# Sticky Chains:

## Spillover Effect of Future Operating Shocks on Supply Chain Network

Chan Kim

*Seoul National University*

Se-jik Kim

*Seoul National University*

Jeong Hwan Lee

*Hanyang University*

Current Draft: December 14, 2018

Preliminary and Incomplete

Abstract

This paper examines how the network of supply chain propagates future operating shocks. For this purpose, we first build a network of supply chain by merging customer information directly extracted from 10-K statements in conjunction with the Compustat customer segment database. We also construct a wide range of similarity indices that identify firm-level shocks to future operating outcomes by comparing year-on-year 10-K/Q filings. We then map these future operating shocks into the supply chain network to empirically test their spillover effects. Our empirical analysis shows significantly negative spillover effects of the future operating shocks on firms' revenues at least two connections away from the origin. Our findings contribute to literature by supporting and quantifying externalities and spillovers along firm connections.

Keywords: Spillover Effect; Supply Chain; Future Operating Outcome Shock

**NOTE**

*Type of work: research paper (work in progress)

*Applicant's contribution to the work: The applicant (Chan Kim) was in charge of the empirical research part in this project. To be specific, he constructed and analyzed a supply chain network from customer information in 10-K, and computed the six similarity measures (including the construction of word embedding from 10-K/Q filings) among 10-K/Q filings using python. Also, he performed the regressions in this study using Stata.

## 1. Introduction

Leading the expansion of world trade, supply chains have formed unprecedentedly long and dense relationships at a global level; such relationships are probably the most comprehensive and complex that has ever existed in world history. These chains go beyond simple transfers of goods and services but act as links conveying substantial interdependency. Acemoglu et al. (2012) show that supply chain can function as a mechanism that propagates and enlarges idiosyncratic shocks throughout the economy. In extreme cases, when disorganization of a supply chain occurs, one economy may suffer from recession as harsh as that experienced by transitional economies in the 1990s (Blanchard, 1996). Hence, whether a company in the supply chain can timely react to such expected crisis is not only a matter of profit but also a key to achieve macroeconomic stability. In this sense, firms should comprehend such risk and cope with it by using appropriate policies.

In this paper, we exploit changes in firms' reporting practices to empirically test how companies in the supply chain network react to future outcome shocks from an upstream firm. We rely critically on a firm's 10-K/Q statements in the EDGAR system for the identification of future outcome shocks and the construction of supply chain network. Most of all, we identify future outcome shocks based on changes to phrasing routines by comparing year-on-year 10-K/Q results. To be specific, Cohen et al. (2018) report that changes in 10-K/Q are related to firms' future returns. These changes are likely to contain useful information on risk factor and litigation but are not given with sufficient attention by most investors. We further clarify the characteristics of changes in filing as a shock hinted beforehand. We adopt a similarity measure that can capture semantic (relating to meaning in language) differences among financial statements. Informational change in filings, which is pertinent to shocks, can be precisely captured by excluding the case of simple rephrasing and restructuring.

We also use customer information reported in 10-K statements to build up a network of supply chain. By adopting the Natural Language Processing techniques, we initially extract the name of customers from the 10-K statements. We then expand the network of the supply chain by adding the customer segment database of Compustat. A total of 121,898 customer–supplier relationships were analyzed from 1995 to 2017 for publicly traded U.S. firms.

Our empirical analysis is conducted as follows. We first confirm the validity of our similarity measure by showing its substantial correlation with future shock components. We then map these changes into the supply chain network and examine if such future outcome shocks have spillover effects on the supply linkage. Then we conduct a set of robustness checks to confirm the spillover effects.

Our empirical results show that the future outcome shocks identified by the similarity measures lead to substantial impact on the firms, mostly up to two connections away from the origin and up to three connections away from the origin in some settings. These results are robust to a wide range of similarity measures and to the restriction of our supply chain network to the customer relationship provided by the Compustat segment database.

This paper contributes to existing literature in a number of aspects. First, our work contributes to the burgeoning literature on empirical identification and measurement of the spillover effect. For instance, Kolay and Lemmon (2011) report that shocks from natural disasters are propagated in the supply chain network. Wu (2016) identifies idiosyncratic shocks from financial statements and other news sources and examine their propagation along the supply chain. Our construction of similarity shock comparing 10-K/Q statements is an idiosyncratic shock in nature, and we directly confirm the propagation of this shock in the supply network.

Our results also quantify the magnitude of the spillover effects inside the supply chain. Traditional theories on production networks predict that firm-specific shocks quickly decay after the first link. Our analysis implies that an outcome shock may propagate up to two or three connections away from the origin, contradictory to existing theories.

Our work also contributes to the literature focusing on information contents in a firm's 10-K/Q statements. In particular, our analysis highlights the possibility of propagation of information contents in a firm's annual/quarterly report via a network of supply chain, which is largely unexamined in literature. We also develop new measures comparing routine phrasing changes in year-on-year 10-K/Q filings and confirm the validity of these measures complementary to the study of Cohen et al. (2018).

This work proceeds as follows. Section 2 describes the data and provides summary statistics. Section 3 presents our main empirical results. Section 4 concludes the paper.

## 2. Data and Summary Statistics

We derive the main empirical inferences from a variety of data sources. This section briefly describes the sources and methodology used to construct a supply chain network and a dataset of future operating shocks.

We download all "10-K," "10-K405," "10-KSB," "10-KSB40," and "10-Q," "10-QSB"' filings from 1994 to 2017 with data on “Filing Date” and “Period of Report” from the SEC’s Electronic Data Gathering, Analysis, and Retrieval website. Similar to the study of Cohen et al. (2018), we extract the textual content from the 10-K/Q filings and remove all the tables. To check whether the main idea of their paper is robust enough to hold even in slightly different settings, we define a table as a sentence whose numeric character content is greater than 10% following the method of Loughran and McDonald (2011). In addition, we take preprocessing steps including bigram and lemmatization for improved textual analysis. We obtain data on monthly and quarterly stock returns from the Center for Research in Security Prices (CRSP) and firms’ financial indicator from Compustat.

To empirically examine the spillover effect of firm-level shocks hinted in 10-K/Q filings, we construct a supply chain from two sources. Compustat offers supplier–customer relations among publicly traded firms in the U.S. However, the coverage of the dataset is limited, especially in the years between the 1990s and early 2000s. To supplement the dataset, we extract information on business relationships from 10-K filings. Many companies state customer relationships to inform investors in addition to major customer information, which is mandatorily disclosed in 10-K as required by Securities and Exchange Commission (SEC). However, the form of customer information documented in 10-K is highly heterogeneous among firms. We train a custom Neural Network-based Natural Language Processing (NLP) pipeline to scrape customer relationship. Applying the word embedding built from 10-K/Q filings, we train a sentence classifier that can sort out the sentences that contain customer information and a named entity recognizer that can identify the customer name from these sentences. For instance, Whirlpool Corporation (CIK: 106640) reported in its 10-K filing at 2010: “The loss of or substantial decline in sales to any of our key trade customers, which include Lowe’s, Sears, Home Depot, Casas Bahia, Ikea, major buying groups, and builders, could adversely affect our financial performance.” Our custom NLP pipeline classifies this sentence as one that contains customer information and recognizes Lowe’s (CIK: 60667), Home Depot (CIK: 354950), Casas Bahia, and Ikea as Whirlpool Corporation’s customers. Our model matches these names with company identifier. Given that Casas Bahia and Ikea are not U.S.-based companies, our model generates two supplier–customer relationships: “Supplier: 106640, Customer: 60667, Year: 2010” and “Supplier: 106640, Customer: 354950, Year: 2010.” However, considering the restrictions on computational resources, we only add about 1,743 relationships of 1,000 companies each year on average in this paper. The resulting dataset contains 121,898 unique relations, covering 7,669 publicly traded firms in the United States from 1994 to 2017. Details of extracting major customer information can be found in Online Appendix.

