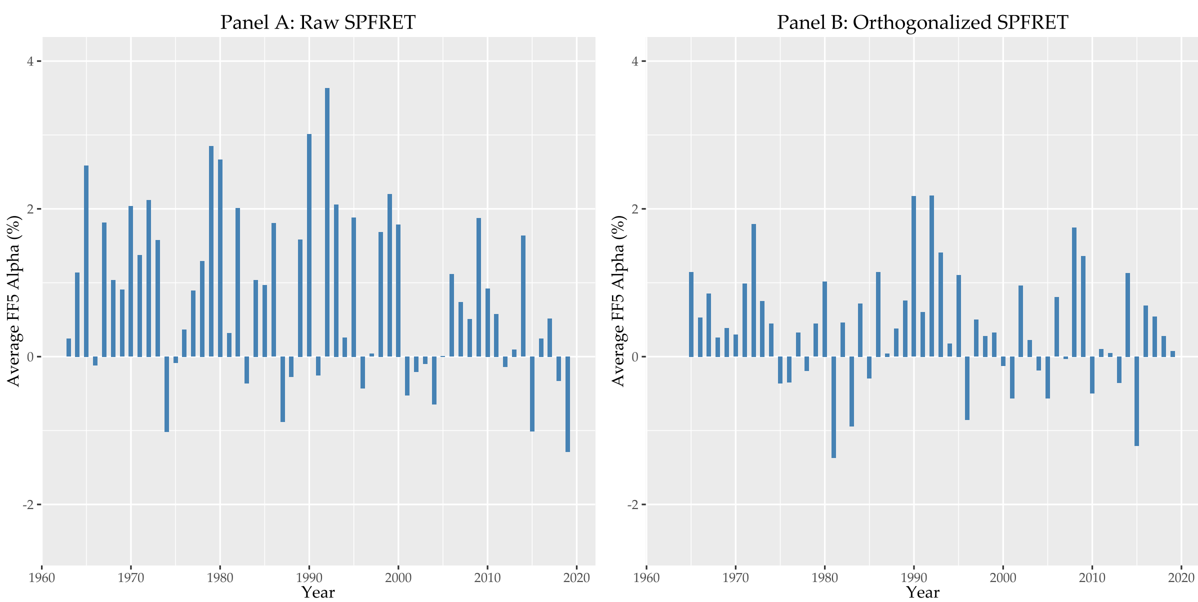


Figure 3: Decomposing *SPFRET* Portfolio Returns

The figure decomposes the average return of the *SPFRET*-sorted value-weighted LS portfolio into its various components. Each bar shows the amount of the average return that is explained by the variable(s) indicated below it. *INDRET* represents the contribution of *INDRET*, the equal-weighted Fama-French 48-industry return. *GEORET* represents the contribution of geographic momentum. *CONGRET* represents the contribution from the complicated firm effect. *CRET* represents the contribution from the customer return. *J2JRET* represents the contribution from the return predictability based on job-to-job movement across states. *TNICRET* represents the contribution of the text-based industry return of Hoberg and Phillips (2018). *TECHRET* represents the contribution of the technological spillover effect. *CFRET* represents the contribution of the analyst-linked firm return. *FF5* represents the combined contribution of the Fama-French five factors. The last bar shows the alpha, the portion of the average return not explained by variables shown in the plot. Red indicates negative values.

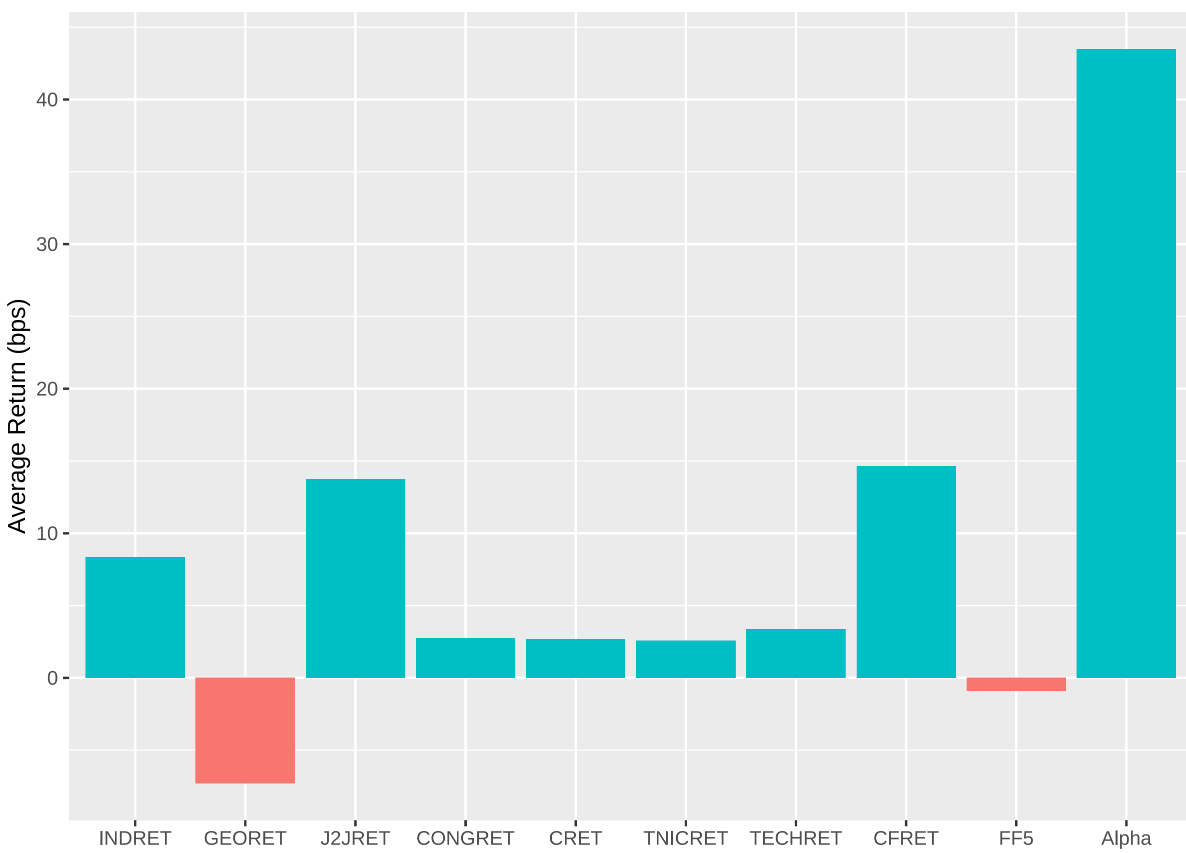


Figure 4: Relative Variable Importance

The figure depicts the relative importance of lead-lag predictors (*SPFRET*, *INDRET*, *GEORET*, *CFRET*, *CRET*, *TECHRET*, *TNICRET*, *J2JRET*, and *CONGRET*) in a partial least squares (PLS) model that uses these predictors along with all the other controls (i.e., *INDMOM*, *RET*, *MOM*, *SIZE*, *BM*, *BMKT*, *ILLIQ*, *SKEW*, *COSKEW* and *MAX*). Starting from July 1994, a new PLS model is trained each month using all the data available up to that point. For each training set, the difference between the full-model R^2 and the R^2 obtained by dropping one of the nine predictors while keeping the coefficient estimates for the rest of the predictors fixed. These marginal changes in R^2 are then normalized so that their sum equals one. The relative variable importance in each panel is then calculated by taking the average over the normalized values over all the training sets. *CFRET* refers to shared analyst coverage returns. *SPFRET* refers to social peer firm returns. *INDRET* is the industry return. *CRET* is the customer return. *GEORET* is the geographic momentum. *TECHRET* is the technology-linked firm return. *TNICRET* is the text-based industry return. *J2JRET* is the job flow–based return. *CONGRET* is the pseudo-conglomerate return.

