Scope, Scale and Concentration: The 21st Century Firm

Gerard Hoberg and Gordon M. Phillips*

October 4, 2023
conditionally accepted Journal of Finance

Abstract

We provide evidence using firm 10-Ks that over the past 30 years, U.S. firms have expanded their scope of operations. Increases in scope were achieved largely without increasing traditional operating segments. Scope expansion significantly increases valuation and is primarily realized through acquisitions and investment in R&D, but not through capital expenditures. Traditional concentration ratios do not capture this expansion of scope. Our findings point to a new type of firm that increases scope through related expansion, which is highly valued by the market.

Keywords: Firm scope, economies of scope, products, concentration, firm size.

JEL Codes: O31, O34, D43, F13

*University of Southern California Marshall School of Business and Tuck School of Business at Dartmouth College and National Bureau of Economic Research, respectively. Hoberg can be reached at hoberg@marshall.usc.edu. Phillips can be reached at gordon.m.phillips@tuck.dartmouth.edu. We thank the editor, Thomas Philippon, and two anonymous referees. We also thank Benoît Durand (DCI discussant), Jon Garfinkel, Kai Li (WFA discussant), Andrey Golubov, Michael Roberts, Rene Stulz, Chad Syverson (NBER discussant), and seminar participants at Aarhus University, Australia National University, Georgia State University, Goethe University Frankfurt, Iowa State University, Ohio State University, Swiss Economic Institute, Tulane University, University of Amsterdam, University of Iowa and the University of Pennsylvania (Wharton), and also conference participants at the Dynamics and Competition Initiative (DCI), MARC conference at Bayes Business School, NBER, Western Finance Association conferences. We also thank Christopher Ball at metaHeuristica for advice on our doc2vec implementation, and Himanshu Rawlani for excellent research assistance. All errors are the authors' alone.

"Product diversification came from opportunities to use existing production, marketing, and research facilities ... Such expansion was based on organizational capabilities that had been developed by exploiting economies of scope." *p. 38 Chandler and Hikino (1994).*

The interplay between scope, scale and competition has been the focus of numerous authors, including business historians and economists.[1] A principle focus of authors has been defining firms by the basket of products firms produce and the industries to which these products belong. Recent authors have also documented the rise in firm size, a rise in traditional industry concentration measures, and a drop in the number of U.S. listed firms.[2]

This paper provides a new perspective on the 21st Century version of a multi-product firm and how it produces in different-but-related markets. It differs markedly from the concept of a diversified conglomerate with a multi-division organizational structure producing products across unrelated industries that was the focus of the early corporate finance literature.[3] Instead of multiple distinct segments, firms might have flexible production and redeployable assets that allow them to pursue multi-sector production without the potential negative consequences of a complex multi-division organization.

We develop new firm-year measures of product market scope using the text describing the product markets in which firms operate. We document that the average firm's scope in related industries has increased steadily and dramatically (roughly 60%) during our sample period from 1989 to 2017. Moreover, firms have increased scope without increasing the number of operating segments they report during our sample period. Our results are consistent with multi-product firms having synergies across related products and with unrelated diver-

[1]See Chandler and Hikino (1994), Hart and Moore (1990), Panzar and Willig (1977) and Williamson (1975).

[2]See Autor, Dorn, Katz, Patterson, and Van Reenen (2020), Doidge, Karolyi, and Stulz (2017), Grullon, Larkin, and Michaely (2019), Kwon, Ma, and Zimmermann (2021) and Philippon and Gutierrez (2017) for aggregate trends. Matvos, Seru, and Silva (2018) examine the link between scope expansion and episodes of financial market frictions.

[3]This perspective from the literature has an agency foundation (Jensen 1986) and suggests that complex division-management structures can create intra-firm agency problems that can create a diversification discount (Lang and Stulz 1994 and Berger and Ofek 1995). See also Matvos and Seru (2014) for a more recent treatment using a structural model.

sification not being a major consideration for the increases in scope we document. Indeed, using spatial representations of multi-industry firm operations, we confirm that scope is increasing strongly for operations that span highly related industries but not for operations that span likely unrelated industries.

An important question is how do corporate investment policies react to incentives to increase scope? To shed light on this question, we examine how plausibly exogenous variation in the ex ante incentives to increase scope impact ex post investment decisions and measures of firm performance. We consider two sources of such variation. The first measures each firm's scope-expansion opportunity set using the diversity of related markets served by each firm's distant peers. When these peers have operations that are distributed across multiple well-defined markets, it suggests that the focal firm itself sees a larger opportunity set of potential scope-enhancing projects, all else equal. Our second measure considers the cost-side and asset redeployability. We examine the asset portfolios of the firm's closer industry peers as compared to the firm's more distant industry peers (where distance is in product market space). If the closer industry peers' assets are easily redeployed to more distant peers' industries, it follows that the focal firm likely faces a low relative cost to expanding scope as its assets can be redeployed to more potential markets with lower adjustment costs.

Both variables measure a focal firm's incentives to increase scope, and neither is measured using any data about the focal firm itself. Instead, both are based on the characteristics of more distant industry peers, whose characteristics are not easily influenced by the focal firm and vice-versa. Our approach thus follows prior studies in the network econometrics literature that highlight the fact that endogenous effects are mitigated by focusing on distant peers and not the focal firm itself.[4] Although the literature indicates that this approach reduces the scope for endogeneity, we nevertheless interpret these results conservatively and view them as strong tests of our predicted mechanisms.

[4]See Bramoulle, Djebbari, and Fortin (2009) for theory and Cohen-Cole, Kirilenko, and Patacchini (2014) for a recent application in finance. These studies indicate that distant peers can produce exogenous variation that can be used as instruments.

We find that firm scope is significantly related to more acquisitions, fewer divestitures, more spending on research and development, increased outsourcing, and increased vertical integration. In contrast, we find no link to capital expenditures. These results are consistent with an ongoing process of asset redeployment across and within firms, which is reinforced by innovation that facilitates flexible and efficient redeployment of assets for multi-industry production (and in some cases, outsourcing). Importantly, this suggests that acquisitions and innovation facilitate scope increases, but our results are novel and are missed by existing studies because scope increases are accomplished without increasing formal operating segments in the Compustat database (the primary database used in the prior literature).

We examine firm outcomes and find that increased scope is associated with higher valuations, and thus it is unlikely that increases in scope are due to bad governance or private benefit extraction. This evidence of higher value is important given the prior literature and possible empire-building incentives managers would have to increase scope. We also find evidence of higher ex post sales growth and asset growth. However, we do not find any significant impact on profitability in the form of return on assets. These results suggest that scope expansion creates positive net present value and sales growth opportunities, and profit-maximizing firms likely pick the most profitable industries to operate in first and then expand into still-profitable but lower return on assets industries. Our valuation results - which show that firm market-to-book equity ratios increase with scope - are in contrast to the historical conglomerate literature (Berger and Ofek (1995) and Lang and Stulz (1994)) that finds that firm valuations decrease as reported Compustat segments increase.

We also examine how firm scope expansion is likely financed. Firms with higher ex ante incentives to increase scope issue more shares and pay lower dividends ex post. We find no significant link to debt financing. These results favoring equity are consistent with intangibles and the redeployment of existing assets playing an important role in scope expansion. In particular, this expansion method through intangibles and better utilizing existing assets does not create much new collateral, favoring financing with equity.

We note that these conclusions are drawn without using the traditional measure of scope: the number of Compustat segments. In particular, Compustat segment counts have not increased over time. We find that the lack of change in Compustat segments stems from four major facts. First, the accounting standard that governs segment reporting after 1997, SFAS 131, does not require reported segments to be based on the number of industries served, but rather on how managers evaluate performance. Intuitively, managers likely evaluate performance across related industries holistically. Second, we decompose multi-industry operations using a spatial model and confirm that Compustat segments essentially only track operations that are only weakly related (distant in the product market space). This finding strengthens after 1997 when SFAS 131 became effective. Third, we find that our text-based doc2vec segment database is more informative than the Compustat segment database in (A) predicting out-of-sample profitability and (B) in predicting which firms make direct statements indicating high-scope. Finally, we note that scope only increases for highly-related industries, which Compustat segments do not cover well.

We conclude with an analysis of whether increases in scope can provide new insights into why traditional concentration measures are increasing over time. We use two methods to adjust traditional HHI industry concentration ratios to account for increases in scope. Using both methods, we find that horizontal concentration measured using scope-adjusted HHIs is not increasing.

Our results suggest that increases in horizontal concentration are smaller when scope is accounted for, and thus changes in horizontal organization might not explain market power growth. Increases in scope, however, might motivate different concerns. Although scope directly reduces horizontal concentration as firms serve more markets, scope can increase anti-competitive conduct through product bundling or by increasing barriers to entry. Understanding the separate influences of horizontal competition and scope is fruitful for future research as regulatory interventions may differ for scope-induced versus horizontal market power.

1 Literature

Theories of economies of scale and scope were first developed by Panzar and Willig (1977) and Panzar and Willig (1981). Teece (1980) further develops relevant theory and suggests that a multi-product enterprise is particularly likely to emerge when economies of scope are based on a recurrent use of proprietary know-how. This theory therefore illustrates why our finding that R&D spending is increased when scope expansion incentives are high is consistent with theories that examine economies of scope. Henderson and Cockburn (1996) provide empirical support for a link between economies of scope and innovation investment in the pharmaceutical industry. Braguinsky, Ohyama, Okazaki, and Syverson (2020) provide further empirical support for these ideas in a novel historical setting: Japan's cotton spinning industry from 1893 to 1914. In particular, the authors show that technological capability was a major ingredient that fostered horizontal scope expansion and firm success in this early environment. Kwon, Ma, and Zimmermann (2021) focus on how scale economies over a long period of time have been achieved through the use of R&D and information technology. We also show the importance of R&D to increases in scope and how scope needs to be measured before attributing size increases to scale.

More recent theory by Maksimovic and Phillips (2002) postulates an efficiency-based view of multi-industry operations based on neoclassical profit optimization. In the model, a conglomerate discount can still emerge even when governance is aligned with shareholders and optimal policies are undertaken. However, if low-cost scope expansion is possible through economies of scope by sharing a scarce resource such as innovation or managerial talent, this discount may result in a premium. Our results favor this perspective on a few dimensions. In particular, we find that scope expansion with higher R&D brings higher valuations and sales growth, consistent with rational expansion.

Our paper also has implications for the older literature documenting a diversification discount (Berger and Ofek (1995) and Lang and Stulz (1994)). Although recent studies call the discount into question (see Custodio (2012) and Hund, Monk, and Tice (2020) for

example), our paper brings an entirely new perspective: modern multiple-industry firms share scarce resources and serve multiple industries while maintaining an efficient single segment organizational form with a high market valuation. Our study also contributes to the literature on acquisition motives and synergies (see Hoberg and Phillips (2010a), Rhodes-Kropf and Robinson (2008), Bena and Li (2014), and Fresard, Hoberg, and Phillips (2020)).

2 Data and Methods

Our sample begins with the universe of Compustat firm-years with available 10-K filings on the EDGAR system (later years) or scanned 10-Ks from the Dartmouth and Harvard libraries (earlier years of our sample). As the standard TNIC database of Hoberg and Phillips 2016 (HP2016) is based purely on EDGAR filings, its coverage begins in 1996. A contribution of the current study is thus that we back-extend the TNIC database to 1988 using 10-Ks from the aforementioned libraries. To remain in our sample, a firm must have a 10-K filing both in the current year of observation and in the previous year. We exclude firms operating in financial industries and regulated utilities (SIC 6000 - 6999 and 4900 - 4949, respectively) and limit the sample to firm-years with sales and assets of at least $1 million. We are left with 101,535 firm-year observations from 1989 to 2017.

2.1 Measures of Scope

We develop new measures of firm-year product market scope using the 10-K text-based framework of HP2016. These descriptions are updated as products evolve every year, and are required by Regulation S-K to accurately represent the products sold in each fiscal year.

2.1.1 Doc2vec-Scope

Our goal in this section is to use natural language processing to extend the work of HP2016, who model firms as operating in atomistic markets, to identify the extent to which firms

operate in multiple product markets. To do so, we decompose textual business descriptions into sub-dialects indicating separate product markets. We create two new data items: (1) a text-based alternative to the well-known Compustat operating segments database and (2) new firm-year measures of product market scope. Throughout this section we denote firms as i, industries as k, and time as t.

We achieve these objectives using the widely-used doc2vec embedding model (see Mikolov and Dean (2013) for example). We outline our methodology at a high level here and report step-by-step instructions in Appendix B. We train our doc2vec model using all 10-K business descriptions for our base year 1997 (HP2016 uses the same base year).[5] The result is a semantic language model that incorporates information about synonyms and words indicating similar context. For example, doc2vec would treat "couch" and "sofa" as similar whereas baseline cosine similarities would deem them unrelated. Although not the focus of the current study, the doc2vec model also facilitates an improved TNIC industry classification that is roughly 20% more informative than the baseline model in HP2016 (see Online Appendix 3 for details). This improved power is also helpful in the current context as we score each firm's 10-K regarding which product-market dialects are present.

Our doc2vec model maps all firm business descriptions into a 300-dimensional spatial model of the U.S. economy (we choose 300 dimensions following HP2016 who report information criterion tests indicating this dimensionality among U.S. firms). We first identify 450 "candidate industries" using k-means clustering run only on the single segment firm 300-dimensional vectors from the base year (the use of single segment firms ensures industries are well-identified). We additionally use the word2vec tool that is built into doc2vec to develop dialects for each of the doc2vec industries. The dialects allow us to score 10-Ks on each dialect, purge boilerplate content, and provide intuitive labels for each industry. We thus prune the 450 candidate industries to obtain 300 final "D2V industries" by purging 150 boilerplate clusters and redundant clusters (see Appendix B for details). The result is

[5] We thank Christopher Ball at metaHeuristica for suggesting doc2vec parameters: pv-dbow specification, 300 dimensions, a 15-word window, and 40 epochs.

a text-based analog to the Compustat segments database. We then measure scope for each firm in each year as the firm's number of doc2vec segments. Finally, we note that although we document strong validation results in Section 3.1, a limitation is that all NLP methods (including ours) are susceptible to noise, which motivates future research to improve power.

2.1.2 Alternative: NAICS-Scope

Our main alternative scope measure is based on NAICS industry definitions, as defined in the 2017 NAICS Manual, which is a 963-page document providing detailed verbal descriptions of each NAICS industry. We use the four-digit NAICS granularity, and group vocabulary for all industries having the same four digits of their NAICS code into each NAICS code's total vocabulary (see Online Appendix 4 for stop word methodology). There are 311 four-digit NAICS that we capture using this approach. We score each firm-year based on how much of each industry's vocabulary it uses. We avoid standard pairwise cosine calculations as that would overly penalize firms that have large product descriptions that cover many industries (because the fraction of each industry in the overall vector would be small resulting in low similarities for industries it actually operates in). Instead, parallel to the approach we use for D2Vscope, we compute the fraction of each industry's vocabulary that appears in each firm-year's product description as the following overlap ratio:

$$Q_{i,j,t,NAICS} = \frac{\#\text{words overlapping in } D_{NAICS,j} \text{ and } V_{i,t}}{\#\text{words in } D_{NAICS,j}} \tag{1}$$

Finally, to compute how many NAICS industries a given firm might operate in, we identify a fixed threshold $\bar{Q_{NAICS}}$ above which we deem a firm having $Q_{i,j,t,NAICS} > \bar{Q_{NAICS}}$ to be operating in industry j in the given year t. As above, we hold $\bar{Q_{NAICS}}$ fixed and do not allow it to vary with time, as otherwise, we might create false inferences in our time series analysis. Following HP2016, we use our base year of 1997 to compute $\bar{Q_{NAICS}}$ as the threshold such that 2% of all firm-industry combinations are operating-pairs in 1997. The

2% threshold is also used above and in HP2016 as it matches the granularity of three-digit SIC industries. Our alternative scope variable NAICS-scope is then the number of industries the given firm likely operates in above $\bar{Q_{NAICS}}$:

$$NAICS - Scope_{i,t} = \sum_{j=1,...K} Indicator\{Q_{i,j,t,NAICS} > \bar{Q_{NAICS}}\} \quad (2)$$

2.2 Local Asset Redeployability and Outward Scope Expansion

We develop two ex-ante measurable shifters of firm scope, which allow us to examine the extent to which plausibly exogenous incentives for firms to increase scope predict ex-post corporate strategies for increasing scope and subsequent performance. This approach provides the foundation for strong tests of our predicted mechanisms. Our first instrument is based on asset redeployability, specifically, each firm's ability to redeploy its assets in product markets that are nearby in the product space. Intuitively, a firm that can easily redeploy assets into spatially proximate product markets likely has strong incentives to increase scope because the cost of doing so is likely to be low. We focus on spatially local redeployability rather than broader measures of redeployability to increase the power of our instrument. This approach is motivated by the language theory of Crémer, Garicano, and Prat (2007) and tests in Hoberg and Phillips (2018) showing that firms expand scope by operating in groups of highly related industries and not distant industries.

Our approach to measuring local redeployability follows Kim and Kung (2017) (KK2017), which use the capital flows tables from the Bureau of Economic Analysis to compute measures of broad asset redeployability. This approach is also motivated by Boehm, Dhingra, and Morrow (2022) who show that input complementarities are important for product entry decisions in India. We refine the KK2017 methodology to focus on localized redeployability. In particular, the BEA tables indicate the extent to which each of the 123 BEA industries (which can be mapped to NAICS codes) utilizes a set of 180 assets. Intuitively, if two BEA industries utilize the 180 assets in similar proportions, we conclude that a firm operating in

one faces a high degree of asset redeployability and a low cost of entry into the other. This is an incentive to expand scope in that direction. To the extent that the asset allocation vectors across different industries are exogenous from a firm's perspective, it would follow that a firm facing high levels of asset redeployability for nearby industries faces exogenously higher incentives to increase scope in the future.

Of course, as innovation can change the distribution of asset allocations within an industry, it follows that asset-allocation vectors are not fully exogenous. We thus take additional precautions to further improve exogeneity. First, following KK2017, we use a single BEA table from 1997 for our entire sample and fix the asset allocation vectors in time. Second, we compute local redeployability for each firm using an approach that examines the product market around the firm, thus strictly avoiding using data from the focal firm itself. We compute the redeployability of the firm's close peers (based on the TNIC3 classification) to expand into the industries covered by the focal firm's more distant peers (those in the focal firm's TNIC2 classification but not those in its TNIC3 classification).[6]

We compute local asset redeployability by first mapping each BEA industry to a four-digit NAICS code (following KK2017), and representing the underlying assets used by each NAICS industry as a 180-element vector, which we denote as A_j for a given NAICS-4 industry j. Each vector is obtained from the 1997 capital flows table, which reports dollar amounts for 180 assets tracked by BEA. Next, for each focal firm in each year, we obtain two sets of peers. Close peers are those in the focal firm's TNIC-3 industry (excluding the focal firm itself). Distant peers are those in the focal firm's TNIC-2 industry but not in its TNIC-3 industry. There are no overlapping peers in these two sets. Next, we compute the fraction of each set of peers in each NAICS-4 industry.[7] $F_{i,t,j,near}$ is the fraction of focal firm i's close peers that are in 4-digit NAICS industry j in year t. $F_{i,t,j,distant}$ is analogously defined for distant peers.

[6]TNIC2 industries are the text-based industry classification from HP2016 that is calibrated to be as granular as two-digit SIC industries. TNIC3 is finer and is as granular as SIC-3. Thus firms that are in TNIC2 but not TNIC3 are "distant peers" as they are still in nearby markets but they are not in the focal firm's current market.

[7]We use NAICS and not TNIC industries for this part of the calculation because BEA tables are linked to NAICS.

These two industry distributions reflect likely paths that firms would take if they outwardly expand scope into the nearest distinct markets. Our key variable, Outward-focused local asset rededployability, is then the weighted average asset-complementarity (cosine similarity of the asset vectors for the two industries in a pair j,k) summed over the distribution of industry pairs spanned by the two sets of peers:

$$LocalAssetRedep._{i,t} = \sum_{j,ks.t.j\neq k\in NAICS-4} F_{i,t,j,near} F_{i,t,k,distant} < \frac{A_j}{A_j \cdot 1} \cdot \frac{A_k}{A_k \cdot 1} > \qquad (3)$$

We note that the above calculation does not depend on the focal firm itself and instead focuses on the industries served by peers that are more distant. This helps to reduce potential channels for violation of the exclusion requirement, as is reinforced by econometric theories of network identification (Bramoulle, Djebbari, and Fortin, 2009). Cohen-Cole, Kirilenko, and Patacchini (2014) is a related application in finance. Importantly, when local asset redeployability is high, it indicates that the focal firm likely has many ways to increase scope with relatively low adjustment costs. These strong ex-ante incentives are thus a shifter of the focal firm's ex-post scope expansion strategy.

