International Review of Financial Analysis 83 (2022) 102225

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

Innovation links, information diffusion, and return predictability: Evidence from China

Kailin Zeng [a,*], Ting Tang [a], Fangbiao Liu [b], Ebenezer Fiifi Emire Atta Mills [c]

[a] *School of Economics and Management, Jiangxi University of Science and Technology, Ganzhou, China*
[b] *Department of Planning and Finance, Wenzhou Medical University, Wenzhou, China*
[c] *School of Mathematical Sciences, Wenzhou-Kean University, Wenzhou, China*

ARTICLE INFO

JEL:
G12
E44
D81

Keywords:
Innovation link
Patent citation
Information diffusion
Return predictability

ABSTRACT

Based on the activities of patent citation in China, a novel type of cross-firm innovation links is generated to investigate the gradual diffusion of information along the innovation chain via tests of cross-sectional return predictability. Various signals are created to represent the value of the information contained in the innovation links; these signals are demonstrated to have robust cross-predictability for stock returns in both the cross-sectional regression model and portfolio strategies. The effect of predictability is found to be stronger for stocks with high institutional ownership and analyst coverage. Considering the minimum number of steps required to establish the cross-firm linkage, innovation links are further partitioned to represent different proximity of the linked firms. It is then found that information diffuses faster across closely-linked firms than across distantly-linked firms. Sophisticated investors are found to be able to properly process the relevant information and benefit from innovation links.

1. Introduction

How information is valued and impounded into stock prices has attracted extensive attention in the past two decades. Under the framework of the limited-information model, investors suffer from cognitive limitedness, and are not aware of the salient signals from related assets. Moreover, investors are usually lacking in necessary skills and techniques to precisely derive these signals from complicated sets of information. These two premises have led to a testable implication of the cross-sectional predictability of asset prices.

To decode the exact process of gradual information diffusion, some studies have exclusively concentrated on economic links as a channel with which to investigate cross-firm return predictability. For instance, Cohen and Frazzini (2008) investigated how a shock to one firm transfers to the firm's customers and suppliers, which are publicly reported in the financial statement of the given firm. Based on the input-output tables recorded by the U.S. Bureau of Economic Analysis, Menzly and Ozbas (2010) found cross-industry evidence of gradual information diffusion along customer-supplier links. Cohen and Lou (2012) further extended these previous studies by considering a type of economic link formed by overlapping segments; they demonstrated that a shock in one segment will first be incorporated into the prices of standalone firms before the impounding of the information into the prices of peer firms operating in multiple segments. More recently, innovation-based economic links have emerged as an alternative channel through which to conduct cross-firm information diffusion.[1] For example, Bekkerman, Fich, and Khimich (2021) employed textual analysis to calculate the similarity of patents as a measure of cross-firm innovation links. Alternatively, Lee, Sun, Wang, and Zhang (2019) used the correlation of patent distribution across the United States Patent and Trademark Office (USPTO) technology classes to identify technologically linked firms.[2] These studies both focused exclusively on U.S. data, and documented

* Corresponding author.
E-mail address: kailinzeng12@163.com (K. Zeng).

[1] Some early studies employed innovation links to explore cross-firm spillovers to patenting, market value, and profitability. For instance, Jafe (1996) used the similarity of patent classification as a proxy of innovation proximity, and investigated the cross-firm spillover to patents, market value, and profits. Bloom et al. extended the study of Jafe (1996) to consider the spillover effects across different patent classes. There are also alternative approaches to the measurement of innovation proximity, such as collaboration on research projects (Ahuja, 2000) and patent citations (Mowery, Oxley, & Silverman, 1996; Stuart & Podolny, 1996).

[2] Based on U.S. patent data, Fung (2003) calculated innovation proximity via the backward citation and overlap of patent citations at the industry level, and documented a positive relationship between innovation proximity and contemporaneous return co-movement.

https://doi.org/10.1016/j.irfa.2022.102225
Received 24 November 2021; Received in revised form 20 May 2022; Accepted 31 May 2022
Available online 3 June 2022

similar evidence featuring robust return predictability based on innovation links. The present study provides new insights into innovation links based on data from China, and employs the frequency of the cross-firm citation of invention patents as a novel tool to measure the innovation proximity of linked firms.

To calculate the strength of innovation links, comprehensive data provided by CnOpenData and the China Stock Market & Accounting Research Database (CSMAR) are considered, which record 683,021 pair-wise patent citation activities covering about 83% of the exchange-listed companies in China. Next, the strength of innovation links is calculated as the frequency of bidirectional invention patent citations. This methodology directly captures the real spillovers of technology to reflect cross-firm innovation activities; in contrast, text-based and patent classification-based measurements commonly adopted by the extant literature are subject to the inadequacies of the text recognition and patent taxonomy, and both serve as indirect measures of innovation linkage.

The economic mechanism of this study is related to technology segmentation and the complex nature of corporate research and development (R&D) activities. Although innovation-linked firms may belong to different industries, they can share the fundamentals of technology and follow the same technology roadmap. A negative technology shock,[3] e.g., the failure of a critical R&D project or a longstanding suspension of R&D activities due to an abrupt incident to a focal firm, can substantially affect the development and market sales of the firm's product, thereby releasing a negative signal to investors and leading to a drop in its stock price. However, due to the complexity and long duration of R&D projects, negative technology shocks do not influence innovation-linked firms contemporaneously. Instead, the linked firms must evaluate the severity of the situation caused by the negative shock before they decide to suspend their current R&D project and shift their technology roadmap in the future. The investors first realize the negative signal of related firms after the firms decide to change their innovation strategy. Consequently, there is an observable lead-lag relationship between the stock prices of innovation-linked firms.

Motivated by the research of Burt and Hrdlicka (2021) and Schlag and Zeng (2019), the return predictive signal is generated as the weighted sum of return-based quantities, and the weight is the relative strength of the innovation linkage. Both excess returns and the residuals of the Fama-French five-factor model are considered as the return-based quantities; the latter was proposed by Burt and Hrdlicka (2021) as an alternative measure to isolate the potential model misspecification originating from the underlying asset pricing model.

Correlated fundamentals are well recognized as a necessary condition for return predictability under the framework of limited information (Hong, Torous, & Valkanov, 2007). The return on assets (ROA) and return on equity (ROE) are employed as proxies for the measurement of profitability, and panel regression is executed to verify whether the firms along the innovation chain have correlated fundamentals. The positive and statistically significant coefficients of the regression models provide preliminary evidence of the co-moving fundamentals across innovation-linked firms.

Furthermore, Fama-MacBeth (see Fama and MacBeth (1973)) cross-sectional regression and portfolio formation tests are performed to investigate whether signals based on innovation links generate robust cross-firm return predictability. In the cross-sectional analysis, an increase of 100 basis points in the return (residual) signal leads to an approximate basis point increase of 15.2 (14.1) in the next month's returns. The magnitude of the predictability is found to be both economically meaningful and statistically significant. Moreover, by including standard control variables in the regression, it is found that predictability is not subject to firm characteristics, such as short-term reversal, momentum, the industry lead-lag effect, and the firm-specific return betas of the factor model (see Menzly and Ozbas (2010)). In the portfolio formation test, the firms are sorted into tertile portfolios according to the signal values. A long-short portfolio is created based on the extreme quantiles, i.e., longing the last tertile with the highest signal value and shorting the first tertile with the lowest signal value. Should the signals carry substantial economic value, the average portfolio returns would monotonically increase along with the signal value, and the long-short portfolio would have positive and significant returns. The empirical findings strongly confirm this suspicion. For example, the long-short portfolios are found to yield positive and significant average returns, as well as alphas from the five-factor pricing model, regardless of which type of signal is used to construct the portfolios.

In a closer examination of the long-short portfolios performance, it is found that signals derived from the innovation chain can exert long-lasting effects on the prediction of stock returns. The returns of long-short portfolios are found to steadily increase over time without an obvious sign of reversion for up to 12 months.

Intuitively, the direction of patent citation flows can substantially affect the information content of the predictive signals. For example, a firm in the upstream of the citation materially relies on the technology and product of the firm in the downstream, while a firm in the downstream can also benefit from its technology transfer via patent licensing and the extra demand of its products from the upstream firms. The firms in the downstream and upstream of the citation flow are closely related, but they exert economical influence on each other through different channels. Consequently, robustness tests of cross-predictability based on upstream citation flow and downstream citation flow are executed separately to reveal the information diffusion via different channels. Signal based on upstream citation flow is found more robust than the signal based on downstream flow. It is found that downstream companies consistently attract a lower degree of attention from informed investors. Because predicting the stocks in the downstream involves using upstream citation flow, this finding exactly explains why the signal based on upstream of the citation flow generates more robust cross-sectional return predictability.

In contrast to the direct innovation links generated according to observable patent citations, indirect innovation links are created to represent more distant linkages that are hidden in the citations. Three types of innovation linkages representing at least one to three steps needed to form a connection between two firms are considered; an innovation link formed in a minimum of three steps is considered a distant linkage, and an innovation link formed in a minimum of one step is considered a close (direct) linkage. It is investigated whether information diffuses faster across firms connected with a shorter distance. The observed heterogeneity in the performance of the long-short portfolios provides substantial evidence of the slower diffusion of information across more distantly-linked firms.

Inspired by the research of Badrinath, Kale, and Noe (1995) and Barber and Odean (2008), institutional investors and financial analysts usually have information advantages over individual investors. Thus, the stock returns of firms with large institutional ownership and extensive financial analyst coverage are naturally less predictable. To verify this hypothesis, Fama-MacBeth cross-sectional model is augmented by considering interactive terms between the predictive signals and quintile dummies of institutional ownership and financial analyst coverage. Unsurprisingly, the magnitude of predictability is found to monotonically decrease with the increase of institutional ownership and financial analyst coverage. This finding is robust to the employment of both return and residual signals, as well as to the inclusion of standard control variables in the analysis. This evidence clearly supports that the value of signals based on the innovation chain is substantially weakened by the intervention of informed investors.