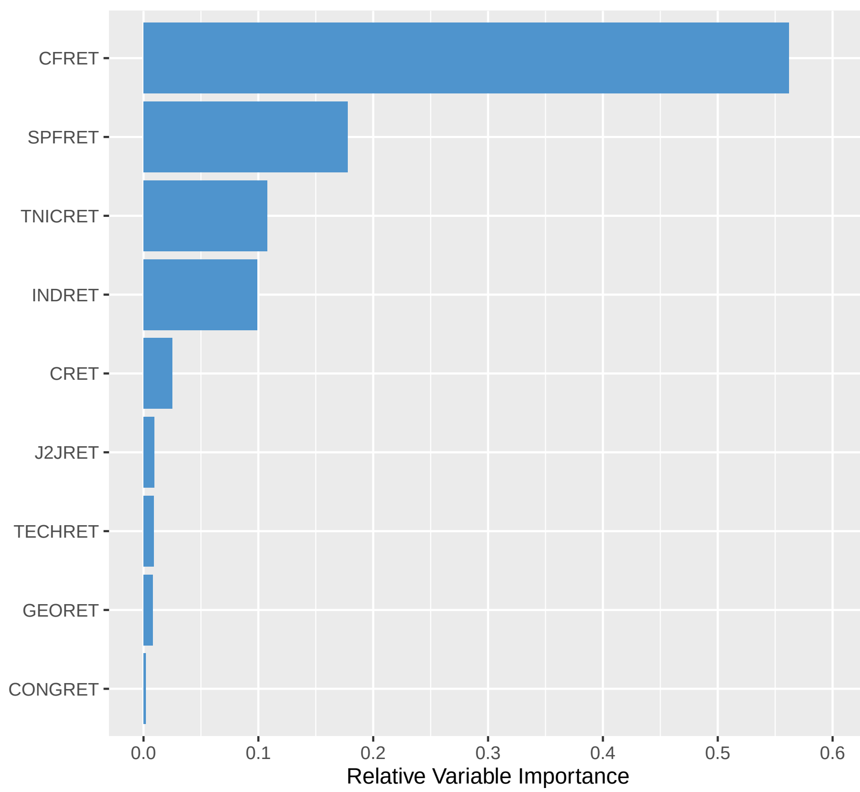
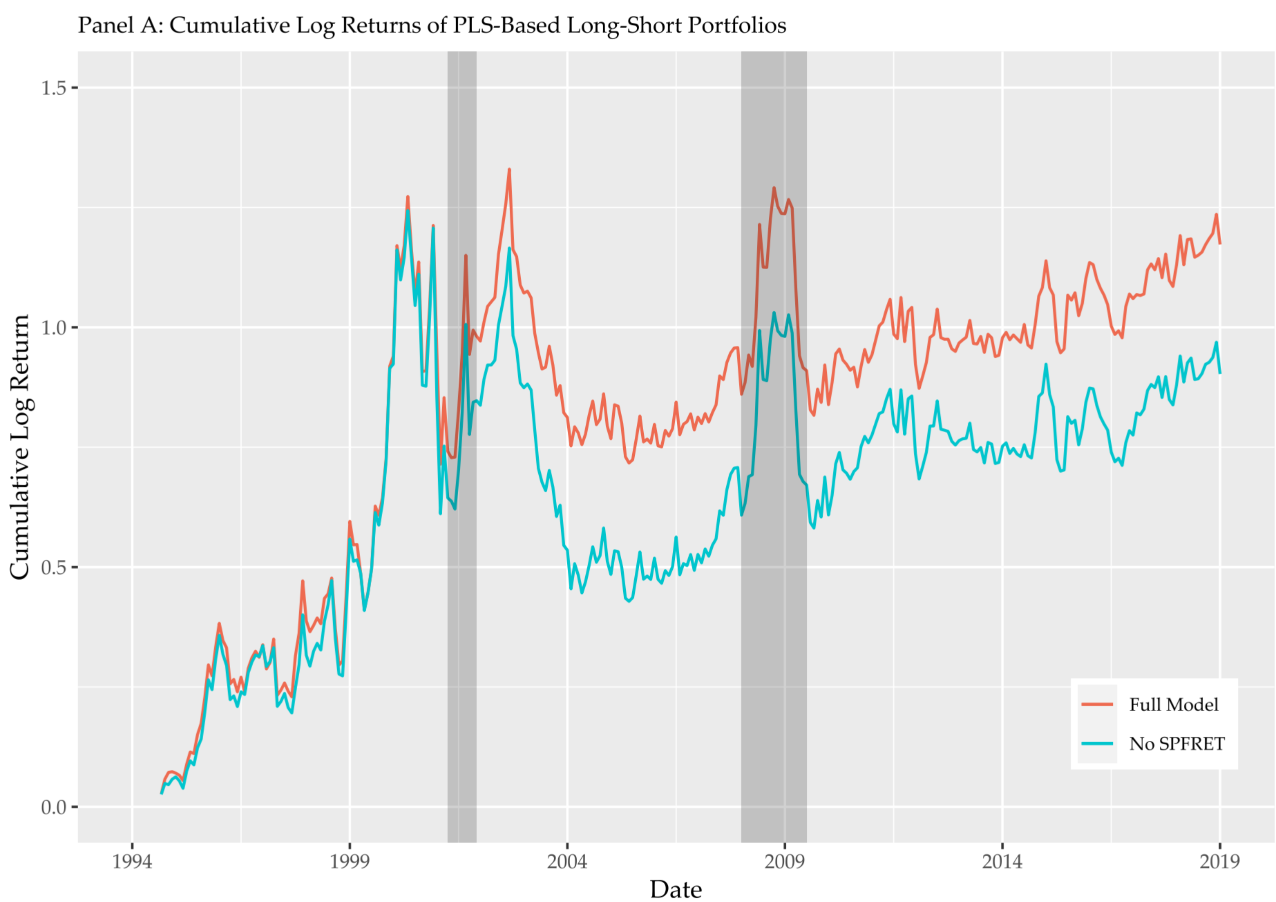


Figure 5: PLS-Based Long-Short Portfolio Analysis

The figure shows the cumulative returns of portfolios that are sorted based on the out-of-sample PLS predictor with and without *SPFRET* from August 1994 through December 2019. Panel A depicts the cumulative log return of the long-short (i.e., decile-10 minus decile-1) portfolios, while Panel B shows the cumulative log returns for long and short legs separately. Portfolios are value-weighted. Shaded areas indicate the NBER recession periods.