2.3 The Local Scope-Expansion Opportunity Set

Our second shifter of scope incentives is based on the size of the outward-scope expansion opportunity set as seen from the focal firm's perspective. As was the case for our above cost-of-entry shifter, we construct our second instrument by focusing on more distant peers and avoiding using the characteristics of the focal firm itself. We thus identify distant peers as rivals that are in the focal firm's TNIC-2 industry but are not in the focal firm's TNIC-3 industry, and we compute the distribution of NAICS-4 industries served by these distant peers as ($F_{i,t,j,distant}$).[8] To compute the local scope expansion opportunity set (our second shifter), we compute one minus the concentration ratio (HHI) based on this distribution:

[8] Although we focus on distant peers to mitigate endogeneity concerns, we note that our results are similar if we instead base this calculation on proximate TNIC-3 peers (see Online Appendix).

$$LocalScopeExpansionOpp.Set_{i,t} = 1 - \sum_{j \in NAICS-4} F^2_{i,t,j,distant} \quad (4)$$

When the magnitude of this variable is high, it indicates that nearby peers serve a wide array of closely related product markets. From the focal firm's perspective, this indicates a high-quality opportunity set for scope expansion. Because this measure is only a function of the firm's distant peers, as noted above, it is plausibly exogenous from the perspective of the focal firm's policies. Focal firms facing a higher value of this variable are likely to increase scope given the wider array of growth opportunities available.

2.4 R&D, Investment and Acquisitions

We examine four investment policies: R&D/assets, CAPX/assets, acquisitions, and divestitures. The R&D (XRD) and CAPX variables are from Compustat. We scale each by the beginning of the period total assets (AT). When R&D is missing, we assume it to be zero.[9] We obtain acquirer and target data using both full-firm and partial-firm asset acquisition data from SDC Platinum. SDC Acquirer is an indicator equal to one if the given firm acquires assets from any seller (public or private) in the given year and is zero otherwise. SDC Target is an indicator equal to one if the given firm sells any assets to any buyer (public or private) in the given year and is zero otherwise. As variables related to investment, we also consider the measure of vertical from Fresard, Hoberg, and Phillips (2020) and the measure of outsourcing based on purchase obligations in 10-Ks from Moon and Phillips (2021).

We also consider four other outcome variables including sales growth and asset growth, which are the log of the ratio of current sales to past-year sales and current assets to past-year assets, respectively. We compute firm valuation ratios as the firm market value (market equity plus book assets minus book equity) scaled by total assets, and we compute profitability as operating income before depreciation scaled by total assets. Finally, we consider four financing policies including equity issuance, debt issuance, equity repurchases, and divi-

[9] If we exclude firms with missing R&D, we obtain similar results.

dends, with all four being scaled by assets. All accounting ratios are winsorized in each year at the 1%/99% level. A complete variable description is in the Appendix.

2.5 Summary Statistics and Correlations

Table 1 displays summary statistics for our 1989 to 2017 panel of 101,535 firm-year observations. The average value of our key D2V-scope and NAICS-scope variables are 7.5 and 6.3, respectively. This suggests that using 2% granularity, the average firm in our sample is operating in markets that are related to roughly six to seven well-defined D2V or NAICS-4 industries, respectively. This is larger than the average number of Compustat Operating segments, which is just 1.4 in our sample. We measure scope in this relatively broad way to ensure there is adequate power to compare firms in the cross-section, and because operating segments likely understate the true girth of the product portfolios offered by public firms in the United States. Notwithstanding that, we also note that our results are robust if we measure scope more narrowly using a 1% threshold or more broadly using a 5% threshold.

We also note that our accounting variables have values that are similar to those in other studies. The average firm in our sample spends roughly 5.5% of its assets each on R&D and CAPX, and 28.5% of our sample firm-years are involved in an acquisition. The average firm's valuation ratio (market to book) is roughly 1.71, and the average firm spends roughly 1.8% of its assets on repurchases and 0.8% on dividend payments.

Table 2 displays Pearson correlation coefficients. Our two scope variables (D2V-scope and NAICS-scope) are about 12% to 20% correlated with the number of Compustat operating segments. This is significant and positive as expected. Yet it is far from unity, illustrating why segment counts are inadequate. Both measures are also roughly 28% correlated with firm size, illustrating that larger firms serve a wider array of product markets. This finding also illustrates why controlling for size is important. For example, both scope measures are positively correlated with the acquisition and target dummy. However, both dummies are even more positively correlated with firm size, as it is well-known that larger firms

are more active in restructuring. Given these facts, it is not surprising that when we run formal regression analysis, we find that scope is associated with more acquisitions but fewer divestitures (targets), which is consistent with the intuition that firms with high incentives to increase scope are indeed net acquirers once size is held fixed.

3 Basic Properties and Time Trends

In this section, we explore and validate the properties of our scope variables and compare them to the Compustat segments database and to firm size. We first illustrate examples. Table 3 displays the markets that CVS operates in over our sample period. The example documents the early focus on variety stores while CVS was a part of Melville in 1990, its shift to the pharmacy and stationery markets by 1999, and its expanding scope into groceries by 2005. Finally, it illustrates the more significant expansion in scope induced by the acquisition of Caremark in 2007, resulting in the more complex array of insurance and medical offerings we observe by the end of our sample in 2017. Notably, CVS had just two segments showing no variation during all of these years except for 1990 (when it had four). Table 4 displays Tesla's segment allocations before and after its merger with Solar City in 2016. The table illustrates Tesla's marginal relevance to the solar industry prior to the acquisition, as its pre-merger capabilities included batteries and energy storage that are relevant to the solar industry. Yet after the acquisition, we observe a dramatic rise in the "solar and photovoltaic" industry as the amount of exposure quadruples from .07 to .28. Note that Tesla had two Compustat segments in both years masking this very rich heterogeneity in the data. These examples document that our text-based approach well-captures variation in scope as firms undergo relevant transformations. In the Online Appendix table IA1, we also show a similar example for SalesForce.com, which illustrates a rich array of markets served despite the Compustat segments database indicating just one segment for this firm throughout its history.

3.1 Compustat Segments and Firm Size

The existing literature uses the Compustat segment tapes when exploring firm scope. It takes the perspective that Conglomerate firms have high product scope, and are diversified as they operate in unrelated product markets. We take the perspective that the segment tapes are problematic for measuring scope, not only due to mismeasurement (see Villalonga 2004), but also because we expect modern firms to increase scope without increasing the number of rigid operating segments. In particular, we expect modern firms to use innovation to increase product scope through more flexible production and by redeploying existing assets. A consequence would be increasing scope but Compustat segments would not increase as managers view these blended multi-industry operations as a single operation.

We begin our analysis by computing summary statistics for subsamples of firms with different numbers of Compustat segments. Panel A of Table 5 displays the results and shows that moving from one segment to two increases D2V-scope and NAICS-scope by one-half to one unit. For example, D2V-scope increases from 7.25 to 7.80. Adding segments beyond two further adds to our scope measures. These results conform to intuition. However, these relationships under-state the true variation in measured scope, which has a standard deviation ranging from five to seven, and has a modest correlation with Compustat segments.

Panel B of Table 5 reports scope statistics versus firm size quintiles. Quintiles are formed by sorting on Compustat assets in each year. The table confirms that all measures of scope increase with firm size. For Compustat segments, the smallest quintile firms have an average of 1.22 segments versus 1.94 for the largest firms. Regarding D2V-scope, the range is from 5.87 to 9.26 product markets. The range is larger for NAICS-scope at 4.01 to 9.23 markets. Yet the growth in scope pales compared to the range in firm size itself as small quintile firms have about $23 million in assets, compared to $11.7 billion for the largest quintile firms. We conclude that some but not all increases in firm size are likely attributed to firm scope. Panel C of Table 5 reports the same statistics for the subset of single-segment firms and illustrates that the large variation in firm scope with firm size is not much diminished. This reinforces

our conclusion that Compustat segments are likely an imperfect source of information about firm scope.

3.2 Compustat Segment Trends and Limitations

Figure 1 plots a number of important stylized facts. The upper panel shows that Compustat segments initially decline from roughly 1.5 in 1989 to 1.3 by 1997. This conforms to the intuition that older conglomerates, which might have been formed in the 1970s and 1980s, were gradually disbanding over time. However, from 1997 to 1999, we observe a major structural break and the number of segments suddenly increases to more than 1.5. This jump can be explained by SFAS 131, in which FASB changed segment disclosure requirements for fiscal years ending after December 15, 1997. This rule required that managers must report segments based on how they internally evaluate operating performance. Before this rule, segment reporting was instead based on an industry approach. The rule change was precipitated by concerns by market participants that segments were being under-reported, perhaps for strategic reasons (see Song (2020) for a detailed summary of events).

The events leading up to SFAS 131, and the rule change itself, suggest that any trend for Compustat segments in Figure 1 should be interpreted with caution. The alleged practice of under-reporting prior to the rule change calls the declining trend from 1989 to 1997 into question. The flat trend after 1997 is also questionable because post-change segment counts are based on how managers internally evaluate performance and not how many product markets the firm actually operates in. A major result we report later is that firms increasing scope choose to operate in industries that are closely related, and intuitively, many managers might prefer to assess the performance of such related product lines internally. If so, the observed number of Compustat segments would understate any increases in scope.

We now directly test our conjecture that managers do not report segments when they are highly related. We start by taking all pairwise permutations of the 300 D2V industries and compute pairwise industry similarities for each industry pair based on the cosine similarity

of each industry's centroid doc2vec vector (see Section 2.1.1). We sort all industry pairs into quintiles based on how spatially distant they are from each other and label pairs in the most similar quintile as "most related," those in the second most similar quintile as "weakly related," and those in the least similar three quintiles as "likely unrelated." (these latter three quintiles have uniformly very low similarity). For each firm in each year, we then compute the number of D2V segment-pair permutations it has in each of the three product distance bins (weighted by the total number of segments). For each firm-year, we thus have an effective number of segments that span "most related", "weakly related" and "likely unrelated" industries. To understand the propensity of managers to report operating segments as a function of industry distance, we then regress the number of Compustat segments reported by the firm on all three distance-based segment counts. We report results separately for each year so that we can report overall results and how these relationships change around the advent of SFAS 131.

Table 6 reports the results. Our results overwhelmingly support the conclusion that managers mainly report segments when their firm's operations span only weakly related industries as this variable is positive and highly significant in every year with an average annual t-statistic of 6.41. We find a similar but weaker positive coefficient (average t-statistic of 2.43) for segment reporting when the firm's operations span likely-unrelated industries. Yet the most striking result is that the number of Compustat segments is in fact negatively related to the number of most-related industries a firm's operations span. This would suggest that firms operating in multiple highly-related industries eschew segment reporting. Instead, these managers likely evaluate performance in these highly related industries in a unified or holistic way. We also note, consistent with SFAS 131, that the negative coefficients for the most similar segments experience a material jump from -.013 to -.023 around 1998 to 1999, which is consistent with the view that SFAS 131 in fact empowered managers to avoid segment reporting for these related operations. At the same time, we see a modest jump in managers reporting segments for weakly related industries at this time, as managers are

much more likely to evaluate performance separately for less related industries, and SFAS 131 would require reporting these weakly related segments. These results reinforce that a key weakness in Compustat segments is that managers do not report segments when they span highly-related industries, and this lack of reporting has become more extreme in the years following the passage of SFAS 131.

3.3 Scope Trends

The middle panel of Figure 1 displays the average D2V-scope and NAICS-scope over our sample. The coverage dating back to 1989 was made possible by our backward extension of the TNIC database. We find that scope was increasing steadily throughout our sample (50% to 70% overall), with the most rapid rate of increase between 1997 and 2013. In contrast, the number of Compustat segments hardly changed during this period. These results are consistent with the view that product market scope increased as firms served more and more related industries (for example selling computers and cell phones) rather than diversified and unrelated industries (such as selling oil and cat food). Compustat segments show no reported increases as managers tend to evaluate related industries together.

The lower panel of Figure 1 shows that both large and small firms experienced similar increases in scope after 1997, although larger firms initially had a period of declining scope in the early 1990s. This latter fact is consistent with the break up of inefficient diversified conglomerates formed in earlier decades. Online Appendix Figure IA1 shows that these trends are robust to using medians instead of means and for large versus small firm subsamples. Perhaps more stark given our paper's overall conclusions about scope increasing mainly for highly related products markets, Figure 2 (and corresponding Online Appendix Table IA3) illustrate that scope indeed increased far more dramatically for the most related product markets (46%), relative to weakly related product markets (29%) and likely unrelated product markets (11%).

A final note is that the increased scale and scope we report might also partly explain why

the length of 10-Ks has increased over time. The upper panel of Online Appendix Figure IA2 plots the average number of words in the 10-K Item 1 over time, which increased in the first half of our sample until 2003. The increasing trend was interrupted around 2005, likely due to the requirement that risk factors be reported in a separate Section 1A of the 10-K, and then grew more slowly thereafter. This suggests that document size is related to scope, and the fact that our graphs in Figure 1 show no structural break around 2005 further illustrates that our scope measures are indeed purged of boilerplate content as was our design.

Regarding scale, Figure 3 plots average firm size over time (based on book assets) both in nominal terms and in inflation-adjusted terms, and shows that average firm size has increased substantially over time. Using the more conservative inflation-adjusted metric displayed, firm size has roughly tripled during our sample. This increase likely partially reflects the increases in firm scope we document above, but it also indicates unique increases in firm scale.

4 Scope but not Diversification

We first examine if the high levels of scope we find are related to companies spanning distant and highly diversified product markets, or more proximate related product markets. We begin by computing the average product market distance between every permutation of pairs of D2V-300 industries. For a given pair of industries in a given year, this is computed as the average TNIC pairwise similarity (see HP2016) between all of the firms in the first industry relative to those in the second industry in the pair. Finally, we sort industry pairs into deciles in each year based on the TNIC similarity of the pair. As this calculation is general for any pairwise relatedness network, we also compute vertical-relatedness deciles for each industry-pair in each year using the firm-pairwise VTNIC vertical-relatedness network from Fresard, Hoberg, and Phillips (2020).

Next, for each pair of industries in each year, we count how many multi-product firms jointly operate in the pair. This is done by counting across permutations. For example, a firm

that operates in three D2V industries {i,j,k} counts as operating in three pairs: {ij,ik,jk}. We then tabulate the number of operating pairs for each pair of industries and obtain the distribution of operating pairs for each year. Table 7 then reports the fraction of all observed operating pairs that are in each industry-pair-similarity decile. We report this distribution for all firms, and separately for single-segment and multi-segment firms. The table shows that firms overwhelmingly operate in industry pairs that are close together in the product space. Almost 53% of all operating pairs are in the highest decile of TNIC industry pairwise similarity. An additional 15.6% are in the next decile. These results indicate that modern multi-industry firms are not the diversified conglomerates portrayed in the early literature. Rather, these firms operate in highly related industries with value-adding synergies (we present evidence of higher valuations later). Regarding vertical relatedness, we observe a U-shaped pattern for operating pairs with many spanning vertically unrelated pairs (31%), many spanning the most vertically related pairs (13.5%), and fewer spanning the middle terciles (5% to 6% for most middle terciles). The results overall indicate that multi-industry firms are more focused on horizontal rather than vertical scope expansion.

We next examine the risk properties of our scope measures. The diversification hypothesis would suggest that firms with spatially distant operating segments should be less risky due to the diversifying effects of unrelated markets. We thus separate D2V-scope components into terciles of "close scope" and"far scope". This is done by computing the average pairwise distance of the D2V-300 industry pairs each firm operates in (using the same operating pair database discussed above) and sorting firms into terciles in each year. Table 8 then reports the results of regressions where measures of firm risk are regressed on D2V-scope separately for each tercile. We include controls for size, age, and year fixed effects. We additionally include firm fixed effects in Panels C and D. The first dependent variable Market Volatility is the standard deviation of the firm's daily stock returns in year t, and Cashflow Volatility is the standard deviation of a firm's quarterly operating income scaled by assets, computed over the 8 quarters of year t and $t+1$.

Panels A and B (without firm fixed effects) show that scope is positively related to both risk measures in all three terciles. Thus, we do not find evidence of diversification, which would predict a negative sign. Rather, these results are consistent with the fact that firms that tend to increase scope in modern times are more innovative as they spend more on R&D for example (documented later), which tends to be risky. Consistent with such a firm fixed effect, Panels C and D (with firm fixed effects) show that scope is not significantly related to either measure of risk in any tercile. These results support our conclusion that diversification is unlikely a major motive for scope increase. This is likely because scope expansions rely on innovation and focus on related markets with less potential for diversification.

4.1 Validation of Scope Measures

We further validate our scope measures using two final tests. Our first aims to assess the relative informativeness of our D2V-segment database compared to the Compustat segment database using out-of-sample profitability (oi/assets and oi/sales) prediction tests analogous to HP2016's validation of the TNIC database. Where $\omega_{i,k,t}$ denotes firm i's exposure to industry k in year t, we do so using a generalized fixed effects model as follows:

$$OIassets_{i,t} = \sum_{k=1 \to 300} \omega_{i,k,t} \cdot \mu_{k,t} + \epsilon_{i,t} \tag{5}$$

For D2V segments, $\omega_{i,k,t}$ is set to the amount of textual exposure each firm has to industry k. This fixed effects model is generalized because it allows each firm to have partial weights for each industry. For Compustat segments, $\omega_{i,k,t}$ is set to the fraction of sales allocated to the given SIC-3 industry k (allowing Compustat to incorporate its main potential advantage as it contains sales data for each segment). The model in equation (5) can be fitted to identify $\mu_{k,t}$ for each industry-year using OLS where the RHS variables include one vector for each industry, which for each firm, is populated by the weights $\omega_{i,k,t}$.

We start by estimating equation (5) in-sample using the set of single-segment firms.

Table 9 reports the results as the first three rows in Panel A (oi/assets) and the first three rows in Panel B (oi/sales). We find that D2V segments have more in-sample explanatory power than Compustat segments (31.6% versus 23.7% adjusted R^2 in Panel A). We find similar conclusions in Panel B for oi/sales, and for both the adjusted R^2 (higher is better) and the Akaike Information Criterion (lower is better). Although D2V segments are more informative, we also find that both segment databases contain unique information, as the adjusted R^2 in Panel A is 33.4% when both sets of generalized fixed effects are included. Finally, the last three rows in each panel use the fitted values from the single-segment firm regressions to fit predicted profitability for multiple-segment firms. This is a pure out-of-sample test that is in the spirit of Berger and Ofek (1995) (who also use single-segment firms to fit a valuation model and then consider multiple-segment firms out-of-sample). The out-of-sample results also confirm that D2V-segments better-explain conglomerate profitability than do Compustat segments (adjusted R^2 of 3.8% in Panel A compared to 2.9%), and that both continue to contain distinct information (adjusted R^2 of 5.2% for both). These findings are robust in both panels and to using the Akaike Information Criterion.

Our second validation test considers four queries of firm 10-Ks to identify direct statements indicating a high degree of scope. We consider three lists:

List A: product lines, product categories

List B: product lines, product categories, service lines, service categories

List C: breadth, broad, broader, wide, multiple, numerous, diverse, categories, divisions

We use the metaHeuristica software to compute four variables of interest. "Product Breadth" is the number of paragraphs in each firm's 10-K that mentions a phrase in List A, scaled by the total number of paragraphs in the firm's 10-K. "Product/Svc Breath" is analogously defined based on List B. "Product Breadth Detail" is the number of paragraphs that contain a phrase in List A and also a word from List C. "Product/Svc Breadth Detail" is analogously defined using List B and List C. Intuitively, when these scores are higher, the firm likely offers a high scope array of products and services. To validate our measures,

we regress the four above variables on candidate scope measures. We additionally include controls for size, age, market to book, and TNIC HHI, and firm and year fixed effects. A quality measure of scope should have a strong positive and significant coefficient.

Table 10 displays the results. The first four rows only include controls as a baseline and illustrate that firm size, not surprisingly, is related to direct statements indicating scope with a t-statistic between 3.5 to 4.0. Rows (5) to (8) add the number of Compustat segments to the regression. This variable is positive and significant at the 1% level for the first two variables (t-statistic of roughly 2.7), but only significant at the 10% level for the latter two more stringent variables. Rows (9) to (12) add D2V-scope to the regression. This variable is much more positive and significant than both firm size and Compustat segments with a t-statistic ranging from 6.0 to 6.7 for all four variables. Additionally, including D2V scope reduces the significance of the Compustat segment by roughly 25%, making it insignificant in the last two rows. Rows (13) to (16) reproduce this test for the NAICS-scope variable, which also performs well but is marginally weaker than D2V-scope. We conclude that our text-based measures dominate Compustat segment counts on this test of direct scope mentions.

5 Scope Incentives and Corporate Finance Policies

We now explore how firms seeking to increase scope modify their corporate finance policies. This question touches upon many issues of importance for understanding corporate finance and also issues of relevance to regulators. For example, is the increase in scope we report related to the boom in acquisition activity reported in media over the past couple of decades? Additionally, does innovation spending increase with scope, or is scope achieved instead through acquisitions and capital expenditures? We examine these mechanisms using plausibly exogenous shifters of the incentives firms have to increase scope. We also assess the link between scope and firm performance and how investments are financed.

5.1 First-Stage Analysis

We first examine the relation between our plausibly exogenous measures of scope incentives and observed levels of scope. This constitutes the first stage in our two-stage least squares analysis in the next section. We use these models to assess the relevance of plausibly exogenous incentives to increase scope on various corporate finance outcomes. This is thus a test of mechanism relevance more than a test of pure causality.

