

Schlag, Christian; Zeng, Kailin

**Working Paper**

## Horizontal industry relationships and return predictability

SAFE Working Paper, No. 256

**Provided in Cooperation with:**

Leibniz Institute for Financial Research SAFE

*Suggested Citation:* Schlag, Christian; Zeng, Kailin (2019) : Horizontal industry relationships and return predictability, SAFE Working Paper, No. 256, Goethe University Frankfurt, SAFE - Sustainable Architecture for Finance in Europe, Frankfurt a. M., <https://doi.org/10.2139/ssrn.3436006>

This Version is available at:

<https://hdl.handle.net/10419/201835>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Christian Schlag – Kailin Zeng

# Horizontal Industry Relationships and Return Predictability

SAFE Working Paper No. 256

**SAFE | Sustainable Architecture for Finance in Europe**

A cooperation of the Center for Financial Studies and Goethe University Frankfurt

House of Finance | Goethe University  
Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main

Tel. +49 69 798 30080 | Fax +49 69 798 33910  
info@safe-frankfurt.de | www.safe-frankfurt.de

# Horizontal Industry Relationships and Return Predictability

Christian Schlag\*

Kailin Zeng<sup>†</sup>

This version: August 9, 2019

---

\* Goethe University Frankfurt and Research Center SAFE, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt/Main, Germany. E-mail: [schlag@finance.uni-frankfurt.de](mailto:schlag@finance.uni-frankfurt.de)

<sup>†</sup> Corresponding author. Jiangxi University of Science and Technology, Hongqi Avenue 86, 341000 Ganzhou Jiangxi, China; Ganzhou Academy of Financial Research, Kejia Avenue 156, 341000 Ganzhou Jiangxi, China. E-mail: [kailinzeng@gmail.com](mailto:kailinzeng@gmail.com).

We gratefully acknowledge research and financial support from the Research Center SAFE, funded by the State of Hessen initiative for research LOEWE. We would furthermore like to thank the participants at the Annual Meetings of the Swiss Finance Association for comments. Special thanks go to Fabian Hollstein.

# Horizontal Industry Relationships and Return Predictability

## Abstract

It has been documented that vertical customer-supplier links between industries are the basis for strong cross-sectional stock return predictability (Menzly and Ozbas (2010)). We show that robust predictability also arises from horizontal links between industries, i.e., from the fact that industries are competitors or offer products, which are substitutes for each other. These horizontally linked industries exhibit positively correlated fundamentals. The signal derived from this type of connectedness is the basis for significant alpha in sorted portfolio strategies, and informed investors take the related information into account when they form their portfolios. We thus provide evidence of return predictability based on a new type of economic links between industries not captured in previous studies.

**Keywords:** Connected industries, information flow, return predictability

**JEL:** G12, E44, D81

# 1 Introduction

In a theoretical model Hong et al. (2007) show that the information contained in industry returns diffuses gradually across markets, which forms the basis for return predictability. More fundamentally, Hirshleifer and Teoh (2003) and Peng and Xiong (2006) study the effects of limited attention on price dynamics and demonstrate cross-sectional implications for return predictability. Recently, Cohen et al. (2018) have documented the slow incorporation of information about changes in firms' reporting practices, which the authors attribute to a lack of attention on the part of investors.

In this paper we investigate the flow of information between related industries over time and the question whether this transmission of information is also visible in financial market prices, i.e., whether it generates return predictability. We focus on those interdependencies between industries, which are generated by economic links, rather than those identified on a purely statistical basis, as, e.g., in a related paper by Rapach et al. (2015). We consider industries to be related when they are operating in similar business environments, so that they are likely to be exposed to similar kinds of economic shocks.

Relative to the existing literature, we add a new dimension to the analysis of information flow between industries over time, and we show empirically that it generates return predictability above and beyond that documented in previous studies. In earlier work, e.g., by Menzly and Ozbas (2010), the focus is on 'vertical' customer-supplier relationships between industries, i.e., industry  $i$  is linked to industry  $j$ , when  $i$  supplies important inputs to  $j$ . For instance, the industries 'Petroleum & Coal Products' and 'Transportation Equipment'<sup>1</sup> are both suppliers for, and thus vertically linked to, the industry 'Transportation by Air'.

We label the new direction of the connection between two industries 'horizontal'. The basic idea is that firms in different industries can be competitors, or their products can be substitutes, where substitute industries can be closely related with respect to the operating

---

<sup>1</sup>The names of the industries are taken from the classification based on two-digit SIC codes.

business. More precisely, equity prices in horizontally related industries are driven by shocks to the demand of joint down-stream customers, to the cost of supply materials purchased from joint up-stream suppliers, and to generally important macroeconomic quantities. As an example, consider the industries 'Railroad transportation', 'Local & Interurban passenger transit', 'Water transportation', and 'Transportation by air', which all provide similar services and use similar input factors like energy. So news about, e.g., energy prices or certain technological innovations are likely to have an effect on all of them. As another example, security brokers are increasingly confronted with competition from other financial intermediaries, such as real estate agencies, insurance companies, or money market funds, which all provide certain types of financial services. The fact that we find strong evidence in favor of horizontally linked industries exhibiting positively correlated fundamentals provides support for this notion of connectedness between industries. On the basis of our empirical analyses, we suggest that this horizontal relationship between industries is a plausible channel for robust return predictability not captured by other types of links.

With respect to the underlying economic mechanism, we argue that there are (at least) two channels through which horizontal lead-lag relations between firms in different industries can arise.<sup>2</sup> The first channel is related to product substitution. Despite the fact that firms belong to different industries, they may still produce goods which are seen as close enough substitutes by customers. However, it may take time until news regarding a given industry are finally also processed by investors of firms in the substitute industry, so that one may observe a lagged return reaction in that substitute industry. For example, increasing demand for the output product of a given industry gradually drives up the price of this product. This will be considered good news by investors so that the stock price will go up. Customers might start looking for substitutes when the price has reached a certain critical threshold, leading to higher profits also for the firms in the substitute industry. This is again considered good news by investors, and taken together this generates a

---

<sup>2</sup>We describe these channels only briefly here and provide more elaborate examples in Section 2.

lead-lag relation between stock returns of the firms in the connected industries.

The second channel is related to what one might call “segment momentum”. Firms belonging to substitute industries will have their own distinct main business segments, but are also likely to have overlaps in segments of lesser importance. News regarding the respective main segment may be processed very quickly by investors due to their pronounced relevance. On the other hand, it may take investors longer to incorporate the same piece of news into the stock price of the firms in substitute industry, since there the news are only relevant for a less important business segment.

For example, assume Firm 1 in industry  $i$  operates only in segment  $s^*$ , while Firm 2 in substitute industry  $j$  does only a small of its business in  $s^*$  and is at the same time active in a number of other segments, which represent the bulk of its sales. In reality the share to which Firm 2 is active in  $s^*$  is even often hard to determine. News regarding segment  $s^*$  can be expected to be processed quickly by investors in case of Firm 1, due to the fact that this firm does business exclusively there. On the other hand, it may take investors longer to incorporate the same piece of news into the stock price of Firm 2, since the relevance for Firm 2’s stock price may be much harder to determine. This will ultimately generate cross-sectional return predictability.

Technically, we generate our return predictor based on horizontal connectivity in two main steps. First, we compute what we call the *combined* predictor, which we then regress on the vertical predictor, yielding the horizontal predictor as the residual.<sup>3</sup>

Our starting point is the paper by Hoberg and Phillips (2017), who note that firms operating in more than one industry (according to Compustat segment data) are likely to do so in industries with similar product language. It appears natural to us to assume that such a pronounced similarity in the product description of two firms will in general

---

<sup>3</sup>In terms of actual computations, we follow the approach suggested by Burt and Hrdlicka (2019), who show that the use of excess returns as predictors can create problems in the presence of a misspecified asset pricing model and suggest to use residuals from a factor model instead. Details concerning this approach will be provided in Section 3.

be implied by a combination of both vertical and horizontal relatedness.

According to Porter (1998), products of substitute industries are those which can perform the same function as or a function very similar to the product of the given industry. This in turn implies that a measure of the degree to which two industries are related derived from Compustat segment data is presumably a good proxy for product language similarity and thus represents both vertical and horizontal, i.e., combined, relatedness.

In our paper, we directly compute the combined predictive signal as a weighted sum of return-based quantities, where the weight is calculated using Compustat segment data. More precisely, we count for a given industry  $i$  how often it is mentioned together with industry  $j$  in the Compustat segment reports for the firms in our sample (and then normalize this number appropriately).

Taking the industry 'Metal containers'<sup>4</sup> for the year 1995 as an example, Figure 1 shows that the combined relatedness derived from Compustat segment data indeed potentially includes both vertical and horizontal links between industries. The black dots represent industries with a (normalized) degree of combined or vertical relatedness to 'Metal containers' greater than 0.05.<sup>5</sup> The largest coefficient for combined relatedness is observed for 'Rubber and miscellaneous plastics products', which is certainly an industry competing with producers of metal containers, as are (supposedly) 'Paperboard containers and boxes' and 'Paper and allied products'. In the sense of our definition above, these two industries are both horizontally related to 'Metal containers'. On the other hand, in the direction of vertical (i.e., customer-supplier-based) relatedness (calculated according to Menzly and Ozbas (2010)), the coefficient is basically zero.

It need not be the case that this vertical coefficient is close to zero, as one can see from the location of the points representing 'Primary nonferrous metals manufacturing',

---

<sup>4</sup>In the Bureau of Economic Analysis (BEA) classification 'Metal containers' is industry no. 32, its 2-digit SIC number is 39.

<sup>5</sup>These coefficients are actually portfolio weights. The exact computations leading to the coefficients of vertical and combined relatedness presented in the graph are explained in detail in Section 3.



'Primary iron and steel manufacturing', and 'Retail and wholesale'. These three industries exhibit relatively large coefficients for combined relatedness with 'Metal Containers', but they are obviously also important suppliers or customers for that industry, indicated by large coefficients for vertical relatedness.

The overall picture of horizontally relations between industries<sup>6</sup> is shown in Figure 2, with the list of industries given in Table 1. One can see from the picture that horizontal links between industries are not a special characteristic of just a few sectors. Furthermore, an industry in the sample is on average horizontally linked to 3.35 other industries so that, overall, horizontal connectedness is not a rare phenomenon in the sphere of BEA industries.

This horizontal return predictability is statistically and economically significant. Following the methodology suggested by Menzly and Ozbas (2010) and adding a set of popular controls to the regression, an increase in the predictive signal of 100 basis points leads to an increase in the return next month of on average 22 basis points with a  $t$ -statistic greater than 4. The analogous result is obtained when we use the alternative methodology suggested by Burt and Hrdlicka (2019) based on factor model residuals as predictive variables. Here, a 100 basis point increase in the predictor yields on average a 13 basis point increase in next month's return, again with a highly significant  $t$ -statistic.

Since Burt and Hrdlicka (2019) show that their approach is very robust to a potential misspecification of the underlying asset pricing model, we perform all the following analyses based on this methodology. For example, to assess the economic value of the horizontal return predictor, we use it to set up a long-short portfolio strategy on the basis of the extreme signal quantiles. This portfolio delivers significant average excess returns as well as alphas with respect to the CAPM and the multi-factor models proposed by Carhart (1997), Fama and French (2015), and Hou et al. (2015). Interestingly, the

---

<sup>6</sup>For a given industry in the BEA classification, the horizontally related industries are those with coefficient for combined relatedness greater than 0.05, and a coefficient of vertical relatedness less than or equal to 0.01.

loadings of all strategy returns are basically zero for all factors, which shows that these returns represent a new dimension relative to the factor models. The economic effects of horizontal predictability are furthermore long-lasting, since we find that the returns of our associated portfolio strategy do not revert over horizons of up to twelve months.

Motivated by the analysis in Menzly and Ozbas (2010), we further investigate the role of informed investors, since they form the group of market participants who presumably recognize the relevance of the predictive signal and form their portfolios accordingly. We proxy the presence of informed investors by the degree of institutional ownership as suggested by Badrinath et al. (1995) and analyze whether this presence has an impact on the predictability relation for our horizontal signal. In line with our expectations, we find that for firms with a very high degree of institutional ownership the predictive signal becomes less important. Furthermore, the interaction term between the quintile dummy for institutional ownership and the predictor actually becomes insignificant for the 20 percent of firms with the highest degree of institutional ownership. This strongly suggests that the information contained in the signal is less valuable for firms with a larger share of informed investors in their ownership.

