Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

When the Tail Wags the Dog: Industry Leaders, Limited Attention, and Spurious Cross-Industry Information Diffusion

Ling Cen, Kalok Chan, Sudipto Dasgupta, Ning Gao

To cite this article:
Ling Cen, Kalok Chan, Sudipto Dasgupta, Ning Gao (2013) When the Tail Wags the Dog: Industry Leaders, Limited Attention, and Spurious Cross-Industry Information Diffusion. Management Science 59(11):2566-2585. http://dx.doi.org/10.1287/mnsc.2013.1722

MANAGEMENT SCIENCE
Vol. 59, No. 11, November 2013, pp. 2566–2585
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2013.1722

When the Tail Wags the Dog: Industry Leaders, Limited Attention, and Spurious Cross-Industry Information Diffusion

Ling Cen
Rotman School of Management and Department of Management (UTSc), University of Toronto, Toronto, Ontario M5S 3E6, Canada, ling.cen@rotman.utoronto.ca

Kalok Chan, Sudipto Dasgupta
Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong {kachan@ust.hk, dasgupta@ust.hk}

Ning Gao
Accounting and Finance Group, Manchester Business School, University of Manchester, Manchester M15 6PB, United Kingdom, ning.gao@mbs.ac.uk

Within an industry, stock returns of larger firms lead those of smaller firms, suggesting an intraindustry information diffusion process. Most industry leaders, however, have business segments in other industries (henceforth, minor-segment industries), whereas most small firms are pure players operating in one industry only. If investors cannot filter out the irrelevant information from the leaders' minor segments, the pure players will be mispriced due to spurious cross-industry information diffusion (SCIID). Consistent with the SCIID hypothesis, we document both a strong contemporaneous and a lead–lag relation in stock returns between firms from industry leaders' minor-segment industries and pure players in the industry leaders' major-segment industry. Our results are not due to potential missing common factors or economic relationships between pure players and firms in the minor-segment industries.

Key words: limited attention; category learning; industry information diffusion
History: Received December 10, 2010; accepted October 19, 2012, by Wei Xiong, finance. Published online in *Articles in Advance* June 14, 2013.

1. Introduction

Understanding the process through which information gets incorporated into stock prices is an important area of research in finance. An extensive literature shows that investors in one stock extract information from the prices of other stocks. Information is more quickly reflected in the stock prices of firms with certain characteristics and diffuses more slowly to other firms, leading to a lead–lag effect in stock returns. For example, Lo and MacKinlay (1990) show that the stock returns of large firms lead those of small firms; Brennan et al. (1993) and Chordia and Swaminathan (2000) show that returns of stocks with more analyst coverage and higher trading volume lead those of stocks with the opposite characteristics; Menzly and Ozbas (2006, 2010) show that stocks that are in economically related supplier and customer industries cross predict each other's returns.

A few recent papers focus on information diffusion within the industry. Hou (2007) shows that the lead–lag effect between stock returns of large and small firms is predominantly an intraindustry phenomenon. In particular, industry-specific information is reflected first in the stock prices of "industry leaders," which are more liquid and are covered by more analysts, and then spreads to other firms, resulting in a lead–lag relationship between the stock returns of industry leaders and industry followers.

Although it is natural for investors to look at the industry leaders in uncovering industry-relevant information, that is easier said than done. Most industry leaders have fairly complicated business segments. Notably, more than 80% of industry leaders have multisegment businesses and their business segments are often not in the same industry. For the proper pricing of the industry leaders' stocks, information from both major and minor segments is relevant. Therefore, their returns reflect both major and minor segment fundamentals. However, if less sophisticated investors price industry pure players[1] based on the industry leaders' returns without being able to differentiate

[1] A pure player is a firm that does not have segments outside a particular industry. "Industry," "leader," and "pure player" are defined in §2. Pure players are typically small firms.

between the major and minor segment fundamentals, the lead–lag relationship of stock returns between industry leaders and pure players may reflect the transmission of some irrelevant information related to the minor-segment industries, leading to mispricing of the pure players. Our main hypothesis is that spurious cross-industry information diffusion (SCIID) occurs because cognitive resource-constrained investors find it costly to process information about pure players in the industry—choosing to focus on the industry leaders' returns instead.[2]

The following example illustrates the SCIID effect. Microsoft is a leader in the computer software industry (two-digit Standard Industrial Classification (SIC) code = 73), and sales from its major business segment (software) accounted for 91.5% of its total sales in 2002. However, Microsoft had other business segments, such as entertainment equipment (two-digit SIC code = 39), in which Microsoft produced set-top boxes (e.g., the Xbox) for users to watch and record TV programs and play video games. In 2002, 8.5% of Microsoft's total sales were generated from the entertainment equipment industry. One of the leaders in the entertainment equipment industry was Callaway Golf, a world leading golf equipment producer. Even though the computer software and entertainment equipment business of Microsoft were unrelated, because Microsoft had a business segment in the entertainment equipment industry, its stock returns were correlated with Callaway Golf's returns. Now consider American Software, a small company specializing in software for supply chain management. If the stock price of Microsoft goes up purely because of strong prospects in the entertainment equipment industry (e.g., U.S. consumers have more leisure time so they play more golf and video games), cognitive resource constrained investors may simply price American Software's stock based on Microsoft's stock price movement. This results in spurious information diffusion from the entertainment equipment industry to American Software—two completely unrelated lines of business.

In this example, Microsoft is considered the leader in the major-segment industry (henceforth, "major-segment industry leader" or simply "industry leader"); Callaway Golf is a leader in the minor-segment industry of the industry leader (henceforth, "minor-segment industry leader"); and American Software is a pure player in the major-segment industry (henceforth, "major-segment industry pure player" or simply "industry pure player"). Our main result shows a very strong contemporaneous relationship between minor-segment leaders' returns and major-segment pure player returns, that is, between Callaway Golf's returns and those of American Software, in the above example. Even though a contemporaneous relationship does not establish a direction of causality, we argue that the only plausible channel through which such a relationship could be observed is SCIID. Indeed, in regressions we also find evidence of a lead–lag relationship between minor-segment leader returns and major-segment pure player returns: even after controlling for contemporaneous returns of minor-segment industry leaders, lagged minor-segment leader returns have smaller but economically and statistically significant effects on major-segment pure player returns.[3] When we extend our analysis to a vector autoregressive (VAR) framework involving stock returns of minor-segment industry leaders, major-segment industry leaders, and major-segment industry pure players, we confirm that the direction of causality is only from minor-segment industry leader returns to major-segment pure player returns.

We carefully address a number of alternative explanations for our results. First, since our analysis is in terms of raw returns, we obviously need to address the extent to which common factors that affect both minor-segment and major-segment industry returns drive our results. We conduct "placebo tests" by replacing the minor-segment industry leaders with a portfolio of randomly selected industry leaders that do not share any industry segments with the major-segment industry leaders. In our main tests, we find that the economic magnitudes of the relationship observed between minor-segment industry leader returns and industry pure player returns exceed those in the placebo analysis in nearly 100% of our 500 trials, suggesting that common factors are unlikely to explain our main findings and that there is indeed an SCIID effect.

Second, it is possible that the major-segment industry leaders operate in their minor segments to exploit economic links, and these links are also relevant for the pure players in the major-segment industry. To examine whether economic links alone could explain our results, we repeat our analysis by excluding minor-segment industries that have customer or

[2] The psychology literature recognizes that attention is a scarce cognitive resource (Kahneman 1973). When faced with the task of processing large amounts of information, individuals can be inattentive to relevant information and use simple heuristics to make decisions. See also Hirshleifer and Teoh (2003), Hirshleifer et al. (2011), and Barber and Odean (2008) for other studies exploring the consequences of investor inattention.

[3] Since the SCIID hypothesis assumes that the "pass-through" from minor-segment industries to major-segment industry pure players occurs via the major-segment leader returns, we do not include the contemporaneous major-segment industry leader returns as an explanatory variable in our regressions. However, we do include lagged major-segment industry leader returns.

supplier relationships with the major-segment industry. Our results continue to hold.

Third, to provide more direct evidence that the cross-industry return relationships are not driven by economic links, we consider an exogenous change to the visibility of an industry leader that is unlikely to be related to a strengthening of economic links between its own industry and a particular minor-segment industry. The SCIID hypothesis assumes that industry leaders are the conduits of spurious information diffusion from the minor segments; this is because less sophisticated investors try to extract industry-relevant information (on the basis of which they trade industry pure players) from the returns of the industry leaders. We would therefore expect that the SCIID from the minor-segment industries of a multisegment firm to the pure players in its major-segment industry would become stronger if that firm is elevated to a special "leader" status that is likely to attract the attention of less sophisticated investors. Similar to Barberis et al. (2005), we examine the consequences of the addition of a stock to the Standard & Poor's (S&P) 500 Index. After a firm's stock is added to the S&P 500, the firm not only attracts more investor attention but also establishes itself as a new industry leader. Focusing on minor-segment industries that are unique to the firm that is newly added to the S&P 500, we find that the returns of the leaders in these minor-segment industries have a much stronger relationship with the returns of the pure players in the major-segment industry of this firm in the three years after its addition to the S&P 500, compared to the three years before.[4]

Fourth, we provide more direct evidence of spurious information diffusion around specific events relating to minor-segment industry leaders. We consider earnings announcement events of minor-segment industry leaders that generate major abnormal price movements and examine the major-segment pure player returns in the same week and subsequent weeks. We exclude confounding events in which there are earnings announcements by any of the major-segment industry leaders, leaders of other minor-segment industries, or leaders in customer or supplier industries in two weeks before or after the earnings event. We benchmark the industry pure player returns corresponding to these events against returns of pure players for the same events in industries unrelated to the minor-segment industry leaders or customer and supplier industries of the major- and minor-segment industries. The placebo analysis shows that the returns of the industry pure players are larger in magnitude than those of the pure players in unrelated industries in 100% of the trials, both in the event week and the two subsequent weeks.

This paper is organized as follows. Section 2 describes the data and sample construction. Section 3 develops our hypothesis and presents the empirical methodology. Section 4 provides the analysis of SCIID effect in all industries, discusses a number of robustness checks, and rules out alternative explanations for our results. Section 5 concludes the paper.

2. Data and Sample Selection

Our sample starts with all firms that are in the intersection of the Compustat Industry Segment File and the Center for Research in Security Prices (CRSP) Daily Security File. In our empirical tests, we exclude all firms with major business segments in financial industries[5] (two-digit SIC code 60–69) or whose major-segment industries cannot be classified (two-digit SIC code 99). Customer and supplier data are obtained from the Compustat Segment Customer File.[6] In the study of S&P 500 additions, the S&P 500 addition information is obtained from the Compustat Index Constituents File and Jeffrey Wurgler's personal website (http://people.stern.nyu.edu/jwurgler/). The main sample period is from January 1986 to December 2008.

2.1. Definition of Major-Segment Industry

Starting in 1977, the Financial Accounting Standards Board's Statement of Financial Accounting Standards No. 14 (SFAS No. 14) and the Securities and Exchange Commission's Regulation S-K required firms to report segment information.[7] Firms must report audited information for business segments whose sales, assets, or profits exceed 10% of consolidated totals. SFAS No. 14 defines an industry segment as "a component of an enterprise engaged in providing a product or service, or a group of related products or services primarily to unaffiliated customers for a profit." We

[4] To ensure that we are not capturing possible economic links between the new S&P addition's major industry and that of its minor segments, we only focus on cases where (a) the minor segment is "unique" in that no other leader in the industry has a minor segment in the same industry three years before and after the S&P addition event, and (b) there are no customer or supplier relationships between any firm in the industry and that of the unique minor segment three years before and after the event.