Our first ex-ante measure of scope incentives is *Sectoral Redeployment Potential*, which we explained in Section 2.2. This variable measures the extent to which the assets owned by a firm's close peers can be easily redeployed for use in the product markets covered by the focal firm's more distant peers. When this variable is high, the focal firm likely has the ability to increase its scope outward at low cost, as its assets are likely redeployable to produce in these nearby product markets. Our measure is *Sectoral Opportunity Set Potential*, which is based on the supply of scope-expansion opportunities rather than the cost of implementing them. This measure is one minus the concentration ratio of the distribution of industries spanned by the focal firm's more distant peers. When this quantity is high, there are many related product markets to expand into (a "thick" opportunity set). Both measures are computed without using the characteristics of the focal firm itself and are weighted on more distant peers. Econometric research suggests that the use of distant peers is more plausibly exogenous due to their second-degree (rather than first-degree) network linkages.

In our first stage analysis, we regress our scope measures on both scope instruments and include all control variables included in our two-stage models, including firm and year fixed effects. The results are in Table 11. Row (1) shows that both scope incentive variables are positively related to D2V-scope. *Sectoral Redeployment Potential* is positive with a t-statistic of 3.7, and *Sectoral Opportunity Set Potential* has a positive t-statistic of 11.9. Results are slightly weaker for NAICS scope and significantly weaker when the # of Compustat segments is the dependent variable. Our scope incentive variables are thus powerful shifters of D2V-scope and NAICS-scope, but not the number of segments. Thus we focus on D2V-scope and

NAICS-scope in our second-stage models.

5.2 Corporate Finance Policies and Scope Expansion

We now consider two-stage regressions where we assess the impact of ex-ante scope incentives on ex-post investments, including acquisitions, divestitures (target of acquisition), R&D, CAPX, vertical integration, and outsourcing. In particular, we instrument our scope measures using our two above instruments based on ex-ante scope incentives. We also control for size, age, and firm and year fixed effects along with a specification that also controls for ex ante market to book and the TNIC HHI.

The results are displayed in Table 12. In Panel A, we find that firms with high ex-ante scope incentives acquire more ($t-statistic$ of 3.5) and divest less ($t-statistic$ of -2.9). Economic magnitudes are also quite large. If we vary predicted scope from the 25th to the 75th percentile, the change in the probability of acquiring another firm increases by 6.55 percentage points, which is 22.7% of the mean and 28.75% of the standard deviation of the probability of making an acquisition. For the probability of divesting, economic significance is smaller but still meaningful with the predicted probability declining by 4.15 percentage points, which is 32.9% of the mean and 12.5% of the standard deviation.

These results indicate that acquisitions are a key way to increase scope, and avoiding divestitures is also helpful to avoid losing previous gains in scope. They also suggest that scope increases might help to explain why acquisitions became so prevalent over the past two decades and inform regulatory debates on excessive acquisitions.

Firms with high incentives to increase scope also increase R&D, but not capital expenditures. The increase in R&D is significant with a t-statistic of 3.4. Economically, if we vary the predicted scope from the 25th percentile to the 75th percentile, firms are predicted to increase their R&D/Assets ratio by .82 percentage points, which is 15.5% of the mean and 7.7% of the standard deviation of R&D / Assets. This finding suggests that innovation likely helps to facilitate scope expansion as do acquisitions. For example, when assets can be

redeployed across product markets, innovation spending can reduce the cost of redeployment and improve productive flexibility. This can also facilitate multi-industry operations that do not need multiple operating divisions. For example, firms might use R&D to develop more universal production sites. Indeed our earlier results suggest that increases in scope were achieved with almost no change in the average number of Compustat segments.

Finally, we examine the extent to which incentives to increase scope also shift vertical integration and offshoring. We find a significant positive relationship for vertical integration, suggesting that vertical integration likely increases scope synergistically. Economically, if we vary predicted scope from the 25th percentile to the 75th percentile, firms are predicted to increase vertical integration by .81 percentage points, which is 71.2% of the mean and 73.7% of the standard deviation. We also find that scope incentives also increase outsourcing, which might further explain why scope increases are possible without higher CAPX.

Online Appendix Table IA4 reruns the tests in Table 12 separately for each Fama-French-5 industry sector. The table shows that firms in the technology sector primarily increase scope by investing in R&D rather than acquisitions. In contrast, firms in manufacturing do the opposite. These results are intuitive, given the focus on intangible assets in the tech industry and the focus on tangible assets in manufacturing. The other three sectors are consistent, with both investment channels (R&D and net acquisitions) being relevant.

Online Appendix Table IA5 reruns these tests using a one-stage model in which the two ex ante scope-incentive measures are included directly as regressors. We find that increased R&D is significantly and positively related to the asset redeployability incentive, and increased acquisitions are most related to the opportunity set incentive. These results suggest that innovation spending indeed might be used to redeploy flexible assets, and acquisitions are more likely in markets with numerous opportunities.

5.3 Ex-post Outcomes, Financing, and Scope

Table 13 reports the results of analogous regressions for ex-post performance. We consider valuations (market to book), sales growth, asset growth, and return on assets. All right-hand-side variables are lagged one period, and we use two-stage least squares using our two ex ante scope incentive variables. We find that firms with high scope-expansion incentives experience higher ex-post valuations, sales growth, and asset growth. In contrast, profitability measured as return on assets (ROA) is not significantly related to scope incentives. Economically, if we vary the predicted scope from the 25th percentile to the 75th percentile, firms' valuation is predicted to increase by .34, which is 19.7% of the mean valuation and 21.4% of the standard deviation. For sales (asset) growth, respectively, these economic magnitudes are 11.5 (15.6) percentage points, which is 43.7% (35.5%) of the standard deviation of sales (asset) growth.

Overall, the higher valuations suggest that scope expansion is a positive net present value investment. The non-result for ROA is consistent with the view that firms focus on the most profitable markets first.[10] This interpretation is consistent with all four of our findings on performance. Our positive valuation results are in contrast to the historical conglomerate literature (Berger and Ofek (1995) and Lang and Stulz (1994)) that finds lower valuations for firms with more Compustat segments. In Online Appendix Table IA6, we consider a one-stage model as before. The results suggest that value creation is strong for both channels, but sales growth and asset growth are highest for the opportunity set channel.

We next explore the importance of scope expansion for venture capital funded private firm entry and product market fluidity (the rate of change in product portfolios of peer existing public firms). We follow Hoberg, Phillips, and Prabhala (2014) to measure both quantities. *VC funding similarity* is the cosine similarity of the focal firm's 10-K product description to the average vocabulary used by all startups in the given year (where the startup vocabulary is obtained from Venture Expert business descriptions of all startups receiving their first

[10]The non-result for ROA is robust to alternative specifications such as (A) truncating ROA at zero when it is negative or (B) scaling profitability by sales instead of assets.

round of financing in the given year). *Product Market Fluidity* is the average market-wide change in the use of the given firm's 10-K product description vocabulary by all other firms.

Table 14 shows that venture capital entry is strongly related to scope increase incentives, and fluidity is also higher when scope incentives are strong. Economically, the predicted impact on the firm is large. If we vary the predicted scope from the 25th percentile to the 75th percentile, venture capital financed entry is predicted to increase by 41.9% of the mean and 114% of the standard deviation of venture capital financed entry in our sample. For product market fluidity these magnitudes are 40% of the mean and 76.8% of the standard deviation. When interpreted alongside our earlier finding that high-scope firms also make more acquisitions, these innovation results are consistent with Phillips and Zhdanov (2013), who show that existing firms have incentives to purchase related start-ups. This suggests existing public firms outsource scope-enhancing R&D to startups (thus the higher VC funding), and they later buy them (thus the increased acquisitions).

Table 15 reports the results of analogous regressions for ex-post financing policies to explore how firms finance scope expansions. Our results suggest that equity is more commonly used than debt. Increased equity appears to accrue both through the issuance of new shares and through lower overall dividend payments. If we vary the predicted scope from the 25th percentile to the 75th percentile, equity issuance is predicted to increase by 3.1 percentage points which is 65.3% of the mean and 23.8% of the standard deviation in our sample. Dividends are predicted to decrease by 32.3% of the mean and 12.4% of the standard deviation. The use of equity is consistent with our earlier results on innovation and asset redeployment, as neither creates a significant amount of new fixed collateral that is traditionally associated with debt financing. The one-stage results in Online Appendix Table IA7 show that these results are rather evenly spread across the two scope incentive variables.

Although we do not claim to establish full causality as our goal is to construct strong tests of mechanism, we nevertheless report statistical tests of the quality of exogenous identification in Tables 12 to 15. We first note that our IV models produce a Kleibergen and Paap

(2006) r-k statistic that is significant at the one percent level, indicating that our instruments have power. Regarding the Sargan-Hansen test of overidentification (see Sargan (1958) and Hansen (1988)), the majority of our dependent variables are not susceptible to overidentification as the Hansen J-test is not significant at the 5% level. These dependent variables include the acquisition dummy, the target dummy, vertical integration, sales growth, asset growth, equity issuance, and dividends. The four variables that have a significant J-test, and thus exclusion is more in doubt, are R&D/assets, M/B ratio, VC funding similarity, and product market fluidity. Overall we believe our results support the conclusion that incentives to increase scope likely explain a wide array of corporate finance policies.

We conclude with a note on robustness. First, our results are robust to adding NAICS-4 x year fixed effects in addition to our firm and year fixed effects (see Table IA8). Second, readers might be concerned that our results are excessively driven by very large firms, like Amazon, which are known to have experienced large scope increases. Online Appendix Table IA9 drops the 50 largest companies from the sample in each year based on lagged assets and our results are fully robust. Third and finally, we define our opportunity set instrument based on distant peers to improve identification, but we also examine robustness to using more proximate peers. The results are in Table IA10, and all results are fully robust with just change: we find positive results for CAPX/assets in this specification. This suggests that more endogenous scenarios might indeed result in the need for more CAPX to increase scope, although the breadth of our results suggests that this is not broadly the case.

5.4 Extent of Scope Expansions

We next examine if our findings regarding scope-induced corporate policies and performance differ as firms become more spread out regarding the markets they serve and as scale increases. These issues are important because, if the gains to adding more scope are largest for firms that have already achieved high scale and scope, it would suggest that the trend toward increased scope is likely to continue. Alternatively, if the gains are largest for smaller

firms, then we might expect the process to slow as scope reaches saturation.

We use annual sorts to split our sample into above and below-median groups based on the extent of scope-based spread (the average spatial distance of the segments each firm covers) firms have achieved ex ante, and also based on firm size using lagged assets. Firms with "near" scope are those with segments that are similar to each other in terms of product market spatial distance (these firms have a narrow market presence and operate in highly related markets). As a result, we have four subsamples once we consider both spread and size: Near-Scope Small Firms, Near-Scope Large Firms, Far-Scope Small Firms, and Far-Scope Large Firms. We then run our two-stage investment, performance, and issuance scope models as we did in the last section separately for each subsample. For parsimony, we only report the coefficients on our instrumented variable of interest (D2V-Scope).

Table 16 shows that most results are strongest for smaller firms with near-scope and weakest for larger firms with far-scope. The smaller and near-scope firms do more acquisitions, have higher valuations and more sales growth. Firms with far scope have weaker results, although they do more R&D. This suggests that far-scope expansions are more difficult and likely require additional investment in innovation. Our stronger results for near scope suggest that there are likely diminishing gains to scope when firms increase scope too far and become too large. This finding is consistent with potential inefficiencies with excessive scope, as suggested by the early conglomerate literature that focused on diversified firms. Further consistent with these views, we documented earlier that few firms actually choose to operate across markets that are spatially distant, especially in more recent years.

6 Implications for Concentration

Although our paper's primary objective is to document the rise in scope and examine the role of scope in corporate finance strategies and performance, we also present some initial evidence that increasing scope has implications for industry concentration. We show that the

trend of increasing industry concentration ratios documented in the literature (see Grullon, Larkin, and Michaely (2019) for example) can be explained by the scope increases we report. Given this evidence is suggestive, we discuss it briefly here and show graphical details of concentration ratios evolving by year after taking into account scope increases. We provide more extensive details in Section 2 of the Online Appendix.

Intuitively, if the number of firms in the economy was held fixed, and every firm expanded its scope to serve twice the number of product markets, it follows that pure horizontal competition would increase economy-wide as more firms would be serving each market, and consumers would have twice as many options in each market. However, if these expanded firms are (incorrectly) assigned only to their historical industry classifications, industry concentration ratios would be overstated. One remedy is to use the Compustat segment tapes (Hoberg and Phillips (2010b), Grullon, Larkin, and Michaely (2019)), which allow the researcher to assign firms to more than one industry. However, as we noted earlier, SFAS 131 decoupled segment reporting from the actual industries firms operate in starting in 1997. Likely as a consequence, Figure 1 shows that segments do not capture increases in scope.

We consider two approaches to adjust concentration ratios for increasing scope. The first method computes HHIs using the scope segments that we construct using text, thus replacing the potentially problematic Compustat Segment database that has been used in the literature. This approach shows no increase in industry concentration since 1997 (see Figure 4). A limitation of this approach is that we don't have sales weights by segment, and therefore, we use textual intensity weights, where the weights are the similarity of the firm's product description to the vocabulary we used to construct each corresponding D2V-300 segment. We also find similar results when we equally weight across segments. This method is discussed in detail in Section 2.1 of the Online Appendix. Our second method does not rely on segment weights. Rather, we compare each firm's product description similarity to narrower SIC-3 product words and broader SIC-2 product words. We justify this approach by showing that, over time, firm product descriptions load more on the broader SIC-2 industry

vocabulary and less on the narrower SIC-3 vocabulary, indicating that firms operate at a coarser granularity in later years. Figure 5 reinforces our above finding using granularity-weighted HHIs and documents that HHI indices do not increase since 1997. This method is discussed in detail in Section 2.2 of the Online Appendix.

Our conclusion is that increases in scope can decrease concentration. A limitation of our concentration analysis common to studies in this area is that we only have scope data for U.S. publicly traded firms and we are unable to account for private or foreign firms. We thus plan to share our data and methods for future researchers to extend. However, we also note that stylized facts suggest that accounting for both foreign competitors and private firms should reinforce our finding that concentration is not strongly increasing over time. For example, globalization has increased and accounting for foreign competition would likely further reduce the growth rate of concentration over time. In addition, studies including Ewens and Farre-Mensa (2020) suggest that larger firms are staying private longer, and accounting for larger private firms should further reduce measured concentration.

Our results suggest that purely horizontal concentration does not increase when scope is taken into account. Increases in scope, however, might motivate different concerns. Although scope increases reduce horizontal concentration as firms serve more markets, scope may increase anti-competitive conduct through product bundling, by increasing barriers to new entry, or it can induce kill zones in the market for innovation (see Kamepalli, Rajan, and Zingales (2020)). Firms achieving high levels of both scale and scope might realize particularly elevated levels of market power through these channels.

7 Conclusions

We use textual analysis of firm 10-Ks to compute novel firm-year measures of firm scope and likely operating segments. Using our new measures, we find that the scope of U.S. firms has increased dramatically during our sample period from 1989 to 2017. Our findings

illustrate the rise of a new 21st-century high-scope firm, which increases product scope using innovation and acquisitions. These firms can serve multiple product markets without increasing the number of operating segments. Indeed, analogous tests using the traditional Compustat segment tapes are generally uninformative regarding the scope of U.S. public firms. These findings support our thesis that modern firms build multiple-industry related product portfolios while maintaining a simple single-segment organizational form.

We find that firms increase scope by acquiring more, divesting less, increasing innovation spending in R&D, and outsourcing more, but they do not increase CAPX. The increased innovation spending is consistent with developing increased flexibility in production. Firms increasing scope also realize higher valuations and higher sales growth. Scope expansion is financed using equity rather than debt, consistent with intangibles and asset redeployment not creating material amounts of new collateral. We document that related-market scope expansion is also highly valued by the market. This finding illustrates the novelty of our results, given the conglomerate discount previously documented in the early literature.

We conclude our analysis with exploratory evidence that the increase in scope that we report might explain why other studies have documented increasing concentration over time. We compute adjusted HHIs that account for the fact that increased scope can increase competition as more firms operate in more overlapping markets. We find that these adjusted HHIs essentially change little since 1997. These results suggest that an extended narrative that considers the growth in scope can help to contextualize prior reports of increasing HHIs and high levels of M&A. In particular, much M&A has been targeted at value creation through increased scope. Yet these results must be interpreted with care, as increased scope can lead to antitrust concerns in the form of product bundling, increased power in supply chains, or decreased entry into markets that larger firms have moved into. Future research examining scope and market power in detailed product areas could be impactful, given the importance of these issues to regulators and society.

References

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, 2020, The fall of the labor share and the rise of superstar firms, *The Quarterly Journal of Economics* 135, 645–709.

Bena, Jan, and Kai Li, 2014, Corporate innovations and mergers and acquisitions, *Journal of Finance* 69, 1923–1960.

Berger, Phillip, and Eli Ofek, 1995, Diversification's effect on firm value, *Journal of Financial Economics* 37, 39–65.

Boehm, Johannes, Swati Dhingra, and John Morrow, 2022, The comparative advantage of firms, *Journal of Political Economy* 130, 311–352.

Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson, 2020, Product innovation, product diversification, and firm growth: Evidence from japan's early industrialization, *American Economic Review* 111, 3795–3826.

Bramoulle, Yann, Habiba Djebbari, and Bernard Fortin, 2009, Identification of peer effects through social networks, *Journal of Econometrics* 113, 41–55.

Chandler, Alfred D, and Takashi Hikino, 1994, *Scale and Scope* (Harvard University Press).

Cohen-Cole, Ethan, Andrei Kirilenko, and Eleonora Patacchini, 2014, Trading networks and liquidity provision, *Journal of Financial Economics* 113, 235–251.

Crémer, Jacques, Luis Garicano, and Andrea Prat, 2007, Language and the theory of the firm, *Quarterly Journal of Economics* 122, 373–407.

Custodio, Claudia, 2012, Mergers and acquisitions accounting and the diversification discount, *Journal of Finance* 69, 219–240.

Doidge, Craig, Andrew Karolyi, and Rene Stulz, 2017, The u.s. listing gap, *Journal of Financial Economics* pp. 464–487.

Ewens, Michael, and Joan Farre-Mensa, 2020, The deregulation of the private equity markets and the decline in ipos, *The Review of Financial Studies* 33, 5463–5509.

Fresard, Laurent, Gerard Hoberg, and Gordon Phillips, 2020, Innovation activities and integration through vertical acquisitions, *Review of Financial Studies* 33, 2937–2976.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely, 2019, Are us industries becoming more concentrated?, *Review of Finance* p. 697–743.

Hansen, Lars Peter, 1988, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

Hart, Oliver, and John Moore, 1990, Property rights and the nature of the firm, *Journal of Political Economy* 98, 1119–1158.

Henderson, Rebecca, and Iain Cockburn, 1996, Scale, scope, and spillovers: the determinants of research productivity in drug discovery, *Rand Journal of Economies* 27, 32–59.

Hoberg, Gerard, and Gordon Phillips, 2010a, Product market synergies in mergers and acquisitions: A text based analysis, *Review of Financial Studies* 23, 3773–3811.

——— , 2010b, Real and financial industry booms and busts, *Journal of Finance* 65, 45–86.

——— , 2018, Industry choice and product language, 64, 3735–3755.

——— , and Nagpurnanand Prabhala, 2014, Product market threats, payout, and financial flexibility, *Journal of Finance* 69, 293–314.

Hund, John, Donald Monk, and Sheri Tice, 2020, A manufactured diversification discount, *Critical Finance Review forthcoming*.

Kamepalli, Sai Krishna, Raghuram Rajan, and Luigi Zingales, 2020, Kill zone, .

Kim, Hyunseob, and Howard Kung, 2017, The asset redeployability channel: How uncertainty affects corporate investment, *Journal of Financial Economics* 30, 245–280.

Kleibergen, Frank, and Richard Paap, 2006, Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics* 133, 97–126.

Kwon, Spencer Yongwook, Yueran Ma, and Kaspar Zimmermann, 2021, 100 years of rising corporate concentration, *Available at SSRN 3936799*.

Lang, Larry, and Rene Stulz, 1994, Tobin's q, corporate diversification, and firm performance, *Journal of Political Economy* 102, 1248–1280.

Li, Feng, Russell Lundholm, and Michael Minnis, 2013, A measure of competition based on 10-k filings, *Journal of Accounting Research* 51, 399–436.

Maksimovic, Vojislav, and Gordon Phillips, 2002, Do conglomerate firms allocate resources inefficiently across industries? theory and evidence, *Journal of Finance* 57, 721–767.

Matvos, Gregor, and Amit Seru, 2014, Resource allocation within firms and financial market dislocation: Evidence from diversified conglomerates, *The Review of Financial Studies* 27, 1143–1189.

——— , and Rui Silva, 2018, Financial market frictions and diversification, *Journal of Financial Economics* 127, 21–50.

Mikolov, Tomas, Chen Kai Corrado Greg, and Jeffrey Dean, 2013, Efficient estimation of word representations in vector, *arXiv:1301.3781, http://arxiv.org/abs/1301.3781*.

Moon, Katie, and Gordon Phillips, 2021, Outsourcing through purchase contracts and firm capital structure, *Management Science* 67, 363–387.

Panzar, J., and R. Willig, 1977, Economies of scale in multi-output production, *Quarterly Journal of Economics* 91, 481–93.

——— , 1981, Economies of scope, *American Economic Review* 71, 268–272.

Philippon, Thomas, and German Gutierrez, 2017, Declining competition and investment in the u.s., *working paper*.

Phillips, Gordon M., and Alexei Zhdanov, 2013, R&d and the incentives from merger and acquisition activity, *Review of Financial Studies* 34-78, 189–238.