Finally, we also investigate if institutional investors actually take the relatedness between firms into account when it comes to portfolio rebalancing. To make our results comparable to the literature, we perform the regression tests suggested by Menzly and Ozbas (2010). We find that the presumably better informed investors indeed adjust their holdings of a certain stock in accordance with either the contemporaneous change of their holdings in the signal-producing firms (i.e., the firms in the related industries) or the contemporaneous signals based on factor model residuals (and after controlling for other return predictors). Moreover, the effect is found to be more pronounced for horizontally than for vertically related firms. Overall, our findings thus complement the literature on cross-industry information diffusion in an important dimension.

In Section 2 we provide examples for the potential channels through which horizontal

connectedness can create return predictability. Section 3 contains a detailed description of our empirical methodology. We present the results of our analysis in Section 4 and our conclusions in Section 5.

## 2 Channels for Horizontal Predictability

In the introduction, we had already introduced two potential channels for horizontal predictability, product substitution and segment momentum, for which we now provide examples. Consider Ball Corporation (SIC code 3411, BEA industry “Metal containers”) and Berry Global Group (SIC code 3089, BEA industry “Rubber and miscellaneous plastics products”). These two firms are operating in horizontally connected industries with respect to the BEA classification scheme and likely have correlated fundamentals, since their respective economic performance depends on the consumers’ demand for packaging products.

Ball is one of the world’s leading suppliers of metal containers. Its packaging products can be used, e.g., for beverages, personal care, and household products. Ball generates around 90% of its net sales in the metal containers industry, with the remaining 10% in the aerospace business. The primary raw material used by Ball in its production is aluminum. The business of Ball depends on a relatively limited number of major customers.

Berry is a leading global supplier of a broad range of innovative non-woven, flexible, and rigid packaging products used every day within consumer and industrial end markets. It offers products in three segments as stated in its 10-K files: “Engineered Materials”, “Health, Hygiene and Specialties”, and “Consumer Packaging”. Berry reports 34%, 35%, and 31% of its sales in the three segments, respectively. The primary raw material of Berry is plastic resin. In addition, Berry also uses other materials such as butyl rubber, adhesives, paper and packaging materials, linerboard, rayon, polyester fiber, and foil. Berry does not have a single customer representing more than 4% of net sales, and the top 10 customers

usually only represent 20% of net sales.

As an example for the first channel described above, suppose there is a negative shock to the production capacity of producers of plastic containers other than Berry, e.g., due to adverse weather conditions or even environmental disasters in other parts of the world. This will at first lead to an increase for the demand of Berry's key products, thereby increasing Berry's sales, profits, and ultimately also its stock price. With overall lower supply, however, prices for plastic containers will go up, and this will make users of these products look for substitutes, such as the metal containers produced by Ball, leading to higher profits and stock prices also for Ball, but with a delay compared to Berry.

Concerning the second channel, remember that Ball's sales are to a very large degree generated in the consumer packaging segment (metal container industry), so it can almost be viewed as a single-segment firm. Berry, on the other hand, operates in a number of segments, with a certain proportion in the consumer packaging segment. It may thus take Berry investors longer to assess the impact of a piece of news relevant for the consumer packaging segment on the future cash flows and profits of the firm than it may take Ball investors, since for Ball the consumer packaging business is basically the only and thus a crucially important business line.

More precisely, Ball reports in its 2018 10-K file that adverse weather and climate changes may result in lower sales. Particularly, unseasonably cool weather can reduce demand for certain beverages packaged in both metal and plastic containers. This unusual climate change would certainly materially and adversely affect Ball's profits. For investors focusing on Ball, this piece of news is relatively straightforward to process. In the case of Berry, with its activities spread out across different segments without even specifically reporting its share of sales in containers for beverages packaging, it may take some time before investors with limited information processing capacity will fully assess the impact of these news for Berry's future cash flows and profits, leading to a delayed reaction in the stock price relative to Ball.

### 3 Empirical Approach

The basic idea of the empirical analysis is to compute return predictors for industry  $i$  from the returns of all the other industries that industry  $i$  is linked to. As discussed above these links can be vertical, horizontal, or a combination of the two. The test of whether vertical connectedness between customers and suppliers creates return predictability has already been performed in Menzly and Ozbas (2010), and we replicate their study for our sample to have a benchmark from the literature against which we can compare our newly proposed horizontal (and combined) links. The degree to which industries are linked as customers and suppliers is measured based on the input-output good and service flow tables compiled by the Bureau of Economic Analysis (BEA).<sup>7</sup> As in Menzly and Ozbas (2010), we represent the relationship between two industries  $i$  and  $j$  by as the sum of the two directional links from  $i$  to  $j$  and from  $j$  to  $i$ .<sup>8</sup>

Hoberg and Phillips (2016) measure the degree to which two industries are related by the similarity of the associated products, which they assess via an analysis of the product description texts in the firms' 10-K filings. We choose to take a different empirical route and proxy product similarity across a pair of industries by the relative frequency with which multiple-industry firms are operating in these two industries, and this can be conveniently measured on the basis of Compustat segment data. Our measure, compared to that by Hoberg and Phillips (2016), is 'low-cost' and much more easily to replicate, since it is only based on publicly available data without complicated further processing. Furthermore, as shown by Hoberg and Phillips (2017), the two measures are indeed closely related, since there is a significant probability that firms doing business in more than one industry, as documented in the Compustat historical segment data, choose to do so in industries with pronounced product language similarity.

---

<sup>7</sup>See <https://www.bea.gov/industry/index.htm>. The BEA changed the industry classification in 1997. Before 1997 SIC codes were used, whereas afterwards industries were identified based on NAIC codes.

<sup>8</sup>The authors show that the intensity of this composite relationship is a better measure for information diffusion between industries than the strength of either unidirectional link.

Compustat provides annual information on the business segments reported by firms together with SIC and NAIC codes. This reporting can be retrospective or non-retrospective. For example, Firm A reports itself as a single-segment firm in year  $t - 1$ . In year  $t$  the firm switches from a single-segment to a two-segment firm, so that it will report twice in year  $t$ , namely its share of business in the two segments in year  $t$  (non-retrospective) and in year  $t - 1$  (retrospective). Since we want to perform the analysis using only the information available to the investor in real time, we do not take retrospective records into consideration.

To meaningfully compare different forms of industry relatedness, industries have to be identified consistently. The vertical links are based on the BEA industry classification, so we also apply this industry classification scheme to the computation of our combined and horizontal relatedness measures, i.e., given the SIC or NAIC codes from the Compustat segment data, we find the industry from the BEA classification corresponding to these codes.<sup>9</sup>

The measurement of strength of the link between industries  $i$  and  $j$  is then straightforward: We count how many firms report business in both industry  $i$  and industry  $j$  in a given year. In more detail, for every firm we look at the Compustat segment data and record all the pairs of industries that this firm has done business in over the previous year. E.g., when the firm has done business in three different industries, let us say industries 2, 5, and 8, in the year 2005, there will be three such pairs of industries, (2,5), (2,8), and (5,8). To assess the strength of the combined connectedness of industries, e.g., 2 and 5 in the year 2005, we count how often we record the pair (2,5) in the Compustat segment files across all firms in that year. For generic industries  $i$  and  $j$  and year  $t$ , we denote this quantity by  $A_t(i, j)$ , i.e., in our example we would obtain  $A_{2005}(2, 5)$  from that procedure.

---

<sup>9</sup>The mapping from the SIC scheme into the BEA classification is unique, but not the other way around. For instance, the BEA industry 'Ordnance and Accessories' has the two 2-digit SIC codes 34 and 37, and 'Food and Kindred Products' has codes 20 and 54. We use BEA 1987 benchmark SIC industry classifications for the years from 1977 to 1996, and BEA 2007 Benchmark NAIC industry classifications for the years from 1997 to 2012.

Alternatively, one could measure the connection between industries also via the respective dollar amount of sales. Chen et al. (2016), however, find that conglomerates tend to divert sales to their segment with the highest investor sentiment. Our approach of simply counting the number of segment pairs rather than considering sales in segments thus mitigates this potential endogeneity concern related to the measurement of industry connectivity.

The actual computation of the predictive signal follows the same procedure for the two sources of information (BEA input-output tables and Compustat segment data) about links between industries  $i$  and  $j$ . Given the strength  $A_t(i, j)$  of a link between industries  $i$  and  $j$  at time  $t$ , the return predictor (or signal) for industry  $i$  (and thus consequently for all individual firms from this industry) at time  $t$ ,  $\tilde{r}_{i,t}$ , is computed as the weighted sum of return-based quantity  $x_{j,t}$  for the industries  $j$  linked to industry  $i$ , i.e.,

$$\tilde{r}_{i,t} = \sum_j w_t(i, j) x_{j,t}, \quad (1)$$

with weights  $w_t(i, j)$  given by

$$w_t(i, j) = \frac{A_t(i, j)}{\sum_k A_t(i, k)}. \quad (2)$$

The weights defined in (2) are non-negative and sum to 1, so that they represent the relative weights in an all-long portfolio. In Equation (1), the return-related quantity  $x_{j,t}$  can be either an excess return or a factor model residual. The former choice is standard in the literature (and is also used by, e.g., Menzly and Ozbas (2010)), while the latter is motivated by the analysis in Burt and Hrdlicka (2019), who show that the use of excess returns can generate spurious results with respect to information transmission.

To obtain the signal  $x$  as a factor model residual at time  $t$ , we regress industry returns over a rolling window of 360 months ( $\tau = t - 359, \dots, t$ ) on the market, size, value, and momentum factors, and take the time- $t$  residual (i.e., the residual from the last

observation in the estimation window) from the regression for industry  $j$  to be  $x_{j,t}$ .

The information necessary to compute the coefficients  $A_t(i, j)$  is updated on a yearly basis, so that the strength of the link between two given industries remains constant over the respective year. The predictor itself, of course, changes from month to month due to variation in return based quantity  $x_j$  over time.

We now explain how we compute our horizontal return predictor. The key idea is to interpret the combined predictor based on Compustat data, as it also appears intuitive, as a combination of the horizontal and the vertical predictor, so that the horizontal part can be extracted via a regression approach. In detail, let  $\tilde{r}_{i,t}^{(c)}$  and  $\tilde{r}_{i,t}^{(v)}$  denote the combined (Compustat-based) and vertical (customer-supplier based) predictors for industry  $i$  at time  $t$ , respectively. Given these two predictors for all industries  $i = 1, \dots, N_t$  in the sample at a given point in time  $t$ , the horizontal predictor  $\tilde{r}^{(h)}$  for industry  $i$  is computed as the residual from the following cross-sectional regression:<sup>10</sup>

$$\tilde{r}_{i,t}^{(c)} = \beta_{0,t} + \beta_{1,t} \tilde{r}_{i,t}^{(v)} + u_{i,t} \quad (i = 1, \dots, N_t) \quad (3)$$

i.e.,  $\tilde{r}_{i,t}^{(h)} \equiv \hat{u}_{i,t} = \tilde{r}_{i,t}^{(c)} - (\hat{\beta}_{0,t} + \hat{\beta}_{1,t} \tilde{r}_{i,t}^{(v)})$ , where hats denote estimated coefficients. The residual is by definition orthogonal to the regressor, so that our horizontal predictor truly represents a dimension in return space completely uncorrelated to vertical connectedness.

The results of the monthly regressions in (3) also deliver information on the degree to which the combined information in the Compustat segment files actually goes beyond that represented by vertical customer-supplier links given in the BEA input-output tables. Should the  $R^2$  of that regression be close to 1, it would imply that the two sources of information are basically the same with respect to the predictor they generate, and consequently, there would be no added value in looking at the combined connectedness between industries. On the other hand, given our economic intuition we would not expect

---

<sup>10</sup>The residual from the cross-sectional regression is a statistical quantity, and as such it is a (probably imperfect) proxy for the true horizontal predictor. At the same time, there may also be other, but most likely less important, channels relevant for firms operating in more than one industry.



to see values for  $R^2$  too close to zero either. We provide information on the goodness of fit for the monthly regressions (3) in Section 4.1 below.