Panel B: Cumulative Log Returns of PLS-Based Long and Short Portfolios

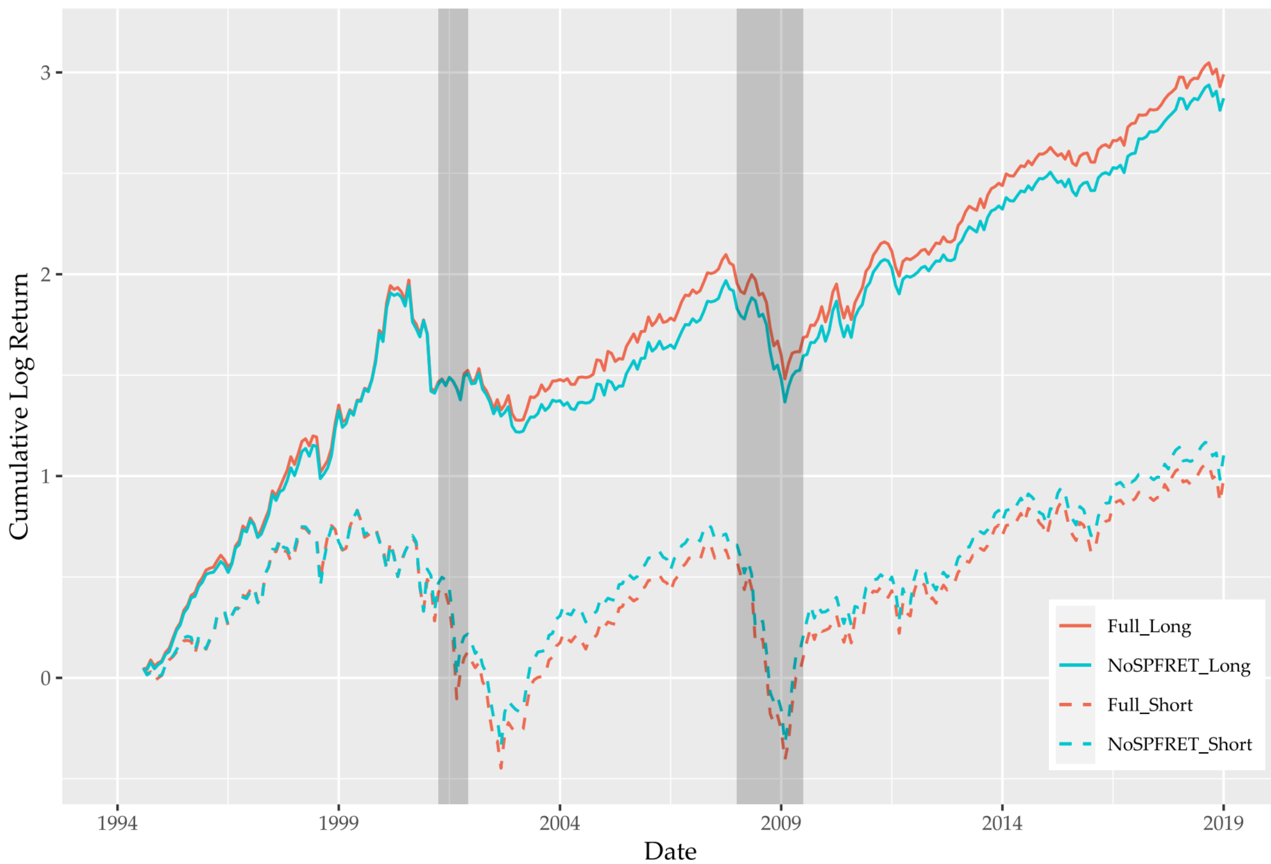


Table 1: Summary Statistics and Correlations

The table reports the summary statistics (Panel A) and time-series averages of cross-sectional correlations (Panel B) for the main variables used in the paper. *i*s the contemporaneous monthly stock return. *RET* is the contemporaneous return. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). $SPFRET_{\perp}$ is *SPFRET* orthogonalized with respect to the equal-weighted industry returns. *SPFMOM* is the cumulative *SPFRET* over months $t-11$ through $t-1$. *INDRET* is the equal-weighted average return of stocks for a given industry. *INDMOM* is the cumulative *INDRET* over months t-11 through t-1. *GEORET* is the equal-weighted average return of peer firms that are in the same economic area as a stock (excluding those in the same industry). *GEOMOM* is the cumulative *GEORET* over months t-11 through t-1. *CFRET* is the average return of peer stocks that share at least one analyst with the focal stock over previous 12 months, weighted by the number of shared analysts between stocks. *CFMOM* is the cumulative *CFRET* over months t - 11 through t - 1. *CRET* is the equal-weighted average stock return of a firm's main customers. *TECHRET* is the weighted average stock return of technology-linked peer firms, where the weights are the technological closeness between the peer firm and the focal firm, determined by the similarities between patent distributions across different technology categories. *CONGRET* is the pseudo-conglomerate return, defined as the sales-weighted return of (value-weighted) single-segment firm portfolios, formed for each segment that a conglomerate firm operates in. *J2JRET* represents the contribution from the return predictability based on job-to-job movement across states. *TNICRET* is the text-based industry return of Hoberg and Phillips (2018). All returns are reported in percentages. Before calculating the correlations, all variables except *RET* are cross-sectionally standardized by mapping them into the $(0, 1]$ interval through scaling the monthly ranks by the number of observations (Gu et al. (2020)).

Panel A: Summary statistics

	N	Mean	Std. Dev.	Min.	25%	75%	Max.
RET	2608	1.741	11.713	-46.511	-4.386	6.569	157.254
SPFRET	2525	1.722	3.837	-14.521	-0.643	3.837	28.001
SPFMOM	2525	19.878	18.671	-32.533	7.887	29.027	164.051
INDRET	2543	1.187	3.215	-8.614	-0.884	3.145	13.377
INDMOM	2543	15.027	15.711	-24.081	4.651	24.479	71.097
GEORET	1612	1.243	1.929	-4.266	0.136	2.270	7.572
GEOMOM	1507	15.059	8.616	-7.843	10.062	19.375	44.855
J2JRET	2360	1.709	2.814	-11.133	0.344	2.951	20.447
J2JMOM	2212	20.882	13.303	-25.329	13.818	26.230	109.765
CONGRET	731	1.255	3.942	-12.948	-1.055	3.459	18.951
CONGMOM	660	14.662	15.743	-30.309	5.112	22.953	90.290
CRET	882	1.221	6.995	-26.711	-2.616	4.757	45.661
CMOM	809	13.878	26.306	-55.013	-1.227	25.900	223.881
TNICRET	2885	1.767	6.163	-33.748	-1.310	4.534	74.585
TNICMOM	2570	22.217	27.174	-60.958	6.172	34.897	330.843
TECHRET	805	1.631	2.401	-9.251	0.339	2.811	17.265
TECHMOM	752	19.788	10.993	-17.764	13.223	25.181	93.483
CFRET	2682	1.481	3.995	-18.192	-0.820	3.647	30.652
CFMOM	2682	17.893	18.928	-39.959	6.766	26.523	228.511

Panel B: Correlation Matrix

	RET	SPFRET	INDRET	GEORET	J2JRET	CONGRET	CRET	TNICRET	TECHRET
SPFRET	0.166								
INDRET	0.143	0.703							
GEORET	0.009	-0.013	-0.029						
J2JRET	0.124	0.513	0.454	0.002					
CONGRET	0.145	0.359	0.351	0.021	0.251				
CRET	0.103	0.141	0.137	0.054	0.082	0.150			
TNICRET	0.167	0.367	0.370	0.062	0.305	0.254	0.179		
TECHRET	0.122	0.250	0.268	0.049	0.062	0.205	0.134	0.278	
CFRET	0.266	0.494	0.482	0.061	0.377	0.412	0.221	0.441	0.370

Table 2: Comovements in Firm Fundamentals and Stock Returns

The table presents the comovements between the focal firm and its industry peers using panel regression analysis. In columns 1-4, the dependent variables are firms' fundamentals, measured annually: ΔEPS is the change in EPS scaled by lagged stock price, $\Delta Sales$ is the percentage growth in sales, $\Delta Employees$ is the percentage growth in the number of employees, and $NewCapital$ is the sum of net equity issuance plus net debt issuance scaled by lagged enterprise value. Only the firms with the same fiscal year-end as the focal firm are included. In column 5, the dependent variable is monthly stock returns. In Panels A and B, the key independent variable is the corresponding fundamentals of the contemporaneous SCI-weighted industry portfolio (excluding firms from the same state) and the contemporaneous equal-weighted industry portfolio (excluding the focal firm), respectively. Panel C includes the corresponding measures of both the SCI- and equally-weighted industry portfolios. We follow Gu et al. (2020) and cross-sectionally rank the independent variables and scale them into the [0,1] intervals. All regressions include time fixed effects. All dependent variables are scaled up by 100. Standard errors are clustered by both time and firm and the corresponding t-statistics are reported in parentheses. For all panels, *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: SCI-Weighted Industry Portfolio