[5] We eliminate financial firms since they hold financial assets of firms from all industries. Therefore, the lead–lag relation between stock returns of financial firms and nonfinancial firms is mechanical and not driven as in our story based on investors' inattention.

[6] The segment customer and supplier data set is extended based on the one used in Banerjee et al. (2008), which has a detailed description of this data set.

[7] SFAS No. 14 was superseded by SFAS No. 131 in 1997. Under SFAS No. 131, firms are required to report more detailed segment information based on how management internally evaluates the operating performance of its business units. We find that our results hold for both sub-periods before and after the regulation change. The results are not tabulated but are available on request.

use the segment primary SIC code (i.e., SSIC1 in Compustat) to decide to which industry a business segment belongs. When the segment primary SIC code is missing, we replace it by the most recent nonmissing SIC code of the segment that shares the same segment name (i.e., SNAME in Compustat). We construct segment-industry-based portfolios (discussed in detail in §2.3) according to the two-digit segment primary SIC code.[8] Therefore, business segments defined by a different four-digit SIC code or three-digit SIC code may appear in the same two-digit SIC segment industry in our sample.

A firm may generate its sales in one or more than one two-digit SIC industries. Its "major-segment industry" is the two-digit SIC industry in which this firm generates more than 50% of its total sales. If a firm does not have any segment industry in which it generates more than 50% of its total sales, its major-segment industry is regarded as "unclassified," and hence this firm will not be included in our sample.

2.2. Industry Leaders and Industry Pure Players

In our empirical tests, we are interested in the information diffusion mechanism between two groups of firms: "industry pure players" and "industry leaders." We identify "industry leaders" as those firms whose major-segment industry sales rank in the top five among all firms in that industry.[9] "Industry pure players" are firms that are not industry leaders and generate all their sales from their major-segment industry.[10] Table 1 shows that the industry pure players are more likely to be small- and medium-sized firms with an average market capitalization around 671.56 million U.S. dollars. On the other hand, industry leaders are much larger firms. Their average market capitalization is 10,299.67 million U.S. dollars, more than 15 times larger than that of pure players. Table 1 also shows that 74.20% of firms in the entire sample are pure players according to our definition.

Besides their major-segment industry, many industry leaders have business segments in other two-digit SIC industries. We call these industries the "minor-segment industries" of industry leaders. Table 1 shows that, on average, each major-segment industry has 3.57 minor-segment industries. Of course, each minor-segment industry also has its own industry leaders according to our definition. These firms will be referred to as "minor-segment industry leaders" throughout this study.

2.3. Portfolio Formation

Daily stock trading data are retrieved from CRSP. Following Chordia and Swaminathan (2000) and Hou (2007), we compute weekly stock returns using Wednesday closing price (with dividend).[11] We employ weekly data to avoid bias due to nonsynchronous trading and other microstructure effects associated with daily data.

For each Wednesday, we construct equal-weighted portfolios of pure players and industry leaders for each two-digit SIC industry at the close price.[12] The status of pure players and industry leaders is obtained from the segment financial information disclosed at the most recent fiscal year end before the portfolio formation day. To make sure that the portfolios of pure players and industry leaders have reasonable diversification, we require that there be five industry leaders and at least five pure players in each two-digit SIC major-segment industry. Small industries that cannot satisfy this requirement are dropped from our sample.

The weights for the portfolio of minor-segment industry leaders are constructed as follows. For each industry, we find the sum of total sales made in

[8] We follow a two-digit SIC classification rather than three- or four-digit classifications to minimize overlap among major- and minor-segment industries. Using a three-digit (four-digit) classification leaves open the possibility of overlap between major and minor segments at the two-digit (two-digit and three-digit) level, thereby making it more likely that economic links between minor and major segments drive our results. To rule out the possibility of economic links driving our results, other than the conservative two-digit classification, we report additional tests below. Our results are also robust under the Fama and French (1997) 48-industry specification (refer to Table O2 in the online appendix, available at http://ssrn.com/abstract=964106).

[9] Our results are not sensitive to the number of industry leaders chosen in each two-digit SIC industry. In untabulated robustness checks, we also define an industry leader as one of the top three firms in terms of major segment sales in the industry or alternatively as one that is a constituent firm of the S&P 500 Index in a given year. The major results of this paper remain unchanged. Selecting fewer firms would put more emphasis on the role of industry leaders, and selecting more firms would improve the diversification of firm-specific factors in portfolio formation. Our choice of five industry leaders within each industry is a compromise between these two concerns.

[10] This specification is slightly different from the "single-segment firm" in the internal capital market literature (Lang and Stulz 1994, Berger and Ofek 1995, Lins and Servaes 1999). A single-segment firm is surely a pure player in our sample. However, a pure player may have many segments, though all of them must be in the same two-digit SIC industry.

[11] Boudoukh et al. (1994) suggest the autocorrelation of weekly stock returns based on Tuesday close are too low and Chordia and Swaminathan (2000) find the autocorrelation of weekly stock returns based on Friday close is too high. Therefore, using Wednesday close has become the convention in studying weekly return dynamics. Previous studies using weekly returns based on Wednesday close include Keim and Stambaugh (1984), Bessembinder and Hertzel (1993), Boudoukh et al. (1994), Chordia and Swaminathan (2000), and Hou (2007).

[12] The value-weighted portfolios of industry leaders are sometimes dominated by the largest firm. Therefore, we use equal weights in portfolio formation. Equal-weighted portfolios have also been adopted in Chordia and Swaminathan (2000) and Hou (2007). However, all our results hold with value-weighted portfolios qualitatively. Refer to Table O1 in the online appendix.

Table 1 Summary Statistics for Industry Leaders and Pure Players

	Industry leaders						Pure players				
Year	Market cap. ($M)	Total sales ($M)	Inst. holding	No. of analysts	No. of minor-segment industries	% sales from minor-segment industries	Market cap. ($M)	Total sales ($M)	Inst. holding	No. of analysts	% pure players among all firms
1986	2,966.81	5,797.20	40.36	13.90	4.04	12.73	177.29	176.30	15.21	2.37	70.07
1987	3,424.82	6,243.78	39.33	13.15	3.89	12.45	184.19	178.73	15.20	2.43	71.93
1988	3,119.07	7,106.51	39.98	13.32	3.98	11.44	172.74	206.05	16.34	2.51	71.96
1989	3,762.48	7,883.50	42.58	14.27	4.09	12.40	213.79	232.10	17.53	2.76	72.05
1990	3,781.51	8,689.30	41.18	13.86	3.78	11.24	221.49	249.89	18.38	2.89	71.75
1991	4,287.75	8,799.15	41.68	12.99	3.78	10.84	270.92	265.43	20.65	2.82	72.50
1992	4,875.02	9,242.42	43.88	13.21	3.87	10.60	320.87	289.18	22.97	2.90	73.73
1993	5,772.06	9,765.32	46.35	14.36	3.69	11.69	333.18	299.82	24.39	2.93	75.61
1994	5,876.67	10,495.71	46.81	14.13	3.47	10.72	332.16	320.94	25.55	3.06	77.00
1995	6,730.88	12,051.13	47.48	13.52	3.50	9.64	412.78	380.13	27.30	3.07	78.08
1996	8,204.53	12,392.91	47.67	13.55	3.63	9.97	490.47	414.10	28.10	3.13	79.84
1997	10,662.63	13,300.17	52.08	13.56	3.17	9.31	566.84	443.61	29.92	3.21	79.90
1998	13,683.02	14,335.12	53.32	13.19	3.37	10.24	671.69	449.93	29.92	3.34	73.93
1999	18,057.17	12,719.29	51.34	13.39	3.23	10.05	828.27	406.07	28.62	3.35	72.30
2000	17,049.53	14,423.52	52.62	12.35	3.69	9.58	1,166.18	470.58	28.89	3.27	72.47
2001	14,204.14	14,972.51	53.73	11.04	3.81	10.40	879.75	539.89	31.64	3.39	71.64
2002	12,135.89	16,123.36	53.53	9.71	3.90	10.59	743.82	585.17	34.19	3.33	71.53
2003	12,632.78	17,928.19	56.67	11.03	3.50	9.91	850.55	705.52	38.31	3.48	71.53
2004	14,215.32	20,771.27	63.58	11.71	3.24	11.45	1,088.66	826.57	39.94	3.86	74.37
2005	15,790.84	22,271.10	60.58	11.14	3.20	11.99	1,243.56	915.11	41.56	4.15	75.54
2006	16,587.56	24,011.57	60.97	10.94	3.26	12.42	1,403.17	998.66	42.55	4.22	76.88
2007	21,259.28	27,010.45	63.83	11.39	2.94	11.35	1,552.22	1,005.51	41.73	4.48	77.51
2008	17,812.61	28,841.92	61.34	10.80	3.13	10.93	1,321.31	1,079.15	43.78	4.49	74.58
All	10,299.67	14,138.06	50.47	12.63	3.57	10.95	671.56	497.32	28.81	3.28	74.20

Notes. Industry leaders in year T are firms that are among the top five firms in the industry in terms of sales in each two-digit SIC industry in year $T-1$. Pure players in year T are firms that were not industry leaders and generated all their sales from only one two-digit SIC industry in year $T-1$. Our sample excludes financial industries (two-digit SIC from 60 to 69), unclassified industries (two-digit SIC 99), and those that have fewer than five pure players. Our sample covers a period from January 1986 to December 2008. We report the annual means of market values, total sales, number of analysts covering the firm, and percentage of ownership held by institutional investors for industry leaders and pure players in this table. In addition, the average number of minor-segment industries for each two-digit SIC industry (major-segment industry), the average proportion of minor segments' contribution in industry leaders' total sales, and the proportion of pure players among all firms in our sample are also presented in this table.

each minor-segment industry by all industry leaders. The weight assigned to each minor segment in the portfolio of minor-segment industry leaders is this sum divided by the total sale by industry leaders in all minor segments. Finally, note that we could have recursive relationships of the following type: industry leader L_X of industry X has a minor segment in industry A, and industry leader L_A of industry A has a minor-segment in industry X. Since our objective is to see if spurious information diffusion occurs from minor-segment industries to major-segment industry pure players, we exclude these pairs of industry leaders in constructing major-segment or minor-segment industry leader portfolios.[13]