We develop new measures that capture changes in the routines of phrasing among 10-K/Q statements. Cohen et al. (2018) construct four types of similarity measures: i.) cosine similarity, ii.) Jaccard similarity, iii.) simple similarity, and iv.) minimum edit distance. These measures do not consider the semantic characteristics of words. For instance, these measures deem "liability" and "debt" as different as "liability" and "asset." In this regard, we construct a new word vector cosine similarity measure (I, II). This measure captures the semantic (and syntactic) relationships among words by constructing word embeddings based on the co-occurrences of words, where every word in the trained model corresponds to a vector sized 1× the size of feature dimensions. For example, the link such as "Man is to Woman as King is to Queen," whose meaning is obvious in literal terms, can be imitated by algebraic operations of vector representations: "King" – "Man" + "Woman" = "Queen."

By adopting the similarity measure that is sensitive to changes in the meaning of documents, we attempt to capture significant changes in the subjects of filings, rather than a simple rephrasing or replacement of wording. Numerous studies on computer sciences have tried to contrive improved text similarity measures, but these state-of-the-art technologies mostly focus on sentences or short documents. Hence, word vector methods would still be a reasonable baseline model for measuring similarity among long documents. We train the custom word embeddings on preprocessed 10-K/Q filings by using the python Gensim library word2vec model. We calculate the word2vec similarity of two documents by calculating the cosine similarity of two vectors representing each of them. We construct a document vector by taking the average of word vectors in the documents. We adopt two sets for the parameters to train the word vectors. For parameter set I, under the conjecture that the main contents of filings are well represented by frequently used words in the financial statement, we include a big feature size with a high minimum count threshold in the vocabulary set (min_count = 400, window = 20, size = 300, CBOW, Gensim (python NLP library) default). For parameter set II, we use the settings asserted to be well-performing on Biomedical NLP context in line with the reports of Chiu et al. (2016).

Each column in Table 1 shows the most similar words to "litigation," "competition," "profit," "customer" in the word embedding trained on 10-K/Q filings with parameter set II. The output words are strongly related to the input words with respect to their meaning. For this reason, using word vector cosine similarity might be superior in capturing changes in meaning compared with other similarity measures proposed by Cohen et al. (2018). For example, suppose a sentence of filing is changed from "The Company may experience substantial competition from other companies" to "The Company may be faced with competitive pressure in the marketplace from large brand name competitors." Considering that these two sentences are fundamentally the same in their meaning, the optimal similarity measure should evaluate the similarity between the two sentences as a value close to 1. The similarity calculated by word vector similarity (0.6859) may not be sufficiently high, but this value is bigger than other similarities [cosine similarity (0.4042), Jaccard similarity (0.2), min edit similarity (0.3333), simple similarity (0.3333)]. This advantage, however, may quickly vanish as the size of the document being compared increases because averaging a hundred thousand vectors does not guarantee to produce a meaningful representation of a document.

For each of these six measures, we calculate similarities between filings in the following ways. First, we set a fiscal year for every firm in our dataset based on its Period of Date information on 10-K. We then compute similarity between a filing and a filing that reported four quarters before. For instance, 2015 Q1 10-Qs are compared with 2014 Q1 10-Qs. However, to remove the irregularity originated from the changes in the filing date or fiscal year, we leave out the samples where fiscal year is modified or where the difference of reporting date between two filings are larger than 14 months or smaller than 10 months. Table 2 provides the descriptive statistics for the set of similarity measures. Panel A reports the mean, standard deviation, quartile values, and minimum and maximum values of the six similarity measures. Panel B presents the pairwise correlation coefficients among the six similarity measures. Sim_Cosine is the cosine similarity; Sim_Jaccard is the Jaccard similarity; Sim_Min is the minimum edit distance; Sim_WV1 is the word vector cosine similarity I (with parameter setting I); and Sim_WV2 is the word vector cosine similarity II. Panel A shows that our similarity measures based on word vector method are generally positioned at higher level, with relatively larger variations, compared with the four measures of Cohen et al. (2018). Such a large variation implies a potentially different role of our similarity measures against their measures. Panel B presents the correlation among the similarity measures. The similarities calculated by two different word embeddings are highly correlated, and strong correlation structures are observed among the six measures. Generally large correlations among the measures argue for the validity of our measure constructions.

## 3. Spillover Effect of Firm-specific Future Operating Shocks in Network

This section presents the result of the empirical test on the spillover effect of future outcome shocks on the supply chain network. In section A, we examine whether our measures of similarity play a role of future outcome shocks as highlighted in the study of Cohen et al. (2018). In section B, we examine the spillover effect of such future outcome shocks on future operating performances of firms that are connected to the origin of shock.

### A. Changes in Reporting Behavior as Firm-specific Future Operating Shocks

Cohen et al. (2018) show that firms' decision to change the language and construction of their SEC filings is related to poor performance/profitability in the future and even increases the probability of bankruptcy. In other words, breaking from routine phrasing and content in 10-K/Q filings is a profound indication of the future operation of firms, predominantly those that bring negative outcomes. In this context, changes in reporting behavior can act as firm-specific idiosyncratic future operating shocks. However, the mechanisms working under these changes are not clearly verified yet. In this section, we illuminate the implication of such phrasing changes. We try to differentiate the effects caused by changes in the actual meaning of 10-K/Q filings from those caused by simple adjustments or rephrases on filings. To isolate changes in meaning from the total variations (changes in meaning, structure or rephrase, etc.) among filings, we adopt a similarity measure based on word vectors, which is known to be sensitive to changes in semantic meaning.

The Fama–MacBeth cross-sectional regression results (Table 3) show the effects of changes in filings on the firms' 12-month cumulative returns. The coefficients of word vector methods seem lower than those of other similarity measures. However, considering their relatively high variations, the magnitudes are bigger than the other measures, except for Jaccard similarity. For example, a one-standard deviation decline in a firm's filings similarity across years measured on word vector lead to 103 basis point lowered stock return after a year. Similar to the findings of Cohen et al. (2018) that a significant change to filing routine is a significant predictor of low future operating performance, Table 4 shows that the changes in meanings captured by similarity measures are related to the firms' future operating performance in a negative direction. To be specific, we define the firm's operating performance by three measures following Cohen et al. (2018): OI, NI, and SA, which represent the operating income before depreciation (Oidbpq), net income (Niq), and sales (Saleq), respectively, each divided by lagged total assets (L1.atq). All regressions in Table 4 include month, industry, and firm fixed effect. The results show that the low similarity with the preceding year's filing calculated by similarity measures is associated with poor performance after two quarters. The relatively considerable magnitude and statistically significant coefficients of word vector similarities in Tables 3 and 4 imply that the changes in 10-K/Q found by Cohen et al. (2018) are likely to capture the actual adjustments in meaning, such as newly noted potential crisis. In this sense, relating a change in 10-K/Q filings that occurred in a firm to a proxy for future operating shock can be further justified.

A potential concern, however, is that the revisions in filings may reflect systematic shocks at the industry or country level. If the shocks captured by low similarity values are related to systematic shocks, such as recession, then their effects are also likely to be systematic. Given that such systematic shocks directly impact other firms, the spillover effect through the supply chain cannot be clearly identified from their systematic impact on other firms. However, we find that the similarity measures do not exhibit correlation with systematic factors, such as business cycle. The four plots in Figure 1 show the averages and quartile values of similarity measures over time. Except for the dip in 2003, the overall similarities simply increase over time. The regular patterns of rise and fall in the monthly plot imply that the firms adopting different fiscal years are likely to be different from others. To control these yearly upward trend and monthly variations, we compare the similarities among filings reported at the same month. In this way, the remaining effects of shocks that commonly affect the mass of companies at the same time are also likely to be canceled out since such systematic shocks are likely to take down all similarity values at once.