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
SCI-Weighted	2.445***	30.848***	13.374***	35.305***	7.912***
	(6.024)	(9.735)	(10.031)	(11.522)	(27.047)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	115,035	124,232	121,090	121,124	1,711,696
R^2	0.018	0.051	0.047	0.041	0.147

Panel B: Equal-Weighted Industry Portfolio

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
Equal-Weighted	1.470***	28.165***	11.076***	28.904***	6.822***
	(3.455)	(9.826)	(8.385)	(10.892)	(25.372)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	115,732	124,983	121,845	121,848	1,711,696
R^2	0.017	0.048	0.043	0.034	0.141

Panel C: Multivariate Analysis

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
SCI-Weighted	3.463***	26.496***	13.526***	36.495***	7.718***
	(7.260)	(7.096)	(10.528)	(8.409)	(23.620)
Equal-Weighted	−1.283**	5.108**	−0.183	−1.371	0.226
	(−2.517)	(2.387)	(−0.179)	(−0.469)	(1.284)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	115,035	124,232	121,090	121,124	1,711,696
R^2	0.018	0.052	0.047	0.041	0.147

Table 3: Return Predictability of Social Peer Firm Returns: Portfolio Analysis

The table reports the results of the univariate portfolio sort based on *SPFRET*, the SCI-weighted average return of a firm's industry peers (excluding those from the same state). For each month, we sort all common stocks into deciles based on *SPFRET*and calculate both the equal-weighted and value-weighted one-month-ahead returns for the decile portfolios, as well as the return of the portfolio that long the decile-10 portfolio and short the decile-1 portfolio. We present the raw returns and the FF5 alphas for the equal- and value-weighted portfolios respectively. All returns are in percentages. The corresponding *t*-statistics based on Newey and West (1994) standard errors are reported in parentheses below the alphas. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1 Low)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10 High)	(10-1)
Raw Return EW	0.277	0.534**	0.799***	0.970***	1.091***	1.219***	1.277***	1.438***	1.630***	1.766***	1.489***
	(1.146)	(2.473)	(3.967)	(4.436)	(4.933)	(5.838)	(5.898)	(6.832)	(7.131)	(7.233)	(8.569)
FF5 Alpha EW	−0.846***	−0.525***	−0.234**	−0.122	−0.036	0.083	0.142**	0.358***	0.610***	0.723***	1.568***
	(−7.043)	(−4.233)	(−2.050)	(−1.403)	(−0.596)	(1.481)	(2.304)	(5.169)	(6.341)	(6.135)	(7.058)
Raw Return VW	0.443**	0.655***	0.935***	0.834***	1.103***	1.032***	1.058***	1.168***	1.098***	1.208***	0.765***
	(2.122)	(3.360)	(5.049)	(4.410)	(5.771)	(5.533)	(5.650)	(7.160)	(5.275)	(6.188)	(4.781)
FF5 Alpha VW	−0.554***	−0.251**	0.007	−0.065	0.146	0.007	0.057	0.199***	0.255**	0.286***	0.840***
	(−5.516)	(−2.546)	(0.067)	(−0.711)	(1.415)	(0.082)	(0.705)	(2.706)	(1.985)	(3.177)	(5.216)

Table 4: Return Predictability of Social Peer Firm Returns: Portfolio Analysis (Bivariate Sort)

The table reports the results of the bivariate portfolio sorts based on industry momentum (*INDRET*) and *SPFRET*. We first sort stocks into quintiles based on *INDRET* and then, within each *INDRET* quintile, we further sort stocks into quintiles based on *SPFRET*. We then report the one-month-ahead FF5 alphas for the 25 equal-weighted (Panel A) and value-weighted (Panel B) portfolios and the portfolios that long in stocks in the *SPFRET* quintile 5 and short in the *SPFRET* quintile 1 stocks. Final rows in both panels report the average alphas for the *SPFRET* quintiles. All returns are in percentages. The corresponding *t*-statistics based on Newey and West (1994) standard errors are reported in parentheses below the alphas. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Bivariate Sort with Equal-Weighted Portfolios

	SPFRET					
	(1 Low)	(2)	(3)	(4)	(5 High)	(5-1)
1 (Low INDRET)	-1.113^{***}	-0.874^{***}	-0.674^{***}	-0.564^{***}	-0.537^{***}	0.576^{***}
	(-7.309)	(-6.326)	(-4.512)	(-4.099)	(-3.773)	(4.184)
2	-0.572^{***}	-0.203^{*}	-0.087	0.015	0.088	0.660^{***}
	(-4.813)	(-1.695)	(-0.659)	(0.127)	(0.908)	(5.043)
3	-0.250^{***}	0.008	0.051	0.078	0.218^{**}	0.468^{***}
	(-2.684)	(0.089)	(0.636)	(0.886)	(2.345)	(4.517)
4	-0.010	0.152^{*}	0.234^{**}	0.379^{***}	0.472^{***}	0.482^{***}
	(-0.126)	(1.681)	(2.446)	(4.482)	(4.249)	(3.318)
5 (High INDRET)	0.373^{***}	0.631^{***}	0.816^{***}	0.767^{***}	1.002^{***}	0.630^{***}
	(3.446)	(4.791)	(6.358)	(5.520)	(6.219)	(4.599)
Average	-0.314^{***}	-0.057	0.068	0.135^{***}	0.249^{***}	0.563^{***}
	(-5.334)	(-1.329)	(1.542)	(3.282)	(5.278)	(7.384)

Panel B: Bivariate Sort with Value-Weighted Portfolios

	SPFRET					
	(1 Low)	(2)	(3)	(4)	(5 High)	(5-1)
1 (Low INDRET)	-0.725^{***}	-0.526^{***}	-0.230	-0.197	-0.240	0.485^{***}
	(-5.289)	(-3.609)	(-1.472)	(-1.507)	(-1.457)	(2.794)
2	-0.618^{***}	-0.102	0.160	0.187	0.146	0.764^{***}
	(-5.214)	(-0.795)	(1.239)	(1.438)	(1.218)	(4.875)
3	-0.057	0.007	-0.127	0.083	0.111	0.168
	(-0.529)	(0.059)	(-1.178)	(0.703)	(0.938)	(1.212)
4	0.004	0.046	0.069	0.245^{**}	0.186^{*}	0.182
	(0.042)	(0.378)	(0.608)	(2.350)	(1.648)	(1.108)
5 (High INDRET)	0.108	0.266^{*}	0.349^{***}	0.358^{**}	0.440^{***}	0.332^{**}
	(0.875)	(1.952)	(2.703)	(2.492)	(3.487)	(2.082)
Average	-0.258^{***}	-0.062	0.044	0.135^{***}	0.129^{**}	0.386^{***}
	(-4.931)	(-1.196)	(0.902)	(2.773)	(2.185)	(4.844)

Table 5: Information Environment and Return Predictability

The table examines how certain firm characteristics affect the return predictability of social peer firm returns by using bivariate portfolio sorting. In Panel A, each month, we sort firms into deciles based on size, institutional ownership, and analyst coverage. We then further sort the firms' with the lowest and highest characteristics into deciles based on *SPFRET*, calculate the value-weighted returns over the next month, and report the FF5 alphas for the decile 1, decile 10, and 10-1 long-short portfolios. In Panel B, for a given Fama-French 48 industry and month, we measure the total industry market capitalization at the county level and define the top 20% (columns 1 to 3) or top 10% (columns 4 and 6) of counties as the industry cluster. We then sort firms into two groups based on whether they are headquartered in their respective industry clusters or not. Within each group, we further apply the same secondary (*SPFRET*-based) sorting that we used in Panel A and report the value-weighted FF5 alphas for decile 1, decile 10, and 10-1 long-short portfolios. All returns are reported in percentages. We report t-statistics based on Newey-West adjusted standard errors in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Size, Institutional Ownership, and Analyst Coverage