3. Hypothesis Development and Empirical Methodology

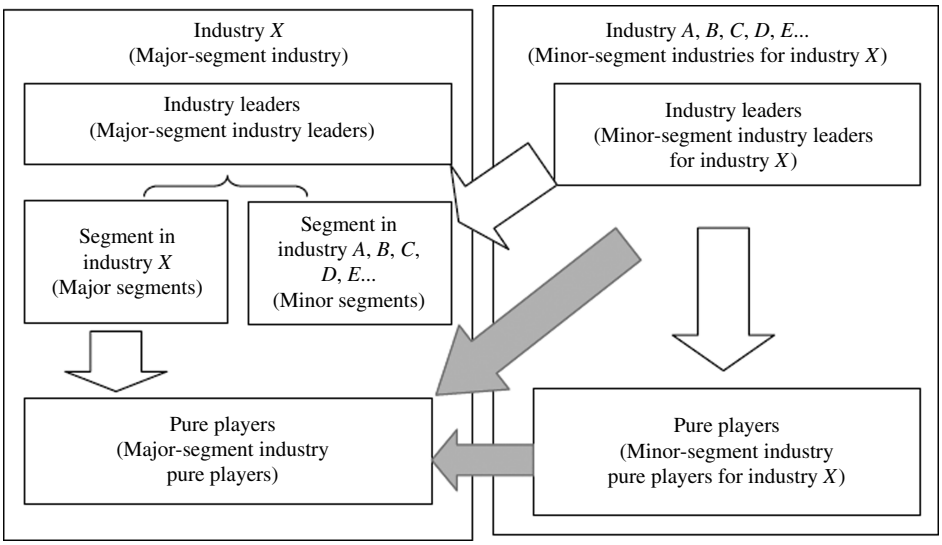
Figure 1 illustrates information diffusion among different groups of firms. We hypothesize that because of limited information processing ability, investors do not observe fundamental information about pure players in an industry—choosing instead to trade pure players based on the returns of industry leaders. However, their cognitive resource constraint also implies that they are unable to distinguish among the different return components of industry leaders and focus on overall return, which reflects information about the leaders' major-segment industry as well as the minor-segment industries. In Figure 1, industry X is the major-segment industry, and industries A, B, C, D and E are the minor-segment industries. Within industry X, we hypothesize the following relation in stock returns between major-segment industry leaders and major-segment industry pure players:

$$Pure\_ret_{t,i} = \sum_{K=1}^{a} \beta_K \, Pure\_ret_{t-K,i} + \sum_{K=0}^{b} \eta_K \, Leader\_ret_{t-K,i} + \varepsilon_{t,i}. \quad (1)$$

Equation (1) suggests that the portfolio returns of major-segment industry pure players for industry X ($Pure\_ret_{t,i}, i = X$) are not only affected by

[13] We provide a more detailed description of portfolio formation in the online appendix.

Figure 1 An Illustration of Information Diffusion Between Industry *X* and Its Minor-Segment Industries



Notes. Industry *X* is a particular two-digit SIC industry. The industry leaders of industry *X* are named major-segment industry leaders. Their segments in industry *X* are called major segments, and their segments in other industries, such as industries *A*, *B*, *C*, *D*, *E*, etc. are called minor segments. Firms that only have business in industry *X* are named major-segment industry pure players. Industry leaders of industries *A*, *B*, *C*, *D*, and *E* are called minor-segment industry leaders for industry *X* and their pure players minor-segment industry pure players for industry *X*. The white arrows show the direction of information diffusion based on fundamentals. Our paper aims to establish the information diffusion due to cognitive biases, which are represented by the grey arrows.

its own lagged return ($Pure\_ret_{t-K,i}, i = X$), but also by the contemporaneous ($Leader\_ret_{t,i}, i = X$) and lagged ($Leader\_ret_{t-K,i}, i = X$) returns of their industry leaders. If attention-constrained investors price industry pure players based on industry leaders' returns, then we expect a relationship as in Equation (1).

Although these industry leaders may be difficult for attention-constrained investors to understand, they are closely watched by analysts and other sophisticated market participants. Therefore, their prices will incorporate fundamental information relatively quickly (Hou 2007). Because leaders of industry *X* have minor segments from industry *A*, *B*, *C*, *D*, and *E*, we expect information from these industries will also be incorporated into stock prices of leaders of industry *X*. We can express this relationship as follows:

$$Leader\_ret_{t,i} = \sum_{K=1}^{b} \rho_K \, Leader\_ret_{t-K,i} + \sum_{K=0}^{c} \upsilon_K \, Mleader\_ret_{t-K,i} + \varepsilon'_{t,i} \quad (2)$$

where $Mleader\_ret_{t-K,i}$ ($i = X$) are the sales weighted portfolio returns of leaders in the minor-segment industries (*A*, *B*, *C*, *D*, and *E* in our example) and proxy for fundamental information in these industries. Equation (2) implies at least a contemporaneous relationship between the industry leader returns and the minor-segment industry leader returns because fundamental shocks in the minor segments where they have business operations are relevant for the major-segment industry leaders; however, if the market is slow to incorporate this information, we would also expect some lead–lag effect. Below, we further discuss why, even though the major-segment industry leaders are expected to be better scrutinized by market analysts, information diffusion from the minor-segment industry leaders to the major-segment industry leaders could be slow, leading to a lead–lag effect.

Substituting Equation (2) into Equation (1) above, we obtain the following:

$$Pure\_ret_{t,i} = \sum_{K=1}^{a} \beta_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{b} \gamma_K \times Leader\_ret_{t-K,i} + \sum_{K=0}^{c} \varphi_K \times Mleader\_ret_{t-K,i} + \varepsilon^*_{t,i}, \quad (3)$$

where

$$\gamma_K = \eta_0 \rho_K + \eta_K, \quad \varphi_K = \eta_0 \upsilon_K, \quad \text{and} \quad \varepsilon^*_{t,i} = \eta_0 \varepsilon'_{t,i} + \varepsilon_{t,i}.$$

Equation (3) shows that there is a contemporaneous and possibly a lead–lag relation in stock returns between firms of industries *A*, *B*, *C*, *D*, and *E* and pure players of industry *X*, even though pure players in industry *X* are fundamentally unrelated to these other industries.

To test our hypothesis, we estimate two-lag and four-lag autoregressive models corresponding to Equation (3):

$$Pure\_ret_{t,i} = \alpha_i + \sum_{K=1}^{k} \beta_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \gamma_K \times Leader\_ret_{t-K,i} + \varphi_0 Mleader\_ret_{t,i} + \sum_{K=1}^{k} \varphi_K \times Mleader\_ret_{t-K,i} + \varepsilon_{t,i}. \quad (4)$$

We recognize that a contemporaneous relationship between minor-segment industry returns and major-segment industry pure player returns does not establish causality. However, other than common factors affecting both sets of returns (an issue we address in our empirical design), SCIID seems to be the only plausible channel through which these returns could be related. In our setup, it is difficult to argue that information diffusion from the major-segment industry to firms in minor-segment industries can occur: this is so because we exclude minor-segment industry leaders that also have minor segments in the major-segment industry. Thus, any contemporaneous relationship could only be attributed to spurious information diffusion from the minor-segment industry leaders to the major-segment industry standalones via the major-segment industry leaders.

It is possible, however, that cross-industry information from the minor-segment industry leaders to the major-segment industry leaders diffuses slowly. In this case, if the SCIID hypothesis is valid, even after controlling for lagged industry leader returns, lagged minor-segment leader returns would affect industry pure player returns. However, there should not be a corresponding effect in the reverse direction; that is, industry pure player returns should not lead minor-segment leader returns. The appeal of the VAR setup is that, absent prior knowledge about the economic structure, as long as information diffusion is not entirely contemporaneous, it is a well-recognized means of establishing causality and the direction of information diffusion (see, e.g., Brennan et al. 1993, Chordia and Swaminathan 2000, Hou 2007).

Thus, it is crucial to understand why information diffusion from the minor-segment industry leaders to the major-segment leaders might be slow enough for us to identify a lead–lag relationship.[14] Our hypothesis is that even though the industry leaders are large firms and attract more analysts specialized in covering firms in their major-segment industry than the smaller standalone firms, they do not have a comparative advantage in attracting analysts specialized in covering their *minor*-segment industries. Thus, the speed of information diffusion from the minor-segment industry leaders to these major-segment industry leaders is likely to be no faster than that to minor-segment industry standalones, after controlling for segment size.

In Table A.1 of the appendix, we show in the first two columns that the minor segments of the major-segment industry leaders are much smaller than the (major segments of) minor-segment industry leaders. On average, they are a quarter of the size (measured in terms of sales and book value of assets) of the minor-segment leaders' major segments and thus comparable to smaller firms in the minor-segment industries. Column (3) compares the numbers of minor-segment industry-specialized analysts that follow the major-segment industry leaders and the minor-segment industry leaders. We define the specialization of financial analysts by their industry coverage; that is, a financial analyst is specialized in one industry in year t if she or he covers at least three firms whose major business segments belong to this industry in year t. Column (3) shows that far fewer minor-segment industry-specialized analysts follow the major-segment industry leaders than the corresponding minor-segment industry leaders.

The above results suggest that even though the minor segments of the industry leaders are parts of large firms, they are more comparable with smaller firms in the minor-segment industries in terms of size and industry-specialized analyst coverage. Hou (2007) finds evidence of intraindustry information diffusion from large firms to small firms, resulting in a lead–lag relationship between their stock returns. Thus, that the minor segments of the major-segment leaders are more comparable in terms of size and industry-specialized analyst coverage to the smaller firms in the minor-segment industries is a possible rationale for a lead–lag return relationship between minor-segment industry leaders and major-segment industry leaders.

In Table A.2 of the appendix, we examine the effect of segment size and segment/firm status on industry-specialized analyst coverage more thoroughly. Our estimation is carried out at the segment level. The dependent variable is the natural logarithm of one plus the number of financial analysts who specialize in the industry that a segment belongs to. Column (1) shows that segment size (log(*Sales*)) is an important determinant of specialized analyst coverage, consistent with extant literature documenting a

[14] We thank an anonymous referee for encouraging us to examine this issue in some detail.

positive relationship between size and analyst coverage. Column (2) shows that, controlling for size, a segment gets more specialized analyst coverage if it is a major segment of an industry leader; however, it gets less specialized coverage if it is a minor segment of an industry leader. Finally, column (3) shows that the degree of undercoverage of minor segments of industry leaders relative to segments of other firms in the same industry increases in segment size. As pointed out by Brennan et al. (1993), stock returns of firms that are followed by many analysts tend to lead those of firms that are followed by fewer analysts even when these firms are of similar firm size. Therefore, our results in Table A.2 offer another explanation of why information from minor-segment industry leaders diffuses slowly to major-segment industry leaders.

We follow Brennan et al. (1993), Chordia and Swaminathan (2000), and Hou (2007) and use vector autoregressions to examine the lead–lag relationships among different portfolios.[15] The VAR system comprises the following equations:

$$\begin{aligned} Pure\_ret_{t,i} = \alpha_i &+ \sum_{K=1}^{k} \beta_K \times Pure\_ret_{t-K,i} \\ &+ \sum_{K=1}^{k} \gamma_K \times Leader\_ret_{t-K,i} \\ &+ \sum_{K=1}^{k} \varphi_K \times Mleader\_ret_{t-K,i} + \varepsilon_{t,1}, \end{aligned} \quad (5)$$

$$\begin{aligned} Leader\_ret_{t,i} = \kappa_i &+ \sum_{K=1}^{k} \delta_K \times Pure\_ret_{t-K,i} \\ &+ \sum_{K=1}^{k} \lambda_K \times Leader\_ret_{t-K,i} \\ &+ \sum_{K=1}^{k} \theta_K \times Mleader\_ret_{t-K,i} + \varepsilon_{t,2}, \end{aligned} \quad (6)$$

$$\begin{aligned} Mleader\_ret_{t,i} = \mu_i &+ \sum_{K=1}^{k} \omega_K \times Pure\_ret_{t-K,i} \\ &+ \sum_{K=1}^{k} \xi_K \times Leader\_ret_{t-K,i} \\ &+ \sum_{K=1}^{k} \chi_K \times Mleader\_ret_{t-K,i} + \varepsilon_{t,3}. \end{aligned} \quad (7)$$

This framework allows us to directly test whether the minor segment information is fully incorporated in leader returns within the same period or with a lag. If the information transmission is within the same period, because we control for lagged industry leader returns, the sum of the coefficients associated with *Mleader_ret* in Equation (6), $\Sigma\theta_k$, will be insignificantly different from zero. However, if the information diffuses slowly, we expect the sum of the coefficients associated with *Mleader_ret* in Equation (5), $\Sigma\varphi_k$, to be significantly positive. To investigate whether the information diffusion is predominately from minor-segment industries to pure players in the major-segment industry, rather than in the other direction, we first test whether or not $\Sigma\omega_k$ in Equation (7) is significantly positive. If $\Sigma\omega_k$ is significantly positive, we can then conduct the cross-equation test as in Chordia and Swaminathan (2000) and Hou (2007) and examine whether or not $\Sigma\varphi_k$ is greater than $\Sigma\omega_k$.