B. **Spillover Effect of Firm-specific Future Operating Shocks in Network**

Theoretically, if a market is free from a set of frictions, then firm-level shocks in the supply chain network quickly perish. However, several recent empirical works find that even idiosyncratic shocks have broad and huge impacts on the supply chain. Considering the costs of searching and changing contracts in real world, especially time costs, these finding may not be entirely unexpected. For instance, in the event of an unanticipated natural disaster, timely responding to such shock is difficult not only for the firms directly affected by it but also for the firms that are linked by the value chain. However, if firms can at least partly predict the shock beforehand, then the potential damage can be decreased by multiple methods, such as adjusting the supply chain or securing more inventories. In this section, we test whether these future operating shocks also have spillover effects in the supply chain network.

Changes in filings provide a decent setting for addressing this question. First, the negative shocks following the changes in 10-K/Q are firm specific. Even after controlling for the industry and fiscal quarters, the negative impact still affects the company in terms of operational outcomes. Second, these changes are predictive rather than ex-post evaluations in the sense that they imply unrealized hazards. Cohen et al. (2018) report the negative impacts implied in these changes occur after a certain time, from months to year. In addition, we present supporting evidence in section A that the actual change in the meaning of filings is one of the potential sources of negative outcomes. In other words, the changes in the filing content can help identify future risks. However, whether such future shocks are actually known beforehand to other companies linked by the supply chain network remains unknown. We explore this issue in related works.

To empirically test the spillover effect of firm-specific projected shock, we construct a dataset by sorting out drastic changes in filings and mapping them to the supply chain as the future operating shocks. The filings that have first quintile similarities with the preceding year's filings are established as shocks. Given that such shocks are firm specific, they can be directly mapped into the value chain. We exclude all firms engaged in financial and personal services (SIC code 6000-7999) as well as those in transportation, communications, electric, gas, and sanitary services (SIC code 4000- 4999) from the sample. All of the accounting variables are winsorized at the 1% level, and the growth variables are trimmed at the 3% level at each year.

We focus on the impact of the shocks at the supplier side. We adopt the following regression used by Wu (2016) to measure the average impacts across all shocks at the supplier side:

$$Y_{it,t+k} = a + \sum_{n=0}^{N} b_n D_{i,t}^{n} + cX_{i,t} + F_{i,t} + \varepsilon_{it}, \tag{1}$$

where $D_{i,t}^{n}$ is a dummy variable that equals to 1 if one of firm *i*'s suppliers from an *n* connection reports a first quintile similarity filing in fiscal quarter *t*. Also, note that we assigned 0 for $D_{i,t}^{n}$ if a firm *i* does not have distance *n* supplier at period *t* in our supply chain, under the assumption that the characteristics of firms that known to have *n* connection supplier are similar to those of the firms without such connections on average, complementing the limited size of our supply chain network. $Y_{it,t+k}$ is the k-quarter growth rate in revenue, operating income, or change in gross margins. $X_{i,t}$ is the vector of lagged controls including market capitalization, book-to-market ratio, P/E ratio, leverage ratio, return on assets, and inventory. $F_{i,t}$ is the set of fixed effects (industry * year, fiscal quarter).

Table 5 reports the summary statistics of shocks generated by suppliers from *n* connections. Given that the customer–supplier relationship data from Compustat and the sampled 10-K major customer information do not fully reflect the supply chain in the real world, the supply chain drawn from them quickly decays as distance grows. However, we find statistically significant negative spillover effects up to distance 2 supplier. Table 6 reports the results of regression (1) with suppliers up to distance 3 (N = 3). The overall estimations on the spillover effects are negative across various similarity measures. On average, a future operating shock from the origin firm affects its distance 1 customers starting from two quarters after, while its impact on distance 2 customers are concentrated on three quarters after.

To further clarify the characteristics of the spillover effects, we first investigate the timeline of these propagations. Table 7 shows the results of similar regressions, which add firm fixed effects instead of control variables and substitute dependent variables by the sales divided by lagged total assets to represent the performance of firms in the market at that period. Similar to the results from Table 6, the impacts of shocks from distance 1 suppliers peak at after two quarters and those of distance 2 shocks reaches the maximum value after three quarters. Furthermore, the spillover effects from distance 3 suppliers are generally negative and even statistically significant for word vector similarities. Secondly, we compare the impacts of shocks from suppliers at different distances by substituting $D_{i,t}^{n}$ in regression (1) with a ratio of the number of firm $i$'s distance *n* suppliers which experience future operating shocks over the total number of firm $i$'s distance *n* suppliers which report filings in that quarter. Table 8 reports the estimated coefficients of $D_{i,t}^{n}$ defined in this way. Assuming that shocks occurred at different distance suppliers are similar and the impact of increased proportion of shocks at each connection are linear, a hundredth of coefficient on $D^{k}$ can be interpreted as the impact on revenue growth from a 1% increase in the ratio of suppliers experiencing future operating shocks at distance *k*. In general, the impacts increase for the shocks that occurred at farther suppliers. Furthermore, we find that the timeline of the spillover effects is consistent with the results from Table 6 and 7, and that the negative spillover effects from distance 3 suppliers are also statistically significant at the 5% level for four of the six similarity measures. This finding implies that the dummy variable $D_{i,t}^{n}$ from regression (1) may over-represent the shocks from faraway suppliers when the number of distance k suppliers increases as *k* increases

Lastly, Table 9 and 10 report the same regressions results based on the supply chain network restricted to the Compustat dataset alone. The signs of coefficients are generally similar to the preceding results from the custom supply chain network. However, given the restricted size of the Compustat dataset, only few of the spillover effects from the suppliers farther than distance 1 remain statistically significant.

## 4. Concluding Remarks

In this study, we empirically quantify the spillover effect of firm-specific future operating shocks along the supply chain linkage. Throughout this research, we find several new results and future research topics.

First, developing the idea of Cohen et al. (2018) that the changes in 10-K/Q filings are mainly associated with negative information on firms' future operation outcomes, we further characterize these changes as proxy for firm-specific future operating shocks. When a future crisis is referred to in a filing, it projects as a change in content or the subject of that filing. By adopting a similarity measure that is more sensitive to semantic likeness among text data, we try to precisely capture such mentions on crisis. We show that the changes identified by the word vector similarity measures are strong predictors of firms' negative long-term outcomes in terms of stock prices and financial indicators, such as revenue. These results support that the changes to reporting behaviors can act as useful proxy for the future operating outcome shocks.

Second, by exploiting these changes in filings, we empirically quantify the theoretical spillover effects on the supply chain. If a profit-maximizing firm expects its contracting party linked within the supply chain to experience shock in the future, then preemptive measures will be conducted to prevent negative externalities. Theoretically, if the market is free from a set of market frictions, then the spillover effect would quickly perish, especially when a shock can be expected. However, we find that statistically significant negative spillover effect exists even for future operating shocks that have the possibility of being predicted. From our custom supply chain network built from supplier-customer information in 10-K and the Compustat dataset, we find supporting evidence that a future operating outcome shock propagates up to two or even three connections away from the origin.

These results indicate that firms are having difficulty preventing the spread of negative effects within the production chain or the transfer of information within the supply chain is limited. Given that more complex supply chains accompany more propagation effects, a subtle crisis in one sector can lead to immense impact at an aggregate level. Hence, we believe that further research is needed to determine how companies respond to shocks within the supply chain and how effective preventive measures can be. We extend this research in related works by adopting updated similarity measures and explicitly taking the form of supply chain into account.

**References**

Acemoglu, D., Carvalho, V.M., Ozdaglar, A. and Tahbaz-Salehi, A., 2012. The network origins of aggregate fluctuations. *Econometrica*, *80*(5), pp.1977-2016.

Ang, E., Iancu, D.A. and Swinney, R., 2016. Disruption risk and optimal sourcing in multitier supply networks. Management Science, 63(8), pp.2397-2419.

Chiu, B., Crichton, G., Korhonen, A. and Pyysalo, S., 2016. How to train good word embeddings for biomedical NLP. In Proceedings of the 15th Workshop on Biomedical Natural Language Processing (pp. 166-174).

Blanchard, O., 1996. Theoretical aspects of transition. *The American Economic Review*, *86*(2), pp.117-122.

Wu, D., 2016. Shock spillover and financial response in supply chain networks: Evidence from firm-level data. Unpublished working paper.