Because these findings suggest a non-trivial influence of informed investors on the value of signals, the next question to investigate is how

[3] A negative technology shock is presented as an example to illustrate the mechanism; alternatively, the use of a positive shock in the illustration would yield the same conclusion.

informed investors realize the implicit value of the signals. For example, do institutional investors trade on signals based on cross-firm innovation linkages, and do they take into account cross-firm innovation linkages during the portfolio rebalancing of innovation-linked stocks? Following the panel analysis[4] proposed by Menzly and Ozbas (2010), it is found that institutional investors do trade on the valuable information inferred from the cross-firm innovation chain.

In summary, the contributions of this study are as follows. First, this study considers invention patent citation data in China as the basis for a novel type of economic link, and provides international evidence of robust return predictability based on innovation links. Second, the effect of informed investors is found applicable to the signals related to innovation links, which is consistent with the findings of previous studies based on various economic links, e.g., Menzly and Ozbas (2010) and Schlag and Zeng (2019). Third, the proximity of the link is identified as an important impact factor of information transmission; information is found to diffuse quickly in closely-linked firms and slowly in distantly-linked firms. Fourth, informed investors are found to recognize the value of the signals inferred from the innovation chain, and to rebalance their portfolios accordingly.

The remainder of this article is organized as follows. In Section 2, a concrete example is provided to illustrate the mechanism of return predictability based on innovation links. Section 3 contains details of the measurement of the strength of innovation links, and provides descriptions of the relevant data. Section 4 presents the summary statistics of the cross-firm patent citations and the main results of this study. Finally, this work is concluded in Section 5.

2. Illustration of innovation links and return predictability

To better understand the mechanism of return predictability based on innovation links, this section provides concrete examples to illustrate the realization of the cross-predictability of returns. Considering three enterprises in China, namely ZTE corporation, Ultrapower, and UNIS Corporation, ZTE is connected to Ultrapower and UNIS via the bilateral citation of their invention patents. More specifically, as shown in Table 1, there is a longstanding mutual citation of invention patents between ZTE and Ultrapower. ZTE also persistently cites the patents of UNIS, and the citation frequency is strikingly high. UNIS has been citing ZTE's patents since 2009, and, given the high citation frequency, this innovation link is also strong.

ZTE, Ultrapower, and UNIS have not only formed strong innovation linkages, but also are fundamentally closely related according to their self-statements in their annual reports. ZTE was founded in 1985, and has been committed to providing integrated solutions in the telecommunications industry, including a series of wireless and wireline devices and professional telecommunications services. Ultrapower was founded in 2001, and has been committed to promoting industry development and social progress with information technology (IT) by forming five business segments, namely operator, Internet of Things and communication, artificial intelligence and big data, mobile game, and innovation businesses. UNIS was founded in 1999, and covers important areas of IT services in terms of hardware, software, and technical services featuring intelligent network equipment, storage systems, and technical consulting. All three companies are devoted to providing products and services in the IT field; ZTE operates as a supplier of information infrastructure, while both Ultrapower and UNIS focus more on the application of IT.

On April 17, 2018, the Bureau of Industry and Security, U.S. Department of Commerce ("BIS"), revoked the suspension of a denial order that was signed on March 23, 2017, and suspended the export privileges of ZTE for a period of seven years, until March 13, 2025. This message was undoubtedly interpreted by the financial market as negative news for ZTE, and after realizing its severe consequences on ZTE's business, ZTE activated a suspension of the trading of its stock on the Shenzhen Stock Exchange.

Although both Ultrapower and UNIS are innovatively linked to ZTE with correlated fundamentals, the stock prices of both Ultrapower and UNIS barely instantaneously reacted to the bad news about ZTE. As shown in Fig. 1, the stock prices of Ultrapower and UNIS did not significantly deviate from their average stock price right before the ZTE event. This phenomenon is not surprising under the limited information model proposed by Hong et al. (2007), which suggests that the market is segmented due to limited attention and information-processing capability. Informed investors specialized in ZTE mainly gather information related to ZTE, and the same is true for the investors of Ultrapower and UNIS. These specialized investors slowly incorporate information from their innovation-linked peers, which generates cross-sectional return predictability.

In the current case, both official and unofficial online interactive communication platforms are thoroughly searched with the keywords of "Ultrapower" and "UNIS," and all the information relevant to the ZTE event are recorded. The shareholders of Ultrapower first posted content expressing apprehension about the influence of the denial order on Ultrapower on April 23, six days after the negative news about ZTE, while the investors of UNIS first expressed concern about the connection between ZTE and UNIS on May 23, more than one month after the occurrence of the ZTE event. Unsurprisingly, the stock prices of both Ultrapower and UNIS began to decrease after the investors realized the connection between the two companies and ZTE, and no other firm-specific news that could have significantly influenced the fundamentals and stock prices of Ultrapower and UNIS had been contemporaneously revealed.

On July 13, BIS announced a superseding settlement agreement with ZTE, and subsequently removed ZTE from the denial list. As of July 13, ZTE's stock price had suffered a more than 80% drop since April 17, and the stock prices of Ultrapower and UNIS had respectively decreased by 40% and 20%. Thus, investors who were able to short sell Ultrapower and UNIS right after the activation of the denial order could have respectively earned abnormal returns of 20% to 40% over the two months after the denial order.

3. Empirical approach

To investigate cross-firm innovation links, all the invention patent citations provided by CnOpenData are scanned by filtering out the citations whose citing party and cited party are not listed in the stock exchange. Given the fact that not all the listed companies are involved in patent citation, the number of firms included in the sample represents only a subset of the entire sample of listed companies. Nevertheless, as shown in the subsequent section, the selected sample represents a nontrivial portion of the entire sample.

Once the patent citation data for the listed companies are properly handled, the calculation of the strength of cross-firm innovation links is straightforward. For example, to calculate the strength of the innovation link between firms i and j in year t, the frequency with which all the invention patents owned by firm i cite patents owned by firm j in year t is first identified, as is the frequency with which invention patents owned by firm i are cited by patents owned by firm j. Next, the strength of the innovation relationship between firms i and j is simply taken as the sum of the frequencies of the two-directional citations.[5] This approach is different from the measurement of technology closeness adopted by Jaffe (1986), Bloom, Schankerman, and Van Reenen (2013), and Lee et al. (2019), who used the similarity between the firms' patent

[4] As shown later, the empirical findings are qualitatively similar when the Fama-MacBeth methodology is used as a robustness check.

[5] In the actual calculation of the strength of the innovation links, natural logarithm of one plus sum of the frequencies of the two-directional citations is used to mitigate the influence of extreme values.

Table 1
The frequency of patent citations between ZTE, Ultrapower, and UNIS.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Panel A: ZTE and Ultrapower										
ZTE citing Ultrapower	2	5	8	7	9	15	19	27	14	11
Ultrapower citing ZTE	3	18	50	38	28	19	15	17	15	9
Panel B: ZTE and UNIS										
ZTE citing UNIS	107	225	210	225	148	362	294	185	123	102
UNIS citing ZTE	8	18	32	55	51	69	83	94	134	42

This table reports the frequency of the bilateral citation of invention patents owned by ZTE, Ultrapower, and UNIS from 2009 to 2018. Panel A (Panel B) presents the frequency of patents owned by ZTE citing patents owned by Ultrapower (UNIS), and vice versa.

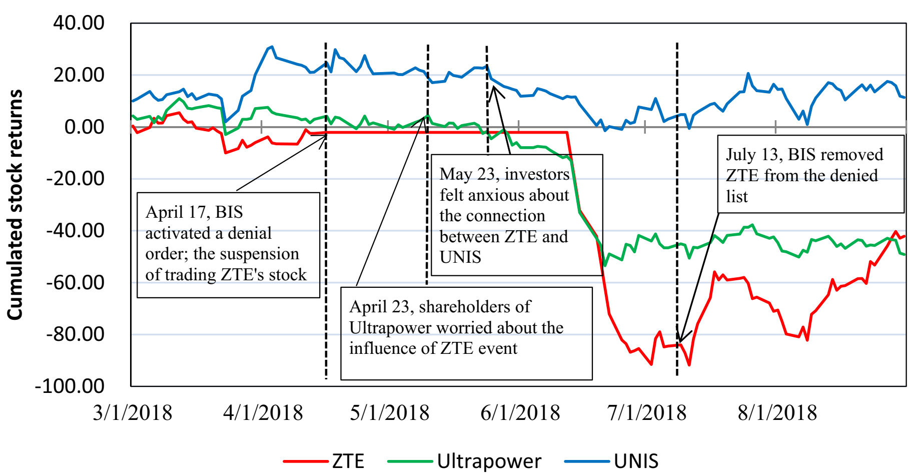

Fig. 1. Stock returns after the activation of a denial order.
This figure reports the cumulative stock returns of ZTE, Ultrapower, and UNIS after the Bureau of Industry and Security, U.S. Department of Commerce ("BIS") activated a denial order on April 17, 2018. The four vertical dashed lines respectively represent the date of activation of the denial order, the date shareholders of Ultrapower expressing apprehension about the impact from ZTE event, the date when investors of UNIS felt anxious about the ZTE event, and the date when ZTE was removed by BIS from the denied list. The text on the figure conveys the key events related to the negative shock encountered by ZTE.

classifications defined by the USPTO as the strength of technology-based linkages. Another strand of literature has adopted the automated text-based measure of innovation proximity based on patent descriptions. For example, Packalen and Bhattacharya (2012), Younge and Kuhn (2016), and Bekkerman et al. (2021) applied the textual methodology to the analysis of innovation similarity, and the last study investigated the process of information diffusion through firms with textual patent similarity. However, these two measures of innovation links have been criticized for the inadequacy of the text recognition method and the subjectivity of the patent taxonomy, respectively.