There are two important issues that we need to address to interpret our results. First, when we do the analyses on raw returns without adjustment for common factors that are likely to affect both minor-segment and major-segment returns, we need to benchmark our coefficient estimates against the null of no spurious information diffusion from the minor-segment industry portfolios to major-segment pure players. To allow the comparison, we conduct the following placebo tests. For each two-digit industry X, we find all other two-digit industries such that none of the top five leaders in these industries have a business segment in X or any of the actual minor-segment industries of the leaders of X. We call these other two-digit industries "irrelevant industries." We then randomly select one of these irrelevant industries to replace each actual minor-segment industry and construct an equally weighted irrelevant industry portfolio. We do this replacement exercise for each industry and then repeat the entire exercise 500 times to obtain the distribution of the coefficient estimates based on irrelevant industry portfolio returns. Finally, we compare the magnitude of the coefficient estimate of the actual minor-segment portfolio returns to a specific percentile value of this distribution.

Second, our estimates are based on a pooled sample that incorporates both time-series and cross-sectional variations. We also estimate our models for each two-digit SIC major-segment industry and report the average coefficient estimates and the corresponding t-statistics across all major-segment industries. Each approach has its advantages. In the former approach, the large number of observations in the panel sample allows us to eliminate noise in the estimations. The latter approach helps us avoid the clustered standard error problem in the panel data and gives us a credible distribution of coefficient estimates across all industries. Consistent with Hou (2007), our results from these alternative approaches are very similar throughout.

Finally, in tests not reported in tables, we replicate our analyses on the basis of residual returns derived from asset pricing models. Again, because we cannot eliminate the possibility of unmodeled common factors, we repeat the placebo analysis as in the case of raw returns. Our results are very similar.

[15] A detailed discussion of the VAR method is provided by Chordia and Swaminathan (2000).

4. Results

4.1. Return Spillover Effects from Minor-Segment Industry Leaders

In panel A of Table 2, we report results corresponding to our estimation of Equation (4). Panel B presents the results of the two-lag and four-lag vector autoregressions. Results in panel A show that contemporaneous minor-segment leader returns have a strong positive effect on industry pure player returns—a 1% higher minor-segment leader portfolio return translates to an almost 0.5% higher industry pure player portfolio return within the same week. However, common factors undoubtedly contribute significantly to the magnitude of this effect. Thus, we postpone an interpretation of the economic significance of the SCIID effect until the next subsection. Lagged industry leader returns also have a significant positive effect on industry pure player returns. We find that lagged minor-segment leader returns have a significant positive effect on industry pure player returns, suggesting that economically relevant information from minor-segment industry leaders to major-segment industry leaders is not fully incorporated within the same week, which causes the former to lead industry pure player returns.

We get further corroboration from the VAR results reported in panel B of Table 2. Minor-segment leader returns lead industry pure player returns both for the four-lag and two-lag specifications, after controlling

Table 2 Regressions with Minor-Segment Industry Leaders' Portfolio

Panel A: Autoregression with contemporaneous *Mleader_ret*

	Four-lag regression						
	RHS variables				Placebo test		
LHS variable	*Pure_ret* $(t-1{:}t-4)$	*Leader_ret* $(t-1{:}t-4)$	*Mleader_ret* (t)	*Mleader_ret* $(t-1{:}t-4)$		*Irr_Mleader_ret* (t)	*Irr_Mleader_ret* $(t-1{:}t-4)$
Pure_ret (t)	0.064	0.117	0.483	0.108	*Pure_ret* (t)	0.424	0.032
F-test/T-test	55.38***	160.32***	121.88***	111.85***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	0.0%
Mean–by industry	0.095	0.088	0.523	0.088	*Mean–by industry*	0.423	0.055
T-test	4.51***	7.06***	19.03***	5.19***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	8.0%
	Two-lag regression						
	RHS variables				Placebo test		
LHS variable	*Pure_ret* $(t-1{:}t-2)$	*Leader_ret* $(t-1{:}t-2)$	*Mleader_ret* (t)	*Mleader_ret* $(t-1{:}t-2)$		*Irr_Mleader_ret* (t)	*Irr_Mleader_ret* $(t-1{:}t-2)$
Pure_ret (t)	0.036	0.076	0.485	0.092	*Pure_ret* (t)	0.426	0.038
F-test/T-test	34.77***	139.75***	122.36***	171.90***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	0.0%
Mean–by industry	0.078	0.052	0.523	0.060	*Mean–by industry*	0.423	0.036
T-test	4.36***	4.50***	19.20***	5.87***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	2.0%

Panel B: Vector autoregressions

	Four-lag regressions				
	RHS variables			Placebo test	
LHS variables	*Pure_ret* $(t-1{:}t-4)$	*Leader_ret* $(t-1{:}t-4)$	*Mleader_ret* $(t-1{:}t-4)$		*Irr_Mleader_ret* $(t-1{:}t-4)$
Pure_ret (t)	0.067	0.132	0.092	*Pure_ret* (t)	0.050
F-test	49.37***	164.31***	65.25***	$\% > \Sigma\varphi$	0.0%
Mean–by industry	0.096	0.094	0.092	*Mean–by industry*	0.038
T-test	4.05***	6.54***	4.21***	$\% > \Sigma\varphi$	0.2%
Leader_ret (t)	−0.006	0.031	0.034	*Leader_ret* (t)	0.013
F-test	0.38	7.94***	7.92***	$\% > \Sigma\theta$	7.6%
Mean–by industry	0.017	−0.006	0.063	*Mean–by industry*	0.011
T-test	0.86	−0.42	3.02***	$\% > \Sigma\theta$	0.8%
Mleader_ret (t)	0.008	0.012	−0.032	*Irr_Mleader_ret* (t)	0.034
F-test	0.71	4.24***	9.27***	$\% > \Sigma\chi$	97.4%
Mean–by industry	0.011	0.016	−0.001	*Mean–by industry*	0.017
T-test	0.60	1.23	−0.05	$\% > \Sigma\chi$	80.4%
$F1$: $\Sigma\varphi = \Sigma\omega$		34.17***			

Table 2 (Continued)

Panel B: Vector autoregressions

	Two-lag regressions				
	RHS variables			Placebo test	
LHS variables	*Pure_ret* ($t-1$:$t-2$)	*Leader_ret* ($t-1$:$t-2$)	*Mleader_ret* ($t-1$:$t-2$)		*Irr_Mleader_ret* ($t-1$:$t-2$)
Pure_ret (t)	0.044	0.094	0.055	*Pure_ret*	0.033
F-test	39.65***	169.45***	48.85***	$\% > \Sigma\varphi$	0.2%
Mean–by industry	0.076	0.063	0.052	*Mean–by industry*	0.025
T-test	3.97***	5.81***	3.75***	$\% > \Sigma\varphi$	1.20%
Leader_ret (t)	0.004	−0.001	0.026	*Leader_ret* (t)	0.015
F-test	0.26	0.05	10.11***	$\% > \Sigma\theta$	17.6%
Mean–by industry	0.024	−0.017	0.048	*Mean–by industry*	0.017
T-test	1.66	−1.49	3.41***	$\% > \Sigma\theta$	1.20%
Mleader_ret (t)	0.005	0.014	−0.007	*Irr_Mleader_ret* (t)	−0.010
F-test	0.25	0.83	0.09	$\% > \Sigma\chi$	56.4%
Mean–by industry	0.018	0.015	−0.023	*Mean–by industry*	−0.009
T-test	1.14	1.66	−1.32	$\% > \Sigma\chi$	82.8%
$F1$: $\Sigma\varphi = \Sigma\omega$		15.65***			

Notes. Industry leaders in year T are firms that were among the top five in terms of industry sales in each two-digit SIC industry in year $T-1$. *Leader_ret* is the return of an equal-weighted portfolio of all industry leaders in each two-digit SIC industry. Pure players in year T are firms that were not industry leaders and generated all their sales from only one two-digit SIC industry in year $T-1$. *Pure_ret* is the return of an equal-weighted portfolio of all pure players in a two-digit SIC industry. Leaders of each two-digit SIC industry, apart from their primary business segment, may also have minor segments in other two-digit SIC industries. We first add up the minor segments sales by all industry leaders in each minor-segment industry each year and compute a sales index showing the total sales by all major-segment industry leaders to each minor segment industry. *MLeader_ret* is the sales-index-weighted average of *Leader_ret* of minor-segment industries. In order to avoid recursive relationship between major-segment industry leaders and minor-segment industry leaders, we exclude those minor-segment industries whose industry leaders also have business segments in the major-segment industry. Our sample excludes financial industries (two-digit SIC from 60 to 69), unclassified industries (two-digit SIC 99), and those that have less than five pure players. All portfolio returns are computed weekly. In panel A, the following four-lag and two-lag autoregressions with the contemporaneous *MLeader_ret* are estimated using weekly returns from January 1986 to December 2008.

$$Pure\_ret_{t,i} = \alpha_i + \sum_{K=1}^{k} \beta_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \gamma_K \times Leader\_ret_{t-K,i} + \varphi_0 \times Mleader\_ret_{t,i} + \sum_{K=1}^{k} \varphi_K \times Mleader\_ret_{t-K,i} + \varepsilon_t. \quad \text{(T.2.1)}$$

In panel B, the following four-lag and two-lag vector autoregressions are estimated using weekly returns from January 1986 to December 2008.

$$Pure\_ret_{t,i} = \alpha_i + \sum_{K=1}^{k} \beta_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \gamma_K \times Leader\_ret_{t-K,i} + \sum_{K=1}^{k} \varphi_K \times Mleader\_ret_{t-K,i} + \varepsilon_{t,1}, \quad \text{(T.2.2)}$$

$$Leader\_ret_{t,i} = \kappa_i + \sum_{K=1}^{k} \delta_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \lambda_K \times Leader\_ret_{t-K,i} + \sum_{K=1}^{k} \theta_K \times Mleader\_ret_{t-K,i} + \varepsilon_{t,2}, \quad \text{(T.2.3)}$$

$$Mleader\_ret_{t,i} = \mu_i + \sum_{K=1}^{k} \omega_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \xi_K \times Leader\_ret_{t-K,i} + \sum_{K=1}^{k} \chi_K \times Mleader\_ret_{t-K,i} + \varepsilon_{t,3}. \quad \text{(T.2.4)}$$