Kolay, M., Lemmon, M.L. and Tashjian, E., 2012. Spillover effects in the supply chain: Evidence from Chapter 11 filings. Unpublished Working Paper.

Cohen, L., Malloy, C. and Nguyen, Q., 2018. Lazy prices (No. w25084). National Bureau of Economic Research.

Loughran, T. and McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), pp.35-65.

Le, Q. and Mikolov, T., 2014, January. Distributed representations of sentences and documents. In International Conference on Machine Learning (pp. 1188-1196).

**Figure 1** Similarities over Time

Figure 1 shows the plots of the time trend of each similarity measure from 1995 to 2017, by year and month of filing date. The top left figure shows the average similarity value for each year. The top right, bottom left, and bottom right figures show the plots of average, 25th , and 75th percentile values of each similarity measure, respectively.

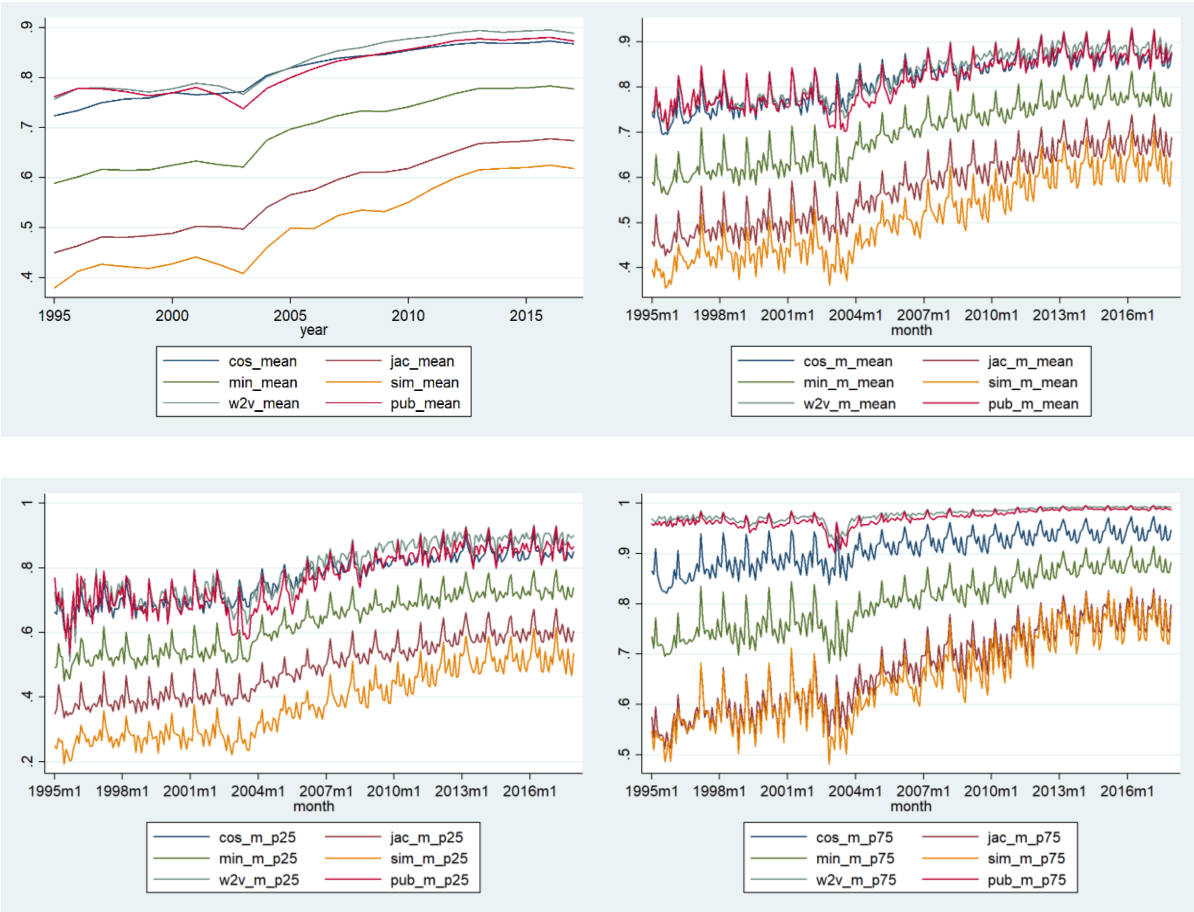

**Table 1** Most similar words calculated from word embedding (Word Vector II)

Table 1 shows the most similar words calculated from Word Embedding II. Each column presents word and its similarity with target word vector (litigation, competition, profit, customer, respectively) measured by cosine similarity.

| Input Word | litigation | | competition | | profit | | customer | |
|---|---|---|---|---|---|---|---|---|
| returns | word | Simi. | word | Simi. | word | Simi. | word | Simi. |
| 1 | lawsuit | 0.761 | competitive | 0.678 | profit_ margin | 0.500 | client | 0.648 |
| 2 | proceeding | 0.688 | compete | 0.676 | income | 0.497 | supplier | 0.617 |
| 3 | legal_ proceeding | 0.662 | competitor | 0.673 | earnings | 0.419 | product | 0.615 |
| 4 | dispute | 0.619 | intense_ competition | 0.661 | proportionate | 0.411 | distributor | 0.552 |
| 5 | matter | 0.576 | competitive_ pressure | 0.648 | profitability | 0.410 | vendor | 0.530 |
| 6 | legal | 0.576 | pressure | 0.599 | operating_ profit | 0.401 | consumer | 0.522 |
| 7 | defend | 0.571 | highly_ competitive | 0.592 | margin | 0.400 | retailer | 0.510 |
| 8 | claim | 0.567 | face_ competition | 0.571 | revenue | 0.400 | user | 0.493 |
| 9 | suit | 0.559 | extremely_ competitive | 0.568 | EVA | 0.381 | oem_ customer | 0.493 |
| 10 | patent_ infringement | 0.544 | competitive_ environment | 0.551 | proportion | 0.355 | subscriber | 0.478 |

**Table 2** Summary statistics on similarity measures

Panel A presents the summary statistics on six different measures used in this paper. Panel B indicates the correlations among different pairs of similarity measures. Sim_Cosine is the cosine similarity, Sim_Jaccard is the Jaccard similarity, Sim_Min is the minimum edit distance, Sim_WV1 is the word vector cosine similarity I (with parameter setting I), and Sim_WV2 is the word vector cosine similarity II (with parameter setting II).

Panel A: Summary statistics on similarity measures

| | Count | Mean | SD | Min | q25 | Median | q75 | Max |
|---|---|---|---|---|---|---|---|---|
| Sim_Cosine | 374,636 | 0.7952 | 0.1563 | 0.2899 | 0.7358 | 0.8454 | 0.9053 | 0.9763 |
| Sim_Jaccard | 374,636 | 0.5531 | 0.1489 | 0.1864 | 0.4504 | 0.5612 | 0.6665 | 0.8417 |
| Sim_Min | 374,636 | 0.6792 | 0.1733 | 0.1434 | 0.5964 | 0.7192 | 0.8069 | 0.9219 |
| Sim_Sim | 374,636 | 0.4736 | 0.1933 | 0.0557 | 0.3337 | 0.4881 | 0.6231 | 0.8417 |
| Sim_WV1 | 375,153 | 0.8091 | 0.2773 | -0.1340 | 0.7659 | 0.9430 | 0.9826 | 0.9983 |
| Sim_WV2 | 375,148 | 0.7957 | 0.2727 | -0.1053 | 0.7352 | 0.9241 | 0.9748 | 0.9976 |

Panel B: Correlation among similarity measures

| | Sim_Cosine | Sim_Jaccard | Sim_Min | Sim_Sim | Sim_WV1 | Sim_WV2 |
|---|---|---|---|---|---|---|
| Sim_Cosine | 1 | | | | | |
| Sim_Jaccard | 0.8222 | 1 | | | | |
| Sim_Min | 0.9245 | 0.9188 | 1 | | | |
| Sim_Sim | 0.7876 | 0.9062 | 0.8816 | 1 | | |
| Sim_WV1 | 0.8829 | 0.6982 | 0.8534 | 0.659 | 1 | |
| Sim_WV2 | 0.8785 | 0.7036 | 0.8491 | 0.6582 | 0.9919 | 1 |