Rhodes-Kropf, Matthew, and David Robinson, 2008, The market for mergers and the boundaries of the firm, *Journal of Finance* 63, 1169–1211.

Sargan, John, 1958, The estimation of economic relationships using instrumental variables, *Econometrica* 26, 393–415.

Song, Shiwon, 2020, The informational value of segment data disaggregated by underlying industry: Evidence from the textual features of business descriptions, *working paper*.

Teece, David J., 1980, Economies of scope and the scope of the enterprise, *Journal of Economic Behavior and Organization* 1, 223–247.

Villalonga, Belen, 2004, Does diversification cause the diversification discount, *Financial Management* 33, 5–27.

Williamson, Oliver E, 1975, Markets and hierarchies, *New York* 2630.

Appendix A. Variable definitions

Table A1: Variable definitions

Table A1

Variable	Definition	Source
D2V-Scope	The number of D2V-300 segments that each firm's 10-K product description is similar to. The classification from firm to segments is based on a 2% granularity, and firm-segments similarities are deemed to be pairs for the 2% highest textual similarities between each firm and the text describing the 300 D2V industries.	
NAICS-Scope	This is computed in a similar way to the D2V-scope variable. The NAICS scope is based on the text describing NAICS industries (using the highly detailed 963 page 2017 NAICS manual) instead of TNIC D2V industries. The classification from firm to segments is based on a 2% granularity, and firm-segment similarities are deemed to be pairs for the 2% highest textual similarities between each firm and the text describing the 311 4-digit NAICS industries.	
Product Breadth	Using metaHeuristica queries, we count the number of paragraphs that mention "product lines" or "product categories". These phrases indicate high levels of product breadth. Product Breadth is this number of paragraphs scaled by the total number of paragraphs in the 10-K.	
Prod/Svc Breadth	This variable is computed analogous to Product Breadth, but we also count paragraphs that mention either "service lines" or "service categories".	
Product Breadth Detail	Same as Product Breadth, as described above, except that we only count paragraphs that additionally mention a specific clarifying term in the following list: {breadth, broad, broader, wide, multiple, numerous, diverse, categories, divisions}	
Prod/Svc Breadth Detail	Same as Product/Svc Breadth, as described above, except that we only count paragraphs that additionally mention a specific clarifying term in the following list: {breadth, broad, broader, wide, multiple, numerous, diverse, categories, divisions}	
Sectoral Redeployment Potential	This is an instrument indicating a shift in the incentives for firms to increase scope. The data draws from information in the Bureau of Economic Analysis, Compustat, and the research paper Kim and Kung (2017). This variable is the average cosine similarity between the asset utilization vector of the focal firm's NAICS industry and that of the NAICS industry of the focal firm's TNIC-2 peers that are not also TNIC-3 peers. This latter step ensures that the measure is based on product market peers that are close but a bit more distant in product space than are near peers, which further increases the extent of exogenous content in the measure. When this value is high, it indicates that expansion in scope for the focal firm is likely to have low cost regarding expansion into neighboring product markets in the given year.	
Sectoral Opportunity Set Potential	This variable is similar to the above except that it is based on product offerings rather than the inputs to production (asset vectors). This variable is computed as the HHI, or concentration ratio, of companies that are in the focal firm's TNIC-2 industry but not in the most proximate TNIC-3 industry. The HHI calculation is based on the NAICS codes of the firms in these near but slightly more distant product markets. When this value is high, it indicates high growth opportunities to scope expansion for the focal firm in the given year.	
Logassets	Natural logarithm of total assets of the firm Compustat	
Log Age	Natural logarithm of one plus the current year of observation minus the first year the firm appears in the Compustat database Compustat	
Valuation Ratio	This ratio is computed as the market value of the firm (book assets minus book equity plus market equity), all divided by book assets. Market equity is Compustat shares outstanding times the share price at the end of the fiscal year PRCC. Book equity is shareholders equity (Compustat SEQ), plus TXDITC minus preferred stock (PSTKRV, and if missing, then PSTKL, and if missing then UPSTK). Shareholders equity iS SEQ, but if missing, is Compustat CEQ plus UPSTK, and if missing, is assets less long term assets.	
10K Size	Natural logarithm of one plus the total number of paragraphs in the focal firm's 10-K report.	
TNIC HHI	The concentration ratio based on TNIC industries as computed in Hoberg and Phillips (2016).	
NAICS HHI	The concentration ratio based on NAICS industries as computed in Grullon, Michaely, and Larkin (2019).	
Acquirer Dummy	A dummy equal to one if the given firm had an acquisition become effective in the current year according to the SDC Platinum database.	

Variable	Definition
Target Dummy	A dummy equal to one if the given firm had a sale of assets or a merger become effective in the current year according to the SDC Platinum database.
R&D/Assets	Compustat XRD divided by total assets AT, winsorized at the 1/99% level. This variable is set to zero if XRD is missing.
CAPX/Assets	Compustat CAPX divided by total assets AT, winsorized at the 1/99% level. This variable is set to zero if it is missing.
Sales Growth	Natural logarithm of total sales in the current year t divided by total sales in the previous year $t-1$.
Asset Growth	Natural logarithm of total assets in the current year t divided by total assets in the previous year $t-1$.
OI/Assets	Compustat OIBDP divided by total assets AT, winsorized at the 1/99% level.
Equity Issuance/Assets	Computed as Compustat (SSTK - PRSTKC) divided by total assets AT, winsorized at the 1/99% level.
Debt Issuance/Assets	Computed as Compustat DLTIS divided by total assets AT, winsorized at the 1/99% level.
Equity Repurchases/Assets	Computed as Compustat PRSTKC divided by total assets AT, winsorized at the 1/99% level.
Dividends/Assets	Computed as Compustat DVC divided by total assets AT, winsorized at the 1/99% level.

Appendix B. Details for Computing D2V-Scope

Our methodology for computing D2V-Scope and generating the D2V segment database consists of the following 5 steps:

Step 1 We use k-means clustering and word2vec to identify 450 "candidate industries" and their word2vec dialects. Doc2vec represents each firm in our base year as a 300-element vector representing its product offerings. We run k-means clustering on the subset of single segment firms in 1997 to extract 450 clusters, which are "candidate industries". Each candidate industry is represented by a 300-element "centroid vector" in our 300 dimensional doc2vec space. To develop dialects for each, we use doc2vec's word2vec tool to identify a 300-element vector for each word that appears in our 10-K business description database (after applying stop-word filters as in HP2016). As the doc2vec centroid vectors and the word2vec word-specific vectors are in the same 300-dimensional doc2vec space, we compute cosine similarities for every permutation of centroids and the product words. For each industry/centroid, we first identify the 2500 words that are spatially most similar to the industry's centroid based on the cosine similarity between the centroid and each word. The result is 2500 x 450 centroid = 1,125,000 word-industry similarity scores. Because some industries have thick vocabularies and others have thin vocabularies, we let the data determine industry vocabulary boundaries using a fixed similarity threshold analogous to the HP2016 approach for selecting peers. In particular, we sort all (2500 x 450) scores using a pooled sort in descending order and deem the top 20% to be the industry dialects. If a given industry has fewer than 50 words in its dialect, we "thicken" its dialect to a minimum of 50 words by simply taking the 50 words with the highest similarity to the given centroid. The result of Step 1 is 450 candidate industries, each with an informative dialect.

Step 2: We then reduce the number of candidate industries as we found that many are redundant to the other industries. For each word in our base year, we count many candidate industries it appears in. We use the fact that each single segment firm is a member of the candidate industry that its own doc2vec vector is most similar to. As many words in 10-Ks are boilerplate or non-industry specific, we initially tag words as "sector specific" if they

are used by firms in no more than 15 candidate industries. We then sort industries by how many sector-specific words each uses and drop the 75 industries using the fewest sector-specific words (dropped industries have six or fewer such words). These dropped industries have too little specificity and their centroids likely reflect non-industry-relevant content. As a second step, we examine the ten most similar dialect vocabulary terms for each of the remaining 375 candidate industries and we drop 20 candidate industries that clearly use boilerplate vocabularies. Finally, we drop 55 more industries that have centroids that are too close to another centroid, making them redundant (these 55 most redundant industries have excessively high pairwise cosine similarities between another centroid ranging from 73.6% to 92.9%). The 300 remaining industries comprise our final set of "D2V industries" used the remainder of our study. They are both purged of boilerplate content and each is relatively unique.

Step 3: We then construct weights for these industry dialects. Economically, a word in a dialect is more important to its industry if it is (A) more spatially close to the industry's centroid and (B) specific to the given industry and not other industries. We thus define our term-specific weights as the following product for each word n in industry k: $w_{k,n}$=(word n's cosine similarity to centroid k as defined above) x (word-specific HHI across industries). To compute word-specific HHI, we first compute how many times each word appears in each industry among single segment firms and divide by the total times the word appears in all industries. The resulting shares sum to one, and hence summing their squares generates a word-specific HHI for each word.

Step 4 We compute each company's exposure to each of the 300 industries. For each firm i in any year t, we define $B_{i,k,n,t}$ as a dummy equal to one if the given firm uses word n of industry k's dialect vocabulary (defined above) in its 10-K business description in year t. A firm i's exposure to an industry k in year t is then denoted as $E_{i,k,t}$, which we computed as the following weighted exposure (weights are from Step 3).

$$E_{i,k,t} = \frac{\sum\limits_{n=1...N} B_{i,k,n,t} \cdot w_{k,n}}{\sum\limits_{n=1...N} w_{k,n}} \tag{6}$$

Step 5: We apply a standard 2% granularity to determine operating segments and D2V-Scope. We tag a firm as likely operating in an industry k in a given year using a fixed exposure threshold determined in the base year. We thus sort all 1997 exposures $E_{i,k,t}$ (there are 300 x N_{firms} such exposures) from highest to lowest and identify the threshold value for the top 2% of these exposures as the "operating threshold" ($\bar{E}$). We then hold this threshold fixed over all years, allowing us to study scope over time. We tag a firm i in year t as likely operating in industry k if its exposure is greater than or equal to the threshold: $E_{i,k,t} \geq \bar{E}$. The set of firm-industry-year observations exceeding the threshold is our D2V segment database, and the count of how many segments a given firm is attached to is our primary measure of firm scope in each year (henceforth "D2V-Scope").

Table 1: Summary Statistics

Summary statistics are reported for our sample of 101,535 observations based on annual firm observations from 1988 to 2017. Our main variables of interest, D2V-scope and NAICS-scope, are based on scoring each firm's Item 1 business description based on how similar it is to the product text of specific fixed industries. For D2V-scope, fixed industries are based on 300 clusters formed using K-means in our 300-dimensional doc2vec space based on Item 1s in firm 10-Ks, and for NAICS-scope, it is based on 4-digit NAICS industries. All variables are described in detail in the variable list in Appendix A and in Section 2 of the paper.

Variable	Mean	Std. Dev.	Minimum	Median	Maximum	# Obs
			Panel A: Scope and Segment Variables			
D2V-Scope	7.546	5.800	1.000	6.000	32.000	101,535
NAICS-Scope	6.288	7.602	0.000	4.000	47.000	100,522
# Compsutat Segments	1.450	0.860	1.000	1.000	11.000	101,535
			Panel B: Accounting Variables			
R&D/Assets	0.053	0.108	0.000	0.000	0.856	101,535
CAPX/Assets	0.057	0.064	0.000	0.037	0.441	101,535
Acquisition Dummy	0.285	0.451	0.000	0.000	1.000	101,535
Target Dummy	0.126	0.332	0.000	0.000	1.000	101,535
Valuation (M/B)	1.713	1.575	0.166	1.202	15.876	100,895
Sales Growth	0.109	0.440	-6.177	0.076	9.383	101,107
Asset Growth	0.076	0.358	-4.294	0.049	5.529	101,499
Equity Issuance	0.048	0.131	0.000	0.004	1.035	101,535
Debt Issuance	0.104	0.210	0.000	0.002	1.590	101,535
Dividends/Assets	0.008	0.020	0.000	0.000	0.259	101,428
Equity Repurchase	0.018	0.062	-0.008	0.000	3.469	93,998
Log Assets	5.429	2.130	0.694	5.312	13.590	101,535
Log Age	2.621	0.765	0.693	2.565	4.220	101,535

Table 2: Pearson Correlation Coefficients

Pearson Correlation Coefficients are reported for our sample of 101,535 observations based on annual firm observations from 1988 to 2017. Our main variables of interest, D2V-scope and NAICS-scope, are based on scoring each firm's Item 1 business description based on how similar it is to the product text of specific fixed industries. For D2V-scope, fixed industries are based on the D2V-300 industries, and for NAICS-scope, it is based on 4-digit NAICS industries. All variables are described in detail in the variable list in Appendix A and in Section 2 of the paper.

Row	Variable	D2V-Scope	NAICS-Scope	# CS Segments	Log Assets	Log Age	R&D/ Assets	CAPX/ Assets	Acquisition Dummy	Target Dummy	Sales Growth
	NAICS-Scope	0.640									
	# Compsutat Segments	0.120	0.199								
	Log Assets	0.291	0.272	0.312							
	Log Age	-0.034	-0.045	0.308	0.398						
	R&D/Assets	0.098	-0.048	-0.170	-0.263	-0.159					
	CAPX/assets	0.002	0.046	-0.042	0.033	-0.096	-0.111				
	Acquisition Dummy	0.095	0.087	0.133	0.302	0.075	-0.093	-0.016			
	Target Dummy	0.053	0.057	0.167	0.248	0.167	-0.062	-0.006	0.156		
	Sales Growth	0.040	0.055	-0.031	-0.010	-0.165	0.031	0.099	0.104	-0.065	
	TNIC HHI	-0.374	-0.298	0.091	-0.182	0.171	-0.163	-0.127	-0.042	-0.011	-0.071

Table 3: CVS Scope

The table displays the D2V scope segments of CVS over time.

Year	Amount	Word List
		Panel A: CVS Scope Allocations in 1990 (CVS was owned by variety store Melville)
1990	0.035	merchandise,store,assortments,stores,merchandising,markdowns,layaway,assortment,markdown,everyday,
1990	0.032	petco,superstore,superstores,petsmart,stores,store,assortment,merchandise,merchandising,retailer,
1990	0.030	roberds,mactavish,store,merchandise,rental,apro,stores,furniture,heilig,alrenco,
1990	0.029	filene,macy,neiman,federated,bloomingdales,lazarus,burdines,dillard,merchandise,fashion,
		Panel B: CVS Scope Allocations in 1999 (CVS was independent)
1999	0.041	prescription,pharmacists,pharmacies,pharmacy,pharmacist,prescriptions,generic,formulary,pharmerica,drugs,
1999	0.029	stationers,stationer,superstores,officemax,envelopes,superstore,skus,lagasse,catalog,stationery,
		Panel C: CVS Scope Allocations in 2005
2005	0.038	supermarkets,supermarket,groceries,grocery,nonfood,delicatessens,perishable,delicatessen,perishables,store,
2005	0.034	prescription,pharmacists,pharmacies,pharmacy,pharmacist,prescriptions,generic,formulary,pharmerica,drugs,
		Panel D: CVS Scope Allocations in 2007 (CVS acquires Caremark)
2007	0.223	hmos,enrollees,ppos,capitated,capitation,enrollee,tpas,payor,copayment,wellpoint,
2007	0.184	behavioral,mental,psychiatric,inpatient,psychologists,psychotherapy,psychiatrist,hoskin,psychiatrists,champus,
2007	0.183	payor,envoy,hboc,billing,submitters,quadramed,healthcare,payors,electronically,clearinghouses,
2007	0.144	physician,physicians,referring,medicare,referral,medicaid,payor,stark,inpatient,outpatient,
2007	0.098	prescription,pharmacists,pharmacies,pharmacy,pharmacist,prescriptions,generic,formulary,pharmerica,drugs,
2007	0.076	worksite,employer,peos,napeo,fica,vincam,futa,employers,payroll,workers,
2007	0.063	policyholder,annuity,persistency,annuities,policyholders,policyowners,fpdas,spdas,surrenders,surrender,
2007	0.060	practitioners,audiology,audiologists,optometrists,ophthalmologis,optometric,audiologist,chiropractic,optometry,audiological,
2007	0.054	tokheim,dispensers,dispensing,dispenser,dispense,sprayers,dispensed,pump,closures,toiletries,
2007	0.048	insurer,policyholders,lawyers,casualty,reinsure,reinsurer,domiciliary,reinsured,domiciled,insurers,
2007	0.039	staffing,clients,workers,professionals,employer,assignments,recruits,employers,paralegals,insureds,
2007	0.035	nutritional,dietary,herbs,supplements,vitamins,dshea,herbal,usana,rexall,phytonutrients,
2007	0.034	reinsurance,casualty,ceded,reinsurers,insurer,writings,policyholders,insureds,reinsurer,insurers,
2007	0.034	syndicates,lloyd,underwrite,reinsurance,reinsurers,reinsureds,brokerage,intermediaries,reinsure,ceding,
2007	0.031	allegiance,stockless,stocking,invoicing,mrop,maxxim,fulfillment,replenishment,skus,baxter,
2007	0.031	lynch,merrill,barney,ilco,salomon,grenfell,witter,morgan,lehman,executions,
2007	0.030	supermarkets,supermarket,groceries,grocery,nonfood,delicatessens,perishable,delicatessen,perishables,store,
		Panel E: CVS Scope Allocations in 2017
2017	0.197	hmos,enrollees,ppos,capitated,capitation,enrollee,tpas,payor,copayment,wellpoint,
2017	0.182	payor,envoy,hboc,billing,submitters,quadramed,healthcare,payors,electronically,clearinghouses,
2017	0.154	prescription,pharmacists,pharmacies,pharmacy,pharmacist,prescriptions,generic,formulary,pharmerica,drugs,
2017	0.127	physician,physicians,referring,medicare,referral,medicaid,payor,stark,inpatient,outpatient,
2017	0.115	behavioral,mental,psychiatric,inpatient,psychologists,psychotherapy,psychiatrist,hoskin,psychiatrists,champus,
2017	0.062	allegiance,stockless,stocking,invoicing,mrop,maxxim,fulfillment,replenishment,skus,baxter,
2017	0.054	worksite,employer,peos,napeo,fica,vincam,futa,employers,payroll,workers,
2017	0.052	practitioners,audiology,audiologists,optometrists,ophthalmologis,optometric,audiologist,chiropractic,optometry,audiological,
2017	0.045	tokheim,dispensers,dispensing,dispenser,dispense,sprayers,dispensed,pump,closures,toiletries,
2017	0.042	living,assisted,nursing,residents,frail,adls,elderly,seniors,resident,congregate,
2017	0.037	supermarkets,supermarket,groceries,grocery,nonfood,delicatessens,perishable,delicatessen,perishables,store,
2017	0.035	checkout,check,scannable,keypad,coded,instructions,keyboard,electronically,readable,registers,
2017	0.034	databases,database,append,metromail,speh,lists,mailing,smartbase,acxiom,list,
2017	0.033	sears,jcpenney,roebuck,srac,mervyn,penney,mervyns,moody,kmart,homelife,
2017	0.032	staffing,clients,workers,professionals,employer,assignments,recruits,employers,paralegals,insureds,
2017	0.032	nutritional,dietary,herbs,supplements,vitamins,dshea,herbal,usana,rexall,phytonutrients,
2017	0.028	client,clients,mainframe,payroll,dataworks,peoplesoft,outsourcing,billing,updates,invoices,

Table 4: Tesla Scope

The table displays the D2V scope allocations of Tesla over before and after its Solar City acquisition.