	Size			Inst. Own.			Analyst Cov.		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Low	-0.930^{***}	0.882^{***}	1.811^{***}	-1.037^{***}	0.360	1.397^{***}	-0.635^{***}	0.582^{**}	1.218^{***}
	(−6.035)	(5.776)	(6.815)	(−3.593)	(1.259)	(3.727)	(−3.448)	(2.382)	(3.827)
High	-0.340^{***}	0.231^{*}	0.571^{***}	-0.674^{***}	0.086	0.760^{***}	-0.418^{***}	0.252	0.670^{***}
	(−3.404)	(1.878)	(3.773)	(−2.922)	(0.463)	(2.649)	(−2.851)	(1.571)	(2.836)

Panel B: Out-Cluster and In-Cluster

	Top 20%			Top 10%		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Out-Cluster	-0.732^{***}	0.531^{***}	1.263^{***}	-0.734^{***}	0.417^{***}	1.150^{***}
	(-5.862)	(4.516)	(6.150)	(-6.684)	(3.446)	(5.806)
In-Cluster	-0.437^{***}	0.209^{**}	0.646^{***}	-0.405^{***}	0.263^{**}	0.669^{***}
	(-4.119)	(2.130)	(4.011)	(-3.436)	(2.164)	(3.599)

Table 6: Return Predictability in the Long Run

The table reports the long-run performance of calendar-time portfolios sorted by social peer firms' returns. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). *SPFMOM* is the cumulative *SPFRET* over months $t-11$ through $t-1$. Panel A (B) reports the FF5 alpha of the equal-weighted (value-weighted) portfolio return for deciles 1 and 10, and the 10-1 return differences for the following months relative to month t: 1–3, 4–6, 7–12, and 13–24. t-statistics based on Newey-West standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: FF5 Alphas (Equal-Weighted)

	SPFRET			SPFMOM		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Month 1	-0.846^{***}	0.722^{***}	1.568^{***}	-0.340^{**}	0.354^{***}	0.694^{***}
	(-7.043)	(6.123)	(7.052)	(-2.322)	(3.048)	(2.907)
Months 2–3	-0.170	0.066	0.236	-0.315^{**}	0.297^{***}	0.611^{***}
	(-1.367)	(0.725)	(1.238)	(-2.317)	(2.766)	(2.866)
Months 4–6	-0.174	0.089	0.262^{*}	-0.331^{***}	0.220^{**}	0.550^{***}
	(-1.525)	(1.250)	(1.674)	(-2.645)	(2.054)	(2.760)
Months 7–12	-0.150	0.128^{**}	0.279^{***}	-0.202^{**}	0.098	0.300^{*}
	(-1.476)	(2.268)	(2.597)	(-2.091)	(0.877)	(1.921)
Months 13–24	-0.029	0.082	0.110	-0.078	0.039	0.117
	(-0.497)	(0.976)	(1.636)	(-0.972)	(0.319)	(0.743)

Panel B: FF5 Alphas (Value-Weighted)

	SPFRET			SPFMOM		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Month 1	-0.554^{***}	0.287^{***}	0.840^{***}	-0.098	0.201^{*}	0.299
	(-5.526)	(3.162)	(5.203)	(-0.671)	(1.778)	(1.353)
Months 2–3	0.057	-0.012	-0.069	-0.163	0.243^{**}	0.405^{**}
	(0.486)	(-0.116)	(-0.351)	(-1.244)	(2.484)	(2.069)
Months 4–6	-0.052	-0.007	0.045	-0.230^{**}	0.151	0.382^{**}
	(-0.558)	(-0.091)	(0.299)	(-2.289)	(1.589)	(2.272)
Months 7–12	-0.140^{**}	0.072	0.213^{**}	-0.216^{***}	0.009	0.225^{*}
	(-1.967)	(1.355)	(2.161)	(-2.702)	(0.120)	(1.810)
Months 13–24	-0.024	0.052	0.076	-0.038	0.014	0.052
	(-0.664)	(0.924)	(1.280)	(-0.577)	(0.156)	(0.426)

Table 7: Predicting Future Earnings Surprises, Analyst Forecast Errors, and Announcement Returns

The table presents the panel regression results of using social peer firm returns to predict a firm's future earnings surprises, analyst forecast errors, and cumulative abnormal returns around earnings announcements. The dependent variables are the focal firm's standardized unexpected earnings (SUE) (columns 1 to 4), analyst forecast errors (FE) (columns 5 to 8), and the three-day abnormal returns around the earnings announcements (columns 9 to 12) for the next four quarters. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). *SPFMOM* is the cumulative *SPFRET* over months $t-11$ through $t-1$. Panel B includes short- and long-term industry momentum (*INDRET* and *INDMOM*). All dependent variables are scaled up by 100. Missing values of independent variables are imputed with the monthly medians. We cross-sectionally standardize all independent variables by mapping them into the $[0,1]$ interval through scaling the monthly ranks by the number of observations (Gu et al. (2020)). All regressions include time and firm fixed effects. t-statistics are computed with standard errors clustered by month and stock and are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Predictability of Social Peer Firm Returns

	Earnings Growth				Forecast Error				Earnings Return			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
SPFMOM	34.067***	25.649***	18.217***	11.812***	14.655***	14.286***	14.327***	11.194***	0.141***	0.096*	−0.003	−0.018
	(21.318)	(16.372)	(11.822)	(7.753)	(7.151)	(6.212)	(6.076)	(5.109)	(2.819)	(1.850)	(−0.053)	(−0.338)
SPFRET	10.925***	12.916***	11.340***	9.178***	4.830***	3.588**	5.874***	5.741***	0.140***	0.013	0.046	−0.0004
	(9.823)	(11.968)	(9.986)	(8.031)	(3.412)	(2.296)	(3.486)	(3.504)	(3.315)	(0.325)	(1.077)	(−0.009)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,457,322	1,462,274	1,467,432	1,472,775	817,742	737,599	673,727	621,535	1,490,579	1,463,703	1,438,009	1,412,416
R^2	0.174	0.175	0.175	0.174	0.132	0.151	0.161	0.169	0.052	0.056	0.059	0.058

Panel B: Controlling for Industry Momentum

	Earnings Growth				Forecast Error				Earnings Return			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
SPFMOM	16.973***	12.622***	9.360***	7.949***	6.214***	7.624***	9.788***	6.425***	0.138**	0.155**	0.119**	0.103*
	(9.837)	(7.431)	(5.357)	(4.539)	(2.898)	(3.118)	(4.040)	(2.639)	(2.359)	(2.538)	(2.033)	(1.671)
SPFRET	2.991***	4.754***	4.548***	4.121***	1.742	−0.041	3.696**	4.079**	0.127***	0.027	0.065*	−0.003
	(3.252)	(5.178)	(4.670)	(4.511)	(1.294)	(−0.030)	(2.272)	(2.369)	(3.334)	(0.735)	(1.664)	(−0.082)
INDMOM	22.306***	16.767***	11.266***	4.678**	10.834***	8.425***	5.765**	6.116**	0.003	−0.077	−0.162**	−0.164**
	(12.367)	(9.192)	(6.024)	(2.464)	(4.989)	(3.544)	(2.408)	(2.211)	(0.041)	(−1.224)	(−2.441)	(−2.501)
INDRET	11.175***	11.647***	9.762***	7.353***	4.145***	4.932***	2.945	2.217	0.019	−0.018	−0.024	0.008
	(9.674)	(9.711)	(7.661)	(6.260)	(2.760)	(2.850)	(1.569)	(1.108)	(0.436)	(−0.410)	(−0.526)	(0.162)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,457,322	1,462,274	1,467,432	1,472,775	817,742	737,599	673,727	621,535	1,490,579	1,463,703	1,438,009	1,412,416
R^2	0.175	0.176	0.176	0.174	0.132	0.151	0.161	0.169	0.052	0.056	0.059	0.058