The advantage of the proposed approach is the ability to obtain a direct and economically meaningful connection based on real activities of corporate innovation. Two firms sharing similar technologies do not necessarily guarantee a technological inheritance and bilateral dependence on their products. Therefore, the proposed approach provides new insights to disentangle the genuine innovation-based connections between firms.

Given the strength of the innovation links now available for all pairs of firms in the sample, the corresponding predictive signals for the firms in the sample are computed in a straightforward manner. More specifically, the predictive signals for firm i can be expressed as the weighted sum of the return-based quantities of its innovation-linked peers, i.e.,

$$signal_{t,i} = \sum_{j} w_{t,i,j} q_{t,j} \tag{1}$$

where $q_{t,\,j}$ is the return-based quantity of firm j at time t, and $w_{t,\,i,\,j}$ is the relative strength of the innovation link between firms i and j at time t, which is given by the following equation.

$$w_{t,i,j} = \frac{L_{t,i,j}}{\sum_k L_{t,i,k}} \tag{2}$$

In Eq. (2), $\sum_k L_{t,\,i,\,k}$ is the sum of innovation link strength for all the firms linked to firm i. The method of computing innovation link $L_{t,\,i,\,j}$ guarantees a non-negative and yearly updated weight, while the sum of the weights in a given period is always equal to 1. However, the update frequency of return predictive signals depends on not only the weights, but also the frequencies of the selected return-based quantities i.e., it is the higher of the two frequencies. In this study, monthly stock returns and the monthly Fama-French five-factor model residuals are used as return-based quantities, so the return predictive signals are updated along with the return-based quantities every month. It is standard to use stock returns as the input of the signal (e.g., Cohen and Frazzini (2008), Menzly and Ozbas (2010)), and the choice of the five-factor model residuals is motivated by Burt and Hrdlicka (2021), who found an upward bias of return predictability by using the returns as the input of the signals.

The patent citation data are obtained from CnOpenData. The stock returns, pricing factors, and information related to the shareholdings of institutional investors are sourced from the CSMAR. Stocks with the "ST" marker and stocks with missing returns in the last 12 months are excluded to mitigate the impact of illiquidity on the empirical results of return predictability.

4. Results

4.1. Summary statistics

To obtain an overview of the cross-firm patent citations, Table 2 provides summary statistics to demonstrate the key dimensions of the data. Due to data availability, the sample period of the patent citation data is from 2009 to 2018. The number of exchange-listed companies in China steadily increases during the sample period, from 1604 firms at the beginning of 2009 to 3584 firms by the end of 2018. The number of firms involved in the activities of patent citation varies across the sample

Table 2
The summary statistics of patent citations.

	Min	Median	Max
Number of stocks	659	2171	2754
Panel A: Statistics of connections in the cross-firm patent citation network			
Frequency of patent citation (being cited)	10% Quantile	Median	90% Quantile
	1	3	21
	Mean	Skewness	Kurtosis
	16.25	22.30	678.47
Frequency of patent citation (citing)	10% Quantile	Median	90% Quantile
	0.8	2	20
	Mean	Skewness	Kurtosis
	16.25	27.34	994.57
Panel B: Proportion of cross-industry citations			
	Min	Median	Max
Cited/citing	0.40	0.45	0.48

This table reports the minimum, median, and maximum numbers of the stocks, as well as the descriptive statistics of the patent citation activities. In Panel A, the frequency of a given patent citing or being cited by other patents is investigated, and basic statistics are reported. Panel B investigates whether the firms involved in the patent citations belong to the same industry, and the minimum, median, and maximum proportions of citations with the patent owners across different industries are reported. The industries are identified by the China Securities Regulatory Commission (CSRC) to represent 90 classifications.

period, yet generally remains above a certain proportion of the listed companies. The median number of firms appearing in the patent citation data is 2171, while the median number of total listed firms is 2613. The former is about 83% of the latter, which represents a sufficiently large share of the entire sample of listed companies.

Panel A of Table 2 reports the statistics of the frequency of cross-firm patent citation. It is clear that the frequencies of a firm's patents cited by and citing patents owned by other firms do not follow a normal distribution. The average frequency of cross-firm patent citation is found to be 16.25,[6] which is far beyond the median of 2. The skewness and kurtosis are also found to be much larger than the corresponding measures of a normal distribution. This evidence implies that some firms are located at the center of the citation network (these firms conduct a relatively large number of patent citations), while other firms are positioned at the periphery of the citation network (these firms receive few citations of their patents, and also rarely cite patents owned by other firms).

Intuitively, firms in the same industry usually share the same technology, so their innovation activities are more likely closely related. Consequently, a natural question is whether the prevalence of cross-firm patent citations primarily originates from the firms within the industry. To answer this question, the proportion of cross-firm links originating from firms across different industries is calculated, and the industries are identified by the China Securities Regulatory Commission (CSRC) to represent 90 classifications. Panel B of Table 2 reports the minimum, median, and maximum values of the proportion of cross-industry citations across the sample period. Unsurprisingly, about half of the cross-firm citations are found to originate from firms within the industry. Nevertheless, the median proportion of cross-industry citations is found to be 45%, with a minimum proportion of 40% and a maximum proportion of 47%, indicating that close to half of the citations convey important information beyond the common laws of the industry.

[6] Self-citations have already been excluded in the construction of cross-firm linkages. For instance, citations between two patents owned by the same company are not counted.

4.2. Evidence of correlated fundamentals

The limited-information model derived from Hong et al. (2007) is the foundation of the cross-predictability of stock returns. In the model, two necessary conditions are needed to guarantee cross-firm return predictability. The first condition is that firms have correlated fundamentals with their innovation-linked peers, e.g., technology shocks can simultaneously influence the cash flows of firms with strong innovation links. Under the second condition, investors and financial analysts are specialized at evaluating firm-specific information, and due to their limited cognition and restricted information processing capability, they are unable to infer the consequences of a firm's innovation-linked peers. Because previous studies (e.g., Menzly and Ozbas (2010)) have consistently confirmed the specialization among informed investors, only the first condition for innovation-linked firms is verified in the present study.

Motivated by the methodology proposed by Menzly and Ozbas (2010) and Schlag and Zeng (2019), both return on assets (ROA) and return on equity (ROE) are considered as proxies for the measure of profitability. A panel regression is then used to verify whether firms along the innovation chain have correlated fundamentals, i.e., ROA and ROE.

$$ROA_{t,i} = \alpha_i + \beta_{market} ROA_{t,market} + \beta_{linked} ROA_{t,i,linked} + \varepsilon_{t,i} \tag{3}$$

$$ROE_{t,i} = \gamma_i + \theta_{market} ROE_{t,market} + \theta_{linked} ROE_{t,i,linked} + \epsilon_{t,i} \tag{4}$$

In the panel regression model, $ROA_{t,\ i}$ and $ROE_{t,\ i}$ are respectively the ROA and ROE of firm i in year t. $ROA_{t,\ market}$ and $ROE_{t,\ market}$ are respectively the aggregated ROA and ROE over the entire market in year t, which is the weighted sum of the firm-level ROA and ROE using the firm assets as the weights. $ROA_{t,\ i,\ linked}$ and $ROE_{t,\ i,\ linked}$ respectively represent the aggregated ROA and ROE of the firms innovatively linked to firm i using the strength of innovation link $w_{t,\ i,\ j}$ in Eq. (2) as the weight. To address the potential heterogeneity across firms, firm-level fixed effects are included in the regression models, with α_i and γ_i respectively representing the firm-level fixed effects for the panel regression on ROA and ROE. Inspired by Thompson (2011) and Petersen (2009), robust standard errors that are double-clustered by firm and year are utilized.

The summary statistics and correlation matrix of the profitability measures are reported in Table 3. It clearly shows that both firms' ROA and ROE have non-trivial correlation coefficients of 0.278 and 0.206 with the corresponding measures of their innovative peers. And the firm level ROA and ROE also co-move tightly with the aggregated measures over the market. The correlation coefficients between firm level ROA (ROE) and market ROA (ROE) is 0.672 (0.701), and statistically significant at 1% confidence level. However, the correlations between market measures and innovatively linked measures are only moderate, indicating that certain information in linked measures can not be captured by the market based measures. The last two rows report the mean and standard deviation of the profitability measures. Unsurprisingly, the mean of ROA related measures is found to be lower than that of the ROE related measures. The market based measures are on average always larger than the firm level measures due to the size effect that larger firms generally outperform the smaller firms in China.

The coefficients of the panel regressions reported in Table 4 confirm the proposition of correlated fundamentals along the innovation chain. In Model 1, the regression coefficient for the ROA of innovation-linked firms is found to be 0.211 and statistically significant. When the aggregated ROA over the entire market is included in the model, the coefficient slightly decreases to 0.205, yet remains statistically significant. Analogous evidence can be found in Model 2 with ROE as the measure of profitability. The regression coefficient for the ROE of linked firms is found to be 0.176, and experiences a slight decrease to 0.156 when the full model is considered. The magnitudes of the coefficients reported in Table 4 are also economically meaningful. For instance, an

Table 3
Summary statistics and correlation matrix of corporate fundamentals.