Year	Amount	Word List
		Panel A: Tesla Scope Allocations in 2015 (Prior to Solar City Acquisition)
2015	0.116	powertrain,engines,engine,servo,motors,brakes,aftermarket,cummins,clutches,brushless,
2015	0.111	chrysler,automotive,airbag,automakers,windshield,defroster,motors,airbags,inflators,ford,
2015	0.096	axle,wheels,wheel,automotive,axles,brake,aftermarket,brakes,chrysler,clutch,
2015	0.083	trim,automakers,lear,automotive,headliners,visors,volkswagen,chrysler,fiat,mazda,
2015	0.071	solar,photovoltaic,photovoltaics,electrificatio,sunlight,inverters,renewables,lanterns,nimh,battery,
2015	0.066	chassis,buses,motorhomes,freightliner,motorhome,coaches,rvia,gillig,coach,vans,
2015	0.057	brake,aftermarket,brakes,clutches,remanufactured,clutch,alternators,automotive,braking,friction,
2015	0.047	batteries,battery,rechargeable,lithium,saft,nicd,varta,yuasa,eveready,hydride,
2015	0.040	teleservices,telemarketing,outbound,inbound,telemarketers,teleservicing,tcpa,sitel,dialers,teleservice,
2015	0.039	avis,thrifty,rental,hertz,rentals,alamo,renters,reservations,reservation,travel,
2015	0.038	carmax,dealerships,dealership,valumax,automobile,vehicle,haggle,dealer,here,autonation,
2015	0.028	actuators,servo,instron,sensors,actuator,sensing,vibration,precision,aerospace,machined,
		Panel B: Tesla Scope Allocations in 2016 (After Solar City Acquisition)
2016	0.277	solar,photovoltaic,photovoltaics,electrificatio,sunlight,inverters,renewables,lanterns,nimh,battery,
2016	0.115	powertrain,engines,engine,servo,motors,brakes,aftermarket,cummins,clutches,brushless,
2016	0.087	chrysler,automotive,airbag,automakers,windshield,defroster,motors,airbags,inflators,ford,
2016	0.066	axle,wheels,wheel,automotive,axles,brake,aftermarket,brakes,chrysler,clutch,
2016	0.063	batteries,battery,rechargeable,lithium,saft,nicd,varta,yuasa,eveready,hydride,
2016	0.057	trim,automakers,lear,automotive,headliners,visors,volkswagen,chrysler,fiat,mazda,
2016	0.052	brake,aftermarket,brakes,clutches,remanufactured,clutch,alternators,automotive,braking,friction,
2016	0.048	chassis,buses,motorhomes,freightliner,motorhome,coaches,rvia,gillig,coach,vans,
2016	0.043	cogeneration,megawatt,steam,electricity,energy,megawatts,fired,purpa,geothermal,cogenerators,
2016	0.040	carmax,dealerships,dealership,valumax,automobile,vehicle,haggle,dealer,here,autonation,
2016	0.035	avis,thrifty,rental,hertz,rentals,alamo,renters,reservations,reservation,travel,
2016	0.030	itron,cellnet,meter,airlink,metretek,meters,metering,rtus,scada,remote,
2016	0.029	voltage,capacitors,voltages,inductors,capacitor,mosfet,resistors,transients,uninterruptibl,transformers,
2016	0.029	siebel,functionality,microsoft,netscape,groupware,implementation,server,symantec,desktop,wirelessly,
2016	0.028	fpsc,imprudently,tampa,florida,electric,ncuc,utilities,kilowatthour,hillsborough,manatee,

Table 5: Scope Statistics vs Segment Counts and Firm Size

The table reports scale and scope statistics separately for firms based on how many operating segments the firm reports in the Compustat database (Panel A) or sorted into size quintiles (Panels B and C). Sorts are annual and are based on Compustat assets (variable AT). Our main variables of interest, D2V-scope and NAICS-scope, are based on scoring each firm's Item 1 business description based on how similar it is to the product text of specific fixed industries. For D2V-scope, fixed industries are based on the D2V-300 industries, and for NAICS-scope, it is based on 4-digit NAICS industries. Assets are from Compustat (variable AT). For size sorts, we report statistics for the full sample (Panel B) and separately for single segment firms only (Panel C).

Panel A: Scope vs Compustat Segments

# CS Segments	D2V-Scope	NAICS-Scope	Assets	# Obs.
1 segment	7.25	5.57	1358	72,374
2 segments	7.80	7.04	3239	18,107
3 segments	8.47	8.61	6175	7,479
4 segments	9.36	10.28	10053	2,361
5+ segments	12.39	15.42	31005	1,214

Panel B: Scope vs Firm Size Quintiles (All Firms)

Row	Size Quintile	# CS Segments	D2V-Scope	NAICS-Scope	Assets	# Obs.
	Small Firms	1.22	5.87	4.01	23	20,296
	Quintile 2	1.26	6.89	5.02	100	20,315
	Quintile 3	1.35	7.61	5.94	301	20,311
	Quintile 4	1.49	8.10	7.18	927	20,315
	Big Firms	1.93	9.26	9.23	11679	20,298

Panel C: Scope vs Firm Size Quintiles (Single Segment Firms Only)

Row	Size Quintile	# Segments	D2V-Scope	NAICS-Scope	Assets	# Obs.
	Small Firms	1	5.81	3.87	18	14,464
	Quintile 2	1	6.73	4.70	73	14,481
	Quintile 3	1	7.46	5.39	200	14,481
	Quintile 4	1	7.81	6.18	598	14,481
	Big Firms	1	8.43	7.66	5906	14,467

Table 6: Spatial Relationship between Compustat Segments and D2V Segments

The table reports annual OLS regressions where the number of Compustat segments is regressed on the number of spatially near versus spatially distant D2V segment pairs. The goal is to assess whether, as hypothesized, Compustat segments are spatially more distant than are D2V segments. We thus take all permutations of the 300 D2V industries and compute pairwise industry similarities for each industry based on the pairwise similarities of the k-means centroids from which each industry was estimated. We then sort all industry pairs into quintiles based on how spatially distant they are from each other (product market distance). We label those industry-pairs in the most similar quintile as "most related" segments and those in the second most similar quintile as "weakly related segments". We group the least similar three quintiles into a single group of "likely unrelated" industry pairs as these last three quintiles uniformly have very low similarity and also have far fewer operating pairs as noted in Table 6. We then tabulate how many pairs are operating in each similarity bin for each firm, where counts are weighted such that each segment gets a total weight of unity. Thus we can regress the number of Compustat segments on all three D2V segment counts for each of the three groups where one observation is one firm in one year. Regressions are annual and average coefficients and t-statistics are reported at the bottom of the table. As regressions are run separately in each year, they are by nature clustered by firm, and time effects are absorbed by the annual intercepts.

Row	Year	Most Similar Segments	Weakly Similar Segments	Likely Unrelated Segments	Obs.
(1)	1990	-0.011 (-1.97)	0.220 (7.16)	0.090 (4.28)	3,246
(2)	1991	-0.012 (-2.17)	0.245 (7.88)	0.009 (0.44)	3,285
(3)	1992	-0.022 (-4.12)	0.231 (7.65)	0.040 (2.03)	3,258
(4)	1993	-0.014 (-2.68)	0.204 (7.07)	0.038 (2.05)	3,439
(5)	1994	-0.016 (-3.33)	0.150 (6.12)	0.030 (1.80)	3,753
(6)	1995	-0.013 (-3.00)	0.102 (4.33)	0.077 (4.79)	4,063
(7)	1996	-0.016 (-3.96)	0.106 (4.80)	0.101 (6.41)	4,508
(8)	1997	-0.013 (-3.65)	0.105 (5.42)	0.073 (5.76)	5,075
(9)	1998	-0.013 (-3.18)	0.149 (6.40)	0.071 (4.10)	5,054
(10)	1999	-0.023 (-5.25)	0.177 (7.12)	0.060 (3.20)	4,811
(11)	2000	-0.022 (-4.98)	0.162 (6.56)	0.046 (2.63)	4,542
(12)	2001	-0.022 (-5.40)	0.194 (7.92)	0.021 (1.30)	4,389
(13)	2002	-0.020 (-4.74)	0.182 (7.23)	0.056 (3.16)	4,032
(14)	2003	-0.020 (-4.61)	0.169 (6.74)	0.043 (2.36)	3,715
(15)	2004	-0.019 (-4.49)	0.162 (6.54)	0.030 (1.61)	3,517
(16)	2005	-0.020 (-4.70)	0.192 (7.58)	0.029 (1.51)	3,436
(17)	2006	-0.017 (-3.96)	0.182 (7.30)	0.032 (1.68)	3,340
(18)	2007	-0.020 (-4.54)	0.176 (7.19)	0.037 (1.98)	3,229
(19)	2008	-0.016 (-3.78)	0.142 (5.69)	0.047 (2.56)	3,171
(20)	2009	-0.017 (-3.82)	0.141 (5.60)	0.039 (2.04)	3,040
(21)	2010	-0.020 (-4.48)	0.155 (5.85)	0.025 (1.29)	2,903
(22)	2011	-0.020 (-4.27)	0.143 (5.31)	0.039 (1.94)	2,793
(23)	2012	-0.020 (-4.29)	0.155 (5.74)	0.042 (2.08)	2,711
(24)	2013	-0.024 (-5.16)	0.137 (5.20)	0.057 (2.79)	2,652
(25)	2014	-0.029 (-6.36)	0.162 (6.32)	0.031 (1.57)	2,677
(26)	2015	-0.029 (-6.22)	0.181 (6.95)	0.012 (0.60)	2,711
(27)	2016	-0.027 (-5.67)	0.147 (5.43)	0.031 (1.48)	2,638
(28)	2017	-0.027 (-6.11)	0.162 (6.33)	0.011 (0.56)	2,556
	Average	-0.019 (-4.32)	0.166 (6.41)	0.043 (2.43)	3,519

Table 7: Scope Statistics vs Industry-Pair-Relatedness

The table reports the distribution of the industries spanned by single firms (scope) across all industry pairs, sorted by how similar are the industries in the given pair regarding horizontal relatedness using TNIC similarities (Panel A) or vertical relatedness using vertical TNIC (VTNIC) relatedness (Panel B). We explain the methodology for panel A based on horizontal relatedness but note that the methodology for Panel B is exactly parallel but uses pairwise vertical relatedness scores from Fresard, Hoberg, and Phillips (2020) instead of horizontal relatedness scores from Hoberg and Phillips (2016). For each pair of doc2vec-based 300 industries in each year from our k-means clusters, we first tabulate the number of firms that operate in both industries in the pair based on the D2V-scope variable's construction. A firm is thus designated as operating in both industries if the given firm's business description is highly similar to the text of both industries in the pair. The result is a panel database of industry-pair-years indicating the number of firms operating in each pair. We then sort industry pairs into deciles based on the average horizontal TNIC similarity score of all firms in the first industry relative to those in the second. Industries that score highly are spatially close in the horizontal sense in the TNIC space (Panel B is similar but is based on vertical relatedness). Finally, we sum the firm-operating-pairs in each decile and report the fraction of operating pairs in each decile. We report this fraction for all firms, only for single segment firms and only for multi-segment firms. Finally, we report the average TNIC distance of the industry pairs in each decile and the number of industry pairs in each group in the final columns.

Industry-Pair Similarity Decile	Fraction Scope Pairs (All Firms)	Fraction Scope Pairs (Single-Seg)	Fraction Scope Pairs (Multi-seg)	Average TNIC-pair Similarity	Ind-Pair x Year # Obs.
			Panel A: Horizontal Relatedness		
Least Similar	0.017	0.015	0.022	0.001	261,361
Decile 2	0.019	0.017	0.026	0.003	261,369
Decile 3	0.023	0.021	0.031	0.005	261,358
Decile 4	0.028	0.026	0.037	0.006	261,377
Decile 5	0.037	0.034	0.046	0.008	261,361
Decile 6	0.045	0.042	0.058	0.010	261,379
Decile 7	0.059	0.055	0.076	0.013	261,374
Decile 8	0.088	0.083	0.109	0.017	261,372
Decile 9	0.156	0.150	0.185	0.025	261,371
Most Similar	0.528	0.556	0.409	0.068	261,355
			Panel B: Vertical Relatedness		
Least Similar	0.311	0.356	0.115	0.002	260,972
Decile 2	0.108	0.117	0.068	0.003	260,989
Decile 3	0.072	0.074	0.062	0.004	260,993
Decile 4	0.061	0.060	0.062	0.004	260,983
Decile 5	0.055	0.053	0.064	0.005	260,993
Decile 6	0.054	0.050	0.069	0.006	260,990
Decile 7	0.058	0.053	0.079	0.007	260,990
Decile 8	0.066	0.060	0.094	0.008	260,986
Decile 9	0.080	0.071	0.119	0.010	260,996
Most Similar	0.135	0.105	0.268	0.015	260,980

Table 8: Risk Profile Regressions

The table reports the results of OLS regressions where measures of risk are regressed on measures of scope decomposed into components that are spatially near versus far from the focal firm. The first dependent variable is stock market risk used in Panels A and C, which is computed as the standard deviation of daily stock returns in year t. The second is cashflow volatility used in Panels B and D, which is computed as the standard deviation of a firm's quarterly operating income scaled by assets, computed over the 8 quarters of year t and $t+1$. We express both as percentages for ease of interpretation. All RHS variables are computed using data from year $t-1$. We report results for three tercile-based subsamples within each panel and as noted in the first column. The subsamples are constructed in each year by sorting firms based on the average product market distance of the D2V industry segments they are assigned to when constructing our D2V-scope variable (firms assigned to just one D2V segment have a distance of zero). Firms in the close-scope tercile subsample are operating across industries that are highly related, whereas firms in the far-scope tercile are operating across industries that are more distant in the product space. Our main variable of interest, D2V-scope, is based on scoring each firm's Item 1 business description based on how similar it is to the product text of specific fixed industries. The fixed industries are based on the D2V-300 industries. All regressions include year fixed effects, and controls for firm size, and firm age. The regressions in Panels C and D additionally include firm fixed effects. t-statistics for both panels are clustered by firm and shown in parentheses.

Row	Dependent Variable	D2V-Scope	Log Assets	Log Age	ARSQ	# Obs
			Panel A: Dependent Variable is Stock Volatility (OLS with Year Fixed Effects)			
(1)	Close-Scope Tercile	0.024	–0.567	–0.487	0.37	32,295
		(4.390)	(–40.810)	(–14.890)		
(2)	Medium-Scope Tercile	0.022	–0.540	–0.361	0.39	29,582
		(5.910)	(–46.420)	(–13.060)		
(3)	Far-Scope Tercile	0.022	–0.542	–0.323	0.38	30,007
		(7.060)	(–41.840)	(–11.360)		
			Panel B: Dependent Variable is Cashflow Volatility (OLS with Year Fixed Effects)			
(4)	Close-Scope Tercile	0.019	–0.438	–0.297	0.17	31,427
		(2.850)	(–27.740)	(–8.910)		
(5)	Medium-Scope Tercile	0.011	–0.526	–0.200	0.18	28,849
		(2.290)	(–29.790)	(–5.860)		
(6)	Far-Scope Tercile	0.023	–0.508	–0.138	0.19	29,251
		(6.190)	(–30.170)	(–4.030)		
			Panel C: Dependent Variable is Stock Volatility (OLS with Firm and Year Fixed Effects)			
(7)	Close-Scope Tercile	–0.008	–0.324	–0.342	0.66	32,295
		(–0.950)	(–9.230)	(–3.600)		
(8)	Medium-Scope Tercile	–0.008	–0.240	–0.331	0.68	29,582
		(–1.460)	(–7.400)	(–4.110)		
(9)	Far-Scope Tercile	0.006	–0.183	–0.366	0.67	30,007
		(1.570)	(–5.280)	(–4.300)		
			Panel D: Dependent Variable is Cashflow Volatility (OLS with Firm and Year Fixed Effects)			
(10)	Close-Scope Tercile	0.007	–0.351	–0.049	0.62	31,427
		(0.750)	(–8.210)	(–0.520)		
(11)	Medium-Scope Tercile	0.007	–0.346	–0.182	0.61	28,849
		(1.110)	(–7.430)	(–1.920)		
(12)	Far-Scope Tercile	0.006	–0.336	–0.265	0.65	29,251
		(1.460)	(–7.850)	(–3.250)		

Table 9: Predictive Power Validation: D2V Segments vs Compustat Segments

The table reports profitability regressions that compare the informativeness of D2V-segments to Compustat segments. We run regressions in which oi/assets (Panel A) and oi/sales (Panel B) is the dependent variable and we assess segment database informativeness using generalized fixed effects. For firm i in industry k in year t, a basic fixed effects model with year x industry fixed effects would take the form $OIassets_{i,t} = \mu_{k,t} + \epsilon_{i,t}$. Our generalized fixed effects model accounts for the fact that a firm can have partial operations in multiple industries, as defined by a set of weights $\omega_{i,k,t}$ that sum to one (i.e., a segment database is a mapping of each firm to the set of industries with weights in $[0, 1]$ that sum to unity). A generalized fixed effects model is analogous to a basic fixed effects model but would be fitted with ω-weights rather than binary weights as follows:

$$OIassets_{i,t} = \sum_{=1 \to 300} \omega_{i,k,t} \cdot \mu_{k,t} + \epsilon_{i,t}$$

This model can be fitted using basic OLS where the RHS variables include one vector for each industry, which for each firm, is populated by the weight of the given firm-year observation to the given industry. In each panel below, we first fit the generalized fixed effects model in-sample using the set of single segment firms as indicated by the Compustat database. For the D2V industries, the weight of each firm in each industry is proportional to the amount of text in its 10-K corresponding to the given industry. For Compustat, we weight each segment by its reported sales (allowing Compustat to incorporate its main potential advantage as it contains sales data for each segment). The first three columns of each Panel below report the results of the above model using single segment firms only for D2V segments, Compustat segments, and both together. Finally, the next three rows use the fitted values from the single segment firm regressions to fit predicted values of profitability for multiple segment firms. This is a pure out-of-sample test that illustrates the ability of each segment database to predict conglomerate profitability using only information in single segment firms. We also report the adjusted R-squared and Akaike Information Criterion for these tests.

Row	Segments Tested	Sample	Adj RSQ	Akaike Information Criterion	# Obs
			Panel A: oi/assets as dependent variable		
(1)	D2V Segments	Single-Seg Firms (In-Sample)	0.316	-343160	114902
(2)	CSTAT SIC Segments	Single-Seg Firms (In-Sample)	0.237	-330678	114902
(3)	Both Segments	Single-Seg Firms (In-Sample)	0.334	-346159	114902
(4)	D2V Segments	Mutli-Seg Firms (Out-of-Sample)	0.038	-151784	38801
(5)	CSTAT SIC Segments	Mutli-Seg Firms (Out-of-Sample)	0.029	-151424	38801
(6)	Both Segments	Mutli-Seg Firms (Out-of-Sample)	0.052	-152341	38801
			Panel B: oi/sales as dependent variable		
(7)	D2V Segments	Single-Seg Firms (In-Sample)	0.354	-57486	108954
(8)	CSTAT SIC Segments	Single-Seg Firms (In-Sample)	0.267	-43752	108954
(9)	Both Segments	Single-Seg Firms (In-Sample)	0.363	-59155	108954
(10)	D2V Segments	Mutli-Seg Firms (Out-of-Sample)	0.033	-83664	38628
(11)	CSTAT SIC Segments	Mutli-Seg Firms (Out-of-Sample)	0.028	-83443	38628
(12)	Both Segments	Mutli-Seg Firms (Out-of-Sample)	0.045	-84150	38628

Table 10: High Product Breadth Validation Regressions

The table reports validation regressions in which the dependent variable is a direct text-based measure of companies indicating that their products are broad. We consider four query-based measures obtained using the metaHeuristica software platform, based on product and service breadth. "Product Breadth" is the number of 10-K paragraphs containing the phrases {product lines, product categories}. "Prod/Svc Breadth" is analogously defined based on the search phrases {product lines, product categories, service lines, service categories}. "Product Breadth Detail" runs the same query as "Product Breadth" but is more stringent and additionally requires that the paragraph include one word from the following list: {breadth, broad, broader, wide, multiple, numerous, diverse, categories, divisions}. "Prod/Svc Breadth Detail" is a parallel more stringent version of the baseline "Prod/Svc Breadth" query. All four variables are scaled by the number of paragraphs in the 10-K overall. All regressions include firm and year fixed effects, and controls for firm size, age, 10-K size, M/B, and the TNIC HHI. Results are robust to dropping any of the controls. All regressions include firm and year fixed effects, coefficients are multiplied by 100 for ease of viewing, and t-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	D2V-Scope	NAICS-Scope	# Segments	Log Assets	Log Age	Log 10K Size	M/B	TNIC HHI	# Obs
(1)	Product Breadth				1.661	0.332	–8.445	–0.295	–2.094	72,277
					(3.650)	(0.220)	(–7.320)	(–2.310)	(–1.740)	
(2)	Prod/Svc Breadth				1.786	0.572	–8.892	–0.296	–2.038	72,277
					(3.850)	(0.370)	(–7.480)	(–2.280)	(–1.660)	
(3)	Prod Breadth Detail				0.756	–1.076	–3.321	0.058	–1.555	72,277
					(3.780)	(–1.630)	(–7.260)	(1.140)	(–3.250)	
(4)	Prod/Svc Breadth Detail				0.785	–0.871	–3.454	0.057	–1.494	72,277
					(3.880)	(–1.300)	(–7.280)	(1.110)	(–3.060)	
(5)	Product Breadth			1.379	1.524	0.076	–8.524	–0.286	–2.099	72,277
				(2.790)	(3.340)	(0.050)	(–7.380)	(–2.250)	(–1.740)	
(6)	Prod/Svc Breadth			1.296	1.657	0.331	–8.966	–0.288	–2.043	72,277
				(2.570)	(3.560)	(0.210)	(–7.540)	(–2.220)	(–1.660)	
(7)	Prod Breadth Detail			0.432	0.713	–1.156	–3.346	0.060	–1.557	72,277
				(1.930)	(3.550)	(–1.740)	(–7.300)	(1.190)	(–3.250)	
(8)	Prod/Svc Breadth Detail			0.418	0.743	–0.949	–3.477	0.060	–1.496	72,277
				(1.830)	(3.660)	(–1.410)	(–7.320)	(1.160)	(–3.060)	
(9)	Product Breadth	0.525		1.042	1.122	–0.156	–9.257	–0.292	–0.436	72,277
		(6.630)		(2.100)	(2.470)	(–0.100)	(–7.950)	(–2.290)	(–0.360)	
(10)	Prod/Svc Breadth	0.541		0.950	1.243	0.092	–9.722	–0.293	–0.330	72,277
		(6.700)		(1.880)	(2.680)	(0.060)	(–8.100)	(–2.260)	(–0.270)	
(11)	Prod Breadth Detail	0.205		0.301	0.556	–1.247	–3.632	0.058	–0.907	72,277
		(5.930)		(1.340)	(2.830)	(–1.880)	(–7.830)	(1.150)	(–1.870)	
(12)	Prod/Svc Breadth Detail	0.212		0.282	0.581	–1.043	–3.773	0.057	–0.826	72,277
		(5.980)		(1.230)	(2.930)	(–1.550)	(–7.840)	(1.120)	(–1.670)	
(13)	Product Breadth		0.305	1.061	1.268	0.318	–9.094	–0.294	–0.687	72,277
			(5.980)	(2.140)	(2.800)	(0.210)	(–7.790)	(–2.310)	(–0.570)	
(14)	Prod/Svc Breadth		0.323	0.961	1.386	0.588	–9.569	–0.296	–0.550	72,277
			(6.200)	(1.900)	(3.000)	(0.380)	(–7.960)	(–2.280)	(–0.450)	
(15)	Prod Breadth Detail		0.130	0.298	0.604	–1.053	–3.588	0.057	–0.957	72,277
			(5.230)	(1.320)	(3.060)	(–1.590)	(–7.740)	(1.120)	(–1.970)	
(16)	Prod/Svc Breadth Detail		0.135	0.278	0.630	–0.842	–3.729	0.056	–0.872	72,277
			(5.300)	(1.210)	(3.160)	(–1.250)	(–7.750)	(1.100)	(–1.760)	