In terms of actual data, all the necessary stock-related information is taken from CRSP. We restrict the analysis to common stock, i.e., to CRSP sharecodes 10 and 11. To mitigate a potential impact of illiquidity on our results we exclude firms with a stock price below US-\$5.<sup>11</sup> Industry returns are computed as value-weighted averages of the respective stock returns, and for the computation of returns on industry quintile portfolios, the individual industry returns are equally weighted. We obtain the pricing factors (market excess-return, SMB, HML, and UMD) from Ken French's data library.<sup>12</sup>

## 4 Results

### 4.1 Descriptive statistics

Table 2 provides an overview of our sample. In the first part of the sample period there are 69 industries represented in the BEA tables, later this number reduces to 65 due to a change in the industry classification by the BEA. There are around 3,000 stocks in our sample on average, and even in periods with fewer stocks this number is still well above 2,000. The typical industry exhibits around four vertical links to other sectors. Not surprisingly, the number of combined links based on Compustat data is about 50% larger, and this is true across the whole range of the data, as shown by the respective values for the 10th, the 50th, and the 90th percentile.

As a first step in our empirical analysis, we want to make sure that the links between industries represented by the Compustat segment data are indeed different from the vertical customer-supplier links derived from the BEA tables. Otherwise, we would

---

<sup>11</sup>A stock is excluded when its price drops below \$5. Should the price go up above \$5 again, it is reincluded in the sample.

<sup>12</sup>See [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

not expect substantial additional informational value coming from horizontal relative to customer-supplier relatedness.

The evidence in the data speaks strongly against the null hypothesis of no additional information. In each year and for each industry we compute the vector representing the degree to which the given industry is related to all other industries, once on the basis of BEA input-output data, and once using the Compustat segment reporting data. Table 2 shows time-series statistics for the cross-sectional correlation coefficient between the two measures. It is certainly not surprising that they are significantly related to each other, but the correspondence between them is far from perfect. For both the more standard (Pearson) and the (Spearman) rank correlation, the median is around 0.4, with 10% and 90% quantiles around 0.2 and 0.7, respectively. We take this as evidence supporting our claim that the BEA input-output tables and Compustat segment data do not simply contain the same information when it comes to measuring connectedness of industry pairs.

Links between industries are very persistent and change only at a very low pace. The cross-sectional median of the autocorrelation coefficient is almost equal to one over the one-year horizon for the vertical links based on the BEA tables, and it decreases very slowly for longer lags. It is also high, but somewhat lower for the combined connectedness, which indicates that this measure is more dynamic and thus has the potential to reflect changes in the relatedness between industries better than the BEA data.

The regression of combined on vertical signals is a second implicit test of the hypothesis that vertical and horizontal connectedness represent different dimensions of links between industries. Again, one would not expect the two types of predictors to actually be orthogonal to each other, but for our purposes it is important that they indeed represent sufficiently different information, i.e., the correlation between combined and vertical signals should not be too high. The time-series average of the cross-sectional correlation is around 0.5, as is the estimated slope coefficient. The average  $R^2$  is less than 30%, and in more than 80% of the cases the null hypothesis of vertical connectedness being an

unbiased predictor of combined relatedness (meaning a regression intercept equal to zero and a slope equal to one) is rejected. This is true for using both excess returns and factor model residuals as the basis for the computation of the signal.

At a basic level, these results provide evidence that industries can be linked in more general ways than just vertically, and we consider these findings a solid justification for our overall empirical strategy.

## 4.2 Correlated fundamentals

According to Menzly and Ozbas (2010), a necessary condition for cross-predictability in the context of a limited-information model is that related industries exhibit correlated fundamentals. To see if this is actually the case in our sample, we follow the methodology proposed by these authors and use the return on assets (ROA) as a profitability measure to proxy for firm fundamentals. We then estimate the following panel regression on the firm and on the industry level:

$$ROA_{i,t} = \alpha_i + \theta_{market}ROA_{t,market} + \theta_xROA_{t,x} + u_{i,t}. \quad (4)$$

In Equation (4),  $ROA_{i,t}$  denotes the return on assets of firm or industry  $i$  in year  $t$ , while  $ROA_{t,market}$  and  $ROA_{t,x}$  represent the corresponding numbers for the market as a whole and aggregated across industries with respect to connectedness of type  $x$  (combined, horizontal, or vertical).

Firm-level  $ROA$  is measured as the ratio of net income to total assets, taken from Compustat.  $ROA$  for a given industry or the market as a whole is calculated by aggregating firm-level  $ROA$ , using firm assets as weights. Combined and vertical  $ROA$  are calculated by weighting the industry-specific  $ROA$  by the strength of links derived from Compustat segment data (see Section 3) and the inter-industry input-output flows from the BEA survey, respectively. Finally, horizontal  $ROA$  is computed as the residual from

the cross-sectional regression of combined on vertical *ROA*. In regression (4),  $\alpha_i$  represents firm- or industry-level fixed effects to control for potential heterogeneity across firms and industries. Standard errors are double-clustered by firm (industry) and year for the firm-level (industry-level) specification (see Thompson (2011) and Petersen (2009)).

In Table 3, columns (1) to (3) show the results for the firm-level specifications, while columns (4) to (6) present the coefficients for the industry-level regressions. Looking at the firm-level results first, we find strong evidence that the fundamentals of a given firm are significantly correlated with the analogous quantities for the market as a whole and for industries connected via combined, horizontal, and vertical links. The associated coefficients are all positive and their *t*-statistics indicate significance at a level of 1% or lower. The industry-level regressions produce qualitatively the same results. Again the coefficients are positive and strongly significant for the market as a whole and for all types of connectedness with respect to industries.

Overall, our results strongly indicate that the necessary condition for return predictability is satisfied, i.e., that the fundamentals of firms in connected industries are contemporaneously correlated.

## 4.3 Return predictability

### 4.3.1 Signals based on excess returns

Our basic claim in this paper is that types of industry connectedness other than vertical are the basis for return predictability. To have a meaningful basis for comparison, we will also include the vertical connectedness measure suggested by Menzly and Ozbas (2010) in our analyses. Furthermore, for the sake of completeness, we will discuss the combined Compustat-based signal as well.

In our first set of analyses we use signals based on excess returns, i.e., the term  $x_{j,t}$  in Equation (1) corresponds to the excess return of industry  $j$  at time  $t$ . The results based

on factor model residuals will be presented in the next subsection. To test the hypothesis that a certain signal is a useful predictor, we follow the approach suggested by Menzly and Ozbas (2010) and run Fama and MacBeth (1973) regressions of monthly *individual stock returns*  $r_{i,t}$  on the return predictors  $\tilde{r}_{i,t}^{(k)}$ , where  $k \in \{v, c, h\}$ .<sup>13</sup> The regression equation is

$$r_{i,t} = \alpha_t + \lambda_t \tilde{r}_{i,t-1}^{(k)} + \Delta'_t Z_{i,t-1} + \epsilon_{i,t}. \quad (5)$$

Here  $Z_{i,t-1}$  is a vector of controls representing characteristics of firm  $i$  (measured in month  $t-1$ ), which are commonly used to predict stock returns. To make our results comparable to the literature the first two components of  $Z_i$  are the same as in Menzly and Ozbas (2010), representing short-term reversal (the return on the stock over the previous month, see Jegadeesh (1990)), and momentum (the average return on the stock stock from month  $t-12$  to month  $t-2$ , see Jegadeesh and Titman (1993, 2001)). Da et al. (2014) document that reversal within industries is stronger than return reversals at the level of individual firm. Therefore we also include the industry return over the previous month as a control (see Moskowitz and Grinblatt (1999)).

We also want to look at potential size effects in the predictability regressions. The motivation for this is that Cohen and Lou (2012) find limits to arbitrage to be more binding for smaller firms, which would *ceteris paribus* imply a stronger effect of return predictability, as sophisticated investors are less able to update the stock prices of lagging smaller firms. To capture this effect the authors suggest to use a dummy variable to indicate whether a stock belongs to the upper or lower half of the sample with respect to market capitalization. We thus add an interaction term between the return predictor and a size dummy set to one for firms with size below the median. From regression equation (5), the total impact of the return predictor for small firms is then equal to the sum of  $\lambda_t$  and the coefficient of the interaction term, while for large firms it is just equal to  $\lambda_t$ . In

---

<sup>13</sup>Since our predictors are formed on the basis of industry returns this implies that every firm in a given industry is assigned the same predictive signal in a given month.

case the coefficient on the interaction term turns out to be significant, this represents a rejection of the null hypothesis that the informational value of the given predictor is the same for small and large firms.

Table 4 shows the results for this first set of analyses. We find all types of return predictors to be significant when used as the only regressor. Without any further controls an increase of 100 basis points in the combined, horizontal, and vertical signals at  $t - 1$  generate an average effect of 32, 34, and 20 basis points of additional return for the given stock in month  $t$ , respectively, as shown in specifications [1] to [3].

Specification [4] features the combined and the vertical predictor together. Under the hypothesis that the combined signal derived from Compustat data subsumes information about both vertical and horizontal connectedness (and not just noise in addition to the vertical signal), we expect to see the coefficient for the former significant, while the latter should not provide additional information, so that its explanatory power should be relatively low. This is what we find, with a large and highly significant coefficient for the combined signal (which furthermore remains basically unchanged compared to specification [1]) and an insignificant one for the vertical predictor.

Not surprisingly, when we include the set of firm characteristics described above as controls in specifications [5] to [7], the coefficients of predictors decrease uniformly. For the combined and the horizontal signal, the coefficient remains statistically significant, and in case of the combined predictor, there is also a pronounced interaction with size. For smaller firms the total impact of the signal is about 60 percent stronger than for large stocks. This interaction with size is not found for the horizontal signal, where the coefficient for the interaction term is small and insignificant. For both the combined and the horizontal signal it is worth noting that the coefficients for these predictors are substantially larger than the coefficients for the controls. This comparison is meaningful, since the control variables are returns as well. The results for the vertical predictor investigated by Menzly and Ozbas (2010) are qualitatively different. There is no significant effect of the signal

itself anymore, but all the explanatory power goes to the interaction term with the size dummy. That we do indeed observe this size effect with respect to the vertical signal is in itself not too surprising, since Cohen and Frazzini (2008) show that the supplier is often small in terms of market capitalization relative to the customer.

In summary, we consider these results as rational evidence that return predictability arises from the returns computed from the combined or the horizontal connectedness between industries, while the vertical signal is potentially weaker with respect to its own informational value.

#### **4.3.2 Signals based on factor model residuals**

As indicated above there is a discussion in the literature about the true sources of return predictability based on economic links between industries or firms. Burt and Hrdlicka (2019) investigate this issue in detail and show that in some cases a large part of the return predictability is due to the fact that industries share alpha with respect to some asset pricing model, so that the observed predictability is due to this common alpha rather than to the type of economic links under investigation.

To see if this potential misspecification problem is relevant in the case of our analysis, we follow the suggestion in Burt and Hrdlicka (2019) and compute the predictive signals based on the residuals of industry returns relative to a factor model, in our case featuring the market, size, value, and momentum as factors. As described in Section 3 we take the preceding 360 months of industry returns and regress them on the factors. The residual for industry  $j$  in the last month of the estimation period (i.e., month  $t - 1$ ) is then taken as the quantity  $x_{j,t-1}$  entering the computation of the signal according to Equation (1).

The results for the analysis, which is otherwise analogous to the one presented for excess returns in the previous section, are shown in Table 5. The findings for specifications [1] to [4] are qualitatively similar to those presented for the case of excess return signals. The coefficients for all predictors are significant, and the vertical information is subsumed

by the combined. As a tendency, we observe that the coefficients have decreased compared to Table 4, which indicates a lower informational for residuals than for excess returns, consistent with the Burt and Hrdlicka (2019) argument.

The most important results are of course those for specifications [5] to [7], where we check whether the signals based on residuals provide additional information relative to standard predictors. Here we find pronounced differences between the horizontal (or combined) and the vertical predictor. For the former, the coefficient for the signal itself is significant and the one for the interaction term with size is not, while for the latter, the coefficients for both the signal itself and the interaction term with size have turned insignificant. This means that, once the impact of a potential misspecification in the forecasting model is taken care of, the informational value of customer-supplier links seems to vanish to a large degree in our sample.

#### **4.4 Portfolio strategies**

The analysis in Table 4 structurally represents a replication of Menzly and Ozbas (2010). The results based on the Burt and Hrdlicka (2019) approach in Table 5, however, show that there is potentially an issue with model misspecification, when excess returns are used to compute the signals.