Variables	ROA	ROA_linked	ROA_market	ROE	ROE_linked	ROE_market
ROA						
ROA_linked	0.278***					
ROA_market	0.672***	0.336***				
ROE	0.421***	0.020	0.353***			
ROE_linked	0.063	0.476***	0.087	0.206***		
ROE_market	0.523***	0.045	0.638***	0.701***	0.175**	
Mean	0.042	0.040	0.053	0.064	0.056	0.120
SD	0.010	0.012	0.009	0.022	0.028	0.020

This table reports the time-series mean, standard deviation and correlation coefficient of corporate fundamentals in Eq. (3) and Eq. (4), i.e., the annual return on assets (ROA) and return on equity (ROE) of a focal firm, the aggregated ROA and ROE over the entire market and the firms innovatively linked to the focal firm. Firm assets are used as weights to calculate the market-based measures. The calculation of the innovation-linked measures is analogous to the computation of return signals, and the relative strength of the innovation links is used as the weight. ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2009 to 2018.

Table 4
Test of correlated fundamentals.

Model 1	ROA		Model 2	ROE	
	[1]	[2]		[1]	[2]
ROA_linked	0.211*** (5.24)	0.205*** (5.00)	ROE_linked	0.176*** (4.24)	0.156*** (3.67)
ROA_market		0.520*** (2.90)	ROE_market		0.613*** (3.07)
Fixed effects	Yes	Yes	Fixed effects	Yes	Yes
Clustered Standard errors	Yes	Yes	Clustered standard errors	Yes	Yes
R^2	0.278	0.357	R^2	0.215	0.282

This table reports the panel regressions of the annual return on assets (ROA) and return on equity (ROE) of a focal firm on the contemporaneous ROA and ROE over the entire market and the firms innovatively linked to the focal firm. Firm assets are used as weights to calculate the market-based measures. The calculation of the innovation-linked measures is analogous to the computation of return signals, and the relative strength of the innovation links is used as the weight. The standard errors are adjusted for heteroskedasticity and double-clustered at the stock-year level, and the t-statistics are reported in parentheses. All specifications in the panel regression include stock and quarter fixed effects. Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2009 to 2018.

increase of 100 basis points in the ROA (ROE) of the innovation-linked firms is found to correspond to an increase of 20.5 (15.6) basis points in the ROA (ROE) of the focal firm.

4.3. Cross-firm return predictability

4.3.1. Cross-sectional regression

In this section, it is verified whether the signals created in the Empirical Approach section generate return predictability under the standard cross-sectional test. To relieve the concern of model misspecification bias due to the common alpha of the excess stock returns (see Burt and Hrdlicka (2021), Schlag and Zeng (2019)), both the signals based on excess returns and on factor model residuals are considered in the test.

Following the methodology proposed by Cohen and Frazzini (2008), a cross-sectional regression initiated by Fama and MacBeth (1973) is conducted to justify the return predictability due to innovation links. The regression has the following form.

$$r_{t,i} = a_t + \lambda_{t,signal} signal_{t-1,i} + \Delta_t Z_{t-1,i} + \mu_{t,i} \quad (5)$$

In the model, $r_{t,\ i}$ and $signal_{t-1,\ i}$ respectively represent the excess returns of firm i at month t and the predictive signals based on the excess returns and factor-model residuals with a one-month lag. The generation of the signals is described in Section 3. To eliminate predictability due to other firm characteristics and to isolate the return predictability from innovation-based links, a set of control variables $Z_{t-1,\ i}$ is included in the regression model. Following a standard setting in the literature, the control variables include the stock return over the previous month to represent the short-term reversal proposed by Jegadeesh (1990), the average return over the previous 11 months from $(t-1)$ to $(t-12)$ as a proxy of the medium-term momentum (see Jegadeesh and Titman (1993, 2001)), and the industry return over the previous month to capture the lead-lag effect within industries (see Moskowitz and Grinblatt (1999) and Da, Liu, and Schaumburg (2014)). For the model based on return signals, the inclusion of the firm-specific return betas of the Fama-French five-factor model is found to hardly affect the results; thus, the details of the return betas are omitted here. Regarding the model based on residual signals, the effect of pricing factors is already eliminated by using the factor model residuals as the sources of the signals. Thus, for brevity and simplicity, the coefficients of the return betas are not reported here.

Table 5 reports the the correlation coefficients and the summary statistics of the stock return, signal based on returns, and signal based on factor model residuals. For simplicity, we call the latter two variables 'return signal' and 'residual signal', respectively. It shows that both return signal and residual signal significantly correlated with contemporaneous stock returns, the correlation coefficients are 0.294 and 0.278, respectively. This finding provides preliminary evidence that the financial market participants are moderately aware of the connection between the focal firm and their innovative peers. The correlation between return signal and residual signal is surprisingly high, with a statistically significant correlation coefficient 0.759. This implies the effect of model misspecification due to the common alpha of the factor pricing model (see Burt & Hrdlicka, 2021) does not significantly affect the information content in the signals. The means of raw return and return signal are 12.9% and 10.6%, while the average value of the residual signal is close to zero because the residual signal is calculated as a linear aggregation of the factor model residuals which is supposed to be zero according to the procedure of OLS regression.

Table 6 presents the time-series averages of the cross-sectional

Table 5
Summary statistics and correlation matrix of predictive signals.

Variables	Return	Return signal	Residual signal
Return			
Return signal	0.294***		
Residual signal	0.278***	0.759***	
Mean	0.129	0.106	0.005
SD	0.431	0.239	0.186

This table reports the time-series mean, standard deviation and correlation coefficient of the predictive signals based on raw returns and Fama-French five-factor model residuals. Creation of the signals is illustrated in Section 3. ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

Table 6
The cross-sectional test of return predictability.

Model	[1]	[2]	[3]	[4]
Constant	1.078	1.097	1.059	1.085
	(1.38)	(1.42)	(1.38)	(1.40)
Return signal	0.152**	0.125***		
	(2.58)	(2.81)		
Residual signal			0.141***	0.131***
			(2.65)	(2.88)
Short-term reversal		−0.046***		−0.046***
		(−5.42)		(−5.41)
Momentum		−0.026		−0.025
		(−0.67)		(−0.64)
Lagged industry return		0.083***		0.083***
		(3.93)		(3.90)
R^2	0.009	0.034	0.007	0.033

This table reports the time-series averages of the Fama-MacBeth cross-sectional regression coefficients for both the return and residual signals generated in Section 3. The control variables included in the regression are the stock return over the previous month (short-term reversal), the average return over the previous 11 months from (t − 1) to (t − 12) (momentum), and the industry return over the previous month (lagged industry return). The t-statistics are presented in parentheses with readjustment for autocorrelation and heteroskedasticity (Newey & West, 1987). Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

regression coefficients for both types of signals and the control variables. The time-series averages of the coefficients can also be interpreted as risk premiums because the regressor is, by construction, simply a tradable long-only portfolio with the sum of weights equal to one.

The empirical findings based on the cross-sectional regressions strongly support the hypothesis of return predictability. As shown in Column 1 of Table 6, the return signal as the standalone regressor in the regression is found to have a statistically significant average coefficient. The magnitude of the average coefficient is economically meaningful, e.g., an increase of 100 basis points in the return signal is found to lead to an additional increase of 15.2 basis points in the stock returns for the focal firm. When the controls are included in the cross-sectional regressions (Column 2 of Table 6), the coefficient is found to slightly decrease, yet is even more statistically significant. Columns 3 and 4 report the average coefficients of the signals based on the factor model residuals, which are found to be qualitatively similar to the case with signals based on excess returns. The average coefficient of the residual signal is slightly smaller than that of the return signal. This finding is reasonable, given that information related to pricing factors in the returns is already removed to obtain the residual signals. In the full model, an increase of 100 basis points in the residual signal is found to correspond to an increase of approximately 13 basis points in the stock returns for the focal firm. In addition, the signs of the control variables are consistent with findings of previous studies, excluding that the average coefficient of momentum is found to be statistically insignificant. The R-squares in Table 6 is relatively low due to some omitted variables and the fact that monthly stock returns are very volatile and hard to predict.

4.3.2. *Portfolio formation*

Portfolio formation is employed as another standard methodology for the verification of return predictability. In the portfolio formation test, the stocks at month t are sorted into three groups according to the two types of signals generated by the information up until month t. For instance, the first portfolio (Q1) contains the stocks with the lowest signal value, while the third portfolio (Q3) contains the stocks with the highest signal value. Then, the returns of the portfolios over the next month are recorded. A long-short portfolio is created by longing the stocks in the third portfolio and shorting the stocks in the first portfolio. In the calculation of the portfolio returns, the stock returns are aggregated using both equal weights and market value-based weights to address the concern of the size effect on the return predictability. Should the cross-sectional predictability prevail, the long-short portfolios would have positive and significant excess returns and factor model alphas, the latter of which is considered to isolate the effect of factor exposure on predictability.

Table 7 presents the excess returns and five-factor model alphas for the equally-weighted and value-weighted portfolios based on the return signals. The performance of these portfolios provides robust evidence of return predictability based on innovation links. As presented in panel A of Table 7, the equally-weighted long-short portfolio is found to have a monthly average excess return of 0.335, which is statistically significant at the 5% level. The five pricing factors are found to have only trivial influences on the returns of the long-short portfolio, e.g., the factor alpha of the portfolio is not substantially deviated from the returns. Panel B provides the excess returns and factor alphas for the value-weighted portfolios. It demonstrates that both the excess returns and factor alphas are close to those of the equally-weighted portfolios, indicating that this return predictability is not due to the size effect.