Table 8: Spanning Regressions

The table presents the results of spanning regressions where the long-short (LS) portfolio return of *SPFRET* is regressed against LS portfolio returns obtained using other variables that capture economic linkages between firms. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). The other variables that are used to form LS portfolios include industry momentum (*INDRET*), geographic momentum (*GEORET*), customer return (*CRET*), the technology-linked firm return (*TECHRET*), the pseudo-conglomerate return (*CONGRET*), the text-based industry return (*TNICRET*), the technology-linked firm return (*TECHRET*), the analyst-linked firm return (*CFRET*), and the job flow–based peer returns (*J2JRET*). We control for Fama-French five-factor model but do not report those coefficients for brevity. The results are based on a common sample for which all explanatory variables are available and all returns are reported in percentages. Newey-West heteroskedasticity and autocorrelation-corrected *t*-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Alpha	0.915***	0.765***	0.749***	0.597***	0.574***	0.539***	0.513***	0.509***	0.435**
	(3.004)	(3.996)	(3.972)	(3.478)	(3.107)	(3.132)	(3.104)	(3.095)	(2.434)
INDRET		0.766***	0.758***	0.583***	0.538***	0.523***	0.462***	0.433***	0.412***
		(11.313)	(11.636)	(11.073)	(9.879)	(8.876)	(8.576)	(9.158)	(7.935)
GEORET			0.024	0.001	−0.018	−0.018	−0.057	−0.077*	−0.111***
			(0.441)	(0.031)	(−0.417)	(−0.420)	(−1.229)	(−1.890)	(−2.833)
J2JRET				0.323***	0.272***	0.275***	0.273***	0.277***	0.244***
				(8.972)	(6.791)	(7.155)	(6.500)	(6.916)	(5.021)
CONGRET					0.130***	0.115***	0.091**	0.076*	0.059*
					(4.248)	(3.319)	(2.458)	(1.876)	(1.767)
CRET						0.057	0.046	0.036	0.044
						(1.328)	(1.054)	(0.729)	(0.866)
TNICRET							0.130***	0.106***	0.042
							(3.276)	(2.796)	(1.076)
TECHRET								0.088	0.074
								(1.297)	(1.111)
CFRET									0.113**
									(1.973)
FF5	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	276	276	276	276	276	276	276	276	276
R^2	0.006	0.691	0.691	0.739	0.755	0.757	0.763	0.767	0.775

Table 9: Predicting Returns: Fama-MacBeth Regressions

The table reports the results of Fama-MacBeth regressions where stock returns are regressed on lagged social peer firm returns. *SPFRET* is the SCI-weighted average return of a firm's industry peers (excluding those from the same state). *INDRET* is the short-term industry momentum. *GEORET* is the geographic momentum, *CFRET* is the analyst-connected firm return, *CRET* is the customer return, *TECHRET* is the technology-linked firm return. *CONGRET* is the pseudo-conglomerate return. *TNICRET* is the text-based industry return. *TECHRET* is the technology-linked firm return. *J2JRET* is the industry peer returns weighted by the average job flow between focal and peer firms (see Section 6.1). We also include the following control variables: *RET*, *SIZE*, *BMKT*, *BM*, *MOM*, *IVOL*, *ILLIQ*, *MAX*, *SKEW*, and *COSKEW*. Section 2.1 provides detailed descriptions. All returns are reported in percentages. Missing values of independent variables are imputed with the monthly medians. We cross-sectionally standardize all independent variables by mapping them into the $(0, 1]$ interval through scaling the monthly ranks by the number of observations (Gu et al. (2020)). t-statistics are computed with Newey and West (1994) standard errors and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

					RET_{t+1}				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SPFRET	1.400***	0.787***	0.781***	0.679***	0.656***	0.644***	0.617***	0.708***	0.493***
	(9.914)	(7.669)	(7.652)	(7.075)	(5.515)	(5.470)	(4.468)	(4.539)	(3.321)
INDRET		1.040***	1.043***	0.990***	0.856***	0.847***	0.721***	0.878***	0.693***
		(9.057)	(9.035)	(8.828)	(6.214)	(6.179)	(4.258)	(4.723)	(3.723)
GEORET			0.205***	0.203***	0.220***	0.214***	0.205***	0.201***	0.165**
			(3.959)	(3.953)	(3.772)	(3.779)	(2.895)	(2.613)	(2.159)
J2JRET				0.297***	0.318***	0.320***	0.233***	0.318***	0.219**
				(4.640)	(4.435)	(4.421)	(2.674)	(3.276)	(2.293)
CONGRET					0.146**	0.139**	0.016	0.055	−0.035
					(2.496)	(2.384)	(0.239)	(0.690)	(−0.457)
CRET						0.364***	0.329***	0.384***	0.319***
						(4.441)	(4.209)	(4.209)	(3.583)
TNICRET							0.593***	0.791***	0.569***
							(4.368)	(5.422)	(4.627)
TECHRET								0.166	0.043
								(1.510)	(0.375)
CFRET									1.301***
									(8.185)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Periods	672	672	672	672	509	509	365	276	276
# Stocks	2,634	2,634	2,634	2,634	3,193	3,193	3,633	3,867	3,867
R^2	0.071	0.074	0.075	0.076	0.065	0.066	0.063	0.067	0.069

Appendix

Table A1: Variable Descriptions

Variable	Definition
Social Connectedness Index (SCI)	Number of Facebook friends links between firms' headquarters counties, scaled by the product of populations of the two counties
Social Peer Firm Return (SPFRET)	SCI-weighted returns based on all firms in the focal firm's industry, excluding same-state firms. $SPFRET_{ALL}$ is the SCI-weighted returns based on all the firms in the focal firm's industry except for the focal firm itself. SPFRET_{DIST} is an alternative social peer firm return measure for which social ties are measured with the inverse distance between firms headquarters locations. $SPFRET_{\perp}$ is *SPFRET* orthogonalized against industry momentum and/or other control variables, depending on the context.
Social Peer Firm Momentum (*SPFMOM*)	The compounded *SPFRET* from months $t-11$ to $t-1$.
Non-social peer Firm Return (*NSPFRET*)	The SCI-weighted average return of stocks that share neither the Fama-French 48-industry nor the state of the focal stock
Industry Return (INDRET)	The equal-weighted average return of stocks with the same Fama-French 48 industry classification as the focal stock, *INDMOM* is obtained by compounding *INDRET* from month $t-11$ to $t-1$.
Geographic Return (GEORET)	The equal-weighted average return of all stocks from the same economic area (EA) as the focal stock but from a different FF48 industry. *GEOMOM* is obtained by compounding *GEORET* from month $t-11$ to $t-1$.
Analyst Momentum (CFRET)	The weighted average return of stocks that share at least one analyst with the focal stock over the previous 12 months, where weights are the number of shared analysts between stocks. *CFMOM* is obtained by compounding *CFRET* from month $t-11$ to $t-1$.
Customer Return (CRET)	The equal-weighted average stock return of the main customers of the focal firm, where a six-month gap is required between the fiscal year-end of the supplier and stock returns. *CMOM* is obtained by compounding *CRET* from month $t-11$ to $t-1$.
Technology-Linked Firm Return (TECHRET)	The weighted average stock return of technology-linked peer firms, where the weights are the technological closeness between the peer firm and the focal firm, determined by the similarities between patent distributions across different technology categories. *TECHMOM* is obtained by compounding *TECHRET* from month $t-11$ to $t-1$.
Pseudo-Conglomerate Return (CONGRET)	The sales-weighted return of value-weighted, single-segment firm portfolios, formed for each segment that a conglomerate firm operates in. *CONGMOM* is obtained by compounding *CONGRET* from month $t-11$ to $t-1$.
Text-Based Industry Momentum (TNICRET)	Equal-weighted stock return of peer firms identified through 10-K product text. *TNICMOM* is obtained by compounding *TNICRET* from month $t-11$ to $t-1$.
Job-Flow-Based Return (J2JRET)	Weighted average stock return of same-industry peer firms, where the weights are the job-to-job flows. *J2JMOM* is obtained by compounding *J2JRET* from month $t-11$ to $t-1$.
Monthly Return (RET)	The monthly stock return. Following Shumway (1997), we adjust stock returns for delisting to avoid survivorship bias.
Firm Size (SIZE)	The logarithm of the market capitalization (in million dollars) as measured at the end of the previous June.
Market Beta (BMKT)	The CAPM beta computed using a 60-month window with a minimum window of 24 months.
Book-to-market (BM)	Computed as the book value of stockholders' equity, plus deferred taxes and investment tax credit (if available), minus the book value of the preferred stock for the last fiscal year, scaled by the market value of equity at the end of December of $T-1$. Depending on availability, the redemption, liquidation, or par value (in that order) is used to estimate the book value of the preferred stock.