From now on we will thus only consider the setup where signals are computed based on factor model residuals, but we even go one step further. The results concerning return prediction presented above show that popular control variables take away a certain part of the predictive effect of the return signals based on economic links. Since we are exclusively interested in the part of the informational value of a signal not already explained by other variables, we remove the impact of these variables on our predictors. In detail, we perform monthly cross-sectional regressions of the signals on reversal, momentum, and the lagged industry return, take the residuals from that regression, and then run all further tests



using only what we call 'residual signals'.<sup>14</sup> This approach of using only the part of a signal which is not explained by other predictors is similar to the methodology used by, e.g., Hou and Loh (2016). It represents a pretty tough test of the informational value of economic links between different industries, since any part of our newly proposed signal that is also reflected in control variables is taken out.

A natural measure for the economic value of the information concerning industry connectedness is then to sort stocks into portfolios based on the predictor (in our case the residual signal) and to run a strategy which is long one extreme portfolio and short the other. At the end of every month we sort the industries (not the individual stocks) into quintiles with respect to the predictor. The industry return is then computed as the value-weighted average of the returns of individual stocks in the given industry, and the return on the quintile portfolio of industries is given by the equally-weighted industry returns. For the zero cost long-short strategy we go long the quintile of industries with the highest values of the signal ('Q5') and short the one with the lowest ('Q1'). The resulting strategy portfolio is labeled 'Q5-Q1'.

Table 6 presents average excess portfolio returns and portfolio alphas for Q1, Q5, and Q5-Q1, based on the three different types of connectedness. The CAPM and the four-factor alpha are the intercepts of the regression of the portfolio returns on just the market excess return and on the market, SMB, HML, and the momentum factor UMD, respectively.<sup>15</sup>

Again, the results show pronounced differences between vertical and the other two types of signals, since sorting on the residual vertical signal (Panel C) does not generate a significant return spread between the two extreme portfolios. Together with the evidence for the return prediction exercise in Table 5 this shows that for our sample period the

---

<sup>14</sup>We also run all the tests using size as yet another control in the computation of the residual signal. All the results of the analyses to follow remain qualitatively unchanged. The quantitative differences are also minor.

<sup>15</sup>The robustness of these results with respect to the choice of the factor model is discussed in Section 4.6.1 below.

vertical signal does not seem to exhibit robust predictive power. Economically this indicates that market participants seem to incorporate the additional information coming from customer-supplier links into prices rather quickly.

This changes when we look at combined (Panel A) and horizontal links (Panel B). Sorting on the signals generated by these types of economic links across industries generates significant excess returns and alphas for the respective Q5-Q1 portfolios. It is striking that for the horizontal signal the alphas with respect to either of the two factor models are very close to the excess returns. The reason for this is that the Q5-Q1 returns for this signal do not load on the factors in a substantial fashion, as shown in Table 7. Neither the coefficient on the market alone nor any of the four factor coefficients in the extended Fama-French model is significant. For the combined signal, the result is somewhat different with significant loadings on HML and UMD and an alpha, which is consequently lower than the pure excess return of the portfolio.

It is of interest to see if this predictability has an impact also over return horizons longer than one month. To this end, we examine the cumulative returns on the Q5-Q1 portfolios over the twelve months following the formation month. Figure 3 presents the graphs for these excess returns of each of the three Q5-Q1 portfolios based on residual signals for holding periods of one to twelve months. As one can see, the cumulative long-short portfolio returns increase gradually over the subsequent year, without a sign of pronounced reversal. In line with the results discussed previously, the portfolio formed on the basis of the horizontal signal persistently outperforms that using vertical links. Overall, the evidence from Figure 3 implies that information contained in the horizontal residual signals has long-lasting impact on stock prices and returns.

The average returns of long-short portfolios in our sample are somewhat lower than those presented by Menzly and Ozbas (2010). We see two main reasons for this. First, we find that the effect of a potential model misspecification in the sense of Burt and Hrdlicka (2019) is strong, accounting for 75%, 45%, and 51% of the original performance of long-

short portfolios formed on the basis of excess return signals for the vertical, horizontal, and combined case, respectively.<sup>16</sup>

Second, the performance of the long-short portfolios has decreased over time. In case of the vertical signal, the average return in the second half of our sample period even turns negative. For the signal based on horizontal connectedness, the average return in the second half is also smaller than in the first, but still positive. One potential reason behind this downward trend is that capital markets seem to be (or to have become) more efficient than previously recognized (see also Hou et al. (2018)). This increasing efficiency is accompanied by a tendency on the part of quantitative asset managers to take anomalies documented in the empirical asset pricing literature as the scientific foundation for their portfolio strategies, as argued by Ang (2014).<sup>17</sup>

We can conclude from the previous analyses that Compustat segment data are a valuable source of information about connected industries. The return signals derived from these data exhibit predictive power for future stock returns even after using factor model residuals instead of excess returns as the basic predictor. Furthermore, sorting stocks on residuals from a regression of these signals on standard control variables still produces significant excess portfolio returns for the long-short strategy. We take the return prediction and the portfolio results together as sound evidence for information flowing across industries in a horizontal direction over time.

---

<sup>16</sup>The average monthly return on the long-short portfolios formed on the basis of excess return signals is 0.41% (vertical), 0.45% (horizontal), and 0.62% (combined) for our sample, respectively. These numbers are similar in magnitude to those reported by Menzly and Ozbas (2010).

<sup>17</sup>For example, over the sample period of our paper, hedge funds have experienced spectacular growth in assets under management (AUM). BarclayHedge, a global leader in research on alternative asset management, reports that hedge fund AUM amounted to 2,878 billion U.S. dollars as of 2018 Q4, among which strategies labeled “Equity Long/Short” accounted for 215 billion U.S. dollars.

## 4.5 Informed investors

The flow of information across economically connected industries is the main force behind the return predictability documented in the previous section. The first question is if there is a group of investors, who supposedly have an informational advantage and adjust their portfolios accordingly. The share of institutional ownership has been suggested as a good proxy for the presence of informed investors in a firm by, e.g., Badrinath et al. (1995). In our framework, where we again follow Menzly and Ozbas (2010), we would thus expect to see the informational value of the signals to be the lower the higher the degree of institutional ownership. As indicated above the analysis here is based on signals computed as residuals from a time series regression of industry excess return on the market, size, value, and momentum factors, and in subsequent cross-sectional regressions we control for any informational value generated by reversal, momentum, and lagged industry returns and consider only residual signals.

Table 8 shows the results of a regression of stock returns on interaction terms between the different residual signals and dummy variables for quintiles of institutional ownership measured at the end of the previous quarter. The stocks with the lowest degree are assigned to the first quintile labeled 'Q1', those with the highest to 'Q5' accordingly. If our basic hypotheses, namely that the predictive signals have informational value and that institutional investors are better informed than the rest of the investor population, we should see large coefficients for Q1 and lower estimates for Q5.

This is exactly what we find. As can be seen from Panel A of Table 8 for the regressions in the full sample, the coefficient for the interaction term is indeed larger for Q1 than for Q5 for all three types of signals. For example, in the case of the horizontal predictor a 100 basis point increase in the signal implies on average a (significant) increase of 14 basis points in the return next month for a stock in the quintile with the lowest institutional ownership, while for Q5 this value is only 6 basis points with an insignificant *t*-statistic of only 0.91. The pattern is not exactly monotonic across the quintiles for institutional own-

ership, but the important result is that the extreme portfolios are substantially different. The findings for combined and vertical predictors are similar, although it is striking that in case of the vertical signal the effect is significant only and strongest for the portfolio of stocks with the smallest share of institutional ownership.

As we had seen before in Table 4, the vertical signal is only significant for the smaller stocks even when excess returns are used to compute the signal. Given this it seems natural to investigate whether small stocks drive the large and positive coefficient for the quintile of lowest institutional ownership. Note that institutional ownership is not simply a proxy for size or vice versa. We find that the unconditional correlation between size and institutional ownership is only around 0.16 over our total sample period with even a decreasing trend, represented by correlations of about 0.3 in 1980 and 0.05 in 2012, respectively. This implies that the result of this additional test with the smallest stocks excluded is not a priori clear.

We repeat the regression analyses from Panel A in Table 8 for a sample where the respective 20 percent of smallest stocks are excluded for every month. The results are presented in Panel B of the same table. For the horizontal and to a certain degree also for the combined signals we now observe an even more pronounced difference between the coefficients for Q1 and Q5, i.e., a higher share of institutional investors reduces the informational role of the signal more, but for the lowest quintile of institutional ownership this informational value is still there. In contrast, this is no longer true for the case of vertical customer-supplier links. Here none of the interaction terms is significant anymore, and this mirrors the results of the return prediction exercises presented in Table 5.

The second question is if institutional investors take the potential informational value of the signals into account and increase (reduce) their holdings of a certain stock together with their holdings of stocks in related industries or when the residual signal for this stock is particularly positive (negative). We analyze these two questions by running a panel regression of the change in institutional ownership for a given stock in a given

quarter (data on institutional ownership are available only at quarterly frequency) on the contemporaneous change in institutional ownership in the related industries and on the residual signal, respectively.

This change in institutional ownership in related industries is computed analogously to the return signals, i.e, they are given as weighted averages of the changes in institutional ownership of related industries, where the weights are computed on the basis of either the BEA input-output tables (vertical) or the Compustat segment data (combined). The change in institutional ownership induced by horizontal connectedness is again obtained the residual from a cross-sectional regression of the combined on the horizontal values.

The results are shown in Table 9. The coefficients for both the changes in institutional ownership and the residual return signals are always positive and statistically significant. However, in line with previous analyses we again find the impact of the vertical signal to be smaller than that of the horizontal and the combined variable, be it the change in institutional ownership in Panel A or the residual return signal in Panel B. The differences are also large, with the coefficient for the vertical variables always at most half the size of those for the horizontal and combined linkages.

Our findings thus support the notion that informed investors act according to the information derived from the different types of connectedness between industries, but that at the same time there are differences in the informational content between these different types, with vertical connectedness exhibiting a generally lower informational value than the other two.

Finally, since the previous findings provide support for the notion that institutional investors consider the connectedness between industries in the asset allocation process, we are interested in testing explicitly whether there is a higher likelihood that institutional investor portfolios contain pairs of horizontally connected industries than pairs of sectors not linked in this particular way. As before, for a given industry in the BEA classification, we consider those sectors horizontally related to it, which exhibit a coefficient of combined

relatedness greater than or equal to 0.05, while the vertical relatedness should be less than or equal to 0.01.

Figure 4 shows the relative frequency of institutional investors holding pairs of horizontally related industries (solid line) versus the average relative frequency of holding such pairs of sectors if the institutional investor's industry holdings were selected randomly (dashed line). To compute the former number, we identify different industries with which the stocks held by a given institutional investor  $i = 1, \dots, I$  belong to in year  $t$ . With this number denoted by  $N_{i,t}$ , there are  $\binom{N_{i,t}}{2}$  pairs of industries. We then compute the relative share of pairs representing horizontally related industries, and average this share across all investors in a given quarter. To see if these relative shares of horizontally connected industry pairs observed in the data are different from what one is likely to obtain by pure chance, we draw  $N_{i,t}$  industries randomly for every combination of  $i$  and  $t$  and perform the same computation as before, i.e., check how many of the pairs represent horizontally connected sectors and average across all investors. We repeat this random selection of industries 1,000 times and take the average across all simulation runs.

For basically all of our sample period, the relative frequency of institutional investors holding pairs of horizontally related industries is higher than what we would obtain from purely random sampling. Additionally, based on a Wilcoxon rank-sum test, the null hypothesis of no difference between the observed value and the value we would expect to obtain from a purely random selection of industries is rejected at the 1%-level in over 91% of the cases.<sup>18</sup>

---

<sup>18</sup>This result is not specific to horizontally connected sectors. We obtain qualitatively the same result for vertical links, i.e., analogous to horizontally connected sectors, the relative frequency with which informed investors hold stocks from vertically related industries is significantly higher than what one would expect from a purely random selection.

## **4.6 Robustness checks**

### **4.6.1 Alternative factor models**

Recently, new factor models with alternative sets of return drivers have been proposed and employed in the empirical asset pricing literature. For instance, Fama and French (2015) develop a model featuring, in addition to the market, factors representing size, value, profitability, and investment. Alternatively, motivated by “investment-based asset pricing”, Hou et al. (2015) propose the market, size, investment, and the return on equity as return drivers. Hou et al. (2018) also show that this model is able to explain a wide range of cross-sectional anomalies and represents a substantial improvement over the Carhart (1997) model.