As mentioned previously, the true sources of cross-sectional predictability have triggered extensive discussion, e.g., Burt and Hrdlicka (2021) demonstrated that predictability can originate from the common component of stock returns related to pricing factors. Moreover, they provided further evidence that the long-short portfolios formed by some well-known economic link based signals suffered a substantial drop in their returns and factor model alphas after controlling for the effect of model misspecification. To verify whether model misspecification is the main source of the positive returns of the long-short portfolios, the methodology suggested by Burt and Hrdlicka (2021) and Schlag and Zeng (2019) is followed, and portfolios are created based on the residual signals.

As shown in Table 8, the performance of the portfolios based on residual signals strongly rejects the argument that model misspecification is the main source of innovation link based return predictability. The magnitudes of the returns and alphas of the long-short portfolios are found to be even slightly larger than those of the portfolios based on raw returns, indicating that the common alpha with respect to asset pricing factors only produces noises irrelevant to the return predictability derived from the signal generation process.

So far, only short-term return predictability is considered in the analysis of portfolio formation. A natural question regarding portfolio performance is whether long-short portfolios can perform persistently

Table 7
Portfolio strategies based on return signals.

Based on return signal	Q1	Q2	Q3	Q3-Q1
Panel A: Equally-weighted				
Excess return	0.826	1.049*	1.161**	0.335**
	(1.51)	(1.74)	(2.18)	(2.32)
Five-factor alpha	0.147	0.303*	0.396**	0.249**
	(0.95)	(1.93)	(2.57)	(2.44)
Panel B: Value-weighted				
Excess return	0.259	0.324*	0.589**	0.330**
	(0.86)	(1.68)	(2.53)	(2.57)
Five-factor alpha	0.066	0.143	0.391**	0.325**
	(0.58)	(1.12)	(2.11)	(2.40)

This table reports the excess returns and alphas of the monthly tertile portfolios created by sorting the stocks according to the one-month lagged return signals introduced in Section 3. The long-short portfolio (Q3-Q1) is constructed by longing the portfolio with the highest signal value and shorting the portfolio with the lowest signal value. The alphas are the intercepts of the five-factor pricing model proposed by Fama and French (2015). The results of equally-weighted and value-weighted portfolios are respectively reported in Panel A and Panel B, and the latter of which is considered to mitigate the well-known size effect on portfolio performance. The standard errors are reported in parentheses. Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

Table 8
Portfolio strategies based on residual signals.

Based on residual signal	Q1	Q2	Q3	Q3-Q1
Panel A: Equally-weighted				
Excess return	0.815	1.038*	1.170**	0.355***
	(1.47)	(1.68)	(2.18)	(2.67)
Five-factor alpha	0.043	0.274**	0.308***	0.265**
	(0.37)	(2.43)	(2.74)	(2.52)
Panel B: Value-weighted				
Excess return	0.244	0.341*	0.603***	0.359**
	(0.58)	(1.71)	(2.72)	(2.46)
Five-factor alpha	−0.049	0.149*	0.291**	0.340**
	(−0.37)	(1.69)	(2.21)	(2.57)

This table reports the excess returns and alphas of the monthly tertile portfolios created by sorting the stocks according to the one-month lagged residual signals introduced in Section 3. The long-short portfolio (Q3-Q1) is constructed by longing the portfolio with the highest signal value and shorting the portfolio with the lowest signal value. The alphas are the intercepts of the five-factor pricing model proposed by Fama and French (2015). The results of equally-weighted and value-weighted portfolios are respectively reported in Panel A and Panel B, and the latter of which is considered to mitigate the well-known size effect on portfolio performance. The standard errors are reported in parentheses. Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

over time and do not reverse in the long term. This is not necessarily the case for all portfolio strategies. For example, event-driven portfolios and portfolios formed by investment sentiment usually cannot yield positive returns over a long period. However, the innovation link is a unique type of economic link through which information related to fundamentals is transferred; therefore, portfolios based on innovation links should generate persistent abnormal returns.

To verify the persistence of the performance of the long-short portfolios, the cumulative returns over the next 1 to 12 months after portfolio formation are calculated. Fig. 2 reports the return patterns of four types of long-short portfolios with holding periods varying from 1 to 12 months. As shown in the figure, the cumulative returns of all types of long-short portfolios steadily increase over the next 4 months after portfolio formation. Subsequently, the performance of long-short portfolios begins to diverge over longer periods. The equally-weighted long-short portfolio-based return signals are found to continue to yield positive returns up to 7 months after portfolio formation, while the other types of long-short portfolios do not show clear patterns regarding their return performance. Most importantly, the cumulative returns of all types of long-short portfolios are found to exhibit no sign of reversion, which implies that the signals based on innovation links can have long-lasting effects on stock prices.

4.3.3. *Innovation proximity and information diffusion over time*

The long-horizon performance of long-short portfolios presented in the previous subsection provides preliminary evidence of gradual information diffusion over time. Specifically, the evidence provided by the long-horizon analysis suggests a slow transmission of the innovative shock over time as the average returns of long-short portfolios slowly but steadily increased for up to 4 to 5 months. In this section, it is further analyzed how the proximity of innovation connectedness influences information transmission over time. More precisely, it is investigated whether a piece of information diffuses more quickly across closely-linked firms than across distantly-linked firms. A distant linkage is a type of indirect connection that cannot be easily identified without patient inference; thus, it takes a longer amount of time for investors to impound information from distantly-linked firms.

To identify the proximity of cross-firm innovation links, the minimum number of steps required to form a connection between two firms is considered. For instance, if patents of firm i are found to cite or to be cited by patents of firm j, then firm i is connected to firm j in one step, i.e., forming a direct linkage, and the distance between the two firms is 1. If firm i is connected to firm k in one step, if firm k is connected to firm j in one step, and if no direct connection between firms i and j is found, then firm i is connected to firm j in two steps, and the distance between the two firms is 2. Generally, if firm i can be connected to firm j in at least n steps, then the distance between the two firms is n. It should be noted that all types of innovation links generated here represent different distances between the links, and they do not overlap by definition.

It is hypothesized that information diffuses faster across firms with shorter distances than across firms with longer distances. To validate this hypothesis, the dynamic performances of long-short portfolios based on innovation links with distances from 1 to 3 are examined.[7]

Table 9 reports the cumulative returns of long-short portfolios based on the three types of innovation links described previously. The holding period of the portfolios is between 1 and 6 months after portfolio formation. As robustness checks, different weighting schemes are also considered, and both return and residual signals are used in the construction of the portfolios. The results reported in Table 9 clearly demonstrate that the first-month returns account for about one-half of the cumulative returns over 6 months in the case of one-step linkages, while in the case of two-step linkages, the first-month returns are much smaller in value. For the case of three-step linkages, the first-month returns are even close to zero. In fact, two-step (three-step) cumulative returns over first two months are found to account for less than 50% (25%) of the cumulative returns over 6 months. It is not until 3 months after portfolio formation that long-short portfolios based on two-step and three-step linkages are able to yield significant returns. This evidence clearly suggests that investors take a longer amount of time to impound the flow of information from more distantly-linked peers.

4.3.4. *Effect of informed investors*

In the limited-information model, limited cognitive capacity and information processing ability of investors are two key elements contributing to the cross-firm return predictability. For instance, Kahneman (1973) suggested that individuals generally have more difficulty in allocating cognitive resources to multitask. This is true for a large number of non-professional investors who mainly devote themselves to their own career development, instead of exclusively concentrating on processing firm-specific information from the financial market. Moreover, the information processing ability of non-professional investors is also generally lower than that of professional investors. Although the data indicate that activities of cross-firm patent citation are persistent over time to a large extent, given the large number of firms and citations from different patents, the identification of the firm-specific activities of citation and the calculation of the weighted signals remain time-consuming and complicated tasks.

Consequently, the cognitive capacity and information processing ability of shareholders can exert substantial influences on the magnitude of the return predictability. In contrast to individual investors, institutional investors are considered as sophisticated agents that are better able to handle a vast amount of salient information and complicated tasks in a timely manner (e.g.,Brennan, Jegadeesh, and Swaminathan (1993), Badrinath et al. (1995), and Barber and Odean (2008)). Thus, in this study, institutional ownership is used as a proxy for informed investors,[8] and a firm's institutional ownership is calculated as the firm's shareholding ratio of institutional investors in a given quarter. Should

[7] The authors also investigate the performance of long-short portfolios with distance from 4 to 6, the result is qualitatively similar to the case with distance from 1 to 3. For brevity, only the result with distance from 1 to 3 is presented here.

[8] Financial analyst coverage and analyst reports of the companies are also used as the proxies of informed investors, and the analytical results are found to be qualitatively similar to those of the analysis based on institutional investors. The details of these robustness checks are reported in Section 4.5.2.

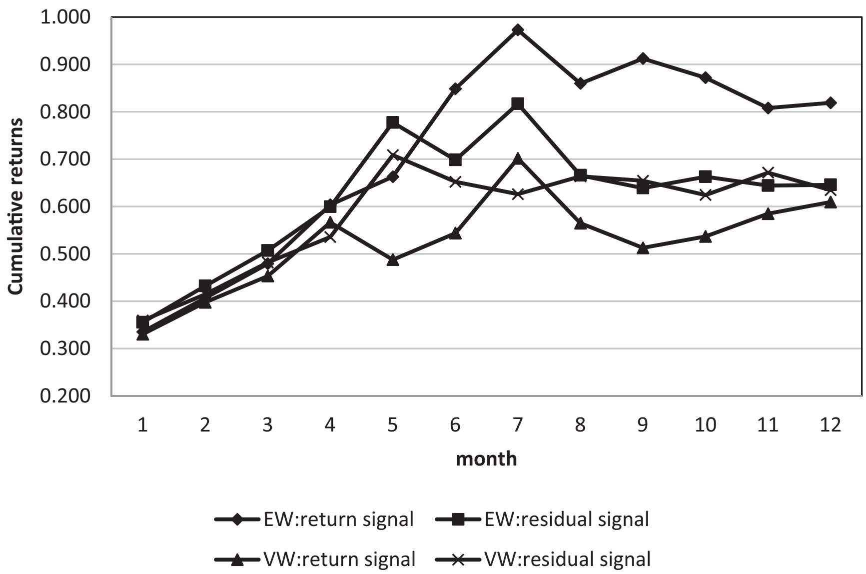

Fig. 2. Cumulative returns over a long horizon.
This figure presents the cumulative returns of the long-short portfolios between 1 and 12 months after portfolio formation. The long-short portfolio is constructed by longing the portfolio with the highest signal value and shorting the portfolio with the lowest signal value. The underlying portfolio is created by sorting the stocks according to the return and residual signals introduced in Section 3. For each type of signal, both equally-weighted (EW) and value-weighted (VW) portfolios are investigated; for the latter, the market capital size of the stock is used as the portfolio weight to relieve the impact of size effect on the performance of the portfolios.