In both panels, the first and the second rows report results for the pooled sample. The estimated sum of coefficients for lagged returns and the corresponding F-statistic, and the estimated coefficient for the contemporaneous *MLeader_ret* (t) and its t-statistic, are reported. The third and fourth rows report the average estimated coefficients and the corresponding t-statistics across all two-digit SIC industries. In panel B, *Pure_ret* ($t-1$:$t-k$), $k=2$ or 4, reports $\Sigma\beta$ from Equation (T.2.2), $\Sigma\delta$ from Equation (T.2.3), or $\Sigma\omega$ from Equation (T.2.4), depending on the LHS variables. Similarly, *Leader_ret* ($t-1$:$t-k$) reports $\Sigma\gamma$ from Equation (T.2.2), $\Sigma\lambda$ from Equation (T.2.3), or $\Sigma\xi$ from Equation (T.2.4) and *MLeader_ret* ($t-1$:$t-k$) reports $\Sigma\varphi$ from Equation (T.2.2), $\Sigma\theta$ from Equation (T.2.3), or $\Sigma\chi$ from Equation (T.2.4). $F1$ refers to the F-statistics for the null hypothesis: $\Sigma\varphi = \Sigma\omega$ in the pooled vector autoregressions. In the placebo test, each minor-segment industry is replaced by a randomly selected irrelevant industry, that is, an industry such that the major-segment industry leaders do not have minor business segments in these irrelevant industries. *Irr_Mleader_ret*, is the equal-weighted portfolio of $Leader\_ret_t$ of all randomly selected irrelevant industries. *MLeader_ret* in (T.2.1–T.2.4) are replaced by *Irr_Mleader_ret*. Each simulation test is repeated 500 times. The means of $\Sigma\varphi$, φ_0, $\Sigma\theta$, and $\Sigma\chi$ from 500 simulated vector regressions are reported in the table. We also report the proportions of simulated coefficients for *Irr_Mleader_ret* that are larger than those for *MLeader_ret* in (T.2.1–T.2.4). RHS (LHS) means right- (left-) hand side.

*, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

for lagged industry leader returns. This is possible under the SCIID only if minor-segment leader returns lead industry leader returns. Panel B indicates that the sum of coefficients $\Sigma\theta_k$ in Equation (6) is significantly positive in the four-lag and two-lag regressions, for both the panel VAR estimates as well as the by-industry estimates. In contrast, the evidence of a bi-directional effect, that industry leaders lead the minor-segment industry leaders, is much weaker. The sum of the coefficients $\sum\zeta_k$ in Equation (7) is insignificant in three of the four estimates reported in panel B of Table 2. Finally, the lead–lag relationship between the minor-segment industry leaders and the industry pure players is asymmetric: the lagged terms of *Mleader_ret* have a strong effect on current *Pure_ret*, but there is an insignificant effect in the reverse direction. In the four-lag VAR estimation, the sum of the four lagged terms ($\Sigma\varphi_k$) is 0.09 (significant at the 1% level), and in the two-lag regressions, it is 0.05 (significant at the 1% level), for both pooled and by-industry estimates. On the other hand, the lagged terms of *Pure_ret* have a negligible effect on *Mleader_ret* in both two-lag ($\Sigma\omega_k = 0.005$, F-statistics $= 0.025$) and four-lag ($\Sigma\omega_k = 0.008$, F-statistics $= 0.71$) VAR estimations. The null hypothesis that $\Sigma\varphi_k = \Sigma\omega_k$ can be easily rejected based on an F-statistics of 34.17 (15.65) in the four-lag (two-lag) VAR estimation. Overall, the results in panel B show that leaders of the minor-segment industries lead pure players of the major-segment industry but not conversely. This confirms our conjecture that information diffuses from minor-segment industry leaders to major-segment industry pure players and not the other way around.

4.2. Placebo Tests: Are Our Results Due to (Observable or Unobservable) Common Factors?

Because all portfolio returns in our specifications contain some common factors, the return spillover from the minor-segment industries to the industry pure players may reflect the effect of these common factors. We therefore conduct additional placebo tests by constructing portfolios of irrelevant industry leaders. These tests are more general than controlling for known common asset pricing factors because they incorporate all possible common factors.[16] Furthermore, this method can partially address the "joint hypothesis" problem raised by Fama (1970). If our results are due to a misspecified asset pricing model, we should observe similar estimates with irrelevant portfolio returns under the same model.

For each two-digit SIC industry X (the major-segment industry), we find all other two-digit SIC industries such that none of the top five leaders from that industry has a business segment in X or any minor-segment industry of X. We call these industries the irrelevant industries. We then randomly select one irrelevant industry to replace each minor-segment industry. We equally weight the *Leader_ret* of irrelevant industries to construct a new portfolio and we call the return of this portfolio *Irr_Mleader_ret*. *MLeader_ret* in Equations (4)–(7) is replaced by *Irr_Mleader_ret* and coefficients are estimated in the four-lag and two-lag specifications. We repeat this replacement test 500 times and obtain 500 sets of coefficients for the *Irr_Mleader_ret*. Means of 500 sets of estimated coefficients are reported in panels A and B of Table 2.

The placebo test gives us an empirically generated distribution of the coefficients for *Irr_Mleader_ret*. The null hypothesis that common factors alone accounting for the relationship is rejected in favor of the SCIID hypothesis at, say, the 5% level of significance if the coefficient estimate for *MLeader_ret* in Table 2 exceeds the 95th percentile of the *Irr_Mleader_ret* coefficient distribution. We see from panel A that φ_0 estimated from Equation (4) exceeds the estimate for *Irr_Mleader_ret* in 100% of the trials for both the pooled and by-industry estimates, for both the two- and four-lag specifications. The sum of the coefficients of lagged *Mleader_ret* also exceed those of lagged *Irr_Mleader_ret* in 100% of the trials for the pooled estimates and in all but 8% (2%) of the trials for the by-industry estimates for the four-lag (two-lag) specifications. Results in panel B are very similar.[17]

The magnitudes of the coefficients of *Irr_Mleader_ret* also allow us to gauge the economic significance of the SCIID effect. In panel A of Table 2, the coefficient φ_0 is about 14% (24%) higher than the corresponding average coefficient from the placebo tests for the pooled (by-industry) regressions. This is economically quite a substantial effect. The sum of the coefficients of the lagged *Mleader_ret* are two to three times larger in magnitude than those of *Irr_Mleader_ret*. When we add the coefficients of the contemporaneous and lagged terms, in panel A of Table 2, the part attributable to the SCIID effect is 28% of the effect due to common factors for both the pooled and by-industry regressions.

[16] In tests not reported in the tables, we replicate our results based on abnormal returns from the market model. Placebo tests (to account for the presence of other common factors) generate results very similar to the ones discussed here.

[17] The results also show that the evidence from our VAR analysis that the minor-segment leader returns lead the industry leader returns is not spurious. The sum of the coefficients $\Sigma\theta_k$ in Equation (6) exceed those when *Mleader_ret* is replaced by *Irr_Mleader_ret* in 100% of the cases for the within-industry estimates and in all but 7.6% (17.2%) of the cases for the pooled estimates corresponding to the two-lag (four-lag) specifications.

4.3. Customer and Supplier Industry Portfolios

An important concern with our previous analysis is that the minor-segment industries may include industries that have vertical relationships with a majority of the industry pure players. This cannot be ruled out because the industry leaders may have chosen to operate in these segments to exploit fundamental economic links between these industries. In that case, the information diffusion from minor-segment industries to major-segment industry pure players is essentially caused by economic links as suggested by Menzly and Ozbas (2006, 2010) and Cohen and Frazzini (2008). In this section, we explicitly control for customer and supplier industry effects and show that the information diffusion from minor-segment industries to major-segment industry pure players remains significant.

We retrieve the information on principal customers for all listed U.S. firms from the Compustat Segment Customer File for our sample period. Prior to 1997, SFAS No. 14 required firms to disclose the names and industries of their principal customers if the revenue generated from the sales to a particular customer exceeded 10% of the firm's total revenue. In 1997, SFAS No. 131 replaced SFAS No. 14 and redefined business segments.[18]

We hypothesize that if the return relationships between minor-segment industry leaders and major-segment pure players are mainly driven by the overlap of customer (supplier) industries for a majority of industry pure players and minor-segment industries of industry leaders, such effect should disappear after we screen out all customer and supplier industries from the minor-segment industries. We first search the Compustat Segment Customer File and find all customer and supplier firms for major-segment industry firms, including leaders and pure players. We name the major-segment industries of these customer (supplier) firms the "customer industries" ("supplier industries"). Second, we eliminate the customer industries and the supplier industries from the minor-segment industries and irrelevant industries. Finally, we reconstruct the portfolio for the minor-segment industry leaders based on the remaining minor-segment industries. The results of the autoregressive and VAR specifications based on these remaining minor-segment industries are reported in Table 3.

After we control for the vertical economic links, the magnitude of φ_0 (the coefficient of contemporaneous *Mleader_ret*) is lower than what we observe in panel A of Table 2. However, the placebo analysis in Table 3[19] shows that these coefficients are still larger than what could be obtained if *Mleader_ret* is replaced with *Irr_Mleader_ret* in 100% of the cases. The placebo analysis indicates that the sum of the coefficients of lagged *Mleader_ret* ($\Sigma\varphi_k$) is larger than that for the coefficients of the irrelevant portfolio returns (*Irr_Mleader_ret*) in at least 95% of the cases. In terms of economic magnitudes, the sum of the contemporaneous and lagged coefficients of *Mleader_ret* is 30% (35%) higher than that of *Irr_Mleader_ret* for the pooled (by-industry) regressions. Therefore, the return relationships between minor-segment industry leaders and major-segment industry pure players cannot be accounted for by vertical economic relations alone.

4.4. A Study Based on S&P 500 Index Additions

The crux of the SCIID hypothesis is the idea that attention-constrained investors follow more visible firms in an industry to glean information about less visible firms in the same industry. However, since they are unable to disentangle industry-relevant information from industry-irrelevant information originating in the minor segments of the visible firms, spurious information diffusion occurs.

Industry leaders (in terms of sales or market capitalization) are the largest firms in the industry and therefore most likely to get attention from investors. In unreported results, we find that consistent with the SCIID hypothesis, the spillover from minor-segment leader returns to industry pure player returns is stronger when the leaders are of larger size. We also find that the spillover is stronger if the minor segments contribute more to the leaders' sales, confirming the idea that industry leaders are the conduits of spurious information diffusion.

In the previous section, we raised and addressed the possibility that economic links between minor-segment industries and major-segment industry pure players could explain the return relationship between minor-segment industry leaders and major-segment industry pure players. Our results survive after explicitly controlling for customer–supplier links. In this section, we address this issue in a different way. In particular, motivated by Vijh (1994) and Barberis et al. (2005), we argue that inclusion in the S&P Index elevates the visibility of a firm (even one that is already an industry leader according to our criteria) and attracts more investor attention. Consequently, according to the SCIID hypothesis, if such a firm is the only leader in the industry to have a minor segment in a particular other industry, there should be stronger spurious information diffusion from that unique minor-segment industry to pure players in the major segment of the new S&P addition after the event (i.e., the S&P 500 addition) than before. It is easy to distinguish the SCIID hypothesis from

[18] Our results are robust to truncating the sample period at 1997.

[19] In this placebo analysis, the customer and supplier industries are also eliminated when we pick the irrelevant industries.