Table 11: First-Stage Regressions

The table reports the results of first-stage regressions where measures of scope (D2V-scope and NAICS-scope) are regressed on our two instruments in addition to all controls and fixed effects. Our first instrument is "Sectoral Redeployment Potential" which is a product market spatially localized version of the asset redeployability measure in Kim and Kung (2017). In particular, we use the BEA capital flows table and represent the assets of each 4-digit NAICS industry as a vector. For each focal's local product market, we compute the average redeployability between the focal firm's nearest peer NAICS industries and the focal firm's distant peer NAICS industries. Intuitively, when the assets of near peers are easily redeployed to the market of more distant peers, the focal firm faces a low cost to expanding its market outward in space (lower cost of increasing scope). The second instrument "Sectoral Opportunity Set Potential" is simply the concentraiton ratio of 4-digit NAICS industries that the moderately distant peers in the focal firm's product market reside in. When this concentration ratio is high, it indicates that the focal firm's peers all tend to operate in the same market, which in turn implies there are few opportunities for scope expansion by the focal firm (as there are fewer related product markets that are spatially close). For success in the first stage, we predict that the former measure will be positively related to our scope variables, and the second will be negatively related. All regressions include firm and year fixed effects, and controls for firm size, age, 10-K size, M/B, and the TNIC HHI. All regressions include firm and year fixed effects, and t-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	Sectoral Redeployment Potential	Sectoral Opportunity Set Potential	Log Assets	Log Age	# Obs
(1)	D2V-Scope	1.156	2.124	0.978	–0.196	99,514
		(3.740)	(11.940)	(19.170)	(–1.380)	
(2)	NAICS-Scope	0.834	3.859	1.302	–0.911	99,513
		(1.740)	(14.540)	(17.800)	(–4.580)	
(3)	# Segments	0.002	0.112	0.107	0.162	99,514
		(0.030)	(3.380)	(12.370)	(6.310)	

Table 12: Investment Regressions

The table reports the second stage results of 2-stage instrumental variable regressions where the dependent variable is a firm investment policy such as acquisitions, divestitures (target of an acquisition), R&D/assets or CAPX/assets. Our instrumented variable of interest is a measure of scope (D2V-Scope or NAICS-Scope) as indicated in the panel headers. The first-stage regressions are displayed in Table 11 and include two instruments for scope (explained in detail in Table 11). The first is a measure of the extent to which the broader product market surrounding a focal firm is characterized by a high degree of outward-directed asset redeployability indicating a low cost to scope expansion by existing firms. The second is a measure of the size of the focal firm's outward-expansion opportunity set. We also include controls for size, age, and in Panel C, we additionally include controls for market to book and the TNIC HHI. All regressions include firm and year fixed effects, and *t*-statistics are clustered by firm.

Row	Dependent Variable	Scope Variable	Log Assets	Log Age	Mkt/Book	TNIC HHI	# Obs
		Panel A: D2V-Scope is Scope Variable					
(1)	Acquirer Dummy	0.019	–0.001	–0.047			98,205
		(3.450)	(–0.170)	(–5.040)			
(2)	Target Dummy	–0.012	0.041	0.044			98,205
		(–2.890)	(8.460)	(6.580)			
(3)	R&D/Assets	0.002	–0.014	0.007			98,205
		(3.400)	(–13.070)	(4.380)			
(4)	CAPX/Assets	0.000	–0.001	–0.013			98,205
		(–0.060)	(–1.680)	(–10.470)			
(5)	Vertical Integration	0.002	–0.002	0.001			98,125
		(12.880)	(–7.820)	(3.360)			
(6)	Outsourcing	0.025	0.012	–0.003			16,478
		(1.980)	(0.660)	(–0.090)			
		Panel B: NAICS-Scope is Scope Variable					
(7)	Acquirer Dummy	0.012	0.002	–0.040			98,204
		(3.590)	(0.360)	(–4.060)			
(8)	Target Dummy	–0.007	0.039	0.040			98,204
		(–2.730)	(9.220)	(5.810)			
(9)	R&D/Assets	0.001	–0.013	0.007			98,204
		(2.770)	(–13.330)	(4.630)			
(10)	CAPX/Assets	0.000	–0.001	–0.013			98,204
		(0.000)	(–1.960)	(–10.150)			
(11)	Vertical Integration	0.001	–0.001	0.002			98,124
		(14.550)	(–7.730)	(6.290)			
(12)	Outsourcing	0.016	0.021	0.025			16,477
		(1.990)	(1.410)	(0.730)			
		Panel C: D2V-Scope is Scope Variable (extra controls added)					
(13)	Acquirer Dummy	0.023	0.004	–0.035	0.018	0.097	97,633
		(2.970)	(0.530)	(–3.790)	(15.340)	(3.200)	
(14)	Target Dummy	–0.016	0.042	0.042	–0.004	–0.054	97,633
		(–2.690)	(7.270)	(6.110)	(–5.800)	(–2.350)	
(15)	R&D/Assets	0.003	–0.015	0.006	–0.001	0.004	97,633
		(2.920)	(–12.380)	(3.770)	(–2.930)	(0.950)	
(16)	CAPX/Assets	–0.001	0.000	–0.010	0.004	–0.005	97,633
		(–1.090)	(0.540)	(–8.110)	(19.970)	(–1.540)	
(17)	Vertical Integration	0.003	–0.002	0.001	0.000	0.010	97,562
		(10.500)	(–7.090)	(2.100)	(–2.090)	(8.360)	
(18)	Outsourcing	0.032	0.009	–0.016	0.002	0.096	16,418
		(1.970)	(0.460)	(–0.420)	(0.390)	(1.960)	

Table 13: Outcomes Regressions

The table reports the second stage results of 2-stage instrumental variable regressions where the dependent variable is a firm outcome variable such as the market to book ratio (market value of firm divided by total assets), sales growth, asset growth or profitability. Our instrumented variable of interest is a measure of scope (D2V-Scope or NAICS-Scope) as indicated in the panel headers. The first-stage regressions are displayed in Table 11 and include two instruments for scope (explained in detail in Table 11). The first is a measure of the extent to which the broader product market surrounding a focal firm is characterized by a high degree of outward-directed asset redeployability indicating a low cost to scope expansion by existing firms. The second is a measure of the size of the focal firm's outward-expansion opportunity set. We also include controls for size, age, and in Panel C, we additionally include controls for market to book and the TNIC HHI. All regressions include firm and year fixed effects, and t-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	Scope Variable	Log Assets	Log Age	Mkt/Book	TNIC HHI	# Obs
			Panel A: D2V-Scope is Scope Variable				
(1)	Valuation	0.100	–0.489	–0.363			97,634
		(5.220)	(–18.900)	(–9.640)			
(2)	Sales Growth	0.034	–0.104	–0.201			97,834
		(5.960)	(–14.370)	(–21.390)			
(3)	Asset Growth	0.046	–0.210	–0.067			98,202
		(8.550)	(–28.650)	(–6.750)			
(4)	OI/Assets	–0.001	0.009	0.005			98,004
		(–0.440)	(3.010)	(1.230)			
			Panel B: NAICS-Scope is Scope Variable				
(5)	Valuation	0.056	–0.464	–0.333			97,633
		(5.070)	(–20.900)	(–8.750)			
(6)	Sales Growth	0.021	–0.098	–0.189			97,833
		(6.210)	(–15.760)	(–19.640)			
(7)	Asset Growth	0.028	–0.201	–0.051			98,201
		(8.970)	(–32.350)	(–5.160)			
(8)	OI/Assets	0.000	0.008	0.005			98,003
		(–0.330)	(3.230)	(1.150)			
			Panel C: D2V-Scope is Scope Variable (extra controls added)				
(9)	Valuation	0.086	–0.375	–0.122	0.323	0.310	97,400
		(4.300)	(–16.870)	(–4.250)	(34.160)	(3.940)	
(10)	Sales Growth	0.041	–0.091	–0.167	0.049	0.170	97,276
		(5.290)	(–11.110)	(–17.200)	(25.360)	(5.580)	
(11)	Asset Growth	0.050	–0.187	–0.017	0.070	0.174	97,633
		(6.900)	(–23.590)	(–1.660)	(38.240)	(6.190)	
(12)	OI/Assets	–0.002	0.012	0.012	0.008	–0.005	97,439
		(–0.700)	(3.780)	(2.860)	(9.980)	(–0.410)	

Table 14: Venture Capital Funding Similarity and Fluidity Regressions

The table reports the second stage results of 2-stage instrumental variable regressions where the dependent variable is a measure of early-stage startup financing or innovative activity in the given firm's product market. Both dependent variables are developed in Hoberg, Phillips, and Prabhala (2014). VC funding similarity is the cosine similarity of the given focal firm's 10-K product description to the average vocabulary used by all startups in the given year (where the startup vocabulary is obtained from Venture Expert business descriptions of all startups receiving their first round of financing in the given year). Product market fluidity is the average market-wide change (fluidity) in the use of the given firm's 10-K product description vocabulary by all other firms. A high value indicates a large amount of product innovation by competing firms in the focal firm's product markets. Our instrumented variable of interest is a measure of scope (D2V-Scope or NAICS-Scope) as indicated in the panel headers. The first-stage regressions are displayed in Table 11 and include two instruments for scope (explained in detail in Table 11). The first is a measure of the extent to which the broader product market surrounding a focal firm is characterized by a high degree of outward-directed asset redeployability indicating a low cost to scope expansion by existing firms. The second is a measure of the size of the focal firm's outward-expansion opportunity set. We also include controls for size, age, and in Panel C, we additionally include controls for market to book and the TNIC HHI. All regressions include firm and year fixed effects, and t-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	Scope Variable	Log Assets	Log Age	Mkt/Book	TNIC HHI	# Obs
			Panel A: D2V-Scope is Scope Variable				
(1)	VC Funding Similarity	1.758	–1.161	–0.364			98,186
		(16.070)	(–8.250)	(–1.550)			
(2)	Product Market Fluidity	0.794	–0.416	–0.610			97,100
		(14.900)	(–6.440)	(–5.860)			
			Panel B: NAICS-Scope is Scope Variable				
(3)	VC Funding Similarity	1.012	–0.758	0.186			98,185
		(18.690)	(–7.920)	(0.950)			
(4)	Product Market Fluidity	0.508	–0.303	–0.318			97,099
		(17.970)	(–6.340)	(–3.520)			
			Panel C: D2V-Scope is Scope Variable (extra controls added)				
(5)	VC Funding Similarity	2.265	–1.406	–0.497	0.050	6.912	97,614
		(12.140)	(–7.140)	(–1.690)	(1.700)	(9.400)	
(6)	Product Market Fluidity	0.935	–0.464	–0.645	0.046	2.312	96,533
		(11.610)	(–5.620)	(–5.420)	(3.800)	(7.190)	

Table 15: Financing Regressions

The table reports the second stage results of 2-stage instrumental variable regressions where the dependent variable is a firm financing policy such as equity issuance, debt issuance, dividends, or equity repurchases. Our instrumented variable of interest is a measure of scope (D2V-Scope or NAICS-Scope) as indicated in the panel headers. The first-stage regressions are displayed in Table 11 and include two instruments for scope (explained in detail in Table 11). The first is a measure of the extent to which the broader product market surrounding a focal firm is characterized by a high degree of outward-directed asset redeployability indicating a low cost to scope expansion by existing firms. The second is a measure of the size of the focal firm's outward-expansion opportunity set. We also include controls for size, age, and in Panel C, we additionally include controls for market to book and the TNIC HHI. All regressions include firm and year fixed effects, and t-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	Scope Variable	Log Assets	Log Age	Mkt/Book	TNIC HHI	# Obs
			Panel A: D2V-Scope is Scope Variable				
(1)	Equity Issuance	0.009	–0.047	–0.016			98,205
		(7.230)	(–25.040)	(–6.160)			
(2)	Debt Issuance	0.002	–0.012	0.009			98,205
		(0.730)	(–3.590)	(1.840)			
(3)	Dividends/Assets	–0.001	0.001	0.002			98,106
		(–2.220)	(1.830)	(4.120)			
(4)	Repurchases/Assets	–0.002	0.006	0.007			90,689
		(–1.640)	(4.650)	(6.010)			
			Panel B: NAICS-Scope is Scope Variable				
(5)	Equity Issuance	0.005	–0.045	–0.013			98,204
		(7.310)	(–27.110)	(–5.070)			
(6)	Debt Issuance	0.002	–0.012	0.010			98,204
		(0.980)	(–4.180)	(1.960)			
(7)	Dividends/Assets	0.000	0.000	0.002			98,105
		(–1.920)	(1.400)	(3.560)			
(8)	Repurchases/Assets	–0.001	0.006	0.007			90,688
		(–1.530)	(5.080)	(4.850)			
			Panel C: D2V-Scope is Scope Variable (extra controls added)				
(9)	Equity Issuance	0.010	–0.042	–0.005	0.015	0.031	97,633
		(5.660)	(–21.470)	(–2.080)	(23.560)	(4.480)	
(10)	Debt Issuance	0.001	–0.011	0.010	0.001	–0.002	97,633
		(0.330)	(–2.850)	(2.040)	(3.070)	(–0.110)	
(11)	Dividends/Assets	–0.001	0.001	0.003	0.001	–0.003	97,537
		(–2.310)	(2.400)	(5.130)	(7.720)	(–1.640)	
(12)	Repurchases/Assets	–0.003	0.007	0.009	0.002	–0.007	90,155
		(–1.570)	(4.600)	(7.100)	(7.060)	(–1.080)	

Table 16: Scope Distance and Firm Size Subsamples

The table reports the second stage results of 2-stage instrumental variable regressions for various dependent variables as noted in the first column. These regressions use the same IV specification as in earlier tables such as Table 12, except we now run these regressions using subsamples as indicated in the column headers. The scope distance subsamples are constructed in each year by sorting firms based on the average past-year product market distance of the D2V industry segments they are assigned to when constructing our D2V-scope variable (firms assigned to just one D2V segment have a distance of zero). Firms in the near-scope (far-scope) subsample are operating across industries that are highly related (unrelated). We also sort firms into above and below median size based on lagged assets. As a result, we have four subsamples once we consider both (as noted in column headers): Near-Scope Small Firms, Near-Scope Large Firms, Far-Scope Small Firms and Far-Scope Large Firms. Our variable of interest is instrumented scope (D2V-Scope) as in earlier tables, and we only report this coefficient for parsimony. All regressions include firm and year fixed effects, controls for size and age (not reported) and t-statistics for the D2V-scope coefficient are clustered by firm and are shown in parentheses.

Row	Dependent Variable	Near Scope & Small Size Subsample	Near Scope & Large Size Subsample	Far Scope & Small Size Subsample	Far Scope & Large Size Subsample
(1)	Acquirer Dummy	0.058	0.011	0.029	0.012
		(3.120)	(0.810)	(1.200)	(0.930)
(2)	R&D/Assets	0.003	0.000	0.002	0.001
		(0.870)	(0.130)	(2.330)	(2.280)
(3)	CAPX/Assets	0.000	0.003	0.001	-0.001
		(-0.070)	(1.570)	(0.450)	(-0.640)
(4)	Outsource Dummy	0.092	0.005	0.051	0.012
		(1.570)	(0.130)	(1.140)	(0.810)
(5)	Valuation (M/B)	0.226	0.176	-0.061	0.048
		(3.450)	(2.690)	(-1.080)	(1.740)
(6)	Sales Growth	0.056	0.072	0.030	0.027
		(2.380)	(3.000)	(2.720)	(3.250)
(7)	Asset Growth	0.110	0.081	0.033	0.030
		(4.930)	(3.960)	(2.660)	(3.420)
(8)	Equity Issuance	0.022	0.023	0.001	0.002
		(4.410)	(3.770)	(0.330)	(2.300)
(9)	VC Funding Score	1.077	1.074	1.208	0.914
		(7.420)	(6.290)	(5.470)	(6.540)
(10)	Prod Mkt Fluidity	2.828	1.685	2.795	1.668
		(8.330)	(6.860)	(6.140)	(6.270)

Figure 1: Measures of scope versus time. The upper figure plots the average number of Compustat segments per firm over time. The middle figure plots the average values of D2V-scope and NAICS-scope over our sample period. D2V-scope and NAICS-scope are based on scoring each firm's Item 1 business description based on how similar it is to the product text of specific fixed industries. For D2V-scope, fixed industries are based on the TNIC-based D2V-300 industries using doc2vec and for NAICS-scope, it is based on 4-digit NAICS industries. The lower figure plots D2V-scope over time using three different weights: equal-weighted (our baseline), sales-weighted, and asset-weighted.

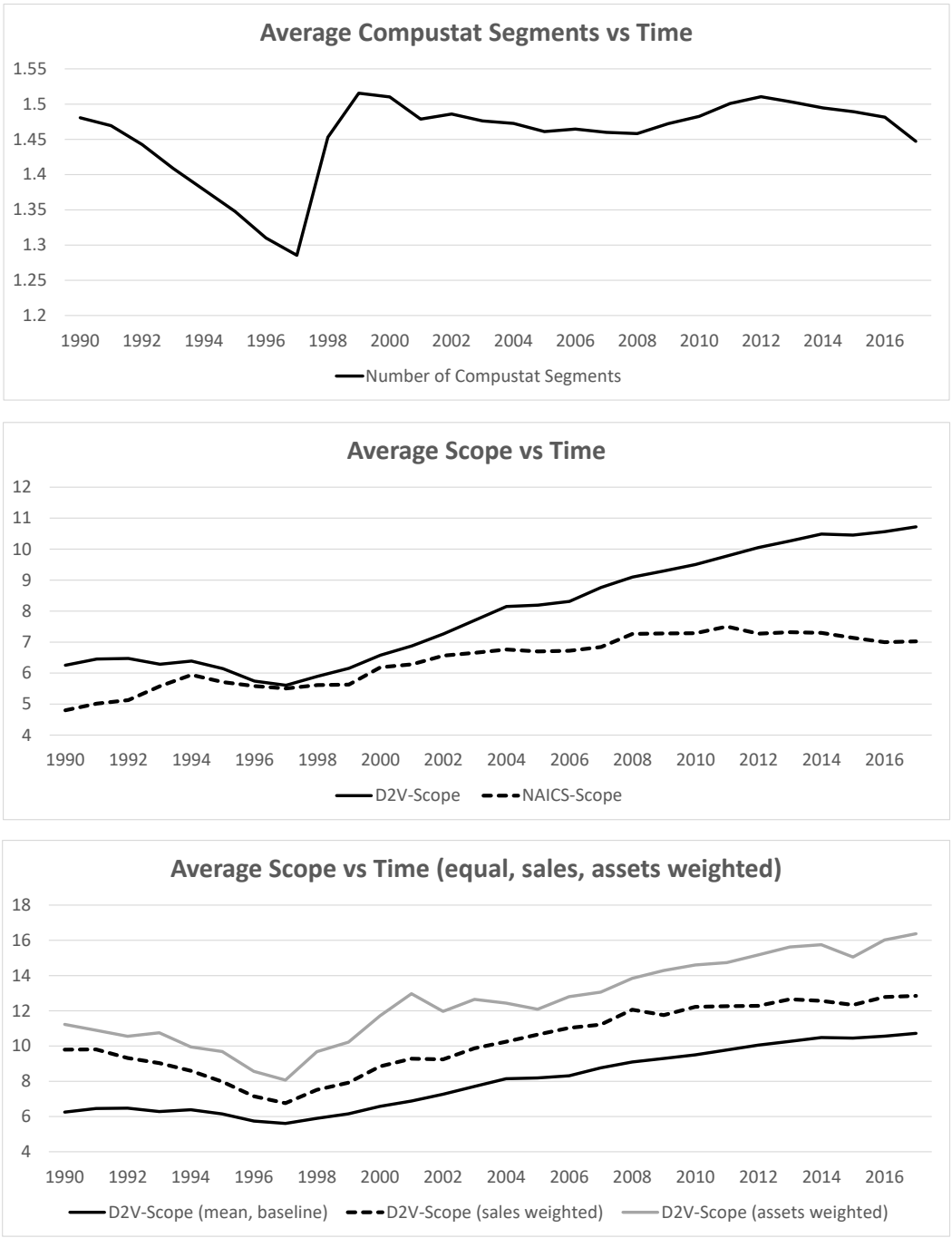


Figure 2: Measures of scope versus time for highly related versus unrelated industry-pairs. The figure decomposes d2vscope into components where firms are operating in the "most related" segments (highest quintile of product similarity among all industry pairs), "weakly related segments" (second quintile) and "likely unrelated segments" (final 3 quintiles). We start by taking all permutations of the 300 D2V industries and compute pairwise industry similarities for each industry based on the pairwise similarities of the k-means centroids from which each industry was estimated. We then sort all industry pairs into quintiles based on how spatially distant they are from each other (product market distance). We label those industry-pairs in the most similar quintile as "most related" segments and those in the second most similar quintile as "weakly related segments". We group the least similar three quintiles into a single group of "likely unrelated" industry pairs as these last three quintiles uniformly have very low similarity and also have far fewer operating pairs as noted in Table 6. We then tabulate how many pairs are operating in each similarity bin for each firm, where counts are weighted such that each segment gets a total weight of unity. In each year, we then average all segments counts in each quintile bin across all firms in each year to generate an average number of segments firms have in each bin in each year. As our goal is to illustrate how segment counts are growing or declining in each bin, we then normalize each by the first year of this analysis 1990 so that each reported time series in the figure indicates total growth relative to the base year. For example, the 1.47 in 2017 in the first quintile indicates that near-segments grew by 47% during our sample period.