Table 9
The proximity of innovation links and information diffusion.

Long-short portfolios	Distance of the innovation linkage	Holding period (months)					
		1	2	3	4	5	6
Equally-weighted: return signal	1	0.335	0.407	0.479	0.603	0.663	0.849
	2	0.203	0.316	0.945	1.078	1.098	1.231
	3	0.052	0.110	0.751	0.945	1.054	1.052
Equally-weighted: residual signal	1	0.355	0.432	0.507	0.599	0.777	0.698
	2	0.229	0.378	0.708	0.896	1.130	1.179
	3	0.049	0.183	0.571	0.753	1.009	1.195
Value-weighted: return signal	1	0.330	0.397	0.453	0.566	0.488	0.544
	2	0.112	0.235	0.529	0.607	0.710	0.736
	3	0.058	0.077	0.546	0.663	0.678	0.757
Value-weighted: residual signal	1	0.359	0.414	0.482	0.535	0.709	0.652
	2	0.152	0.286	0.510	0.571	0.830	0.789
	3	0.047	0.095	0.532	0.662	0.735	0.744

This table reports the cumulative returns of the long-short portfolios based on three types of innovation links up to 6 months after portfolio formation. The types of innovation links have different linkage proximity. For instance, the most distant linkage here means at least three steps are required for a firm to make a connection to the target firm, while the closest linkage means only one step is needed to form a cross-firm connection. The underlying portfolio is created by sorting the stocks according to the return and residual signals. The signals are generated by the same methodology introduced in Section 3, but by using innovation links with different strengths. For each type of signal, both equally-weighted and value-weighted portfolios are investigated in the analysis; for the latter, the market capital size of the stock is used as the portfolio weight to relieve the impact of size effect on the performance of the portfolios.

companies have higher institutional ownership, their salient information from innovation-related firms would be collected in a more timely manner and processed more accurately, and would therefore be impounded more quickly into their stock prices. In this case, the stock returns of companies with higher institutional ownership should be less predictable.

In practical operation, the aforementioned cross-sectional regressions are extended by including interaction terms between the predictive signals and the quintile dummy of institutional ownership.

$$r_{t,i} = a_t + \sum_{k=1}^{5} \lambda_t^k Q_{t-1}^k signal_{t-1,i} + \Delta_t Z_{t-1,i} + \mu_{t,i} \tag{6}$$

In the augmented model, Q_k is equal to 1 if the institutional ownership of firm i belongs to the k^{th} quintile; otherwise, it is equal to 0. Moreover, λ_t^k represents the magnitude of the return predictability at time t for the group of stocks whose institutional ownership lies in the k^{th} quintile of the sample. Table 10 reports the time-series averages of the magnitude of the return predictability for firms in different quintiles of institutional ownership. The results clearly confirm the hypothesis that companies with more informed investors are less likely to generate predictable returns.

Regardless of whether return or residual signals are used in the augmented model, the average magnitude of return predictability λ_t^k is found to strictly decrease along with the increase in the companies' institutional ownership. Furthermore, not only the magnitude presents a downward trend; the statistical significance is also on the wane. For example, the interaction term based on return signals (residual signals) is found to become statistically insignificant for firms with institutional ownership in the 5th (4th and 5th) quintile.

As reported in Table 10, these findings of the magnitude and statistical significance of the predictability are found to be robust after the inclusion of the control variables introduced in Section 4.3.1 in the regression model.

4.4. Innovation links and institutional trading

The empirical findings presented in Section 4.3.4 have already demonstrated that institutional investors are sophisticated agents who are able to digest a substantial amount of information from firms innovatively linked to the firm whose stock they hold. Naturally, the

Table 10
The effect of informed investors.

	Return signal		Residual signal	
	Without control	With controls	Without control	With controls
Constant	1.069	1.094	1.062	1.091
	(1.42)	(1.42)	(1.42)	(1.40)
Signal *Q1	0.204***	0.188**	0.237***	0.233***
	(2.67)	(2.51)	(2.81)	(2.85)
Signal *Q2	0.184**	0.166**	0.211**	0.201***
	(2.43)	(2.47)	(2.52)	(2.64)
Signal *Q3	0.171**	0.145**	0.157**	0.124*
	(2.06)	(1.98)	(2.01)	(1.81)
Signal *Q4	0.129*	0.110*	0.107	0.079
	(1.76)	(1.67)	(1.57)	(1.34)
Signal *Q5	0.062	0.042	0.048	0.017
	(0.88)	(0.58)	(0.96)	(0.63)
R^2	0.009	0.037	0.007	0.035

This table reports the time-series averages of the Fama-MacBeth cross-sectional regression coefficients for the interactive terms between the predictive signal and a dummy variable representing the quintiles of institutional ownership. The predictive signals include both the return and residual signals calculated in Section 3. The control variables in the regression include the stock return over the previous month (short-term reversal), the average return over the previous 11 months from (t − 1) to (t − 12) (momentum), and the industry return over the previous month (lagged industry return). The t-statistics are presented in parentheses with readjustment for autocorrelation and heteroskedasticity (Newey & West, 1987). Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

next question is whether informed investors consider innovation-based signals when they change their stock holdings. Taking institutional investors as a proxy of informed investors, this question yields two testable hypotheses, the first of which is whether institutional investors simultaneously trade on innovation-linked stocks by taking cross-firm innovation links into account, and the second of which is whether institutional investors trade on the value of information inferred from innovation-based signals. For instance, should the signals be positive (negative), the institutional investors would accordingly increase (reduce) the shareholding of the linked firm. To investigate the two hypotheses, the following panel regressions are estimated:

$$\Delta IO_{i,q} = \omega_i + \tau_q + \beta^{IO}\Delta IO^{linked}_{i,q} + \eta_{i,q} \tag{7}$$

$$\Delta IO_{i,q} = \omega_i + \tau_q + \beta^{signal}Signal_{i,q} + \eta_{i,q} \tag{8}$$

where $\Delta IO_{i,\ q}$ and $\Delta IO^{linked}_{i,\ q}$ are the changes in the percentages of the institutional ownership of firm i and firms innovatively connected to firm i, respectively. The methodology of computing the signals is analogous to the methodology used in Section 3, except that the return-based quantity considered here is the cumulative returns or the cumulative residual returns of the linked firms over the given quarter. Moreover, ω_i is the stock-level fixed effects used to control the estimation bias due to the heterogeneity across the firms. The year-quarter fixed effects τ_q are also included to isolate the systematic influence of abnormal flows of funds on institutional ownership.

Panel A of Table 11 reports the magnitude and statistical significance of the coefficients of the change in the institutional ownership of the linked firms and the coefficients of the two types of predictive signals. The results provide strong evidence that institutional investors increase their ownership of a stock together with their holdings in its innovation-related stocks. The coefficient of the change in the institutional ownership of the innovatively linked firms is found to be 0.174, which is positive and statistically significant at the 1% level. The positive and statistically significant coefficients for both return signals (0.018) and residual signals (0.032) in the regression affirm a positive association between the change in the institutional ownership of a certain stock and the signal of the stock. As a robustness test, the Fama-MacBeth

Table 11
Innovation links and institutional trading.

Panel A: panel regression			
ΔIO_{linked}	0.174***		
	(34.50)		
Residual signal		0.032***	
		(7.74)	
Return signal			0.018***
			(4.87)
R^2	0.118	0.037	0.025
Observations	80,024		
Panel B: Fama-MacBeth regression			
ΔIO_{linked}	0.129***		
	(12.99)		
Residual signal		0.081***	
		(3.66)	
Return signal			0.053**
			(2.39)
R^2	0.053	0.015	0.007

This table reports the results of regressions in which the quarterly changes in a firm's institutional ownership are regressed on the contemporaneous changes of institutional ownership in innovation-linked firms and on the return and residual signals. The results of the panel regression (Fama-MacBeth regression) are presented in Panel A (Panel B). The change of institutional ownership is the change in the proportion of shares outstandingly held by institutional investors in the current quarter. The changes of institutional ownership in innovation-linked firms are computed analogously to the procedure of computing return and residual signals. The standard errors are adjusted for clustering at the stock-year level, and the t-statistics are reported in parentheses. All specifications in the panel regression include stock and quarter fixed effects. Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

regression is executed and the hypotheses are re-verified. In the Fama-MacBeth regression, the quarterly change in the institutional ownership of a stock is regressed on the institutional ownership of the related stocks or on the signals. Panel B of Table 11 reports the time-series averages of the coefficients of the regressors. As evidenced by the magnitude and significance of the coefficients of ΔIO_{linked} and the signals, the Fama-MacBeth result is found to be qualitatively similar to the findings in the panel regression, which reaffirms that institutional investors do trade on the valuable information inferred from innovation-linked firms.[9]

4.5. Robustness checks

4.5.1. Different citation flow

Intuitively, firms citing from other firms (forward citation) and firms being cited (backward citation) are different in many aspects. Firms citing from other firms can be attributed to the demand of updating outdated technology and cross-field innovation. In the first case, the firms citing from others substantially rely on the technology of their target firms. In the second case, the firms citing from others usually execute the cross-filed strategy in their product development and R&D. For instance, a firm manufacturing refrigerator would consider adding digital controller and internet devices into its physical products to realize the function of remote and real-time automatic monitoring. The development of the new products corresponds to patent citation from manufacturing industry to computer, communications, and other electronic equipment manufacturing industry.