**Table 3** Fama MacBeth Regressions

This table presents the Fama–MacBeth cross-sectional regressions of firm-level stock returns (cumulative returns from month +1 to +12) on six different similarity measures. Sim_Cosine is the cosine similarity, Sim_Jaccard is the Jaccard similarity, Sim_Min is the minimum edit distance, Sim_WV1 is the word2vec cosine similarity I (with custom parameter setting I), and Sim_WV2 is the word2vec cosine similarity II (with custom parameter setting II). Size is the log(market value of equity), log(BM) is the log(book value of equity over market value of equity), Ret(-1,0) is the return of previous month, and Ret(-12,-1) is the cumulative return from month -12 to month -1.

| VARIABLES | Ret(1, 12) | | | | | |
|---|---|---|---|---|---|---|
| Sim_Cosine | 0.0380** | | | | | |
| | (0.0154) | | | | | |
| Sim_Jaccard | | 0.0911*** | | | | |
| | | (0.0215) | | | | |
| Sim_MinEdit | | | 0.0559*** | | | |
| | | | (0.0159) | | | |
| Sim_Simple | | | | 0.0402** | | |
| | | | | (0.0171) | | |
| Sim_WV1 | | | | | 0.0441*** | |
| | | | | | (0.0131) | |
| Sim_WV2 | | | | | | 0.0397*** |
| | | | | | | (0.0114) |
| Size | -0.0193*** | -0.0197*** | -0.0195*** | -0.0191*** | -0.0278** | -0.0251*** |
| | (0.0045) | (0.0049) | (0.0047) | (0.0045) | (0.0117) | (0.0093) |
| log(BM) | 0.0084 | 0.0077 | 0.0080 | 0.0081 | 0.0001 | 0.0024 |
| | (0.0061) | (0.0064) | (0.0062) | (0.0061) | (0.0109) | (0.0091) |
| Ret(-1, 0) | -0.0076 | -0.0077 | -0.0075 | -0.0078 | -0.0070 | -0.0072 |
| | (0.0264) | (0.0264) | (0.0263) | (0.0263) | (0.0265) | (0.0265) |
| Ret(-12, -1) | -0.0566*** | -0.0583*** | -0.0575*** | -0.0581*** | -0.0579*** | -0.0575*** |
| | (0.0176) | (0.01766) | (0.0176) | (0.0176) | (0.0181) | (0.0180) |
| Constant | 0.3489*** | 0.3355*** | 0.3447*** | 0.3549*** | 0.4692*** | 0.4332*** |
| | (0.0680) | (0.0721) | (0.0710) | (0.0685) | (0.1710) | (0.1356) |
| Observations | 315,476 | 315,476 | 315,476 | 315,476 | 308,241 | 307,776 |
| R-squared | 0.0499 | 0.0503 | 0.0501 | 0.0510 | 0.0494 | 0.0495 |
| Number of groups | 267 | 267 | 267 | 267 | 267 | 267 |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 4** Real effects

This table reports the regression of net income and sales growth on filings' similarity measures. OI2Q, NI2Q, and SA2Q, are Oibdpq/L1atq (operating income before depreciation divided by lagged total assets), Niq/L1atq (net income divided by lagged total assets), and Saleq/L1atq (sales divided by lagged total assets) measured two quarters ahead, respectively. All variables in the table are winsorized at 1% level. Sim_Cos is the cosine similarity, Sim_Jac is the Jaccard similarity, Sim_Min is the minimum edit distance, and Sim_Sim is the simple similarity. Sim_WV1 is the word2vec cosine similarity I (with custom parameter setting I) and Sim_WV2 is the word2vec cosine similarity II (with custom parameter setting II). Similarity All regressions include month, industry, and firm fixed effects.

| VARIABLES | OI2Q | NI2Q | SA2Q | OI2Q | NI2Q | SA2Q | OI2Q | NI2Q | SA2Q |
|---|---|---|---|---|---|---|---|---|---|
| | | *Sim_Cos* | | | *Sim_Jac* | | | *Sim_Min* | |
| Similarity | 0.001*** | 0.002*** | 0.006*** | 0.005*** | 0.006*** | 0.015*** | 0.003*** | 0.004*** | 0.008*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Constant | 0.012*** | -0.01*** | 0.220*** | 0.011*** | -0.011*** | 0.217*** | 0.011*** | -0.011*** | 0.220*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 306,716 | 335,462 | 335,243 | 306,716 | 335,462 | 335,243 | 306,716 | 335,462 | 335,243 |
| R-squared | 0.688 | 0.549 | 0.844 | 0.688 | 0.549 | 0.844 | 0.688 | 0.549 | 0.844 |
| | | *Sim_Sim* | | | *Sim_WV1* | | | *Sim_WV2* | |
| Similarity | 0.004*** | 0.005*** | 0.015*** | 0.001*** | 0.001*** | 0.003*** | 0.001*** | 0.002*** | 0.002*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Constant | 0.011*** | -0.010*** | 0.218*** | 0.012*** | -0.009*** | 0.221*** | 0.012*** | -0.009*** | 0.221*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 306,716 | 335,462 | 335,243 | 298,188 | 326,054 | 325,837 | 299,333 | 327,396 | 327,177 |
| R-squared | 0.688 | 0.549 | 0.844 | 0.69 | 0.55 | 0.844 | 0.689 | 0.55 | 0.844 |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 5** Summary statistics of shocks and network connections

Table 5 reports the summary statistics of shocks and network connections. Origin equals to 1 if the company being analyzed issues filing whose similarity belongs to first quintile (least similar to precedent filing) and 0 if otherwise. Distance n takes 1 as value if one of the suppliers from an n connection reports a first quintile similarity filing and 0 if otherwise.

| VARIABLE | Obs (Compustat) | Mean (Compustat) | Obs (Total) | Mean (Total) | Min | Max |
|---|---|---|---|---|---|---|
| D0 (Origin) | 357,693 | 0.2016 | 357,693 | 0.2016 | 0 | 1 |
| D1 (Distance 1) | 88,129 | 0.2935 | 90,790 | 0.2988 | 0 | 1 |
| D2 (Distance 2) | 21,857 | 0.3349 | 23,813 | 0.3510 | 0 | 1 |
| D3 (Distance 3) | 6,771 | 0.3980 | 7,901 | 0.3965 | 0 | 1 |
| D4 (Distance 4) | 2,396 | 0.4144 | 2,928 | 0.3982 | 0 | 1 |

**Table 6** Spillover effects on revenues

Table 6 reports the coefficients estimates on $b_n$, n = 0, 1, 2, 3 from regression (1) of the text. Each column shows the spillover effects on *k* quarter revenue growth of shocks identified by different similarity measures. REVkQ is the k-quarter revenue growth. Sim_Cos is the cosine similarity, Sim_Jac is the Jaccard similarity, Sim_Min is the minimum edit distance, and Sim_Sim is the simple similarity. Sim_WV1 is the word2vec cosine similarity I (with custom parameter setting I) and Sim_WV2 is the word2vec cosine similarity II (with custom parameter setting II). Dn is a dummy variable that equals to 1 if one of the suppliers from an n connection reports a first quintile similarity filing and 0 if otherwise. All regressions include industry*year, fiscal quarter fixed effects, and are in quarterly frequency from 1995 to 2017. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