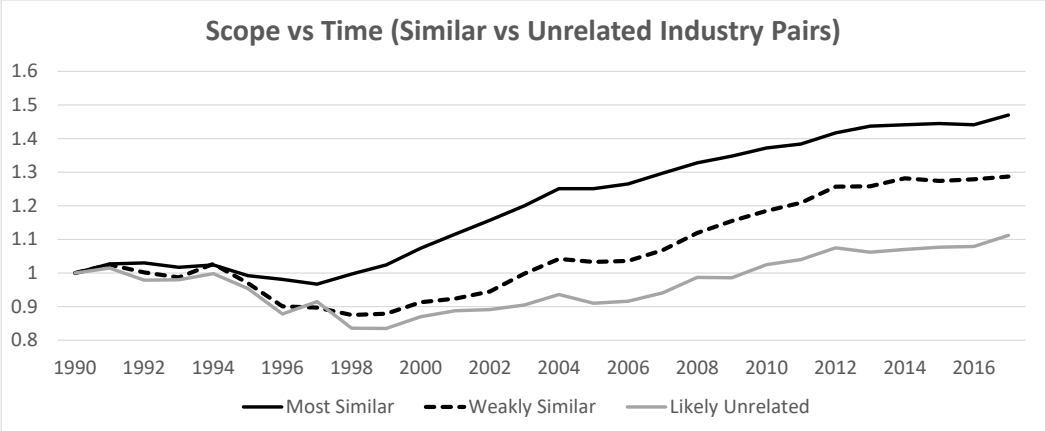


Figure 3: Firm size versus time. The figure displays firm size (measured as Compustat assets, both nominal and inflation adjusted) over time.

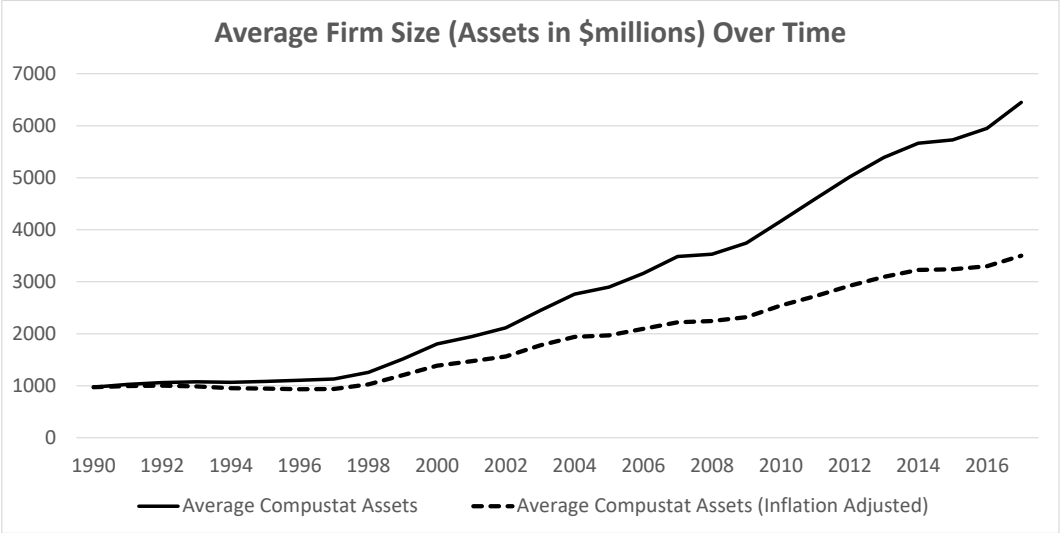
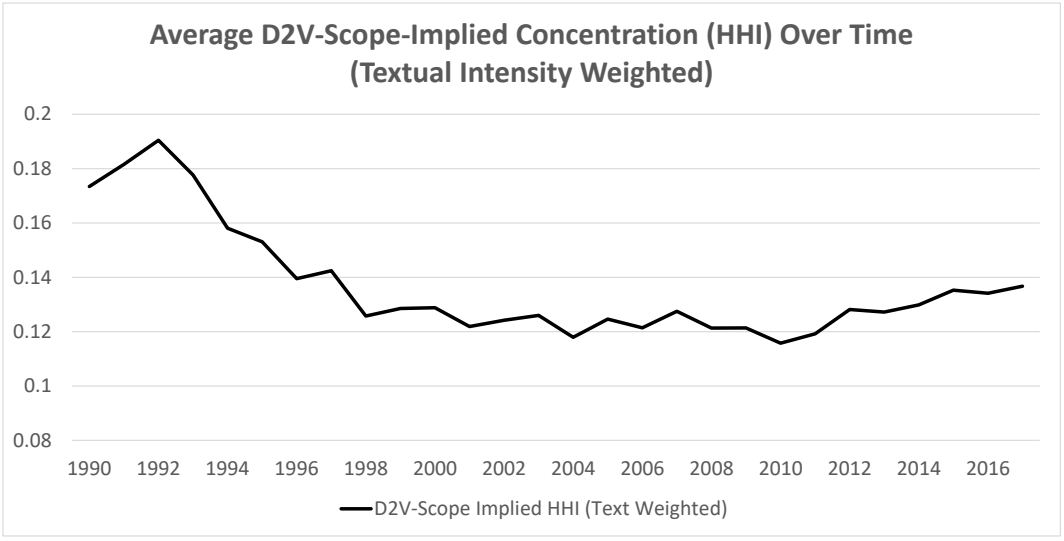


Figure 4: The D2V-Scope Implied HHI computes the HHI after allowing firms to have a presence in multiple industries, as is identified during the derivation of D2V-scope itself. Market shares are based on sales. Each firm's sales are allocated across the multiple sectors each firm is assigned to using similarity weights (similarity weights are defined as $Q_{i,j,t,D2V}$ in equation (1)). HHIs are then computed at the D2V industry level using these allocated sales where firms operate in multiple sectors. We then aggregate these HHIs back to the firm level by computing weighted averages over the sectors each firm operates in (again using weights $Q_{i,j,t,D2V}$). We then aggregate HHIs to the economy-wide annual level by computing a sales weighted average of the firm HHIs or an equal weighted average of the firm HHIs. The upper figure reports the sales-weighted average HHI over time and the lower figure reports the equal weighted average HHI over time.



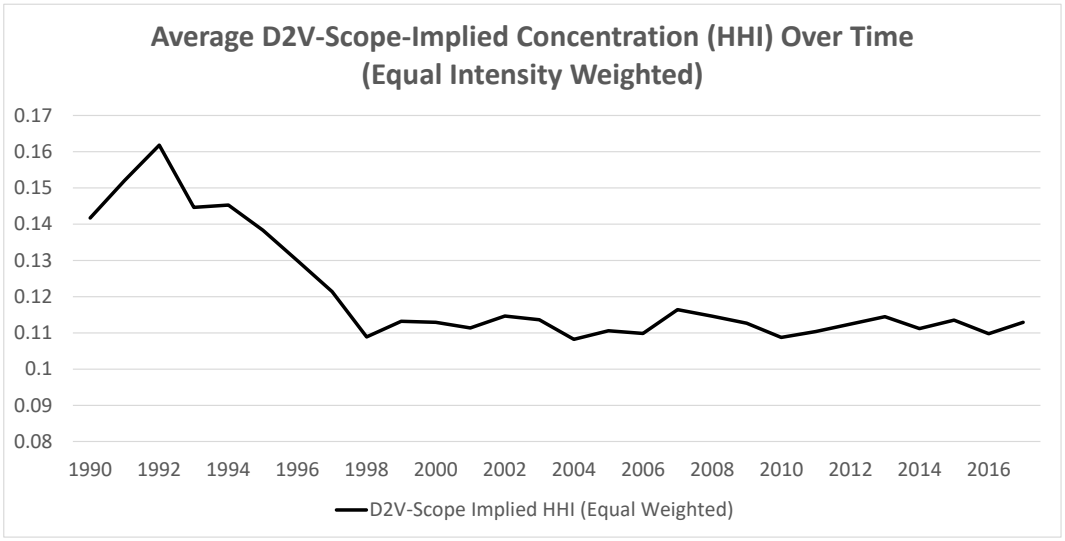
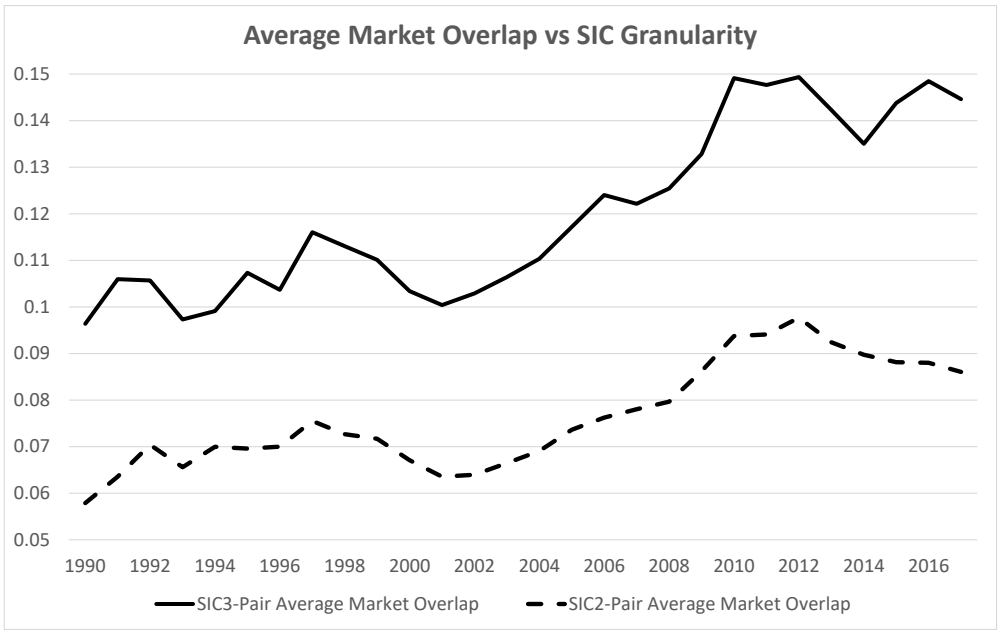
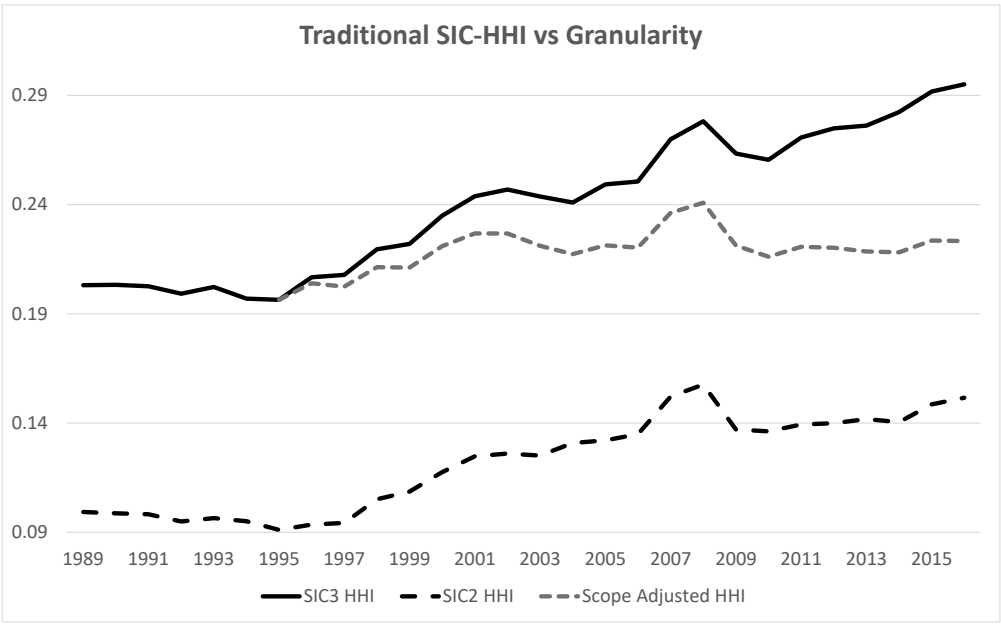


Figure 5: The upper figure reports the average market overlap of pairs of firms that are in the same SIC3 or SIC2 industries. Market overlap at the industry pair level is the average of firm-pair market overlap for all permutations of firms where one firm is in each of the industries being compared at the industry-pair level. Firm-pair overlap is the intersection of industries the two firms in the pair likely operate in divided by the union of industries they operate in (based on the industries assigned to each firm as indicated by the construction of the D2V-scope variable). This market overlap score ranges from zero to unity and is one if the firms operate in exactly the same industry and zero if they have no overlaps. The lower figure reports the two and three digit SIC HHI over time. The scope-adjusted HHI is the average the SIC2 and SIC3 HHI, where the weights start at zero in 1996 and grow linearly until they reach 50% by the end of our sample in 2017.





Online Appendix: Scope, Scale and Competition: The 21st Century Firm

Gerard Hoberg and Gordon M. Phillips

(not for publication)

1. A More Detailed Explanation of Our Scope Measure

Our methodology for modeling multi-industry firms and scope is summarized in detail in Appendix B in the main paper. The formulation first embeds each firm's business description into a 300-dimensional space using doc2vec. Next, K-means clustering is run on the resulting set of 300-dimensional firm vectors, creating cluster centroids that are spatial representations of specific industries in the same 300-dimensional space. These methods are rather technical, and some readers might be less familiar wtih them. This section thus uses an illustrative example based on three fictitious firms to more intuitively illustrate the methodology while providing intuition for what each step contributes to the process.

Consider an economy with 4 industries (furniture, chocolates, autos, paper) and 3 firms. Each firm has a 10-K business description with some firms operating in multiple markets as indicated below.[11] Text corresponding to the four industries are denoted in blue, purple, green, and brown, respectively.

Firm 1: We sell furniture items including chairs and sofas. Our chocolate division sells gift wrapped chocolates including creams and nuts.

Firm 2: We operate in the automotive industry and sell high-performing sports cars.

Firm 3: We sell furniture using Eastern European hardwoods. We sell paper used in high speed photocopiers.

We start by training a doc2vec model on all business descriptions and drop stop words. In our example, stop words might include (and, in, including, items, operate, our, the, we).[12] There are 25 unique words in these three business descriptions after dropping stop words indicating that these business descriptions reside in a 25 dimensional space at their full dimensionality (one unique word corresponds to each specific dimension when representing these business descriptions in standard vector form). However, the doc2vec algorithm developed by Mikolov et al. 2013 (see Hanley and Hoberg 2019 for an application in finance), allows researchers to represent the three firms in a reduced dimension space where the researcher can specify a desired number of dimensions. Suppose the researcher requests doc2vec to model these business descriptions in a six-dimensional space (our actual implementation is based on 300-dimensions given HP2016 indicate 300 industries as a benchmark). The use of dimensionality reduction algorithms such as doc2vec significantly improves economic signals as doc2vec can account for the fact that words like "chair" and "sofa" are highly

[11] Note that a typical business description is roughly 4 pages long and we provide short examples here for illustrative purposes.

[12] In our study, stop words are those appearing in more than 25% of all 10-Ks in the base-year 1997.

related, whereas full-dimension raw vector representations are inflexible and treat all words as unrelated.

The result of using doc2vec as indicated is a separate six-element vector that represents the business description of each firm.

Firm 1: [.25, .20, 0, 0, .91, 0]
Firm 2: [0, 0, .91, .09, 0, 0]
Firm 3: [.26, .19, 0, 0, 0, .88]

The vectors obtained from doc2vec will have the property that firms using similar text will have vectors with higher numbers in common elements. For example, the cosine similarity of the vectors for Firms 1 and 3 is 33.8%. This reflects the fact that both of these firms operate in the furniture industry, and they use some related content in their business descriptions. In contrast, the similarity between Firms 1 and 2, and also Firms 2 and 3, are both zero. This indicates that these pairs do not span any common industry exposures.

Doc2vec spatial representations are highly informative as they bin similar content into specific dimensions making similarity comparisons more informative. Yet the example above illustrates that the dimensions of doc2vec vectors will not generally correspond to specific industries. For example, the positive similarity between Firms 1 and 3 is driven by common exposure to both vector elements 1 and 2. Hence, without additional steps, a taxonomy of specific industries is not possible as there is no clear mapping from these vectors to a set of labelled industries.

1.1 From Firm Vectors to Industry Vectors

The discussions in the prior section illustrate how embedding technologies such as doc2vec can be used to represent entire firm product portfolios. The objective in this section is to process the spatial representations to identify specific industries. We achieve this goal using K-means clustering. We run K-means clustering only on the set of single segment firms in our sample (these are firms that have only one Compustat segment in our full implementation). In the example above, only Firm 2 is a single segment firm given that it only operates in the automotive industry. Suppose there were a large number of single segment firms operating in the furniture, chocolate, automotive and paper industries in our above example. In this case, a researcher could run K-means clustering on the set of 6-element vectors for each of these firms. The likely result would be four clusters, each having a centroid vector that is

defined as the average vector of the firms grouped into each cluster. The result would look similar to the following:

Furniture Industry Centroid: [.25,.19,0, 0, 0, 0, 0, 0, 0, 0]
Chocolate Industry Centroid: [0, 0, 0, 0, .92, 0, 0, 0, 0, 0]
Automotive Industry Centroid: [0,0,.89, .11, 0, 0, 0, 0, 0, 0]
Paper Industry Centroid: [0, 0, 0, 0, 0, 0.90, 0, 0, 0, 0]

K-means clustering solves the problem of identifying specific industries in the vector space because it groups firms using common text. Because the algorithm will group all single-segment furniture firms into one cluster, the average centroid of this cluster would be represented as above, revealing that furniture companies are represented in the doc2vec space as having content in the first two dimension. Chocolate is represented by the 5th dimension, automotive by the second and third dimensions, and paper by the 6th dimension. Our actual implementation runs K-means with 450 clusters, which we later reduce to 300 industries using a boilerplate cleaning procedure discussed below.

1.2 Labeling Industries

As noted above, the K-means procedure run after running doc2vec can generate spatial representations of both firms and industries within the same space. Yet these algorithms do not provide labels for each industry. Labels are important for many research applications as they, for example, would confirm that Tesla is in the automotive industry and CVS is in the pharmacy industry. Please see Tables 3 and 4 in the main paper to see examples of industry labels for Tesla and CVS.

To label our industries, we use the word2vec feature of doc2vec, which allows us to obtain a spatial representation for each individual word. The ability to represent individual words is a feature built into doc2vec, and hence we do not need to run additional tools. We use word2vec to obtain word representations for all words in the corpus (except stop words). We then assign each industry a label consisting of the top 10 words that have spatial vectors that are most similar (i.e., they have the highest cosine similarity) to each centroid vector. Intuitively, the individual words that will be most similar to each industry's centroid vector will be most directly indicative of the industry itself. For example, the automotive centroid would score highly on individual words such as car and vehicle. It might also score high on auto-related features such as windshields and upholstery, and possibly on highly related products such as vans and motorcycles.

1.2 Scoring Firms Regarding Which Industries They Operate In

A naive approach to scoring firms might be to simply compute the cosine similarity between each firm's doc2vec representation (see Firm 1 vector above for example) and each industry's representation (see Furniture Industry Centroid above for example). However, this approach would be severely biased because firms selling products in a large number of industries would score low on each industry they actually operate in due to the fact that only a small fraction of its text is from each industry and cosines are relative similarities. We thus take an approach based on whether each firm's business description actually contains a significant fraction of the label-keywords that strongly associate with each industry centroid as noted above. The direct approach using actual industry keywords (i.e., dialects) is crucial for this specific task, as it does not punish firm scores when they operate in a large number of industries.