Table 3 Regressions That Exclude Minor Segment Industries with Economic Links with the Major-Segment Industry

Panel A: Autoregression with contemporaneous *Mleader_ret*

	Four-lag regression						
	RHS variables				Placebo test		
LHS variable	*Pure_ret* $(t-1{:}t-4)$	*Leader_ret* $(t-1{:}t-4)$	*Mleader_ret* (t)	*Mleader_ret* $(t-1{:}t-4)$		*Irr_Mleader_ret* (t)	*Irr_Mleader_ret* $(t-1{:}t-4)$
Pure_ret (t)	0.060	0.124	0.447	0.104	*Pure_ret* (t)	0.405	0.019
F-test/T-test	47.07***	177.24***	114.71***	110.18***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	0.0%
Mean–by industry	0.102	0.086	0.499	0.089	*Mean–by industry*	0.402	0.035
T-test	4.67***	6.83***	19.20***	5.33***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	0.4%
	Two-lag regression						
	RHS variables				Placebo test		
LHS variable	*Pure_ret* $(t-1{:}t-2)$	*Leader_ret* $(t-1{:}t-2)$	*Mleader_ret* (t)	*Mleader_ret* $(t-1{:}t-2)$		*Irr_Mleader_ret* (t)	*Irr_Mleader_ret* $(t-1{:}t-2)$
Pure_ret (t)	0.035	0.081	0.449	0.089	*Pure_ret* (t)	0.407	0.028
F-test/T-test	30.89***	156.16***	115.22***	171.44***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	0.0%
Mean–by industry	0.085	0.050	0.499	0.060	*Mean–by industry*	0.401	0.028
T-test	4.59***	4.00***	19.40***	5.77***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	4.4%

Panel B: Vector autoregressions

	Four-lag regressions				
	RHS variables			Placebo test	
LHS variables	*Pure_ret* $(t-1{:}t-4)$	*Leader_ret* $(t-1{:}t-4)$	*Mleader_ret* $(t-1{:}t-4)$		*Irr_Mleader_ret* $(t-1{:}t-4)$
Pure_ret (t)	0.062	0.138	0.084	*Pure_ret* (t)	0.046
F-test	36.90***	162.06***	53.10***	$\% > \Sigma\varphi$	0.2%
Mean–by industry	0.097	0.089	0.080	*Mean–by industry*	0.018
T-test	4.18***	6.13***	5.10***	$\% > \Sigma\varphi$	0.2%
$F1$: $\Sigma\varphi = \Sigma\omega$		21.20***			
	Two-lag regressions				
	RHS variables			Placebo test	
LHS variables	*Pure_ret* $(t-1{:}t-2)$	*Leader_ret* $(t-1{:}t-2)$	*Mleader_ret* $(t-1{:}t-2)$		*Irr_Mleader_ret* $(t-1{:}t-2)$
Pure_ret (t)	0.040	0.098	0.049	*Pure_ret* (t)	0.023
F-test	29.58***	166.68***	25.48***	$\% > \Sigma\varphi$	1.4%
Mean–by industry	0.080	0.061	0.042	*Mean–by industry*	0.018
T-test	4.15***	5.21***	4.05***	$\% > \Sigma\varphi$	0.0%
$F1$: $\Sigma\varphi = \Sigma\omega$		7.93***			

Notes. The results reported in this table are based on a test specification similar to that used in Table 2. The improvement is that we eliminate all customer industries and supplier industries (of both major-segment leaders and pure players) from the minor-segment industries and the irrelevant industries. The information on principal customers and dependent suppliers of major-segment industry firms is retrieved from the Compustat Business Segment File. Coefficients and corresponding *F*- and *t*-statistics for Equations (T.2.1) and (T.2.2) are reported in this table. The sample period is from January 1986 to December 2008.

***Denotes statistical significance at the 1% level.

the alternative explanation of "economic link" on the basis of this test because the unique minor-segment industry already existed prior to the addition of the leader to the S&P 500 Index, so it does not represent a new economic link. In order to be even more conservative and ensure that we are not picking up any economic links, we impose the following additional filters: (i) there are no customer or supplier relationships between a firm in the unique minor-segment industry and any firm in the major-segment industry in the three years prior to the S&P addition event, and (ii) none of the existing top five leaders (including incumbent S&P Index members) in the major-segment industry adds a segment in the industry of

Table 4 S&P 500 Additions and Spurious Cross-Industry Information Diffusion

Panel A: Autoregression with contemporaneous *Uleader_ret*

(−156 W, 0 W)

Four-lag regression

	RHS variables				Placebo test		
LHS variable	*Pure_ret* $(t-1{:}t-4)$	*Leader_ret* $(t-1{:}t-4)$	*Uleader_ret* (t)	*Uleader_ret* $(t-1{:}t-4)$		*Irr_Uleader_ret* (t)	*Irr_Uleader_ret* $(t-1{:}t-4)$
Pure_ret (t)	0.237	0.096	0.324	0.025	*Pure_ret* (t)	0.299	0.032
F-test/T-test	45.52***	6.82***	28.49***	0.99	$\% > \varphi_0(\Sigma\varphi)$	28.2%	56.2%
Mean–by event	0.118	0.123	0.400	0.070	*Mean–by event*	0.360	0.089
T-test	2.84***	3.06***	12.21***	1.51	$\% > \varphi_0(\Sigma\varphi)$	18.4%	64.8%

Two-lag regression

	RHS variables				Placebo test		
LHS variable	*Pure_ret* $(t-1{:}t-2)$	*Leader_ret* $(t-1{:}t-2)$	*Uleader_ret* (t)	*Uleader_ret* $(t-1{:}t-2)$		*Irr_Uleader_ret* (t)	*Irr_Uleader_ret* $(t-1{:}t-2)$
Pure_ret (t)	0.189	0.078	0.322	0.021	*Pure_ret* (t)	0.296	0.033
F-test/T-test	56.19***	10.52***	28.75***	0.35	$\% > \varphi_0(\Sigma\varphi)$	28.0%	69.6%
Mean–by event	0.097	0.081	0.403	0.059	*Mean–by event*	0.369	0.053
T-test	3.22***	3.02***	12.50***	1.82*	$\% > \varphi_0(\Sigma\varphi)$	17.4%	42.6%

(0 W, 156 W)

Four-lag regression

	RHS variables				Placebo test		
LHS variable	*Pure_ret* $(t-1{:}t-4)$	*Leader_ret* $(t-1{:}t-4)$	*Uleader_ret* (t)	*Uleader_ret* $(t-1{:}t-4)$		*Irr_Uleader_ret* (t)	*Irr_Uleader_ret* $(t-1{:}t-4)$
Pure_ret (t)	0.136	0.135	0.420	0.055	*Pure_ret* (t)	0.315	0.026
F-test/T-test	20.01***	19.46***	44.13***	5.09**	$\% > \varphi_0(\Sigma\varphi)$	0.0%	14.8%
Mean–by event	0.049	0.127	0.439	0.197	*Mean–by event*	0.333	0.117
T-test	1.26	2.96***	11.81***	4.30***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	1.6%

Two-lag regression

	RHS variables				Placebo test		
LHS variable	*Pure_ret* $(t-1{:}t-2)$	*Leader_ret* $(t-1{:}t-2)$	*Uleader_ret* (t)	*Uleader_ret* $(t-1{:}t-2)$		*Irr_Uleader_ret* (t)	*Irr_Uleader_ret* $(t-1{:}t-2)$
Pure_ret (t)	0.055	0.105	0.420	0.050	*Pure_ret* (t)	0.314	0.024
F-test/T-test	6.54**	26.78***	44.33**	8.80***	$\% > \varphi_0(\Sigma\varphi)$	0.2%	10.0%
Mean–by event	0.050	0.053	0.439	0.135	*Mean–by event*	0.333	0.082
T-test	2.23**	1.99**	11.81***	5.67***	$\% > \varphi_0(\Sigma\varphi)$	0.0%	2.6%

this unique minor-segment in the three years after the event (note that to define the unique minor segment, we already require that no other leader in the industry has a segment in the industry of this unique minor segment three years prior to the S&P addition event).

With these filters, we conduct tests similar to those in Equations (4)–(7).[20] *Mleader_ret* is replaced by *Uleader_ret*, the return of an equally weighted portfolio of industry leaders in the unique minor segment of the new S&P addition. For the placebo analysis, we randomly pick an industry in which no industry leader has a segment. We perform our tests over two windows around the S&P addition event:

[20] In principle, we can also carry out similar tests for S&P 500 deletions if the deletions are not driven by takeovers. If the deleted leader is acquired, the acquirer becomes a new leader, and all minor-segment industries of the deleted leader automatically become minor-segment industries of the new leader. Unfortunately, of the 453 observations with available information of deletion events (nonmissing confounding events code (CECODE) in the data set provided on Jeffrey Wurgler's website), 322 (71%) are associated with mergers and acquisitions (CECODE = 0). The size of the remaining sample does not allow us to find a sufficient number of deletions where the deleted leaders have unique segment industries.

Table 4 (Continued)

Panel B: Vector autoregressions					
	(−156 W, 0 W)				
	Four-lag regressions				
	RHS variables			Placebo test	
LHS variables	*Pure_ret* ($t-1{:}t-4$)	*Leader_ret* ($t-1{:}t-4$)	*Uleader_ret* ($t-1{:}t-4$)		*Irr_Uleader_ret* ($t-1{:}t-4$)
Pure_ret (t)	0.204	0.113	0.039	*Pure_ret* (t)	0.030
F-test	27.24***	7.73***	1.93	% > $\Sigma\varphi$	37.4%
Mean–by event	0.082	0.127	0.042	*Mean–by event*	0.061
T-test	1.68*	2.59**	1.32	% > $\Sigma\varphi$	63.8%
*F*1: $\Sigma\varphi = \Sigma\omega$		0.65			
	Two-lag regressions				
	RHS variables			Placebo test	
LHS variables	*Pure_ret* ($t-1{:}t-2$)	*Leader_ret* ($t-1{:}t-2$)	*Uleader_ret* ($t-1{:}t-2$)		*Irr_Uleader_ret* ($t-1{:}t-2$)
Pure_ret (t)	0.170	0.085	0.025	*Pure_ret* (t)	0.031
F-test	36.81***	10.38***	1.54	% > $\Sigma\varphi$	61.4%
Mean–by event	0.072	0.096	0.046	*Mean–by event*	0.041
T-test	2.08**	2.51**	1.21	% > $\Sigma\varphi$	43.8%
*F*1: $\Sigma\varphi = \Sigma\omega$		0.34			
	(0 W, 156 W)				
	Four-lag regressions				
	RHS variables			Placebo test	
LHS variables	*Pure_ret* ($t-1{:}t-4$)	*Leader_ret* ($t-1{:}t-4$)	*Uleader_ret* ($t-1{:}t-4$)		*Irr_Uleader_ret* ($t-1{:}t-4$)
Pure_ret (t)	0.117	0.162	0.075	*Pure_ret* (t)	0.034
F-test	10.61***	20.20***	7.12***	% > $\Sigma\varphi$	7.6%
Mean–by event	0.083	0.085	0.148	*Mean–by event*	0.061
T-test	1.76*	2.05**	2.60**	% > $\Sigma\varphi$	6.6%
*F*1: $\Sigma\varphi = \Sigma\omega$		5.18**			

from 156 weeks before the event week to the event week, and from the event week to 156 weeks after the event week. The results are in Table 4.