| VARIABLES | REV2Q | REV3Q | REV4Q | REV2Q | REV3Q | REV4Q | REV2Q | REV3Q | REV4Q |
|---|---|---|---|---|---|---|---|---|---|
| | | *Sim_Cos* | | | *Sim_Jac* | | | *Sim_Min* | |
| D0 | -0.0004 | -0.0049*** | -0.0024 | 0.0016 | -0.0023 | -0.0018 | -0.0003 | -0.0044*** | -0.0038** |
| | (0.0014) | (0.0016) | (0.0015) | (0.0014) | (0.0016) | (0.0016) | (0.0014) | (0.0016) | (0.0015) |
| D1 | -0.0213*** | -0.0196*** | -0.0237*** | -0.0187*** | -0.0177*** | -0.0209*** | -0.0194*** | -0.0192*** | -0.0234*** |
| | (0.0025) | (0.0029) | (0.0028) | (0.0025) | (0.0029) | (0.0028) | (0.0025) | (0.0029) | (0.0028) |
| D2 | -0.0091** | -0.0173*** | -0.0122** | -0.0081* | -0.0185*** | -0.0118** | -0.0085* | -0.0172*** | -0.0111** |
| | (0.0045) | (0.0051) | (0.0050) | (0.0045) | (0.0052) | (0.0050) | (0.0045) | (0.0052) | (0.0050) |
| D3 | -0.0010 | 0.0073 | 0.0062 | 0.0011 | 0.0124 | 0.0070 | -0.0045 | 0.0054 | 0.0036 |
| | (0.0070) | (0.0079) | (0.0077) | (0.0070) | (0.0080) | (0.0077) | (0.0071) | (0.0080) | (0.0078) |
| Constant | -0.0132** | 0.0240*** | 0.0592*** | -0.0105* | 0.0263*** | 0.0623*** | -0.0117** | 0.0246*** | 0.0600*** |
| | (0.0055) | (0.0063) | (0.0062) | (0.0055) | (0.0063) | (0.0062) | (0.0055) | (0.0063) | (0.0062) |
| Control Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 132,248 | 128,422 | 124,772 | 132,248 | 128,422 | 124,772 | 132,248 | 128,422 | 124,772 |
| R-squared | 0.1314 | 0.1879 | 0.2453 | 0.1312 | 0.1878 | 0.2452 | 0.1313 | 0.1879 | 0.2453 |
| | | *Sim_Sim* | | | *Sim_WV1* | | | *Sim_WV2* | |
| D0 | 0.0005 | -0.0025 | -0.0042*** | -0.0009 | -0.0048*** | -0.0035** | -0.0011 | -0.0043*** | -0.0033** |
| | (0.0014) | (0.0016) | (0.0016) | (0.0014) | (0.0015) | (0.0015) | (0.0013) | (0.0015) | (0.0015) |
| D1 | -0.0183*** | -0.0173*** | -0.0212*** | -0.0195*** | -0.0186*** | -0.0206*** | -0.0198*** | -0.0196*** | -0.0213*** |
| | (0.0026) | (0.0029) | (0.0029) | (0.0025) | (0.0029) | (0.0028) | (0.0025) | (0.0029) | (0.0028) |
| D2 | -0.0117** | -0.0213*** | -0.0151*** | -0.0054 | -0.0139*** | -0.0126** | -0.0064 | -0.0142*** | -0.0113** |
| | (0.0046) | (0.0052) | (0.0051) | (0.0046) | (0.0052) | (0.0051) | (0.0045) | (0.0051) | (0.0050) |
| D3 | 0.0052 | 0.0121 | 0.0087 | -0.0075 | -0.0017 | 0.0011 | -0.0054 | -0.0006 | -0.0004 |
| | (0.0073) | (0.0082) | (0.0080) | (0.0071) | (0.0080) | (0.0078) | (0.0071) | (0.0080) | (0.0078) |
| Constant | -0.0104* | 0.0257*** | 0.0613*** | -0.0113** | 0.0256*** | 0.0617*** | -0.0116** | 0.0247*** | 0.0612*** |
| | (0.0055) | (0.0063) | (0.0062) | (0.0055) | (0.0063) | (0.0062) | (0.0055) | (0.0063) | (0.0062) |
| Control Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 132,248 | 128,422 | 124,772 | 132,371 | 128,547 | 124,899 | 132,392 | 128,571 | 124,922 |
| R-squared | 0.1312 | 0.1878 | 0.2452 | 0.1311 | 0.1877 | 0.2450 | 0.1311 | 0.1877 | 0.2450 |

**Table 7** Spillover effects on sales

Table 7 reports the coefficients estimates on $b_n$, n = 0, 1, 2, 3 from regression (1) of the text. SAkQ is Saleq/L1atq (sales divided by lagged total assets) measured k-quarters ahead. Sim_Cos is the cosine similarity, Sim_Jac is the Jaccard similarity, Sim_Min is the minimum edit distance, and Sim_Sim is the simple similarity. Sim_WV1 is the word2vec cosine similarity I (with custom parameter setting I) and Sim_WV2 is the word2vec cosine similarity II (with custom parameter setting II). Dn is a dummy variable that equals to 1 if one of the suppliers from an n connection reports a first quintile similarity filing and 0 if otherwise. All regressions include month, industry, and firm fixed effects, and are in quarterly frequency from 1995 to 2017.

| VARIABLES | SA2Q | SA3Q | SA4Q | SA2Q | SA3Q | SA4Q | SA2Q | SA3Q | SA4Q |
|---|---|---|---|---|---|---|---|---|---|
| | | *Sim_Cos* | | | *Sim_Jac* | | | *Sim_Min* | |
| D0 | -0.0005 | -0.0012** | -0.0001 | -0.0019*** | -0.0024*** | 0.0001 | -0.0014** | -0.0017*** | -0.0004 |
| | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) |
| D1 | -0.0062*** | -0.0028** | -0.0034*** | -0.0075*** | -0.0033** | -0.0043*** | -0.0068*** | -0.0038*** | -0.0036*** |
| | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) |
| D2 | -0.0071*** | -0.0080*** | -0.0063*** | -0.0058*** | -0.0093*** | -0.0074*** | -0.0074*** | -0.0091*** | -0.0073*** |
| | (0.0022) | (0.0022) | (0.0023) | (0.0022) | (0.0022) | (0.0023) | (0.0022) | (0.0023) | (0.0023) |
| D3 | -0.0044 | -0.0034 | -0.0023 | -0.0043 | -0.0006 | -0.0001 | -0.0059* | -0.0031 | -0.0032 |
| | (0.0034) | (0.0034) | (0.0034) | (0.0035) | (0.0034) | (0.0034) | (0.0035) | (0.0034) | (0.0034) |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.2882*** | 0.2878*** | 0.2885*** | 0.2885*** | 0.2881*** | 0.2885*** | 0.2884*** | 0.2880*** | 0.2886*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Observations | 181,703 | 175,300 | 169,324 | 181,703 | 175,300 | 169,324 | 181,703 | 175,300 | 169,324 |
| R-squared | 0.8044 | 0.8075 | 0.8078 | 0.8045 | 0.8075 | 0.8078 | 0.8044 | 0.8075 | 0.8078 |
| | | *Sim_Sim* | | | *Sim_WV1* | | | *Sim_WV2* | |
| D0 | -0.0027*** | -0.0023*** | -0.0010* | -0.0006 | -0.0008 | -0.0001 | -0.0007 | -0.0012** | -0.0000 |
| | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) |
| D1 | -0.0066*** | -0.0025* | -0.0029** | -0.0061*** | -0.0038*** | -0.0034*** | -0.0062*** | -0.0036*** | -0.0035*** |
| | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) | (0.0013) |
| D2 | -0.0070*** | -0.0104*** | -0.0074*** | -0.0055** | -0.0074*** | -0.0070*** | -0.0067*** | -0.0085*** | -0.0080*** |
| | (0.0023) | (0.0023) | (0.0023) | (0.0023) | (0.0023) | (0.0023) | (0.0022) | (0.0023) | (0.0023) |
| D3 | -0.0041 | -0.0021 | -0.0024 | -0.0072** | -0.0042 | -0.0038 | -0.0069** | -0.0044 | -0.0030 |
| | (0.0035) | (0.0035) | (0.0035) | (0.0035) | (0.0034) | (0.0035) | (0.0034) | (0.0034) | (0.0034) |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.2886*** | 0.2881*** | 0.2887*** | 0.2882*** | 0.2879*** | 0.2886*** | 0.2883*** | 0.2880*** | 0.2885*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Observations | 181,703 | 175,300 | 169,324 | 181,874 | 175,473 | 169,496 | 181,880 | 175,478 | 169,501 |
| R-squared | 0.8045 | 0.8075 | 0.8078 | 0.8044 | 0.8075 | 0.8078 | 0.8044 | 0.8075 | 0.8078 |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 8** Spillover effects on revenues