Variable	Definition
Momentum (MOM)	Obtained by compounding RET from month $t-11$ to $t-1$.
Idiosyncratic Volatility (IVOL)	Computed as the standard deviation of the daily residuals obtained by regressing the daily excess stock returns on the daily market excess return, small-minus-big (SMB) and high-minus-low (HML) factors over the previous month.
Illiquidity (ILLIQ)	Amihud's illiquidity (Amihud, 2002), defined as the average daily ratio of the absolute stock return to the dollar trading volume within the previous month.
Maximum Return (MAX)	The maximum daily stock return realized over the previous month.
Skewness (SKEW)	The sample skewness of the daily stock returns from the previous month.
Coskewness (COSKEW)	The stock's monthly coskewness constructed following Harvey and Siddique (2000).
Cumulative Abnormal Return (CAR)	Market-adjusted returns cumulated over a three-day window around earnings announcements.
Standardized Unexpected Earnings (SUE)	Calculated as the difference between a stock's quarterly earnings minus the same-quarter value from the previous year, divided by its standard deviation over the past eight quarters.
Analyst Forecast Errors (FE)	Calculated as the difference between the announced earnings and analysts' consensus forecast, scaled by the stock price five trading days before the earnings announcement date.

Table A2: Comovements in Firm Fundamentals and Stock Returns, Value-Weighted Industry Portfolio

The table presents the comovements between the focal firm and its industry peers using panel regression analysis. In columns 1-4, the dependent variables are firms' fundamentals, measured annually: ΔEPS is the change in EPS scaled by lagged stock price, $\Delta Sales$ is the percentage growth in sales, $\Delta Employees$ is the percentage growth in the number of employees, and $NewCapital$ is the sum of net equity issuance plus net debt issuance scaled by lagged enterprise value. Only the firms with the same fiscal year-end as the focal firm are included. In column 5, the dependent variable is monthly stock returns. In Panel A and B, the key independent variable is the corresponding fundamentals of the contemporaneous value-weighted industry portfolio (excluding the focal firm). Panel B includes the corresponding measures of both the SCI- and value-weighted industry portfolios. We follow Gu et al. (2020) and cross-sectionally rank the independent variables and scale them into the [0,1] intervals. All regressions include time fixed effects and we present the coefficient estimates in percentages. Standard errors are clustered by both time and firm and the corresponding t-statistics are reported in parentheses. For all panels, *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Value-Weighted Industry Portfolio

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
Value-Weighted	1.312***	18.083***	6.926***	23.947***	5.021***
	(4.168)	(9.746)	(6.845)	(8.354)	(24.613)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	115,626	120,737	119,748	120,858	1,667,918
R^2	0.017	0.035	0.037	0.029	0.134

Panel B: Multivariate Analysis

	ΔEPS	ΔSales	ΔEmployees	NewCapital	Returns
	(1)	(2)	(3)	(4)	(5)
SCI-Weighted	2.426***	29.034***	14.139***	34.255***	7.201***
	(6.255)	(10.831)	(10.477)	(10.710)	(25.469)
Value-Weighted	0.148	0.497	−0.949	3.279	0.923***
	(0.581)	(0.435)	(−1.577)	(1.420)	(9.322)
Time FE	Yes	Yes	Yes	Yes	Yes
Observations	114,931	120,009	119,003	120,140	1,664,695
R^2	0.018	0.048	0.047	0.044	0.149

Table A3: Return Predictability of Social Peer Firm Returns: Within Industry Univariate Sort

The table reports the results of the univariate portfolio sorts based on *SPFRET*, the SCI-weighted average return of a firm's industry peers (excluding those from the same state), averaged over the FF48 industries. Each month, for each FF48 industry, we sort all common stocks into deciles based on *SPFRET* and calculate both the equal-weighted and value-weighted one-month-ahead returns for the decile portfolios, as well as the return of the portfolio that long the decile 10 portfolio and short the decile 1 portfolio. We then take the average of these portfolio returns over all industries and present both the average (raw) returns and the FF5 alphas. All returns are in percentages. The corresponding *t*-statistics based on Newey and West (1994) standard errors are reported in parentheses below the alphas. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1 Low)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10 High)	(10-1)
Raw Return EW	0.935***	1.042***	0.990***	1.104***	1.121***	1.184***	1.177***	1.199***	1.203***	1.286***	0.351***
	(4.387)	(4.724)	(4.435)	(4.942)	(4.888)	(5.345)	(5.414)	(5.565)	(5.643)	(6.246)	(5.217)
FF5 Alpha EW	−0.293***	−0.151**	−0.233***	−0.137*	−0.109	−0.022	−0.027	−0.059	−0.028	0.058	0.351***
	(−4.577)	(−2.085)	(−3.545)	(−1.912)	(−1.543)	(−0.308)	(−0.428)	(−0.818)	(−0.428)	(0.886)	(4.707)
Raw Return VW	0.914***	1.041***	0.979***	1.066***	1.084***	1.160***	1.106***	1.165***	1.144***	1.229***	0.314***
	(4.510)	(4.901)	(4.686)	(5.121)	(4.973)	(5.744)	(5.392)	(5.743)	(5.742)	(6.499)	(4.607)
FF5 Alpha VW	−0.310***	−0.120	−0.240***	−0.158**	−0.135*	−0.045	−0.086	−0.078	−0.067	0.025	0.335***
	(−4.625)	(−1.617)	(−3.386)	(−2.066)	(−1.713)	(−0.655)	(−1.193)	(−0.908)	(−1.020)	(0.345)	(4.529)