The results show that for both the two- and four-lag specifications, *Uleader_ret* has a higher contemporaneous effect on industry pure player returns in the post-event period than in the pre-event period (panel A). The pooled regression estimates show an almost 30% higher coefficient for the post-event period, and the difference in the coefficient for *Uleader_ret* (t) (as well as the sum of coefficients for *Uleader_ret*($t-1{:}t-4$)) between two periods is statistically significant at the 5% level based on the Chow test.[21] Although the difference is more modest (about 10%) for the by-event estimate, the sum of the contemporaneous and lagged effects is 35% higher in the post-event periods. The placebo analysis reveals that the pre-event period coefficients of the contemporaneous returns from the irrelevant leader portfolios exceeds the ones in the actual data between 17.4% to 28.2% of the time, so we cannot reject the null hypothesis of no SCIID in the pre-event period at conventional levels of significance. In contrast, for the post-event period, the contemporaneous coefficient in the actual data exceeds that of the placebo analysis in almost 100% of the trials. Conclusions from the VAR analysis in panel B are similar.[22] Overall, these results provide strong

[21] For example, the null hypothesis $\varphi_{0,\text{pre-event}} = \varphi_{0,\text{post-event}}$ (i.e., coefficient for *Uleader_ret* (t)) is rejected by an F-statistic of 14.97 ($p < 0.001$) and the null hypothesis $\Sigma\varphi_{K,\text{pre-event}} = \Sigma\varphi_{K,\text{post-event}}$ (i.e., sum of coefficients for *Uleader_ret*($t-1{:}t-4$)) is rejected by an F-statistic of 5.67 ($p < 0.02$).

[22] The VAR results show that the lead–lag effects are stronger after the addition of the new leader to the S&P 500 Index. This is consistent with the new leaders' new role as a conduit of spurious information spillover from its unique minor segment after the addition, which was absent before.

Table 4 (Continued)

	Panel B: Vector autoregressions				
	(0 W, 156 W)				
	Two-lag regressions				
	RHS variables			Placebo test	
LHS variables	*Pure_ret* ($t-1$:$t-2$)	*Leader_ret* ($t-1$:$t-2$)	*Uleader_ret* ($t-1$:$t-2$)		*Irr_Uleader_ret* ($t-1$:$t-2$)
Pure_ret (t)	0.037	0.133	0.063	*Pure_ret* (t)	0.030
F-test	2.11	31.38***	10.09***	% $> \Sigma\varphi$	6.2%
Mean–by event	0.055	0.077	0.115	*Mean–by event*	0.046
T-test	1.59	2.48**	2.78***	% $> \Sigma\varphi$	6.2%
*F*1: $\Sigma\varphi = \Sigma\omega$		3.77**			

Notes. Unlike the previous tests, industry leaders reported in this table are defined according to whether or not a firm is a constituent of the S&P 500 Index. *Leader_ret* is the return of an equal-weighted portfolio of all industry leaders in each two-digit SIC industry. Pure players in year T are firms that were not industry leaders and generated all their sales from only one two-digit SIC industry in year $T-1$. *Pure_ret* is the return of an equal-weighted portfolio of pure players in one two-digit SIC industry. We define an industry as a unique-segment industry if a leader newly added to the S&P 500 has a minor-segment industry in that industry that is neither a minor-segment industry of the incumbent leaders (including both S&P 500 constituent firms and firms with top five industry sales) nor a customer or supplier industry of the major-segment industry in previous three years and subsequent three years. *Uleader_ret* is the *Leader_ret* of a unique-segment industry of the recently added S&P 500 constituent firm. The sample period is from January 1986 to December 2008 and all returns are computed weekly. The following four-lag and two-lag autoregressions (T.4.1) and vector autoregressions (T.4.2–T.4.4) are estimated across all S&P 500 addition events with unique segment industries using weekly returns from January 1986 to December 2008 in two subperiods: (−156 W, 0 W) represents the three-year (156 weeks) period before new S&P constituent firms with unique segment industries are added, and (0 W, 156 W) is the three-year (156 weeks) period after the addition.

$$Pure\_ret_{t,i} = \alpha_i + \sum_{K=1}^{k} \beta_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \gamma_K \times Leader\_ret_{t-K,i} + \varphi_0 \times ULeader\_ret_{t,i} + \sum_{K=1}^{k} \varphi_K \times ULeader\_ret_{t-K,i} + \varepsilon_{t,1}, \quad \text{(T.4.1)}$$

$$Pure\_ret_{t,i} = \alpha_i + \sum_{K=1}^{k} \beta_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \gamma_K \times Leader\_ret_{t-K,i} + \sum_{K=1}^{k} \varphi_K \times ULeader\_ret + \varepsilon_{t,1}, \quad \text{(T.4.2)}$$

$$Leader\_ret_{t,i} = \kappa_i + \sum_{K=1}^{k} \delta_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \lambda_K \times Leader\_ret_{t-K,i} + \sum_{K=1}^{k} \theta_K \times ULeader\_ret + \varepsilon_{t,2}, \quad \text{(T.4.3)}$$

$$ULeader\_ret_{t,i} = \mu_i + \sum_{K=1}^{k} \omega_K \times Pure\_ret_{t-K,i} + \sum_{K=1}^{k} \xi_K \times Leader\_ret_{t-K,i} + \sum_{K=1}^{k} \chi_K \times ULeader\_ret + \varepsilon_{t,3}. \quad \text{(T.4.4)}$$

Coefficients and corresponding *F*- and *t*-statistics for Equations (T.4.1) and (T.4.2) are reported in panel A and panel B, respectively. In both panels, the first and the second rows report results for the pooled sample. The estimated sum of coefficients for lagged returns and the corresponding *F*-statistic, and the estimated coefficient for the contemporaneous *Uleader_ret* (t) and its *t*-statistic are reported. The third and fourth rows report the average estimated coefficients and the corresponding *t*-statistics across all addition events. *F*1 refers to the *F*-statistics for the hypothesis: $\Sigma\varphi = \Sigma\omega$ for the pooled vector autoregressions. In the placebo test, each unique-segment industry is replaced by a randomly selected irrelevant industry; that is, the major-segment industry leaders do not have minor business segments in these irrelevant industries. *Irr_Uleader_ret* is the *Leader_ret* of a randomly selected irrelevant industry. Tests above are repeated after *Uleader_ret* in (T.4.1) and (T.4.2–T.4.4) are replaced by *Irr_Uleader_ret*. Each simulation test is repeated 500 times. The means of coefficients from 500 stimulations are reported in the table. We also report the proportions of simulated coefficients for *Irr_Uleader_ret* that are larger than those for *Uleader_ret* in (T.4.1) and (T.4.2).

*, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

evidence that the return spillover is on account of SCIID and cannot be attributed to economic links.

4.5. Earnings Announcements of Minor-Segment Industry Leaders and SCIID

The results based on vector autoregressions and autoregressions discussed above strongly support our main hypothesis in the full sample. However, regression specifications have one typical weakness: we cannot precisely identify the source of spurious information diffusion under this design. Therefore, in this section, we complement our previous tests by examining the SCIID effect based on earnings shock events in the minor-segment industries.

Although this test is clear and intuitive, there exists one important concern: because earnings announcements cluster over time, if there are contemporaneous earnings shocks from minor-segment industry

Table 5 Spurious Information Diffusion Around Earnings Shocks

Panel A: Market-model-based abnormal returns of minor-segment industry leaders' portfolio at week t (%)		
	Negative shocks	Positive shocks
Mean	−8.301	8.794
Std. dev.	2.550	3.774

Panel B: Raw returns of major-segment industry pure players' portfolio (%)						
Week	*Major_pure_ret*	*Irr_pure_ret*	% > *Major_pure_ret*	*Major_pure_ret*	*Irr_pure_ret*	% > *Major_pure_ret*
t	−0.629	0.018	100.0	1.322	0.946	0.0
$t+1$	−0.328	0.072	100.0	1.035	0.684	0.0
$t+2$	−0.190	0.142	100.0	0.969	0.693	0.0
$t+3$	−0.160	0.163	100.0	0.601	0.564	28.0
$t+4$	0.001	0.151	78.2	0.643	0.515	18.0
N		290			342	

Notes. A positive (negative) earnings shock to a minor-segment industry is defined when at least one minor-segment industry leader makes an earnings announcement in week t and the market-model weekly abnormal return of *Mleader_ret* is higher than 5% (lower than −5%). N is the number of earnings shocks identified. When we identify a positive or negative earnings shock in a minor-segment industry in week t, we require that there are no other earnings shocks to industry leaders in the major-segment industries, customer industries, or supplier industries of the major-segment industries or other minor-segment industries of the major-segment industries in a period between week $t-2$ to week $t+2$. Panel A reports the means and standard deviations of abnormal returns of minor-segment industry leaders' portfolios during the week of the earnings shock. In panel B, *Major_pure_ret* is the weekly return of a sales-index-weighted portfolio of *Pure_ret* for all major-segment industries that are associated with this minor-segment industry. We compute *Major_pure_ret* from week t to week $t+4$. In the placebo test, we replace all major-segment industries associated with a specific minor-segment industry by randomly picked irrelevant industries. Irrelevant industries are those two-digit SIC industries that are not minor-segment industries, customer industries, or supplier industries for either the major-segment or minor-segment industries that we investigate. For each round of random replacement, we compute *Irr_pure_ret*, the equal-weighted portfolio of *Pure_ret* for irrelevant industries for the same period as in the "true" test above. We repeat the procedure of replacing major-segment industries with randomly picked irrelevant industries 500 times. The mean levels of *Irr_pure_ret* for the investigation period are reported. We also provide the percentage of trials for which the *Irr_pure_ret* that is larger than *Major_pure_ret* for both positive and negative earnings shocks. The sample period is from January 1986 to December 2008 and all returns are computed weekly.

leaders and major-segment industry leaders, one cannot really distinguish whether the spillover effect to major-segment industry pure players is rational or spurious. Therefore, in this test, we require that there must be no earnings shocks in major-segment industries, customer industries, and supplier industries between week $t-2$ and week $t+2$ around an earnings shock in a minor-segment industry in week t. Not surprisingly, since most Compustat firms have December fiscal year ends and their earnings announcements are highly clustered, this screening procedure significantly reduces the sample size.

We define earnings shocks in terms of the market-model-based abnormal *Mleader_ret* in week t when at least one minor-segment industry leader makes an earnings announcement. In Table 5, we report results based on a 5% cutoff; that is, when the abnormal return is higher than 5% (lower than −5%) in week t, we define this event as a positive (negative) earnings shock. In untabulated robustness checks, we have tried different cutoff levels and the pattern in our results remains unchanged.

As with placebo tests in previous results, we address the concern of potential missing common factors by using the returns of irrelevant industry pure players as the benchmark in detecting the SCIID effect. More specifically, if the earnings shocks from minor-segment industries contain any market-wide risk factors or are correlated with any unobserved common factors that may also affect returns of major-segment industry firms, these factors should also affect returns of irrelevant industry firms. Therefore, the SCIID effect can be isolated by computing the return differentials between the major-segment pure players and the irrelevant industry pure players. An irrelevant industry corresponding to a major-segment industry is another industry that is not a minor-segment industry, a customer industry, or a supplier industry for the major-segment industry or its minor-segment industries.