In Table 10, we report the alphas of the long-short portfolios based on vertical, horizontal and combined residual signals relative to these two recently introduced factor models. The results of this exercise show that the findings of our analysis in Section 4.4 are not affected by the choice of the benchmark factor models. As discussed above for our original analysis presented in Table 6, the long-short portfolios formed on the basis of combined and horizontal signals produce significant alphas relative to the factor models. Also in line with our previous findings, the alpha generated by the portfolio strategy based on vertical signals is positive, but not significant.

### **4.6.2 Industry classification**

The notion of combined and horizontal connectedness is the key innovation in our analysis. The analysis done so far was based on the BEA industry classification, but the industries in which a firm has reported business in a certain year in the Compustat segment data are coded following the SIC classification. To put combined and vertical connectedness on the same basis in terms of industry classification (so that we could actually extract the horizontal return signal), we had to map the Compustat into BEA classification. This in



turn meant that we were not able to perform the full analysis based on SIC codes instead of on the scheme represented in the BEA tables.

Nevertheless, we can perform a robustness check for our primary claim that the information about linkages between industries derived from the segment reporting provided in Compustat, i.e., what we call “combined” relatedness, generates strong return predictability. To this end, we repeat the analyses from Section 4.3 for the combined signal, but this time based on the 2- and the 3-digit SIC classification for industries. Here we do not consider signals based on excess returns any more, but use factor model residuals and residual signals right away.

Tables 11 to 13 present the results concerning return predictability, the performance of the long-short-strategy, and the associated factor loadings, respectively.

As one can see, the results with respect to basic return predictability are not qualitatively affected by the choice of industry classification. A comparison of specification [1] in Table 5 with specifications [1] and [3] in Table 11 shows that the combined signal exhibits a positive and strongly significant coefficient for both versions of the SIC classification. When we add the usual controls to the regression, the significance of the coefficient for the return predictor is not as pronounced anymore for the 2-digit SIC classification, but it remains strong in the case of the 3-digit SIC grouping. Furthermore, the coefficients for the control variables do not exhibit strong variation with respect to the chosen industry classification, neither between the two SIC-based classifications presented in Table 11 itself nor between SIC and BEA, as shown by a comparison to Table 5.

Also concerning properties of the long-short strategy results of the robustness checks are similar for the two basic industry classification schemes. If anything, the performance of the portfolio strategy appears even better for 2-digit and 3-digit SIC classification than based on the BEA tables. The average excess returns and alphas shown in Table 12 are slightly higher and more significant than those presented in Table 6. The sensitivity of strategy returns to the pricing factors is also low, i.e., the entries in Tables 7 and 13 are

very similar to each other. Overall, we can conclude that the actual choice of industry classification scheme does not have a material impact on our main results.<sup>19</sup>

#### 4.6.3 Informed investors

Finally, we want to check the robustness of our findings regarding the changes in institutional ownership with respect to the estimation approach. We had performed the analyses presented above in Table 9 via panel regressions. An alternative approach is to apply the Fama and MacBeth (1973) procedure, i.e., to take the time-series average of the coefficients obtained in the quarterly cross-sectional regressions as the coefficient estimate. The fundamental result obtained for the panel regression does not change in a pronounced fashion when we employ this alternative regression methodology, as can be seen from Table 14. All the coefficients again turn out positive and strongly statistically significant, with the patterns very similar to those obtained in the panel regression, i.e., substantially larger and more significant estimates for the combined and the horizontal linkages than for the vertical customer-supplier connectedness. In summary, our main findings appear rather robust and exhibit only minor variation when alternative approaches are employed.

## 5 Conclusion

The diffusion of information across industries has been a topic of intense financial research over the past couple of years. The fact that industries are vertically connected as customers and suppliers has been shown to represent one such source of information (e.g., by Menzly and Ozbas (2010)), and the fact that this information is not immediately reflected in prices creates short-run return predictability.

---

<sup>19</sup>In additional analyses (unreported) we find that for the 2-digit SIC classification the returns of the long-short portfolio are much larger when the signal is based on excess returns than when we consider the residual signal. The difference is much smaller in the 3-digit SIC case, which is an indication that the use of the finer classification scheme could potentially reduce the problem of model misspecification pointed out by Burt and Hrdlicka (2019).

Our analysis in this paper provides plausible evidence for a new and different pattern of information flow and associated return predictability. Two industries are related in general, because there are firms doing business in both of them, or because their respective products are substitutes. We exploit the information in the Compustat segment data to infer the general degree of connectedness between industries and associate this with a combination of 'vertical' and 'horizontal' relatedness.

In our sample, we find that our newly suggested horizontal signal exhibits substantial and robust predictive power. This is observed independent of whether we compute the signal from industry excess returns or residuals relative to a four factor model, where the latter approach is motivated by a recent paper by Burt and Hrdlicka (2019). The signal remains a significant predictor even in the presence of other variables used by, e.g., Menzly and Ozbas (2010) to predict returns, which emphasizes its informational value even more. In our empirical tests, we take out everything from our signals that is explained by these other variables and then only use these 'residual signals' in performance tests of long-short portfolios and in studying the behavior of informed investors. The horizontal signal retains its predictive power in all of these tests, and this predictive power is even more robust than that of the vertical predictor based on customer-supplier links in our sample.

Overall, we introduce the notion of horizontal connectedness between industries and show via a variety of empirical analyses that it has substantial informational value. This represents an important contribution to the literature regarding the flow of information across sectors of the economy.

## References

- Ang, Andrew, 2014, *Asset Management: A Systematic Approach to Factor Investing* (Oxford University Press).
- Badrinath, Swaminathan G., Jayant R. Kale, and Thomas H. Noe, 1995, Of shepherds, sheep, and the cross-autocorrelations in equity returns, *Review of Financial Studies* 8, 401–430.
- Burt, Aaron, and Christopher Hrdlicka, 2019, Where does the predictability from sorting on returns of economically linked firms come from?, Working Paper.
- Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.
- Chen, Huaizhi, Lauren Cohen, and Dong Lou, 2016, Industry window dressing, *Review of Financial Studies* 29, 3354–3393.
- Cohen, Lauren, and Andrea Frazzini, 2008, Economic links and predictable returns, *Journal of Finance* 63, 1977–2011.
- Cohen, Lauren, and Dong Lou, 2012, Complicated firms, *Journal of Financial Economics* 104, 383–400.
- Cohen, Lauren, Christopher Malloy, and Quoc Nguyen, 2018, Lazy prices, Working Paper.
- Da, Zhi, Qianqiu Liu, and Ernst Schaumburg, 2014, A closer look at the short-term return reversal, *Management Science* 60, 658–674.
- Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.
- Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.
- Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124, 1423–1465.
- Hoberg, Gerard, and Gordon Phillips, 2017, Conglomerate industry choice and product language, *Management Science* 64, 3469–3970.
- Hong, Harrison, Walter Torous, and Rossen Valkanov, 2007, Do industries lead stock markets?, *Journal of Financial Economics* 83, 367–396.
- Hou, Kewei, and Roger K. Loh, 2016, Have we solved the idiosyncratic volatility puzzle?, *Journal of Financial Economics* 121, 167–194.

- Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2018, Replicating anomalies, forthcoming *Review of Financial Studies*.
- Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *Journal of Finance* 45, 881–898.
- Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.
- Jegadeesh, Narasimhan, and Sheridan Titman, 2001, Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* 56, 699–720.
- Menzly, Lior, and Oguzhan Ozbas, 2010, Market segmentation and cross-predictability of returns, *Journal of Finance* 65, 1555–1580.
- Moskowitz, Tobias J., and Mark Grinblatt, 1999, Do industries explain momentum?, *Journal of Finance* 54, 1249–1290.
- Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Peng, Lin, and Wei Xiong, 2006, Investor attention: Overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.
- Petersen, Mitchell, 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.
- Porter, Michael E., 1998, *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, second edition (New York: Free Press).
- Rapach, David E., Jack K. Strauss, Jun Tu, and Guofu Zhou, 2015, Industry interdependencies and cross-industry return predictability, Working Paper.
- Thompson, Samuel B., 2011, Simple formulas for standard errors that cluster by both firm and time, *Journal of Financial Economics* 99, 1–10.

Table 1:

## Industry List

The table presents the industry list under industry classification by Bureau of Economic Analysis (BEA) in 1995.

Industry number	Industry name
1	Livestock and livestock products and other agricultural products
2	Forestry and fishery products
3	Agricultural, forestry, and fishery services
4	Metallic ores mining
5	Coal mining
6	Crude petroleum and natural gas
7	Nonmetallic minerals mining
8	Construction
9	Ordinance and accessories
10	Food and kindred products
11	Tobacco products
12	Broad and narrow fabrics, yarn and thread mills
13	Miscellaneous textile goods and floor coverings
14	Apparel
15	Miscellaneous fabricated textile products
16	Lumber and wood products
17	Furniture and fixtures
18	Paper and allied products, except containers
19	Paperboard containers and boxes
20	Newspapers and periodicals and other printing and publishing
21	Industrial and other chemicals and Agricultural fertilizers
22	Plastics and synthetic materials
23	Drugs, Cleaning and toilet preparations
24	Paints and allied products
25	Petroleum refining and related products
26	Rubber and miscellaneous plastics products
27	Footwear, leather, and leather products
28	Glass and glass products
29	Stone and clay products
30	Primary iron and steel manufacturing
31	Primary nonferrous metals manufacturing
32	Metal containers
33	Heating, plumbing, and fabricated structural metal products
34	Screw machine products and stampings
35	Other fabricated metal products
36	Engines and turbines
37	Farm, construction, and mining machinery

**Table 1 – *continued from previous page***

Industry number	Industry name
38	Materials handling machinery and equipment
39	Metalworking machinery and equipment
40	Special industry machinery and equipment
41	General industrial machinery and equipment
42	Miscellaneous machinery, except electrical
43	Computer and office equipment
44	Service industry machinery
45	Electrical industrial equipment and apparatus
46	Household appliances
47	Electric lighting and wiring equipment
48	Audio, video, and communication equipment
49	Electronic components and accessories
50	Miscellaneous electrical machinery and supplies
51	Motor vehicles, Truck and bus bodies, trailers, and motor vehicles parts
52	Aircraft and parts
53	Other transportation equipment
54	Scientific and controlling instruments
55	Ophthalmic and photographic equipment
56	Miscellaneous manufacturing
57	Railroads, motor freight, water and air transportation, pipelines
58	Communications, except radio and TV
59	Radio and TV broadcasting
60	Electric services, gas distribution, water and sanitary services
61	Retail and wholesale
62	Finance and insurance
63	Owner-occupied dwellings, real estate and royalties
64	Hotels and lodging places, personal and repair services (except auto)
65	Computer, legal, engineering, accounting services and advertising
66	Eating and drinking places
67	Automotive repair and services
68	Amusements
69	Health, educational and social services, and membership organizations

Table 2:

Summary Statistics

The table presents summary statistics for our sample. The number of stocks in the sample is computed every month. The statistics for the number of links between industries represent the 10% quantile, the median, and the 90% quantile of the cross-sectional average computed every year. A link is counted when the relative strength computed according to Equation (2) is greater than or equal to 0.05. The correlations of the industry relatedness measures based on BEA input-output data and Compustat segment reporting data are computed once a year, and the table shows their 10% quantile, median, and 90% quantile. The fourth panel reports the cross-sectional median of the average auto-correlation coefficient of the industry relatedness. For every industry, we compute the correlation coefficient of its industry relatedness at times  $t$  and  $t + \tau$  ( $\tau = 1, 3, 6$  years). Starting in year 1, we then move forward by  $\tau$  years and compute the next correlation coefficient. The table shows the cross-sectional median of the average of these  $\tau$ -period correlation coefficients. Every month we run a cross-sectional regression of the combined on the vertical signal. The table shows the time-series average of the cross-sectional correlation coefficient, of the slope coefficient in the regression and the associated  $R^2$ , and the share of the cases in which the null hypothesis of a zero intercept and a slope equal to one is rejected.