Economically, the firm in the upstream of the citation materially

[9] It seems that the R-squares in Table 11 are relatively low. This is because some important variables are omitted in the panel regression. However, the low R-squares are consistent with the previous studies, e.g., Menzly and Ozbas (2010).

relies on the technology and product of the firm in the downstream, so the progress in the research and product development of the downstream firm will severely influence the business of the firms in the upstream. On the contrary, the firm in the downstream can also benefit from its technology transfer via patent licensing and the extra demand of its product from the upstream firms. Consequently, the firms in the downstream and upstream of the citation flow are closely interdependent, but they exert significant influence on each other via different channels.

Consequently, robustness tests of cross-predictability based on upstream citation flow and downstream citation flow are executed separately to reveal the information diffusion via different channels. More specifically, the signal generating process in Section 3 is used to create two types of signals: one based on the downstream citation flow and the other based on the upstream citation flow.

Table 12 reports Fama-MacBeth cross-sectional regression coefficients for both the return and residual signals based on downstream citation flow and upstream citation flow. The results are qualitatively analogous to those in Table 6 where bidirectional citation flow is used to generate the predictive signal. Unsurprisingly, the regression coefficients for both signals based downstream citation flow and upstream citation flow are slightly lower than the coefficient for signal based on bidirectional citation flow. However, the regression coefficients are still statistically significant at 5% (10%) confidence level for the upstream (downstream) signal. Particularly, the regression coefficient of the upstream signal is larger and statistically more significant than that of the downstream signal. This finding is also confirmed in the robustness test on portfolio strategies, which is shown in Table 13. Panel A (Panel B) of Table 13 shows the excess return and Fama and French (2015) five-factor model alpha for equally-weighted (value-weighted) long-short portfolios constructed by longing the portfolio with the highest signal value and shorting the portfolio with the lowest signal value. The signal here is generated based on both downstream citation flow and upstream citation flow. Consistent with the findings in Table 12, the long-short portfolio based on upstream signal outperforms the corresponding portfolio based on downstream signal in both magnitude and statistical significance of the portfolio return and alpha.

Why the signal based on upstream citation flow matters more for cross-return predictability is both an important and intriguing research question. As illustrated at the beginning of this Section, the firms in the downstream and upstream of the citation flow are closely related, and they can interact through different channels of information transmission. Hence, it can be because the two different channels attract different degrees of attention from the informed investors, while the less concerned channel takes longer time to impound the information from innovatively related firms into the stock prices of the focal firm, resulting in more significant return predictability. This hypothesis is verified by examining the analyst coverage and analyst reports for the stocks frequently appearing in the upstream and downstream of the citation flow. Stocks with frequent appearance in the upstream (downstream) refer to stocks with frequency of citing downstream firms (being cited by the upstream firms) in the first quartile.[10] The left panel of Fig. 3 presents the number of analysts in group who make earnings forecasts on the stocks frequently cited by the upstream stocks and stocks frequently citing downstream stocks in a given year. It shows that stocks in the downstream of the citation flow consistently attract lower degree of attention from the financial analysts than the stocks in the upstream except the year 2018. The gap of attention from the financial analysts is particularly large in the first half of the sample and becomes smaller in the second half of the sample. As shown in the right panel of Fig. 3, similar results are obtained using analyst reports of the stock issued by the analysts in a given year as a proxy of informed investors. Because predicting the stocks in the downstream involves using upstream citation flow, this finding exactly explains why the signal based on upstream of the citation flow generates more robust cross-sectional return predictability.

Table 12
Robustness check on the cross-sectional test of return predictability.

Model	Based on downstream flow		Based on upstream flow	
	[1]	[2]	[3]	[4]
Constant	1.118	1.097	1.059	1.085
	(1.42)	(1.42)	(1.42)	(1.40)
Return signal	0.087*		0.112**	
	(1.74)		(2.36)	
Residual signal		0.095*		0.121**
		(1.86)		(2.48)
Short-term reversal	−0.046***	−0.046***	−0.046***	−0.046***
	(−5.33)	(−5.39)	(−5.43)	(−5.41)
Momentum	−0.025	−0.025	−0.027	−0.025
	(−0.64)	(−0.65)	(−0.70)	(−0.65)
Lagged industry return	0.086***	0.086***	0.084***	0.084***
	(3.98)	(3.99)	(3.95)	(3.91)
R^2	0.033	0.033	0.034	0.034

This table reports the time-series averages of the Fama-MacBeth cross-sectional regression coefficients for both the return and residual signal based on downstream citation flow and upstream citation flow. The control variables included in the regression are the stock return over the previous month (short-term reversal), the average return over the previous 11 months from (t − 1) to (t − 12) (momentum), and the industry return over the previous month (lagged industry return). The t-statistics are presented in parentheses with readjustment for autocorrelation and heteroskedasticity (Newey & West, 1987). Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

Table 13
Robustness check on portfolio strategies.

Long-short portfolio return	Based on downstream flow		Based on upstream flow	
	Return signal	Residual signal	Return signal	Residual signal
Panel A: Equally-weighted				
Excess return	0.177*	0.189*	0.249**	0.205**
	(1.67)	(1.79)	(2.33)	(1.99)
Five-factor alpha	0.161	0.191*	0.215**	0.221**
	(1.58)	(1.82)	(1.99)	(2.26)
Panel B: Value-weighted				
Excess return	0.207*	0.188*	0.296**	0.355**
	(1.85)	(1.86)	(2.23)	(2.47)
Five-factor alpha	0.196*	0.167*	0.275**	0.398***
	(1.69)	(1.73)	(2.01)	(2.87)

This table reports the excess returns and alphas of the long-short portfolios created by sorting the stocks according to the one-month lagged return signals based on both downstream citation flow and upstream citation flow. The long-short portfolio (Q3-Q1) is constructed by longing the portfolio with the highest signal value and shorting the portfolio with the lowest signal value. The alphas are the intercepts of the five-factor pricing model proposed by Fama and French (2015). The results of equally-weighted and value-weighted portfolios are respectively reported in Panel A and Panel B, and the latter of which is considered to mitigate the well-known size effect on portfolio performance. The standard errors are reported in parentheses. Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

4.5.2. *Alternative proxy for informed investors*

In Section 4.3, institutional ownership is used as a proxy for informed investors and it is found that stocks with higher institutional ownership

[10] Using first quintile and first tertile to define the stocks frequently appearing in the upstream and downstream of the citation flow obtain qualitatively similar results to the case using first quartile in the main text. For brevity, only the results using first quartile in the definition of stocks with frequent appearance are reported.

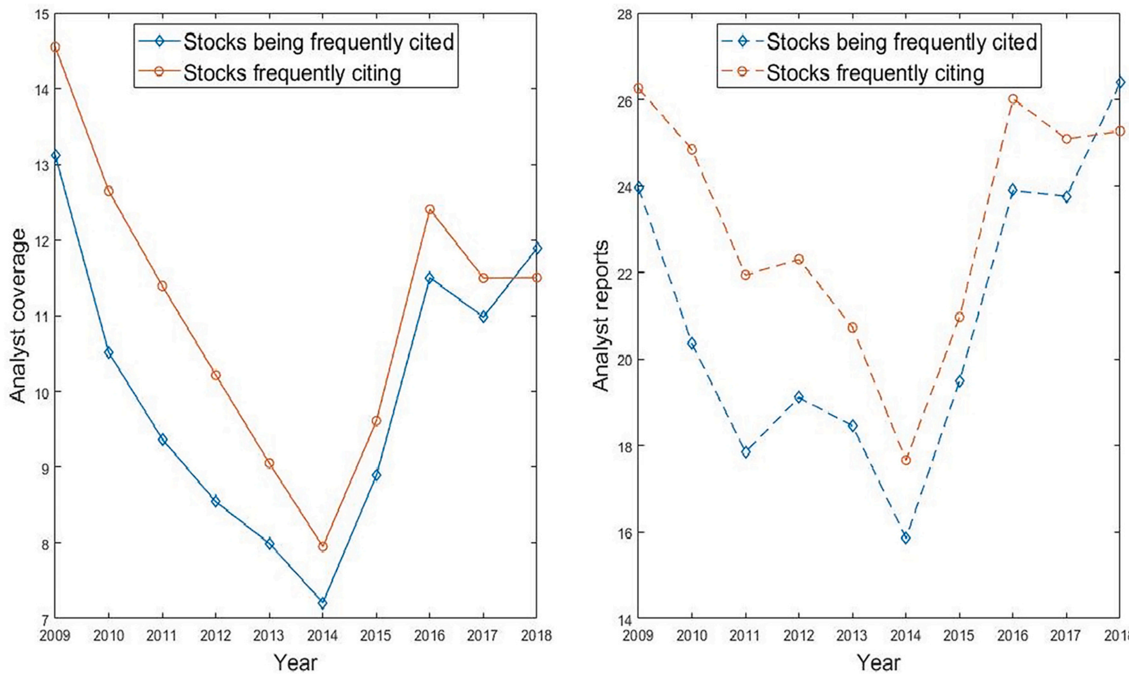

Fig. 3. Attention of informed investors for the stocks frequently appearing in the citation flow.
This figure presents the analyst coverage and analyst reports for the stocks frequently citing from downstream stocks in citation flow and stocks being frequently cited by upstream stocks in citation flow. Analyst coverage is the number of analysts in group who make earnings forecasts on the stock in a given year. Analyst report is the number of research reports of the stock issued by the analysts in a given year. Stocks are sorted into quartile group according to the frequency of citing from downstream stocks and being cited by upstream stocks, respectively. And those in the first quartile are considered as stocks frequently citing or being cited. The sample is from 2009 to 2018.

are less likely to generate predictable returns. The magnitude and statistical significance of the return predictability λ_t^k in Eq. (6) strictly decrease along with the increase of the stocks' institutional ownership. In this section, financial analyst coverage and analyst reports of the company are used as alternative proxies for informed investors. Analyst coverage in a given month is the number of financial analysts in group who made earnings forecasts on the stock within the last 12 months. Analyst report in a given month is the number of research reports of the stock issued by the analysts within the last 12 months.