In an event study setting, the results in Table 5 confirm our previous finding that a strong contemporaneous and lead–lag effect exists between the returns of the minor-segment industry leaders and major-segment industry pure players. In a period between week t and week $t+2$, the major-segment industry pure players outperform (underperform) all irrelevant industry pure players in all 500 replacements when there is a positive (negative) earnings shock in the minor-segment industry. Furthermore, this test provides an estimate of the magnitude of the spurious information diffusion, which is reflected

in the return differentials between the major-segment industry pure players and the irrelevant industry pure players. For example, for positive earnings shocks, the contemporaneous return differential between major-segment industry pure players and irrelevant industry pure players is 0.38 percentage points, and it is 0.79 percentage points in the following four weeks. The cumulative effect of 1.17 percentage points is economically important compared to the 8.79% average weekly abnormal return of the minor-segment industry leaders for positive earnings shocks defined by the 5% cutoff.[23]

5. Conclusion

We find that information from unrelated industries spuriously diffuses to the pure players in an industry via the industry leaders. Both the contemporaneous and lagged returns of firms in the minor-segment industries significantly affect the returns of pure players in the major-segment industry. VAR analysis shows that the returns of leaders in the minor-segment industries affect the returns of pure players in the major-segment industries more than in the reverse direction. This result survives a number of robustness checks.

We find that when a firm is added to the S&P 500 Index and becomes a more prominent industry leader, the effect of the new S&P 500 firm's unique minor segments on the pure players of its major-segment industry becomes much stronger in the three-year period after the addition than before, consistent with the idea that investor attention on the leaders of the major-segment industries drives the return spillover. We also look at earnings announcement events of minor-segment industry leaders that are associated with large abnormal stock price movements, and we find spillover effects on the returns of pure players of the major-segment industries. These effects are significantly larger in magnitude than those on pure players of industries that are unrelated to the minor-segment industry leaders.

Our results are consistent with a particular type of bias experienced by attention-constrained investors that affects pure player returns. Investors respond to the limited attention problem by categorization, and they focus on industry leaders in order to extract information about an industry category (Peng and Xiong 2006).[24] However, since the industry leaders themselves are conglomerates, these investors are unable to distinguish the effects of shocks that originate in the minor segments of the leaders from those that originate in the major-segment industry of the leaders. As a result, the major-segment industry pure players are mispriced since their valuations are affected by the information from the minor-segment industries that are fundamentally unrelated to the major-segment industry. In subsequent research, we plan to investigate whether spurious information diffusion has real consequences: for example, whether investment by pure players in the major-segment industry is sensitive to returns in minor-segment industries, and whether such effects are stronger when major-segment industry leaders are more visible or when the minor segments contribute more to the leaders' sales.

Acknowledgments

The authors thank John Griffin, Harrison Hong, Michael Lemmon, Juhani Linnainmaa, Mark Seasholes, Sheridan Titman, Siew Hong Teoh, Walter Torous, and especially Wei Xiong (the department editor) and two anonymous referees, whose comments greatly improved the paper. For useful comments, the authors thank seminar participants at the Hong Kong University of Science and Technology; University of New South Wales; University of Sydney; University of Technology, Sydney; 2007 Asia Finance Association–Financial Management Association Conference at Hong Kong; 2007 China International Conference in Finance at Chengdu; 2007 Cornell–Shanghai University of Economics and Finance Conference at Shanghai; 2009 Western Finance Association Conference at San Diego; and 2011 Financial Intermediation Research Society Conference. The authors are grateful to Jeffrey Wurgler for making available the data on addition and deletion of S&P 500 constituents on his personal website.

[23] Even though the industry pure players are temporarily mispriced, one expects that the opportunity for arbitrage and the trading behavior of informed market participants would lead to correction. Identifying return reversal, however, is a challenging exercise since industry pure player returns are likely to become noisier as one moves further away from an initial long-short hedge portfolio formation period. Nonetheless, for the larger industry pure players for whom the limits to arbitrage are likely to be less severe, we document return reversal three weeks after portfolio formation period. We present this result in Table O6 in the online appendix.

[24] Peng and Xiong (2006) show that attention-constrained investors can cause investors to process sector and market-wide information at the expense of firm-specific information, leading to a type of category-learning behavior. Combined with investor overconfidence (another behavioral trait with considerable support from the psychology literature), such behavior can explain many features of observed return comovement. Hou et al. (2009) show that investor inattention causes stock prices to underreact to earnings news, while investor attention results in overreaction-driven price momentum.

Appendix

Table A.1 Differences Between Minor Segments of Major-Segment Industry Leaders and Majors Segment of Minor-Segment Industry Leaders

Year	(1) Ratio (sales)	(2) Ratio (assets)	(3) Ratio (no. of analysts)
1986	0.238	0.222	0.292
1987	0.242	0.238	0.246
1988	0.289	0.286	0.261
1989	0.278	0.260	0.289
1990	0.271	0.251	0.260
1991	0.258	0.269	0.251
1992	0.263	0.294	0.226
1993	0.300	0.316	0.237
1994	0.275	0.253	0.260
1995	0.253	0.249	0.231
1996	0.232	0.235	0.248
1997	0.218	0.235	0.261
1998	0.247	0.251	0.271
1999	0.286	0.234	0.277
2000	0.230	0.216	0.223
2001	0.248	0.213	0.187
2002	0.210	0.192	0.125
2003	0.216	0.203	0.189
2004	0.188	0.191	0.150
2005	0.216	0.187	0.183
2006	0.271	0.242	0.197
2007	0.256	0.245	0.204
2008	0.239	0.251	0.222
All	0.249	0.241	0.230

Notes. This table shows comparisons of sales, book value of assets, and minor-segment-industry-specialized analyst coverage between minor segments of major-segment industry leaders and major segments of industry leaders in these minor-segment industries. In column (1), we report the average ratio of segment sales, defined as the sales of a minor segment of a major-segment industry leader scaled by the average sales of the major segments of the corresponding minor-segment industry leaders. In columns (2) and (3), we report similar ratios of the segment book value of assets and the number of financial analysts specialized in the minor-segment industries. We define the specialization of financial analysts by their industry coverage; that is, a financial analyst is specialized in one industry in year t if she or he covers at least three firms whose major business segments belong to this industry in year t.

Table A.2 Industry Specialization of Financial Analysts, Segment Size, and Industry Leader Status

	(1)	(2)	(3)
Log(*Sales*)	0.199***	0.201***	0.204***
	(41.45)	(55.82)	(68.14)
	[23 | 0]	[23 | 0]	[23 | 0]
Major segment of industry leader		0.112***	0.249***
		(3.87)	(7.18)
		[13 | 0]	[16 | 0]
Log(*Sales*) ∗ *Major segment of industry leader*			0.040
			(1.39)
			[8 | 9]
Minor segment of industry leader		−0.358***	−0.336***
		(−10.02)	(−11.37)
		[0 | 22]	[0 | 22]
Log(*Sales*) ∗ *Minor segment of industry leader*			−0.056***
			(−4.91)
			[0 | 10]
Average adjusted R-squared	0.359	0.365	0.368
Industry fixed effect in each cross section		Yes	
Number of years		23	

Notes. This table reports the results of Fama and MacBeth (1973) regressions based on segment-level data. The dependent variable is the natural logarithm of one plus the number of financial analysts who specialize in the industry of the segment in question and cover the firm that has this segment. We define the specialization of financial analysts by their industry coverage; that is, a financial analyst is specialized in an industry in year t if she or he covers at least three firms whose major business segments belong to this industry in year t. Independent variables are defined as follows: Log(*Sales*) is the natural logarithm of segment total sales (in billions); *Major segment of industry leader* is a dummy variable that equals 1 if this segment is the major segment of an industry leader and 0 otherwise; and *Minor segment of industry leader* is a dummy variable that equals 1 if this segment is a minor segment of an industry leader (i.e., the major segment of this industry leader belongs to another industry) and 0 otherwise. In the first step of this two-pass test, we include the segment industry fixed effects in our cross-sectional regressions and the standard errors of coefficients are also clustered at the segment industry level. The average coefficients in time series (i.e., across 23 years from 1986 to 2008) are reported in the table. Their corresponding Fama–MacBeth t-statistics are reported in parentheses. In addition, we also report in square brackets the number of years where the coefficients in cross-sectional regressions are statistically significant at 5% level; that is, $[X \mid Y]$ indicates that we have observed positive (negative) and statistically significant coefficients in X (Y) years.

*** Denotes statistical significance at the 1% level.

References

Banerjee S, Dasgupta S, Kim Y (2008) Buyer–supplier relationships and the stakeholder theory of capital structure. *J. Finance* 63(5):2507–2552.

Barber B, Odean T (2008) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Rev. Financial Stud.* 21(2):785–818.

Barberis N, Shleifer A, Wurgler J (2005) Comovement. *J. Financial Econom.* 75(2):283–317.

Berger PG, Ofek E (1995) Diversification's effect on firm value. *J. Financial Econom.* 37(1):39–65.

Bessembinder H, Hertzel MG (1993) Return autocorrelations around nontrading days. *Rev. Financial Stud.* 6(1):155–189.

Boudoukh J, Richardson M, Whitelaw RF (1994) A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *Rev. Financial Stud.* 7(3):539–573.

Brennan MJ, Jegadeesh N, Swaminathan B (1993) Investment analysis and the adjustment of stock prices to common information. *Rev. Financial Stud.* 6(4):799–824.

Chordia T, Swaminathan B (2000) Trading volume and cross-autocorrelations in stock returns. *J. Finance* 55(2):913–935.

Cohen L, Frazzini A (2008) Economic links and predictable returns. *J. Finance* 63(4):1977–2011.

Fama EF (1970) Efficient capital markets: A review of theory and empirical work. *J. Finance* 25(2):383–417.

Fama EF, French KR (1997) Industry cost of equity. *J. Financial Econom.* 43(2):153–193.

Fama EF, MacBeth JD (1973) Risk, return, and equilibrium: Empirical tests. *J. Political Econom.* 81(3):607–636.

Hirshleifer D, Teoh SH (2003) Limited attention, financial reporting and disclosure. *J. Accounting Econom.* 36(1–3):337–386.

Hirshleifer D, Lim SS, Teoh SH (2011) Limited investor attention and stock market misreactions to accounting information. *Rev. Asset Pricing Stud.* 1(1):35–73.

Hou K (2007) Industry information diffusion and the lead–lag effect in stock returns. *Rev. Financial Stud.* 20(4):1113–1138.

Hou K, Peng L, Xiong W (2009) A tale of two anomalies: The implications of investor attention for price and earnings momentum. Working paper, Ohio State University, Columbus.

Kahneman D (1973) *Attention and Effort* (Prentice Hall, Englewood Cliffs, NJ).

Keim DB, Stambaugh RF (1984) A further investigation of the weekend effect in stock returns. *J. Finance* 39(3):819–834.

Lang LHP, Stulz RM (1994) Tobin's *Q*, corporate diversification and firm performance. *J. Political Econom.* 102(6):1248–1280.

Lins K, Servaes H (1999) International evidence on the value of corporate diversification. *J. Finance* 54(6):2215–2239.

Lo AW, MacKinlay AC (1990) When are contrarian profits due to stock market overreaction? *Rev. Financial Stud.* 3(2): 175–205.

Menzly L, Ozbas O (2006) Cross-industry momentum. Working paper, University of Southern California, Los Angeles.

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

Peng L, Xiong W (2006) Investor attention, overconfidence and category learning. *J. Financial Econom.* 80(3):563–602.

Vijh AM (1994) S&P 500 trading strategies and stock betas. *Rev. Financial Stud.* 7(1):215–251.