	Industries		
BEA industry classification	Sample 1972-2012	Number of industries before 1997 69	Number of industries since 1997 65
Number of stocks	10% Quantile 2664	Median 3265	90% Quantile 4241
	Number of Links		
	10% Quantile	Median	90% Quantile
BEA inflow-outflow data	3.23	4.75	6.25
Compustat segment data	4.38	6.80	9.32
	Correlation Between Combined and Vertical Relatedness		
	10% Quantile	Median	90% Quantile
Pearson correlation	0.183	0.439	0.771
Spearman correlation	0.250	0.428	0.634
	Persistence of Combined and Vertical Relatedness		
	1 year lag	3 years lag	6 years lag
BEA inflow-outflow data	0.99	0.98	0.96
Compustat segment data	0.95	0.89	0.83
	Combined vs. Vertical Signal		
	Correlation coefficient	Slope coefficient ( $R^2$ )	Test for unbiasedness rejected
... based on excess returns	0.501	0.524(28%)	82.6%
... based on factor model residuals	0.502	0.506(26%)	84.0%



Table 3:

## Correlation in Fundamentals

The table reports the stock- and industry-level panel regressions of annual return on assets ( $ROA$ ) on contemporaneous market-wide  $ROA$  and  $ROA$  in industries which are connected via combined, horizontal, and vertical links. The industry classification scheme is taken from the input-output surveys of the BEA.  $ROA$  is measured as the ratio of net income to total asset from Compustat. Market-wide and industry  $ROA$  is calculated by aggregating firm-level  $ROA$  using firm assets as weights. The  $ROA$  of combined and vertically related industries is calculated by weighting the industry  $ROA$  by the strength of links from Compustat segment data described in Section 3 and by the inter-industry input-output tables in the BEA survey. The  $ROA$  of horizontally related industries is computed as the residual from the cross-sectional regression of  $ROA$  of combined on vertically related industries. Stock and industry fixed effects is included in models with stock- and industry-level specifications in columns 1 to 3 and 4 to 6. The  $t$ -statistics are reported in parentheses. Standard errors are adjusted for heteroskedasticity and double clustered by stock and year in column 1 to 3 and by industry and year in column 4 to 6. \*\*\*, \*\* and \* indicate that the coefficient is different from zero at 1%, 5%, and 10% level of significance, respectively. The sample period is from 1977 to 2012 with 272217 observations in column 1 to 3 and 2420 observations in column 4 to 6.

	Firm level			Industry level		
Model	(1)	(2)	(3)	(4)	(5)	(6)
$ROA_{market}$	0.623*** (6.41)	0.581*** (5.32)	0.753*** (6.87)	0.311*** (5.37)	0.298*** (4.27)	0.333*** (6.20)
$ROA_{combined}$	0.748*** (3.82)			0.648*** (4.32)		
$ROA_{horizontal}$		0.553*** (4.12)			0.441*** (3.34)	
$ROA_{vertical}$			0.841*** (4.67)			0.733*** (3.59)
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Clustered Standard Errors	Yes	Yes	Yes	Yes	Yes	Yes
$R^2$	0.528	0.218	0.362	0.327	0.139	0.228

Table 4:

Return predictability (signals based on excess returns)

The table presents the time-series average of the coefficients estimated from the monthly cross-sectional regression (5)

$$r_{i,t} = \alpha_t + \lambda_t \tilde{r}_{i,t-1}^{(k)} + \Delta_t' Z_{i,t-1} + \epsilon_{i,t},$$

where  $r_{i,t}$  is the excess return on stock  $i$  in month  $t$ .  $\tilde{r}_{i,t-1}^{(k)}$  is the vertical, horizontal, or combined predictive return signal for the industry which stock  $i$  belongs to at time  $t-1$ . The computation of the signals based on industry excess returns is described in Section 3. The industry classification is taken from the BEA input-output tables.  $Z_{i,t-1}$  is a vector of controls containing stock  $i$ 's short-term reversal (the return on the stock over the previous month), its momentum (the average return on the stock from month  $t-12$  to month  $t-2$ ), the one-month lagged return of the industry which firm  $i$  belongs to, and an interaction term between the predictive signal and a dummy variable equal to 1 if the stock's market capitalization is below the median of the sample, and zero otherwise. The  $t$ -statistics are adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\*, \* and \* indicate significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1977 to 2013 with 432 monthly observations.

Model	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Constant	0.008*** (2.97)	0.010*** (4.01)	0.006** (2.48)	0.006** (2.25)	0.007*** (2.60)	0.008*** (3.45)	0.007*** (2.72)
Predictive signal							
combined	0.317*** (5.99)			0.306*** (6.09)	0.146*** (2.89)		
horizontal		0.342*** (6.10)				0.219*** (4.14)	
vertical			0.204*** (4.70)	0.025 (0.68)			0.020 (0.54)
Short-term reversal					-0.040** (-9.98)	-0.041*** (-10.04)	-0.039*** (-9.84)
Momentum					0.059*** (3.67)	0.061*** (3.75)	0.058*** (3.59)
Lagged industry return					0.076*** (7.99)	0.094*** (9.04)	0.099*** (9.83)
Predictive signal * size dummy					0.112*** (3.85)	-0.066 (-1.48)	0.120*** (4.46)
$R^2$	0.0092	0.0066	0.0068	0.0117	0.0327	0.0290	0.0333

Table 5:

Return predictability (signals based on factor model residuals)

The table presents the time-series average of the coefficients estimated from the monthly cross-sectional regression (5)

$$r_{i,t} = \alpha_t + \lambda_t \tilde{r}_{i,t-1} + \Delta_t' Z_{i,t-1} + \epsilon_{i,t},$$

where  $r_{i,t}$  is the excess return on stock  $i$  in month  $t$ .  $\tilde{r}_{i,t-1}$  is the vertical, horizontal, or combined predictive return signal for the industry which stock  $i$  belongs to at time  $t-1$ . The computation of the signals based on factor model residuals is described in Section 3. The industry classification is taken from the BEA input-output tables.  $Z_{i,t-1}$  is a vector of controls containing stock  $i$ 's short-term reversal (the return on the stock over the previous month), its momentum (the average return on the stock from month  $t-12$  to month  $t-2$ ), the one-month lagged return of the industry which firm  $i$  belongs to, and an interaction term between the predictive signal and a dummy variable equal to 1 if the stock's market capitalization is below the median of the sample, and zero otherwise. The  $t$ -statistics are adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\* and \* indicate significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1977 to 2013 with 432 monthly observations.

Model	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Constant	0.010*** (3.82)	0.010*** (4.02)	0.010*** (4.04)	0.010*** (3.97)	0.009*** (3.79)	0.008*** (3.62)	0.009*** (3.77)
Predictive signal							
combined	0.256*** (4.98)			0.246*** (5.02)	0.104** (2.08)		
horizontal		0.274*** (5.13)				0.132** (2.53)	
vertical			0.154*** (3.47)	0.016 (0.42)			0.008 (0.18)
Short-term reversal					-0.040*** (-10.01)	-0.040*** (-9.99)	-0.040*** (-9.96)
Momentum					0.060*** (3.72)	0.060*** (3.70)	0.060*** (3.71)
Lagged industry return					0.098*** (8.68)	0.104*** (8.90)	0.110*** (10.02)
Predictive signal * size dummy					-0.0035 (-0.08)	-0.054 (-1.05)	0.030 (0.70)
$R^2$	0.0077	0.0053	0.0059	0.0103	0.0307	0.0289	0.0304

Table 6:

## Long-short portfolio strategies (residual signals)

The table reports the average excess returns and factor alphas for quintile portfolios formed every month by sorting industries with respect to the vertical, horizontal, or combined residual signal at time  $t-1$ . Q1 (Q5) is the quintile portfolio of industries with lowest (highest) signal value, and Q5-Q1 denotes the portfolio long the highest and short the lowest quintile. The computation of the residual signals is described in Section 4. The industry classification is taken from the BEA input-output tables. The CAPM (four factor) alpha is the intercept from a regression of quintile or high-low portfolio excess returns on the market excess return (on the three Fama-French factors plus the momentum factor UMD). The  $t$ -statistics are adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\* and \* indicate that the coefficient is different from zero at 1%, 5%, and 10% level of significance, respectively. The sample period is from 1977 to 2013 with 432 monthly observations.

	Q1(Low)	Q5(High)	Q5-Q1
A: Combined predictive signal			
Excess return	0.86*** (3.26)	1.12*** (4.26)	0.27*** (2.68)
CAPM alpha	0.21 (1.40)	0.48*** (3.15)	0.27*** (2.66)
Four-factor alpha	0.09 (0.90)	0.27*** (2.78)	0.18* (1.92)
B: Horizontal predictive signal			
Excess return	0.89*** (3.25)	1.10*** (4.32)	0.21*** (2.47)
CAPM alpha	0.24 (1.58)	0.47*** (3.23)	0.23*** (2.68)
Four-factor alpha	0.07 (0.67)	0.27*** (2.96)	0.20** (2.35)
C: Vertical predictive signal			
Excess return	0.96*** (3.62)	1.11*** (4.14)	0.15 (1.41)
CAPM alpha	0.33** (1.98)	0.47*** (3.45)	0.14 (1.35)
Four-factor alpha	0.15 (1.45)	0.28*** (3.04)	0.13 (1.19)

Table 7:

## Factor exposures of Q5-Q1 portfolios (residual signals)

The table reports the factor exposure of the long-short (Q5-Q1) portfolios. Quintile portfolios are formed each month by sorting the industries with respect to the combined, horizontal, or vertical residual signal. The computation of the residual signal is described in Section 4. We use the BEA industry classification. The vertical and the combined predictive signal are derived from the BEA input-output tables and Compustat segment data as described in Section 3. The industries are then sorted into quintile portfolios based on their respective residual signal. Q1 (Q5) denotes the portfolio of industries with the lowest (highest) signal, and Q5-Q1 is the self-financing portfolio long Q5 and short Q1. The Q5-Q1 portfolio returns are regressed on only the market excess return MKT (columns [1], [3], and [5]) and on MKT, SMB, HML, and the momentum factor UMD (columns [2], [4], and [6]). The  $t$ -statistics are adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\* and \* indicate significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1977 to 2013 with 432 monthly observations.

Signal	Combined		Horizontal		Vertical	
Model	[1]	[2]	[3]	[4]	[5]	[6]
Alpha	0.272*** (2.66)	0.179* (1.92)	0.227*** (2.68)	0.199** (2.35)	0.142 (1.35)	0.128 (1.19)
MKT	-0.0007 (-0.33)	0.030 (1.43)	-0.031 (-1.62)	-0.014 (-0.61)	0.006 (0.22)	0.031 (1.09)
SMB		-0.028 (-0.69)		-0.036 (-0.74)		-0.079** (-2.00)
HML		0.119*** (3.27)		0.042 (1.29)		0.049 (1.23)
UMD		0.060*** (3.34)		0.019 (1.07)		0.001 (0.03)
$R^2$	-0.0021	0.0295	0.0039	0.0068	-0.0021	0.0135

Table 8:

## Institutional ownership and return predictability (residual signals)

The table presents the time-series average of the monthly coefficient from a regression of returns on the interaction terms between the predictive return signal and a dummy variable representing the quintile of institutional ownership in the previous quarter. Panel A shows the results for the full sample, while Panel B presents the coefficients for a subsample excluding the 20 percent of smallest stocks. Institutional ownership is defined as the number of shares held by institutional investors divided by the total number of shares outstanding. 'Q1' denotes the dummy variable equal to 1 for the stocks in the quintile of lowest institutional ownership, with an analogous definition for 'Q5'. The residual signal is computed in two steps. First, we take the residuals from a time-series regression of industry returns on a standard four-factor pricing model with the market, SMB, HML and UMD as factors as described in Section 3. Then we cross-sectionally regress these residuals once more on reversal, momentum, and lagged industry returns as described in Section 4. The industry classification is taken from the BEA input-output tables. The  $t$ -statistics are adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\* and \* indicate significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1980 to 2013 with 396 monthly observations.