Table 14 presents the time-series average of the Fama-MacBeth cross-sectional regression coefficients for the interactive terms between the predictive signal and two dummy variables representing the quintiles of lagged analyst coverage and analyst reports. Regardless of whether return or residual signal is used in the regression model, the average magnitude and statistical significance of return predictability exhibit a pattern analogous to that in Table 10, i.e., λ_t^k is found to strictly decrease along with the increase in the degree of companies' analyst coverage and analyst reports. In addition, the average value of the regression coefficients even becomes negative for the extreme quintiles. For example, for the companies with analyst coverage or analyst reports in the 4th and 5th quintile, the average coefficients of the interactive term turn into negative, but statistically insignificant.

Table 14
Robustness check on the effect of informed investors.

	Effect of analyst coverage		Effect of analyst reports	
	Return signal	Residual signal	Return signal	Residual signal
Constant	1.100	1.083	1.116	1.091
	(1.41)	(1.40)	(1.43)	(1.40)
Signal *Q1	0.304***	0.310***	0.287***	0.338***
	(2.74)	(2.63)	(2.69)	(2.86)
Signal *Q2	0.292**	0.303***	0.198**	0.271**
	(2.26)	(2.618)	(2.01)	(2.11)
Signal *Q3	0.182**	0.264**	0.160*	0.243*
	(1.99)	(2.36)	(1.72)	(1.89)
Signal *Q4	−0.103	−0.046	−0.068	−0.053
	(−1.01)	(−0.56)	(−0.54)	(−0.34)
Signal *Q5	−0.162	−0.131	−0.152	−0.143
	(−1.53)	(−1.27)	(−1.34)	(−1.31)
R^2	0.038	0.035	0.039	0.035

This table reports the time-series average of the Fama-MacBeth cross-sectional regression coefficients for the interactive terms between the predictive signal and two dummy variables representing the quintiles of lagged analyst coverage and analyst reports. The predictive signals include both the return and residual signals calculated in Section 3. Analyst coverage in a given month is the number of analysts in group who made earnings forecasts on the stock within the last 12 months. Analyst reports in a given month is the number of research reports of the stock issued by the analysts within the last 12 months. The control variables in the regression include the stock return over the previous month (short-term reversal), the average return over the previous 11 months from (t − 1) to (t − 12) (momentum), and the industry return over the previous month (lagged industry return). The t-statistics are presented in parentheses with readjustment for autocorrelation and heteroskedasticity (Newey & West, 1987). Moreover, ***, **, and * respectively indicate statistical significance at the 1%, 5%, and 10% levels. The sample period is from 2010 to 2019.

5. Conclusion

Based on the activities of invention patent citations in China, the frequency of citations is carefully aggregated to form cross-firm innovation links of exchange-listed companies in China. This measure of innovation linkage is able to capture both vertical and horizontal technology spillover, e.g., new knowledge flowing through the supplier chain and across the firms producing substitute products; this is also distinctive from the indirect measures of innovation links, such as the text-based measure and the measure of the similarity in patent classification. The cross-firm innovation link is also not only an approximation of cross-industry activity; evidence shows that about one-half of the patent citations originate from firms across different industries.

Before the validation of return predictability based on innovation links, preliminary evidence of the strong correlation of the fundamentals of linked firms is found, e.g., the profitability of linked firms is found to co-move contemporaneously. As the activities of corporate innovation usually rely on complex and long-lasting R&D projects, and due to the cognitive limitedness of common investors, innovative shocks are supposed to diffuse throughout the financial market with a wave pattern. Strong evidence is found to support this hypothesis. The returns of innovation-linked firms are found able to predict returns of the focal firm in the standard Fama-MacBeth cross-sectional regression. In the test of portfolio formation, long-short portfolios based on the value of information derived from innovation links are found to yield annual returns between 4% and 4.3%. The implied information from innovation links has long-lasting effects on the stock prices of the related firms, as no sign of reversion is found by holding the portfolio over a long horizon. Further analysis indicates that the return predictability comes from both the upstream and downstream of the patent citation flow, and the former is found more robust than the latter because companies in the downstream citation flow consistently attract a lower degree of attention from informed investors.

In addition, information is found to transmit faster across closely-linked firms than across distantly-linked firms, the latter of which is a type of indirect linkage and more difficult to identify. Intuitively, institutional investors and financial analysts are supposed to be less affected by cognitive capacity and information processing ability; the findings of this study affirm that the stock returns of firms with higher institutional ownership and analyst coverage are less predictable.

Finally, the analyses conducted in this research provide concrete evidence of institutional trading on the information from innovation links. First, institutional investors trade on innovation-related signals, e.g., increasing (reducing) the holding of a focal firm by observing a positive (negative) signal of the firm. Second, institutional investors change the holding of a focal firm in accordance with a contemporaneous change in the holdings of innovation-linked firms.

In conclusion, this study provides international evidence for the relationship between limited information and return predictability based on records of patent citation. The speed of information transmission depends on the degree of difficulty in identifying the innovation links, e.g., information diffuses faster via a direct (close) link and slower via an indirect (distant) link. The sophisticated investors who recognize the value of the information in innovation links can benefit from trading on the related signals.

Acknowledgement

The authors gratefully acknowledge research and financial support from Management Science Project (Grant/Award Number: 20192BAA208018) of Science and Technology Department of Jiangxi Province, China, high-level overseas talents returning project of Ministry of Human Resources and Social Security of the People's Republic of China (Grant/Award Number: 201916010), School of Economics & Management, Jiangxi University of Science & Technology, Ganzhou, China.

References

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly, 45*(3), 425–455.
Badrinath, S. G., Kale, J. R., & Noe, T. H. (1995). Of shepherds, sheep, and the cross-autocorrelations in equity returns. *Review of Financial Studies, 8*, 401–430.
Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies, 21*(2), 785–818.
Bekkerman, R., Fich, E. M., & Khimich, N. V. (2021). *The effect of innovation similarity on asset prices: Evidence from patents' big data* (SSRN working paper).
Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica., 81*(4), 1347–1393.
Brennan, M., Jegadeesh, N., & Swaminathan, B. (1993). Investment analysis and the adjustment of stock prices to common information. *Review of Financial Studies, 6*, 799–824.
Burt, A. P., & Hrdlicka, C. M. (2021). Where does the predictability from sorting on returns of economically linked firms come from? *Journal of Financial and Quantitative Analysis, 56*(8), 2634–2658.
Cohen, L., & Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance, 63*(4), 1977–2011.
Cohen, L., & Lou, D. (2012). Complicated firms. *Journal of Financial Economics, 104*(2), 383–400.
Da, Z., Liu, Q., & Schaumburg, E. (2014). A closer look at the short-term return reversal. *Management Science, 60*(3), 658–674.
Fama, E. F., & French, K. R. (2015). Incremental variables and the investment opportunity set. *Journal of Financial Economics, 117*(3), 470–488.
Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy, 81*(3), 607–636.
Fung, M. K. (2003). Technological proximity and co-movements of stock returns. *Economics Letters, 79*(1), 131–136.
Hong, H., Torous, W., & Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics, 83*, 367–396.
Jaffe, A. (1986). Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. *American Economic Review, 76*(5), 984–1001.
Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance, 45*(3), 881–898.
Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance, 48*(1), 65–91.
Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance, 56*(2), 699–720.
Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice Hall.
Lee, C., Sun, S. T., Wang, R., & Zhang, R. (2019). Technological links and predictable returns. *Journal of Financial Economics, 132*(3), 76–96.
Menzly, L., & Ozbas, O. (2010). Market segmentation and cross-predictability of returns. *The Journal of Finance, 65*(4), 1555–1580.
Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance, 54*(4), 1249–1290.
Mowery, D. C., Oxley, J. E., & Silverman, B. S. (1996). Strategic alliances and interfirm knowledge transfer. *Strategic Management Journal, 17*(S2), 77–91.
Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*, 703–708.
Packalen, M., & Bhattacharya, J. (2012). *Words in patents: Research inputs and the value of innovativeness in invention*. NBER Working Paper (18494).
Petersen, M. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies, 22*(1), 435–480.
Schlag, C., & Zeng, K. (2019). Horizontal industry relationships and return predictability. *Journal of Empirical Finance, 53*, 310–330.
Stuart, T. E., & Podolny, J. M. (1996). Local search and the evolution of technological capabilities. *Strategic Management Journal, 17*(S1), 21–38.
Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics, 99*(1), 1–10.
Younge, K. A., & Kuhn, J. M. (2016). *Patent-to-patent similarity: A vector space model. Working paper* (Available at SSRN 2709238).