Table 8 reports the coefficients estimates on $b_n$, n = 0, 1, 2, 3 from regression (1) of the text, where Dn is defined as a proportion of occurrences of shocks among distance n supplier. Each column shows the spillover effects on *k* quarter revenue growth of shocks identified by different similarity measures. REVkQ is the k-quarter revenue growth. Sim_Cos is the cosine similarity, Sim_Jac is the Jaccard similarity, Sim_Min is the minimum edit distance, and Sim_Sim is the simple similarity. Sim_WV1 is the word2vec cosine similarity I and Sim_WV2 is the word2vec cosine similarity II. Dn is a dummy variable that equals to the ratio of the suppliers from an n connection which report a first quintile similarity filing over total number of distance n suppliers which report a filing at each quarter. All regressions include industry*year, fiscal quarter fixed effects, and are in quarterly frequency from 1995 to 2017.

| VARIABLES | REV2Q | REV3Q | REV4Q | REV2Q | REV3Q | REV4Q | REV2Q | REV3Q | REV4Q |
|---|---|---|---|---|---|---|---|---|---|
| | | *Sim_Cos* | | | *Sim_Jac* | | | *Sim_Min* | |
| D0 | 0.0002 | -0.0027* | -0.0011 | 0.0020 | -0.0009 | -0.0004 | -0.0000 | -0.0019 | -0.0020 |
| | (0.0014) | (0.0016) | (0.0015) | (0.0014) | (0.0016) | (0.0016) | (0.0014) | (0.0016) | (0.0015) |
| D1 | -0.0146*** | -0.0129*** | -0.0164*** | -0.0143*** | -0.0092** | -0.0132*** | -0.0143*** | -0.0138*** | -0.0179*** |
| | (0.0038) | (0.0043) | (0.0043) | (0.0039) | (0.0045) | (0.0044) | (0.0038) | (0.0044) | (0.0043) |
| D2 | -0.0071 | -0.0190** | -0.0090 | -0.0067 | -0.0180** | -0.0093 | -0.0086 | -0.0200** | -0.0125 |
| | (0.0073) | (0.0083) | (0.0081) | (0.0073) | (0.0083) | (0.0081) | (0.0073) | (0.0084) | (0.0082) |
| D3 | -0.0180 | -0.0045 | -0.0289** | -0.0045 | 0.0098 | -0.0234* | -0.0203* | -0.0042 | -0.0284** |
| | (0.0115) | (0.0132) | (0.0128) | (0.0119) | (0.0136) | (0.0132) | (0.0121) | (0.0138) | (0.0134) |
| Constant | 0.0886*** | 0.1403*** | 0.1260*** | 0.0894*** | 0.1425*** | 0.1277*** | 0.0886*** | 0.1402*** | 0.1256*** |
| | (0.0054) | (0.0062) | (0.0061) | (0.0054) | (0.0062) | (0.0061) | (0.0054) | (0.0062) | (0.0061) |
| Control Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 129,058 | 125,276 | 121,794 | 129,058 | 125,276 | 121,794 | 129,058 | 125,276 | 121,794 |
| R-squared | 0.1287 | 0.1855 | 0.2431 | 0.1286 | 0.1854 | 0.2430 | 0.1287 | 0.1855 | 0.2431 |
| | | *Sim_Sim* | | | *Sim_WV1* | | | *Sim_WV2* | |
| D0 | 0.0008 | -0.0010 | -0.0034** | -0.0009 | -0.0025 | -0.0017 | -0.0005 | -0.0024 | -0.0017 |
| | (0.0014) | (0.0016) | (0.0016) | (0.0014) | (0.0015) | (0.0015) | (0.0013) | (0.0015) | (0.0015) |
| D1 | -0.0129*** | -0.0095** | -0.0117*** | -0.0140*** | -0.0148*** | -0.0159*** | -0.0138*** | -0.0143*** | -0.0159*** |
| | (0.0039) | (0.0045) | (0.0044) | (0.0038) | (0.0044) | (0.0043) | (0.0038) | (0.0044) | (0.0043) |
| D2 | -0.0125* | -0.0267*** | -0.0164** | -0.0023 | -0.0150* | -0.0060 | -0.0019 | -0.0159* | -0.0092 |
| | (0.0074) | (0.0084) | (0.0083) | (0.0075) | (0.0086) | (0.0084) | (0.0074) | (0.0085) | (0.0083) |
| D3 | -0.0029 | 0.0134 | -0.0148 | -0.0281** | -0.0170 | -0.0414*** | -0.0260** | -0.0144 | -0.0422*** |
| | (0.0123) | (0.0140) | (0.0136) | (0.0123) | (0.0141) | (0.0137) | (0.0121) | (0.0138) | (0.0134) |
| Constant | 0.0897*** | 0.1417*** | 0.1278*** | 0.0888*** | 0.1404*** | 0.1266*** | 0.0890*** | 0.1405*** | 0.1262*** |
| | (0.0054) | (0.0062) | (0.0061) | (0.0054) | (0.0062) | (0.0061) | (0.0054) | (0.0062) | (0.0061) |
| Control Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 129,058 | 125,276 | 121,794 | 129,196 | 125,422 | 121,940 | 129,202 | 125,428 | 121,945 |
| R-squared | 0.1286 | 0.1854 | 0.2431 | 0.1285 | 0.1853 | 0.2429 | 0.1285 | 0.1854 | 0.2429 |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 9** Robustness - Spillover effects on revenues

Table 9 reports the coefficients estimates on $b_n$, n = 0, 1, 2, 3 from regression (1) of the text, using the supplier-consumer relationships from Compustat alone. Each column shows the spillover effects on $k$ quarter revenue growth of shocks identified by different similarity measures. REVkQ is the k-quarter revenue growth. Sim_Cos is the cosine similarity, Sim_Jac is the Jaccard similarity, Sim_Min is the minimum edit distance, and Sim_Sim is the simple similarity. Sim_WV1 is the word2vec cosine similarity I (with custom parameter setting I) and Sim_WV2 is the word2vec cosine similarity II (with custom parameter setting II). Dn is a dummy variable that equals to 1 if one of the suppliers from an n connection reports a first quintile similarity filing and 0 if otherwise. All regressions include industry*year, fiscal quarter fixed effects, and are in quarterly frequency from 1995 to 2017.