	Signal					
	combined	horizontal	vertical	combined	horizontal	vertical
	A: All stocks			B: Without 20% smallest stocks		
Constant	0.013*** (5.14)	0.013*** (5.18)	0.013*** (5.17)	0.015*** (5.78)	0.015*** (5.84)	0.015** (5.17)
Signal * Q1	0.136* (1.95)	0.144* (1.67)	0.186*** (2.72)	0.153** (2.22)	0.231*** (2.97)	0.074 (0.99)
Signal * Q2	0.117** (2.09)	0.125** (2.30)	0.050 (0.92)	0.119* (1.92)	0.132** (2.41)	0.057 (0.92)
Signal * Q3	0.124** (2.16)	0.172*** (3.18)	-0.005 (-0.1)	0.142** (2.27)	0.205*** (3.40)	-0.006 (-0.11)
Signal * Q4	0.141** (2.33)	0.183*** (3.16)	0.024 (0.48)	0.142** (2.31)	0.167*** (2.99)	0.041 (0.79)
Signal * Q5	0.091 (1.45)	0.055 (0.91)	0.076 (1.34)	0.102 (1.55)	0.074 (0.22)	0.077 (1.29)
$R^2$	0.006	0.0054	0.0050	0.006	0.0048	0.0051

Table 9:

## Change in institutional ownership

The table reports the results of panel regressions in which the quarterly changes in institutional ownership at the stock level are regressed on the contemporaneous institutional ownership in related industries (Panel A) and on the combined, horizontal, and vertical residual signals (Panel B). The change of institutional ownership is the current quarter change in the relative holdings of institutional investors, where these relative holdings are defined as the number of shares held institutionally divided by the total number of shares outstanding. The procedure to compute the combined, horizontal, and vertical change in institutional ownership is fully analogous to the one applied to compute the return predictors. The residual signals are computed in two steps. First, we take the residuals from a time-series regression of industry returns on a standard four-factor pricing model with the market, SMB, HML and UMD as factors as described in Section 3. Then we cross-sectionally regress these residuals once more on reversal, momentum, and lagged industry returns as described in Section 4. Vertical relatedness is measured based on BEA input-output tables, combined relatedness is computed on the basis of Compustat segment data, and horizontal relatedness is obtained as the residual from a regression of the combined on the horizontal measure. All specifications include stock and year-quarter fixed effects. *t*-statistics are reported in parentheses. The standard errors are adjusted for clustering at the industry-year level. \*\*\*, \*\* and \* indicate significance at the 1%, 5% and 10% level, respectively. The sample period is from 1980 to 2012.

Panel A: regressing on change in institutional ownership			
Regressor			
$\Delta IO_{\text{combined}}$	0.503*** (12.06)		
$\Delta IO_{\text{horizontal}}$		0.592*** (7.10)	
$\Delta IO_{\text{vertical}}$			0.260*** (5.02)
$R^2$	0.0119	0.0032	0.0102
Obs	365,506		
Panel B: regressing on residual signals			
Regressor			
combined	0.092*** (4.48)		
horizontal		0.089*** (3.97)	
vertical			0.048*** (2.70)
$R^2$	0.0010	0.0008	0.0002
Obs	352,216		

Table 10:

## Robustness: Alphas of long-short portfolio strategies (residual signals)

The table reports alphas for quintile portfolios formed every month by sorting industries with respect to the vertical, horizontal, or combined residual signal. Q1 (Q5) is the quintile portfolio of industries with lowest (highest) signal value, and Q5-Q1 denotes the portfolio long the highest and short the lowest quintile. The computation of the residual signals is described in Section 4. The industry classification is taken from the BEA input-output tables. The alphas of the portfolios are computed relative to the models proposed by Fama and French (2015) (five factors) and Hou et al. (2015) (four factors) and are obtained as the intercepts of regressions of the respective portfolio (excess) returns on the factors featured in the respective model. The numbers in parentheses represent  $t$ -statistics adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\* and \* indicate that the coefficient is different from zero at 1%, 5%, and 10% level of significance, respectively. The sample period is from 1977 to 2013 with 432 monthly observations.

Alpha relative to ...	Q1(Low)	Q5(High)	Q5-Q1
A: Combined			
Fama and French (2015)	-0.08 (-0.94)	0.16** (2.02)	0.24** (2.48)
Hou et al. (2015)	-0.06 (0.90)	0.17** (2.18)	0.23** (2.42)
B: Horizontal			
Fama and French (2015)	-0.04 (-0.58)	0.15** (1.98)	0.19** (2.26)
Hou et al. (2015)	0.01 (0.23)	0.21** (2.32)	0.20** (2.24)
C: Vertical			
Fama and French (2015)	-0.02 (-0.38)	0.12 (1.39)	0.14 (1.45)
Hou et al. (2015)	0.03 (0.37)	0.18* (1.69)	0.15 (1.43)



Table 11:

Robustness: Return predictability for SIC codes (signals based on factor model residuals)

The table presents the time-series average of the coefficients estimated from the monthly cross-sectional regression (5)

$$r_{i,t} = \alpha_t + \lambda_t \tilde{r}_{i,t-1}^{(k)} + \Delta_t' Z_{i,t-1} + \epsilon_{i,t},$$

where  $r_{i,t}$  is the excess return on stock  $i$  in month  $t$ .  $\tilde{r}_{i,t-1}^{(k)}$  is the combined predictive return signal for the industry which stock  $i$  belongs to at time  $t - 1$ . The computation of the signals based on factor model residuals is described in Section 3. The industry classification is given by 2-digit SIC (Panel A) or 3-digit SIC (Panel B).  $Z_{i,t-1}$  is a vector of controls containing stock  $i$ 's short-term reversal (the return on the stock over the previous month), its momentum (the average return on the stock from month  $t - 12$  to month  $t - 2$ ), the one-month lagged return of the industry which firm  $i$  belongs to, and an interaction term between the predictive signal and a dummy variable equal to 1 if the stock's market capitalization is below the median of the sample, and zero otherwise. The  $t$ -statistics are adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\* and \* indicate significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1977 to 2013 with 432 monthly observations.

	Panel A: 2-digit SIC		Panel B: 3-digit SIC	
	[1]	[2]	[3]	[4]
Constant	0.011*** (4.24)	0.009*** (3.88)	0.012*** (4.58)	0.010*** (4.07)
Predictive signal	0.234*** (5.02)	0.072* (1.67)	0.147*** (5.63)	0.127*** (4.97)
Short-term reversal		-0.034*** (-8.80)		-0.037*** (-9.71)
Momentum		0.053*** (4.95)		0.053*** (4.98)
Lagged industry return		0.105*** (8.31)		0.069*** (9.82)
Predictive signal * size dummy		-0.007 (-0.15)		-0.041 (-1.55)
$R^2$	0.0065	0.0272	0.0045	0.0247

Table 12:

Robustness: Long-short portfolio strategy based on SIC codes  
(residual signals)

The table reports the average excess returns and factor alphas for quintile portfolios formed every month by sorting industries with respect to the vertical, horizontal, or combined residual signal at time  $t - 1$ . Q1 (Q5) is the quintile portfolio of industries with lowest (highest) signal value, and Q5-Q1 denotes the portfolio long the highest and short the lowest quintile. The computation of the residual signals is described in Section 4. The industry classification is based on 2-digit SIC (Panel A) and 3-digit SIC (Panel B).

The CAPM (four factor) alpha is the intercept from a regression of quintile or high-low portfolio excess returns on the market excess return (on the three Fama-French factors plus the momentum factor UMD). The  $t$ -statistics are adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\* and \* indicate that the coefficient is different from zero at 1%, 5%, and 10% level of significance, respectively. The sample period is from 1977 to 2013 with 432 monthly observations.

	Q1(Low)	Q5(High)	Q5-Q1
Panel A: 2-digit SIC			
Excess return	0.96*** (3.44)	1.32*** (4.95)	0.35*** (2.96)
CAPM alpha	0.33* (1.83)	0.67*** (4.32)	0.34*** (2.95)
Four factor alpha	0.20 (1.61)	0.46*** (4.85)	0.27** (2.16)
Panel B: 3-digit SIC			
Excess return	0.95*** (3.41)	1.43*** (5.31)	0.48*** (5.88)
CAPM alpha	0.31* (1.82)	0.79*** (5.05)	0.48*** (5.96)
Four factor alpha	0.15 (1.57)	0.57*** (5.32)	0.42*** (4.97)

Table 13:

Robustness: Factor exposures of Q5-Q1 portfolios based on SIC codes  
(residual signals)

The table reports the factor exposure of the Q5-Q1 portfolios. Quintile portfolios are formed every month by sorting the industries with respect to the combined industry return predictor computed according to Equation (1) using 2-digit (regressions [1] and [2]) or 3-digit SIC codes (regressions [3] and [4]). The strength of the links between the industries is derived from Compustat segment data. The Q5-Q1 portfolio returns are regressed on only the market excess return MKT (regressions [1] and [3]) or on MKT, SMB, HML, and the momentum factor UMD (regressions [2] and [4]). The  $t$ -statistics are adjusted for autocorrelation and heteroskedasticity following Newey and West (1987). \*\*\*, \*\* and \* indicate significance at the 1%, 5%, and 10% level, respectively. The sample period is from 1977 to 2013.

	Panel A: 2-digit SIC		Panel B: 3-digit SIC	
	[1]	[2]	[3]	[4]
Alpha	0.341*** (2.95)	0.267** (2.16)	0.480*** (5.96)	0.419*** (4.97)
MKT	0.017 (0.68)	0.036 (1.37)	-0.003 (-0.12)	0.014 (0.65)
SMB		-0.033 (-0.75)		-0.018 (-0.48)
HML		0.014 (0.34)		0.032 (0.76)
UMD		0.098** (2.25)		0.067* (1.96)
$R^2$	-0.0013	0.0245	-0.0023	0.0195
Obs	432	432	432	432

Table 14:

## Robustness: Change in institutional ownership (Fama-MacBeth regression)

The table reports the results of Fama and MacBeth (1973) regressions in which the quarterly changes in institutional ownership at the stock level are regressed on the contemporaneous institutional ownership in related industries. The change of institutional ownership is the current quarter change in the relative holdings of institutional investors, defined as the number of shares held institutionally divided by the total number of shares outstanding. The procedure to compute the combined, horizontal, and vertical change in institutional ownership is fully analogous to the one applied to compute the return predictors. Vertical relatedness is measured based on BEA input-output tables, combined relatedness is computed on the basis of Compustat segment data, and horizontal relatedness is obtained as the residual from a regression of the combined on the horizontal measure.  $t$ -statistics are reported in parentheses. \*\*\*, \*\* and \* indicate significance at the 1%, 5% and 10% level, respectively. The sample period is from 1980 to 2012.

Panel A: regressing on change in institutional ownership			
Regressor			
$\Delta IO_{\text{combined}}$	0.701*** (13.15)		
$\Delta IO_{\text{horizontal}}$		0.754*** (13.49)	
$\Delta IO_{\text{vertical}}$			0.391*** (9.00)
$R^2$	0.0048	0.0040	0.0022
Panel B: regressing on residual signals			
Regressor			
$r_{\text{combined}}$	0.046*** (3.97)		
$r_{\text{horizontal}}$		0.040*** (3.41)	
$r_{\text{vertical}}$			0.027** (2.32)
$R^2$	0.0012	0.0008	0.0009

Figure 1:

An example for vertical versus combined industry relatedness

The figure shows the degree to which various industries were related to the 'Metal containers' industry in 1995. On the horizontal axis, we plot the coefficient for vertical (customer-supplier) relatedness, as measured by the strength of the input-output good and service flows contained in the BEA tables. The vertical axis represents the degree of combined relatedness derived from the Compustat segment data. The coefficients are normalized such that they sum to one for both vertical and combined relatedness. The industries shown in the graph are those with combined or vertical coefficients greater than or equal to 0.05.