Table A4: Return Predictability of Peer Firms' Returns: Robustness Checks

The table presents robustness checks for Table 9, column 9 with alternative specifications. Section 2.1 provides detailed variable descriptions. Column 1 is the same as Table 9, column 9. In column 2, we use a panel regression with month fixed effects and standard errors double-clustered by month and by stock. Column 3 replaces the equal-weighted industry return with its value-weighted version. Column 4 considers an alternative social peer firm return measure, $SPFRET_{DIST}$, in which we use the the geographical proximity of firms' headquarters locations as a proxy for social ties. Column 5 replaces county-based *SPFRET* with its ZIP code-based version. In column 6, we replace the Fama-French 48 classification with the text-based industry classification while calculating *SPFRET*. In column 7, we use an alternative standardization method and transform all the explanatory variables by subtracting its sample mean and dividing by its standard deviation. Note that the coefficient in this column is not directly comparable to those in the other columns due to the scale differences. In column 8, we use an orthogonalized *SPFRET*, defined as a regression residual of *SPFRET* on all the other independent variables. In column 9, *SPFRET* is calculated using only the peer firms that are headquartered in states where the focal firm does not have an economic presence. In column 10, we add *SIMRET* to the controls, which is the similarity-weighted peer firm return, where the similarity between the focal county and peer firm county is measured based on four socioeconomic county characteristics: population density, education, political inclination, and income. All returns are reported in percentages. Missing values of independent variables are imputed with the monthly medians. In all columns except column 2, standard errors are adjusted based on Newey and West (1994). *t*-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	RET$_{t+1}$									
	(1)	(2)$_{\text{Panel}}$	(3)$_{\text{VW}}$	(4)$_{\text{DIST}}$	(5)$_{\text{ZIP}}$	(6)$_{\text{TNIC}}$	(7)$_{\text{SD}}$	(8)$_{\perp}$	(9)$_{\text{No Overlap}}$	(10)$_{\text{SIM}}$
SPFRET	0.493***	0.748***	0.899***	0.411***	0.401***	0.742***	0.109**	0.265***	0.317***	0.339***
	(3.320)	(3.860)	(5.112)	(5.152)	(3.578)	(6.143)	(2.516)	(2.811)	(3.047)	(2.819)
INDRET	0.693***	0.666**	0.082	0.802***	0.753***	0.887***	0.221***	0.894***	0.792***	0.649***
	(3.723)	(2.348)	(0.531)	(4.731)	(5.329)	(5.213)	(3.616)	(4.855)	(4.404)	(3.207)
GEORET	0.165**	0.206*	0.150*	0.153**	0.161**	0.166**	0.066***	0.128	0.164**	0.169**
	(2.159)	(1.702)	(1.954)	(2.035)	(2.117)	(2.191)	(2.701)	(1.521)	(2.145)	(2.234)
J2JRET	0.219**	0.196	0.236**	0.279***	0.257***	0.267***	0.095***	0.327***	0.287***	0.218**
	(2.294)	(1.334)	(2.488)	(2.954)	(2.743)	(2.821)	(3.839)	(3.261)	(3.015)	(2.326)
CONGRET	−0.035	−0.105	−0.021	−0.028	−0.029	−0.030	−0.016	−0.047	−0.022	−0.042
	(−0.457)	(−0.939)	(−0.286)	(−0.370)	(−0.360)	(−0.392)	(−0.899)	(−0.577)	(−0.294)	(-0.547)
CRET	0.319***	0.308***	0.333***	0.317***	0.320***	0.315***	0.068***	0.303***	0.322***	0.316***
	(3.583)	(2.770)	(3.790)	(3.583)	(3.590)	(3.538)	(3.659)	(3.340)	(3.605)	(3.562)
TNICRET	0.569***	0.739***	0.599***	0.576***	0.582***	0.144	0.147***	0.597***	0.584***	0.568***
	(4.627)	(3.531)	(4.884)	(4.621)	(4.636)	(1.501)	(4.233)	(4.559)	(4.700)	(4.655)
TECHRET	0.043	0.115	0.065	0.051	0.050	0.045	−0.006	0.032	0.055	0.040
	(0.375)	(0.746)	(0.605)	(0.482)	(0.436)	(0.431)	(−0.204)	(0.262)	(0.490)	(0.359)
CFRET	1.301***	1.544***	1.335***	1.313***	1.322***	1.268***	0.415***	1.335***	1.334***	1.295***
	(8.185)	(7.999)	(8.119)	(8.213)	(8.296)	(8.090)	(7.367)	(7.697)	(8.269)	(8.231)
SIMRET										0.219
										(1.288)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Periods	276	276	276	276	276	276	276	253	276	276
# Stocks	3,867	3,867	3,867	3,867	3,867	3,867	3,867	3,966	3,867	3,867
Month FE	–	Yes	–	–	–	–	–	–	–	–
Observations	–	1,067,332	–	–	–	–	–	–	–	–
R^2	0.069	0.126	0.069	0.069	0.069	0.069	0.072	0.070	0.069	0.070

Table A5: Return Predictability of In-State Peer Firms and Non-Peer Firms

The table presents robustness checks for Table 9, column 9 with two additional controls. $SPFRET_{STATE}$ is the SCI-weighted return of same-state peer firms and *NPFRET* is the SCI-weighted return of firms from other industries (excluding same-state firms). Section 2.1 provides detailed descriptions of the other variables. All returns are reported in percentages. Missing values of independent variables are imputed with the monthly medians. We cross-sectionally standardize all independent variables by mapping them into the $(0, 1]$ interval through scaling the monthly ranks by the number of observations (Gu et al. (2020)). We report *t*-statistics with standard errors based on Newey and West (1994) adjustments in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	RET_{t+1}	
	(1)	(2)
SPFRET	0.467***	0.478***
	(3.218)	(3.413)
SPFRET_STATE	0.331***	
	(4.515)	
NPFRET		−0.139
		(−1.634)
INDRET	0.627***	0.648***
	(3.343)	(3.760)
GEORET	0.159**	0.163**
	(2.113)	(2.314)
J2JRET	0.210**	0.221**
	(2.247)	(2.376)
CONGRET	−0.038	−0.032
	(−0.504)	(−0.423)
CRET	0.314***	0.320***
	(3.548)	(3.611)
TNICRET	0.558***	0.567***
	(4.585)	(4.638)
TECHRET	0.035	0.047
	(0.303)	(0.422)
CFRET	1.273***	1.297***
	(8.188)	(8.195)
Controls	Yes	Yes
# Periods	276	276
# Stocks	3,867	3,867
R^2	0.070	0.070

Table A6: Long-Run Predictability, Earnings Surprises, Analyst Forecast Error, and Earnings Returns (Orthogonalized SPFRET and SPFMOM)

The table presents robustness checks of Tables 6 and 7 with $SPFRET_{\perp}$ and $SPFMOM_{\perp}$, which are the version of *SPFRET* and *SPFMOM* orthogonalized to the other predictive variables, respectively. Panel A replicates Table 6, and Panel B replicates Table 7 Panel A. See Tables 6 and 7 for detailed descriptions. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Calender-Time Portfolio Sort on $SPFRET_{\perp}$ and $SPFMOM_{\perp}$ (FF5 alphas)

	$\text{SPFRET}_{\perp}$			$\text{SPFMOM}_{\perp}$		
	(1)	(10)	(10-1)	(1)	(10)	(10-1)
Months 1–3	−0.204***	0.124**	0.329***	−0.172**	0.215***	0.387***
	(−3.229)	(2.357)	(4.177)	(−2.484)	(2.959)	(3.056)
Months 4–6	−0.104**	0.031	0.136**	−0.095	0.122	0.216
	(−2.080)	(0.714)	(2.437)	(−0.872)	(1.628)	(1.488)
Months 7–12	−0.139**	0.010	0.149***	−0.134	0.016	0.150
	(−2.207)	(0.226)	(2.966)	(−1.513)	(0.263)	(1.529)
Months 13–24	−0.128**	0.004	0.131***	−0.090	0.021	0.110
	(−2.140)	(0.083)	(3.631)	(−1.223)	(0.359)	(1.409)

Panel B: Earnings Growth, Analyst Forecast Errors, and Earnings Announcement Returns

	Earnings Growth				Forecast Error				Earnings Return			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
SPFMOM$_\perp$	7.081***	5.605***	5.478***	3.798**	3.465**	4.058*	4.924**	2.085	0.153***	0.225***	0.142**	0.140**
	(4.641)	(3.694)	(3.569)	(2.483)	(2.004)	(1.952)	(2.512)	(1.106)	(2.681)	(3.857)	(2.547)	(2.349)
SPFRET$_\perp$	0.565	1.439	1.414	1.977**	0.490	0.950	1.756	1.433	0.120***	0.043	0.063	0.028
	(0.646)	(1.604)	(1.459)	(2.282)	(0.487)	(0.939)	(1.459)	(0.940)	(3.180)	(1.227)	(1.552)	(0.717)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	812,690	817,056	821,603	826,164	549,509	492,687	447,513	413,251	921,282	906,370	891,737	876,953
R^2	0.197	0.201	0.202	0.200	0.148	0.169	0.176	0.185	0.064	0.070	0.071	0.068