| VARIABLES | REV2Q | REV3Q | REV4Q | REV2Q | REV3Q | REV4Q | REV2Q | REV3Q | REV4Q |
|---|---|---|---|---|---|---|---|---|---|
| | | *Sim_Cos* | | | *Sim_Jac* | | | *Sim_Min* | |
| D0 | 0.0002 | -0.0054*** | -0.0030* | 0.0027* | -0.0026 | -0.0015 | 0.0005 | -0.0043** | -0.0035** |
| | (0.0015) | (0.0018) | (0.0018) | (0.0016) | (0.0018) | (0.0018) | (0.0016) | (0.0018) | (0.0018) |
| D1 | -0.0256*** | -0.0273*** | -0.0351*** | -0.0257*** | -0.0274*** | -0.0351*** | -0.0257*** | -0.0273*** | -0.0351*** |
| | (0.0040) | (0.0046) | (0.0045) | (0.0040) | (0.0046) | (0.0045) | (0.0040) | (0.0046) | (0.0045) |
| D2 | -0.0069 | -0.0052 | -0.0020 | -0.0068 | -0.0052 | -0.0020 | -0.0069 | -0.0054 | -0.0021 |
| | (0.0105) | (0.0119) | (0.0118) | (0.0105) | (0.0119) | (0.0118) | (0.0105) | (0.0119) | (0.0118) |
| D3 | 0.0103 | 0.0073 | 0.0238 | 0.0104 | 0.0076 | 0.0240 | 0.0103 | 0.0078 | 0.0241 |
| | (0.0165) | (0.0188) | (0.0186) | (0.0165) | (0.0189) | (0.0186) | (0.0165) | (0.0188) | (0.0186) |
| Constant | -0.0296*** | 0.0044 | 0.0414*** | -0.0299*** | 0.0048 | 0.0416*** | -0.0296*** | 0.0047 | 0.0416*** |
| | (0.0065) | (0.0075) | (0.0074) | (0.0065) | (0.0075) | (0.0074) | (0.0065) | (0.0075) | (0.0074) |
| Control Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 114,841 | 111,344 | 108,011 | 114,841 | 111,344 | 108,011 | 114,841 | 111,344 | 108,011 |
| R-squared | 0.1305 | 0.1833 | 0.2398 | 0.1305 | 0.1833 | 0.2398 | 0.1305 | 0.1833 | 0.2398 |
| | | *Sim_Sim* | | | *Sim_WV1* | | | *Sim_WV2* | |
| D0 | 0.0022 | -0.0032* | -0.0037** | -0.0004 | -0.0043** | -0.0031* | 0.0003 | -0.0041** | -0.0028 |
| | (0.0016) | (0.0018) | (0.0018) | (0.0015) | (0.0018) | (0.0017) | (0.0015) | (0.0018) | (0.0017) |
| D1 | -0.0257*** | -0.0273*** | -0.0350*** | -0.0257*** | -0.0274*** | -0.0352*** | -0.0257*** | -0.0274*** | -0.0352*** |
| | (0.0040) | (0.0046) | (0.0045) | (0.0040) | (0.0046) | (0.0045) | (0.0040) | (0.0046) | (0.0045) |
| D2 | -0.0068 | -0.0053 | -0.0022 | -0.0066 | -0.0042 | -0.0019 | -0.0066 | -0.0042 | -0.0019 |
| | (0.0105) | (0.0119) | (0.0118) | (0.0105) | (0.0119) | (0.0118) | (0.0105) | (0.0119) | (0.0118) |
| D3 | 0.0104 | 0.0076 | 0.0239 | 0.0100 | 0.0067 | 0.0239 | 0.0100 | 0.0068 | 0.0240 |
| | (0.0165) | (0.0188) | (0.0186) | (0.0165) | (0.0188) | (0.0186) | (0.0165) | (0.0188) | (0.0186) |
| Constant | -0.0296*** | 0.0045 | 0.0414*** | -0.0297*** | 0.0051 | 0.0417*** | -0.0297*** | 0.0051 | 0.0418*** |
| | (0.0065) | (0.0075) | (0.0074) | (0.0065) | (0.0075) | (0.0074) | (0.0065) | (0.0075) | (0.0074) |
| Control Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 114,841 | 111,344 | 108,011 | 114,982 | 111,490 | 108,160 | 114,982 | 111,490 | 108,160 |
| R-squared | 0.1305 | 0.1833 | 0.2398 | 0.1304 | 0.1831 | 0.2396 | 0.1304 | 0.1831 | 0.2396 |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 10** Robustness - Spillover effects on sales

Table 10 reports the coefficients estimates on $b_n$, n = 0, 1, 2, 3 from regression (1) of the text, using the supplier-consumer relationships from Compustat alone. SAkQ is Saleq/L1atq (sales divided by lagged total assets) measured k-quarters ahead. Sim_Cos is the cosine similarity, Sim_Jac is the Jaccard similarity, Sim_Min is the minimum edit distance, and Sim_Sim is the simple similarity. Sim_WV1 is the word2vec cosine similarity I (with custom parameter setting I) and Sim_WV2 is the word2vec cosine similarity II (with custom parameter setting II). Dn is a dummy variable that equals to 1 if one of the suppliers from an n connection reports a first quintile similarity filing and 0 if otherwise. All regressions include month, industry, and firm fixed effects, and are in quarterly frequency from 1995 to 2017.

| VARIABLES | SA2Q | SA3Q | SA4Q | SA2Q | SA3Q | SA4Q | SA2Q | SA3Q | SA4Q |
|---|---|---|---|---|---|---|---|---|---|
| | | *Sim_Cos* | | | *Sim_Jac* | | | *Sim_Min* | |
| D0 | -0.0007 | -0.0015** | -0.0007 | -0.0020*** | -0.0026*** | -0.0002 | -0.0016*** | -0.0019*** | -0.0009 |
| | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) |
| D1 | -0.0174*** | -0.0119*** | -0.0091*** | -0.0174*** | -0.0119*** | -0.0091*** | -0.0174*** | -0.0119*** | -0.0091*** |
| | (0.0039) | (0.0038) | (0.0034) | (0.0039) | (0.0038) | (0.0034) | (0.0039) | (0.0038) | (0.0034) |
| D2 | -0.0366*** | -0.0212* | -0.0156 | -0.0366*** | -0.0214* | -0.0156 | -0.0367*** | -0.0213* | -0.0157 |
| | (0.0119) | (0.0116) | (0.0103) | (0.0119) | (0.0116) | (0.0103) | (0.0119) | (0.0116) | (0.0103) |
| D3 | 0.0048 | 0.0044 | 0.0223 | 0.0046 | 0.0042 | 0.0222 | 0.0049 | 0.0045 | 0.0223 |
| | (0.0198) | (0.0196) | (0.0185) | (0.0198) | (0.0196) | (0.0185) | (0.0198) | (0.0196) | (0.0185) |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.2818*** | 0.2817*** | 0.2824*** | 0.2820*** | 0.2819*** | 0.2823*** | 0.2820*** | 0.2818*** | 0.2824*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Observations | 156,218 | 150,024 | 144,492 | 156,218 | 150,024 | 144,492 | 156,218 | 150,024 | 144,492 |
| R-squared | 0.7965 | 0.7991 | 0.7992 | 0.7965 | 0.7991 | 0.7992 | 0.7965 | 0.7991 | 0.7992 |
| | | *Sim_Sim* | | | *Sim_WV1* | | | *Sim_WV2* | |
| D0 | -0.0026*** | -0.0024*** | -0.0014** | -0.0005 | -0.0014** | -0.0003 | -0.0009 | -0.0012** | -0.0002 |
| | (0.0006) | (0.0006) | (0.0007) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) |
| D1 | -0.0175*** | -0.0119*** | -0.0092*** | -0.0173*** | -0.0120*** | -0.0090*** | -0.0174*** | -0.0120*** | -0.0088*** |
| | (0.0039) | (0.0038) | (0.0034) | (0.0039) | (0.0038) | (0.0034) | (0.0039) | (0.0038) | (0.0034) |
| D2 | -0.0368*** | -0.0215* | -0.0158 | -0.0353*** | -0.0198* | -0.0157 | -0.0353*** | -0.0198* | -0.0157 |
| | (0.0119) | (0.0116) | (0.0103) | (0.0117) | (0.0115) | (0.0103) | (0.0117) | (0.0115) | (0.0103) |
| D3 | 0.0046 | 0.0043 | 0.0223 | 0.0044 | 0.0036 | 0.0223 | 0.0045 | 0.0037 | 0.0223 |
| | (0.0198) | (0.0196) | (0.0185) | (0.0198) | (0.0196) | (0.0185) | (0.0198) | (0.0196) | (0.0185) |
| Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 0.2821*** | 0.2818*** | 0.2825*** | 0.2818*** | 0.2817*** | 0.2823*** | 0.2819*** | 0.2817*** | 0.2823*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Observations | 156,218 | 150,024 | 144,492 | 156,367 | 150,178 | 144,647 | 156,390 | 150,199 | 144,666 |
| R-squared | 0.7965 | 0.7991 | 0.7992 | 0.7965 | 0.7991 | 0.7993 | 0.7965 | 0.7991 | 0.7992 |

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1