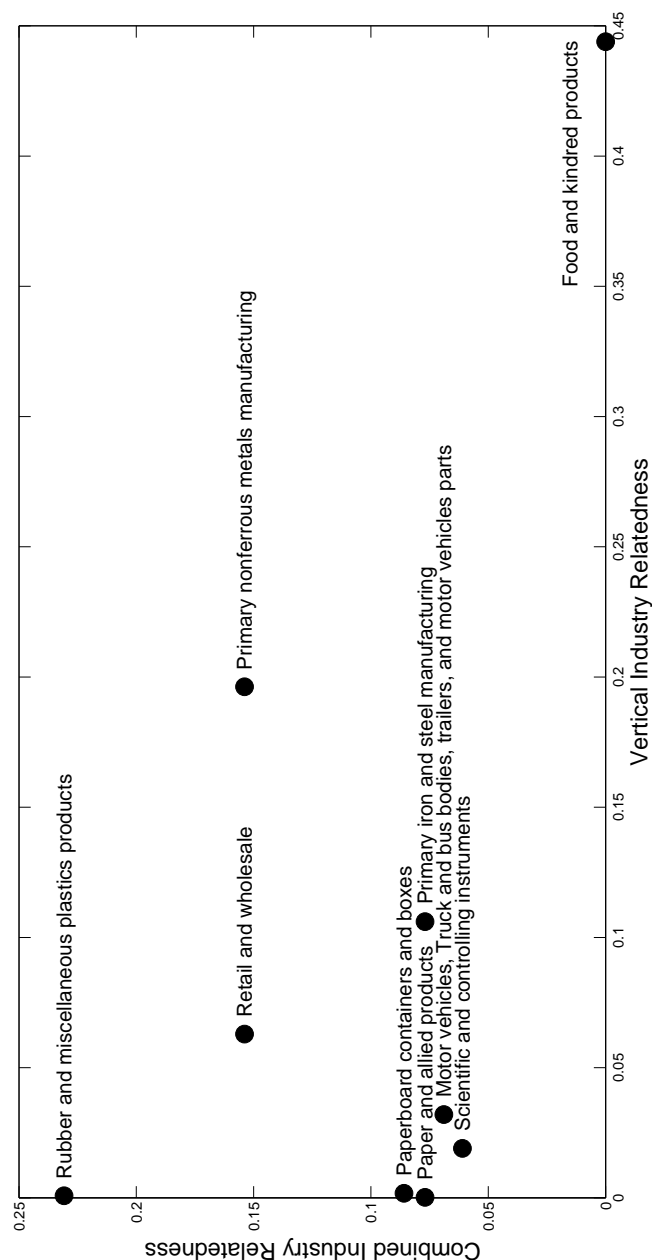


Figure 2:

### Horizontally related industries

The figure shows industries that are horizontally related to each of the industries in a BEA classification in 1995. To identify the horizontally related industries, we examine the degree of combined relatedness derived from the Compustat segment data and vertical relatedness measured by the strength of input-output good and service flows contained in the BEA tables. For a given industry in the BEA classification, the horizontally related industries are those with combined relatedness greater than or equal to 0.05, and vertical relatedness less than or equal to 0.01.

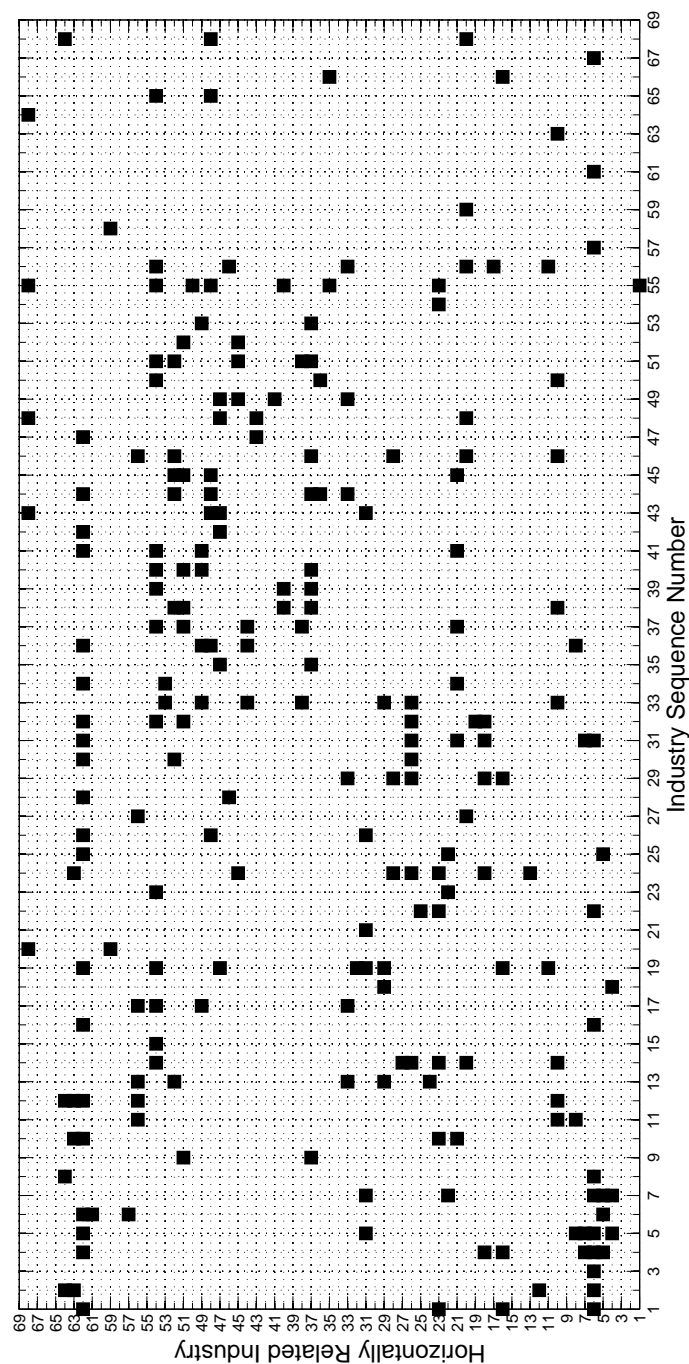


Figure 3:

Average cumulative returns of long-short portfolio strategies over long horizon (residual signals)

The figure shows the average cumulative return for the long-short portfolio strategies based on the vertical, horizontal, or combined residual signal, for 1 month to 12 months after portfolio formation. Long-short strategy denotes the portfolio long the quintile of stocks with highest signal value and short the quintile of stocks with lowest signal value. The computation of the residual signals is described in Section 4. The industry classification is taken from the BEA input-output tables.

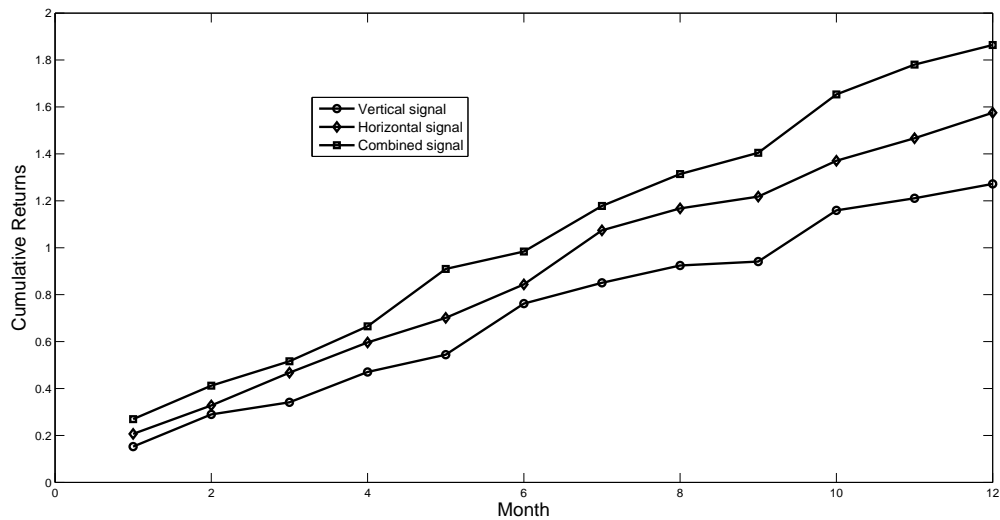
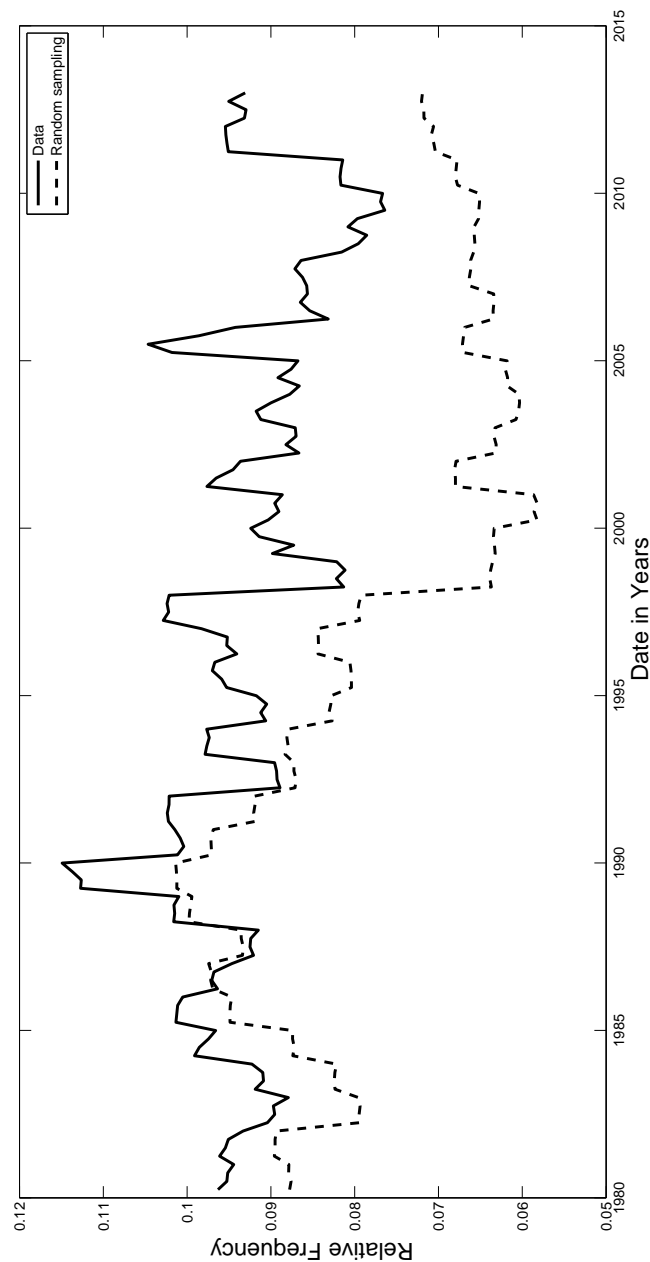


Figure 4:

# Relative frequency of institutional investors holding horizontally related industries

The figure shows the relative frequency of institutional investors holding horizontally related industries over the time. To identify the horizontally related industries, we examine the degree of combined relatedness derived from the Compustat segment data and vertical relatedness measured by the strength of input-output good and service flows contained in the BEA tables. For a given industry in the BEA classification, the horizontally related industries are those with combined relatedness greater or equal to 0.05 and vertical relatedness less or equal to 0.01. In the dashed line, we report the frequency of holding horizontally related industries from random sampling (a bootstrapping method). The number of bootstrapping experiments is 1,000. The data on institutional ownership are available since 1980 and only at quarterly frequency. The sample period is from 1980 to 2012 with 132 observations.





## Recent Issues

No. 255	Silvia Dalla Fontana, Marco Holz auf der Heide, Lorian Pelizzon, Martin Scheicher	The Anatomy of the Euro Area Interest Rate Swap Market
No. 254	Martin R. Goetz	Financing Conditions and Toxic Emissions
No. 253	Thomas Johann, Talis Putnins, Satchit Sagade, Christian Westheide	Quasi-Dark Trading: The Effects of Banning Dark Pools in a World of Many Alternatives
No. 252	Nicole Branger, Patrick Konermann, Christian Schlag	Optimists and Pessimists in (In)Complete Markets
No. 251	Wenhui Li, Christian Wilde	Belief Formation and Belief Updating under Ambiguity: Evidence from Experiments
No. 250	Nathanael Vellekoop, Mirko Wiederholt	Inflation Expectations and Choices of Households
No. 249	Yuri Pettinicchi, Nathanael Vellekoop	Job Loss Expectations, Durable Consumption and Household Finances: Evidence from Linked Survey Data
No. 248	Jasmin Gider, Simon N. M. Schmickler, Christian Westheide	High-Frequency Trading and Price Informativeness
No. 247	Mario Bellia, Lorian Pelizzon, Marti G. Subrahmanyam, Jun Uno, Draya Yuferova	Paying for Market Liquidity: Competition and Incentives
No. 246	Reint Gropp, Felix Noth, Ulrich Schüwer	What Drives Banks' Geographic Expansion? The Role of Locally Non-Diversifiable Risk
No. 245	Charline Uhr, Steffen Meyer, Andreas Hackethal	Smoking Hot Portfolios? Self-Control and Investor Decisions
No. 244	Mauro Bernardi, Michele Costola	High-Dimensional Sparse Financial Networks through a Regularised Regression Model