We thus extract large centroid-specific dialects of roughly 500 words each and simply compute the fraction of these dialect terms that actually appear in a given firm's business description.[13]

In reviewing the list of words associated with each dialect, we observed that some words are more focal to a given industry than others (for example, car is more focal to the auto industry than is motorcycle, tire or door). This similarity-to-the-centroid is captured by the word-centroid cosine similarity discussed above. We also noticed that some terms are more dispersed across centroids than others (for example, "car" is more uniquely used by automotive industry than is "door", which is also used by furniture companies and home builders). We thus score each word by measuring its concentration, which we compute as the word's HHI across industries (the sum of squared shares of the word's usage across industries). We compute a measure of word importance or word quality for each word n and each industry k as the product of our two metrics of quality: similarity-to-centroid and unique-industry-focus-concentration:

$$\text{importance}_{k,n} = \text{Similarity}_{k,n} \cdot \text{word-concentration-HHI}_n \tag{7}$$

[13]More specifically, we take the top 2500 words for each centroid thus having 2500 x 450 centroid-word pairs, and then sort all centroid-word pairs in a pooled sort and take the top 500 x 450 pairs as centroid-specific dialects. We further ensure that no centroid has fewer than 50 words assigned to it by simply taking its top 50 terms if the pooled sort does not assign the centroid at least 50 words. This approach, on average, assigns each centroid an average of just over 500 words, constituting the given centroid's dialect. Some centroids have more than 500 terms if the dialect is verbose and others naturally have fewer if its dialect is Spartan, allowing the data to naturally establish the unique vocabulary of each centroid.

Our measure of each firm i's overall exposure to an industry k is then the importance-weighted average fraction (for each word n) of the industry's dialect that appears in the given firm's business description.

$$\text{Exposure}_{i,k} = \frac{\sum\limits_{n=1...N} B_{i,k,n} \cdot \text{importance}_{k,n}}{\sum\limits_{n=1...N} \text{importance}_{k,n}} \tag{8}$$

We compute separate measures of exposure for each firm-industry pair in each year. A firm is deemed to be operating in industry k in the given year if this pair-specific exposure score is in the highest 2% of all firm-industry pair permutations in the given year.[14]

1.3 Purging Boilerplate and Redundant Industries

Because business descriptions typically contain some boilerplate content as well as some discussions about topics unrelated to product offerings (such as employees, or in more isolated cases, information about geography or financials), we purge 150 of our 450 candidate industries, thus reaching our target of 300 genuine industries based on product offerings. We do so by removing candidate industries that satisfy any of the following three conditions:

- [**Boilerplate via Overly Broad Content**] Some candidate industries only include dialects that are very broad, with their keywords appearing in a large fraction of industries. We deem a word to be "specific" if it appears in no more than 15 candidate industries. We then drop the 75 candidate industries with the fewest specific words. This results in dropping candidate industries with six or fewer specific words.

- [**Boilerplate via Manual Review**] We read the top keywords for each of the remaining 375 candidate industries and identify 20 that have non-product content. These are not valid candidate industries and we drop them, resulting in 355 candidate industries.

- [**Redundant Industries**] The K-means clustering algorithm produces larger numbers of clusters in regions of the product space where more firms are clustered. This is the case, for example, for the banking industry where there are a large number of highly similar publicly traded firms. We compute the cosine similarity of the doc2vec centroid vector for all remaining 355 industries and identify 55 candidate industries that have a near-duplicate candidate industry. These 55 candidate industries have a cosine similarity of 73.6% to 92.9% with their near-duplicates. We drop these 55

[14]The 2% figure obtains from HP2016 and is the granularity of TNIC-3 industries and 3-digit SIC industries.

candidate industries as they are redundant to avoid double counting scope for firms in these markets.

The result is 300 final industries that have a critical mass of words that are specific to each industry (each has its own dialect), that are not boilerplate, and that are unique. As noted above, we compute firm exposures to each industry using the formulation above in equation 8, and we retain the 2% firm-industry pairs that have the highest exposures. The result is then a set of industries each firm is exposed to in each year in a significant way. These firm-industry mappings indicate a set of industries that each firm likely operates in, and this data structure is analogous to the Compustat segment database. Some firms operate in just one or very few industries, and others operate in many. We measure scope simply as the number of these industries that each firm operates in. Intuitively, a firm with high scope is likely selling products to customers in many different industries.

2. Broadening Scope of Businesses

This section provides more details specifically to supplement Section 6 of the main paper. The primary issue with existing classifications is they have fixed granularities, and as firms broaden scope of their operations, concentration in specific product markets cannot be computed using narrow industry assignments of firm sales data. Firms may produce in multiple 3-digit SIC codes and competition is mismeasured if researchers assign their sales to just a small subset of these SIC codes. The extent of this bias could be time-varying. For example, firms may produce in a just one 3-digit SIC industry code early in our sample, while later producing in multiple SIC codes - thus making one or even two 2-digit SIC codes more representative of its scope of production.

We first illustrate this point using a test that provides external validity of this idea. We intentionally avoid using the spatial TNIC representation of industries for this test and instead consider annual OLS regressions where the intensity of managerial competition complaints is the dependent variable (see Li, Lundholm, and Minnis (2013)). We regress this variable on both one minus the Compustat SIC-3 HHI where each firm is assigned to the Compustat 3-digit SIC code it reports and one minus the Compustat SIC-2 HHI - using the 2-digit SIC code the firm reports. We flip the sign on the HHIs for convenience as (1-HHI) is a positive measure of competition. We also use the same sample selection criteria as Grullon, Larkin, and Michaely (2019) for consistency. The results are reported in Table IA11.

The table illustrates an economically large trend toward increasing importance of coarser SIC-2 codes in understanding firm production and competition (reinforcing our conclusion

that firms are operating in more markets over time). In the first year of this sample, 1997, only competition measured using the SIC-3 HHI predicts competition complaints. This early result indicates that 3-digit SIC codes well-represented the appropriate granularity of market boundaries at which competition among firms took place. However, throughout our sample, the relative importance of the HHI measured using 2-digit SIC codes increases and the relative importance of the SIC-3 digit HHI decreases. By the end of our sample, the coefficients for both HHIs are roughly equal in size, suggesting that competition is taking place across multiple 3-digit SIC codes and complaints are arising from multiple 3-digit level product markets. In the next section, we will show that market overlaps using our TNIC scope-based framework will indicate the same conclusion, and adjusting HHIs for broadening scope suggests that horizontal concentration is not rising materially in our sample.

2.1 Scope Adjustment via Market Overlap Analysis

We now develop an intuitive adjustment of HHIs for scope based on examining how market overlap varies with granularity. We define "Firm-Pair Market Overlap" for a pair of firms i and j using the D2V-300 segment database derived in Section 2 of the paper. For example, suppose firm i operates in industry A and B, and j operates in A, C and D. Market Overlap for this pair is the true overlap in their markets, which is $\frac{1}{4}$, as they intersect on just one industry but the union of industries they serve is four. We next define "Industry-Pair Market Overlap" for a pair of industries in any classification as the average Firm-Pair Market Overlap averaged over all permutations of pairs of firms i in the first industry and j in the second industry. If the two industries have high Industry Pair Market Overlap, it follows that the boundary between the two industries is not material and that competition plays out at a more coarse level of industry granularity.

To illustrate the impact of scope on industry boundaries and the role of granularity, the upper graph in Figure 5 plots the average Industry-Pair Market Overlap for all pairs of 4-digit SIC industries that have the same 3-digit SIC code, and also for all 4-digit SIC industries that have the same 2-digit SIC code but not the same 3-digit SIC code. Both statistics have been increasing rather dramatically throughout our sample. This illustrates that the aforementioned rise in scope is indeed rendering narrow industry boundaries less relevant over time, especially for more fine granularities such as three-digit SIC codes. The more important observation, however, is that the average market overlap for the SIC-2 pairs late in our sample is almost as high as the level of market overlap for SIC-3 pairs early in our sample. This indicates that industry boundaries are almost as strong today at the two-digit

SIC level as they were at the three-digit SIC level 25 years ago.

At the start of our sample, SIC-2 market overlap was roughly 6% and SIC-3 market overlap was roughly 10%. By the end of our sample, SIC-2 market overlap rose to 8.6%, and this 2.6% rise is enough to close over half of the ex-ante gap of 4% between SIC-2 and SIC-3. It follows that the granularity at which competition takes place moved over one half of one level of granularity. To show the impact of such a shift on concentration levels, we plot three trend-lines for concentration in the lower graphic of Figure 5. These include the benchmark SIC-2 HHI and the SIC-3 HHI as computed in the existing literature,[15] as well as a mixture of the two that starts at 100% SIC-3 HHI in 1997 and linearly moves to 50% SIC-3 HHI and 50% SIC-2 HHI at the end of our sample. Only the mixed HHI roughly holds market overlap fixed during the crucial post-1996 sample, and hence only this specification is a reasonable scope-adjusted HHI trend line.

The figure illustrates that horizontal concentration is not rising materially when we consider a scope-adjusted HHI. In contrast, we replicate the finding in the existing literature that concentration does appear to be rising dramatically if we do not adjust HHI measures for scope. The scope-adjusted HHI essentially allows granularity to shift with average scope whereas past studies hold granularity fixed over time. Adjusting granularity is necessary because increasing scope broadens industry boundaries, and competition thus occurs over increasingly coarse levels of granularity. The linear adjustment we employ in this section is highly simplified, and we reiterate that our goal here is to show intuition for how scope can impact competitive granularity, which in turn, can impact how competition is changing over time. In the next section, we adopt a more direct scope-adjusted measure of concentration based on our implicit modeling of the multiple industries firms operate in.

2.2 Scope Adjustment via Implied Multi-Industry Assignments

The construction of the D2V-scope variable assigns each firm to multiple industries when its Item 1 is similar to more than one industry. We now use this enhanced data structure to compute new HHIs at the D2V-300 industry level that use this multi-industry-assignment-classification directly. To do so, we first allocate each firm's total sales to the multiple industries it is assigned to using the basic similarity weights (see $Q_{i,j,t,D2V}$ in equation (1))

[15] We compute baseline SIC-2 and SIC-3 HHIs following the sample and weighting scheme used by Grullon, Larkin, and Michaely (2019). We limit the sample to firms with CRSP exchange codes of 1 to 3, CRSP share codes 10 and 11, sales and assets greater than one million, and we exclude financials and utilities. We also compute HHIs based on assigning firms to more than one industry if indicated in the Compustat segment tapes. The annualized average HHIs are also weighted by sales.

that were used to construct the classification itself. HHIs are then computed at the D2V industry level using these allocated sales where firms operate in multiple sectors. We then aggregate these HHIs back to the firm level by computing weighted averages over the sectors each firm operates in (again using weights $Q_{i,j,t,D2V}$). Note that our results are similar if we use equal weights instead.

We then aggregate HHIs to the economy-wide annual level by computing a sales weighted average of the firm HHIs or an equal weighted average of the firm HHIs.[16] We then plot both estimates of the HHI faced by average firm in each year in the Figure 4. The figure illustrates, as was the case with the scope adjustment used in the previous section, that horizontal concentration levels are not rising materially after 1997.

The results in this section suggest that an extended narrative might be relevant to understand the rise in industry concentration reported in the literature. This extended narrative is that scope has been rising rapidly as companies merged and listings declined, and as a consequence, traditional HHIs measured without adjustment are increasing. Yet the rise in these HHIs might not indicate reductions in horizontal competition as they are based on overly rigid classifications that do not account for scope and that assign firms to single industry categories.

3. A more informative TNIC Industry Classification

As noted in the main paper, doc2vec is a more technically advanced technology compared to the simple approach used in HP2016 to develop the original TNIC classification. The original methodology was based on cosine similarities of pairs of companies based on the raw text in each firm's 10-K. This approach, despite its limitations proved more informative than baseline industry classifications such as SIC or NAICS codes. In this section, we note that the baseline HP2016 approach can be improved simply by replacing the raw cosine similarities it employs with a doc2vec alternative.

The extension is simple to implement once the doc2vec models used in the current paper have been trained. Intuitively, in this context, we should view doc2vec simply as a dimensionality reduction algorithm. The raw text of 10-Ks used in HP2016 has a dimensionality of roughly 60,000, which is the number of unique terms in 10-Ks in a given year after dropping stop words. The cosines used in HP2016 are based on 60,000 element normalized vectors. When using doc2vec, we essentially do the calculation but we reduce the 60,000 dimensions to just 300. Hence each firm is represented by a 300 element vector describing its prod-

[16]The sales-weighted approach is used in Grullon, Larkin, and Michaely (2019).

ucts rather than a 60,000 element vector. Pairwise similarities between firms can thus be computed by taking the cosine similarities of these 300 element vectors instead of the raw 60,000 element vectors. Doing so should be more informative as doc2vec incorporate semantics and the relative position of words within a document, whereas raw cosines are simply a bag-of-words method that does not take positioning or semantics into consideration.

Other than using cosines of reduced-dimension 300 element vectors instead of the raw 60,000 element vectors, the methodology in HP2016 is unchanged. In particular, we can develop a D2V-TNIC-3 industry classification that mirrors the standard TNIC-300 classification by mirroring its approach to granularity. In the case of TNIC-3, the classification is calibrated to match 3-digit SIC codes in granularity so the same would be the case for D2V-TNIC-3 industries. In particular, we would deem the highest 2% of pairwise similarities in the base year 1997 to be industry peers, and we would then impose the 1997 threshold to all other years to firm the time-varying D2V-TNIC-3 classification. We also note that HP2016 also provide a FIC-based classification that has the binary-membership and transitivity properties of SIC-industries. In the context of D2V-based industries, the most similar analog to FIC industries would be to use the D2V-segment database discussed in the main paper, and simply assign each firm to the one industry that it is most textually similar to.

Overall basic profitability prediction regressions akin to those used by HP2016 indicate that baseline TNIC-3 generates roughly 19% adjusted R^2 relative to SIC-3 fixed effects, a 46% improvement. We find that D2V-TNIC-3 generates 24% R^2 in this same test, an additional 26% improvement over TNIC-3. This confirms the expected finding that doc2vec appears to be substantially more informative than baseline cosine similarities.

4. Additional NAICS-Manual Stop Words

Following HP2016, we first omit stop words defined as those appearing in more than 25% of all NAICS industry groups. In addition, we also manually reviewed the common words in the NAICS manual and identified a list of additional stop words that we omit. This is list presented below:

THE, AND, OF, COMPANIES, IN, SERVICES, CLASSIFIED, OR, INCLUDES, EXCLUDES, PRODUCTS, PRIMARILY, INCLUDING, NOT, PROVIDING, DIVERSIFIED, OTHER, THAT, TO, ENGAGED, GAS, MANAGEMENT, OPERATORS, RELATED, OWNERS, A, PRODUCERS, CONSUMER, ELSEWHERE, PROVIDERS, ALSO, FOR, COMPONENTS, DEVELOPMENT, PRODUCTION, AS, BUT, CENTERS, WITH, ARE, PRODUCING, LARGE, NON, OPERATING, OPERATIONS, USING, FROM, IT, MULTI,

EITHER, EMPLOYMENT, THREE, UNDER, WHOSE, ACTIVITY, CORPORATE, DO, END, HELD, HIGH, MORE, WHICH, THEIR, WIDE, ACROSS, ASSETS, AT, OPERATE, INDUSTRY, MANUFACTURING, ESTABLISHMENTS, THIS, COMPRISES, EXCEPT, CROSS, REFERENCES, MERCHANT, SUCH, GROUP, EXAMPLES, ALL, PRODUCT, ILLUSTRATIVE, THESE, ACTIVITIES, MAY, NEW, PURCHASED, TYPE, MADE, SUPPORT, SECTOR, ONE, SUBSECTOR, WITHOUT, BASIS, INCLUDED, WORK, KNOWN, PROCESSING, PROVIDE, DIRECT, ORGANIZATIONS, PREPARATION, SELLING, GROWING, INTO, OTHERS, FOLLOWING, BUSINESS, COMBINATION, MISCELLANEOUS, SALE, INDUSTRIES, USE, MAKING, ORDER, PROGRAMS, THEY, BENEFICIATING, SIMILAR, STOCK, CONTRACT, BASED

Table IA1: Salesforce Scope vs Time

The table displays the D2V scope allocations of Salesforce in 2005 (Panel A) and 2017 (Panel B).

Year	Amount	Word List
		Panel A: Salesforce Scope Allocations in 2005
2005	0.074	functionality,enhancements,implementation,enterprise,introductions,evolving,optika,integrate,client,peoplesoft,
2005	0.064	cobol,mainframe,legacy,porting,client,unix,migrations,relational,minicomputer,toolset,
2005	0.062	client,clients,mainframe,payroll,dataworks,peoplesoft,outsourcing,billing,updates,invoices,
2005	0.058	online,website,internet,yahoo,websites,infoseek,chat,netscape,excite,lycos,
2005	0.057	siebel,functionality,microsoft,netscape,groupware,implementation,server,symantec,desktop,wirelessly,
2005	0.055	modeling,aspentech,ansys,functionality,graphical,tools,baan,usability,dataworks,implementation,
2005	0.053	harbinger,commerce,premenos,translation,edifact,templar,internet,geis,extranets,intranets,
2005	0.048	databases,database,append,metromail,speh,lists,mailing,smartbase,acxiom,list,
2005	0.044	dataworks,functionality,hardware,nondisclosure,solutions,knowledge,misappropriati,sophisticated,aided,intergraph,
2005	0.042	isps,internet,netcom,hosting,uunet,psinet,compuserve,earthlink,online,prodigy,
2005	0.040	text,searching,download,databases,search,online,searchable,verity,searches,retrieval,
2005	0.039	internet,online,viewers,interactivity,chat,viewer,content,video,download,supersites,
2005	0.037	engagements,engagement,client,clients,consulting,keane,reengineering,skills,methodologies,outsourcing,
2005	0.033	ticket,tickets,ticketing,keyboard,scanning,barcode,scanners,scanner,intermec,printer,
2005	0.032	document,microfilm,microfiche,retrieval,documents,microfilmed,workflow,micrographic,archiving,micrographics,
2005	0.032	simulations,whiteboard,user,equations,cursor,learn,gamepads,tutorials,solver,touchscreen,
2005	0.031	encryption,authentication,cryptographic,cryptography,encrypted,firewalls,encrypt,tokens,trusted,firewall,
2005	0.030	illuminet,transxpress,lidbs,fastlink,billing,call,interexchange,switching,distance,dial,
2005	0.030	unisys,mainframe,compaq,multivendor,hardware,networking,midrange,microsystems,novell,unix,
2005	0.030	forrester,questionnaires,panelists,respondent,clients,influencers,cognizant,interviewers,insights,cati,
2005	0.029	corel,macromedia,autodesk,flowcharter,publishers,micrografx,corelflow,visio,broderbund,publisher,
		Panel B: Salesforce Scope Allocations in 2017
2017	0.074	cobol,mainframe,legacy,porting,client,unix,migrations,relational,minicomputer,toolset,
2017	0.062	harbinger,commerce,premenos,translation,edifact,templar,internet,geis,extranets,intranets,
2017	0.057	modeling,aspentech,ansys,functionality,graphical,tools,baan,usability,dataworks,implementation,
2017	0.056	functionality,enhancements,implementation,enterprise,introductions,evolving,optika,integrate,client,peoplesoft,
2017	0.049	document,microfilm,microfiche,retrieval,documents,microfilmed,workflow,micrographic,archiving,micrographics,
2017	0.048	databases,database,append,metromail,speh,lists,mailing,smartbase,acxiom,list,
2017	0.047	client,clients,mainframe,payroll,dataworks,peoplesoft,outsourcing,billing,updates,invoices,
2017	0.047	text,searching,download,databases,search,online,searchable,verity,searches,retrieval,
2017	0.044	engagements,engagement,client,clients,consulting,keane,reengineering,skills,methodologies,outsourcing,
2017	0.043	online,website,internet,yahoo,websites,infoseek,chat,netscape,excite,lycos,
2017	0.043	dataworks,functionality,hardware,nondisclosure,solutions,knowledge,misappropriati,sophisticated,aided,intergraph,
2017	0.042	siebel,functionality,microsoft,netscape,groupware,implementation,server,symantec,desktop,wirelessly,
2017	0.040	teletrak,helm,split,effected,inactive,ceased,transferred,definitive,surviving,executed,
2017	0.035	encryption,authentication,cryptographic,cryptography,encrypted,firewalls,encrypt,tokens,trusted,firewall,
2017	0.031	internet,online,viewers,interactivity,chat,viewer,content,video,download,supersites,
2017	0.029	microcomputer,resellers,microcomputers,peripherals,reseller,networking,multivendor,gtsi,merisel,hayes,
2017	0.028	corel,macromedia,autodesk,flowcharter,publishers,micrografx,corelflow,visio,broderbund,publisher,
2017	0.028	encryption,cryptographic,authentication,cryptography,firewall,firewalls,encrypted,trusted,encrypt,untrusted,

Table IA2: General Dynamics Scope

The table displays the D2V scope allocations of General Dynamics over time.

Year	Amount	Word List
		Panel A: General Dynamics Early Scope in 1989
1989	0.143	defense,navy,tracor,weapons,missile,military,warfare,logicon,tactical,nasa,
1989	0.120	defense,military,missiles,missile,warfare,aerospace,navy,radar,subcontracts,spacecraft,
1989	0.116	initiators,pyrotechnic,airbag,initiator,military,ruggedized,defense,pyrotechnics,airbags,missiles,
1989	0.102	reconnaissance,defense,military,warfare,subcontracts,ruggedized,assemblies,procurements,tracor,radar,
1989	0.082	radar,sensing,sensors,military,weapons,reconnaissance,battlefield,radars,explosive,ruggedized,
1989	0.079	precast,paving,concrete,prestressed,runways,gravel,aggregates,roads,highway,excavation,
1989	0.072	radar,dgps,navigation,radars,jamming,military,avionics,rangefinder,missiles,receivers,
1989	0.064	ammunition,firearms,remington,handgun,pistols,incendiary,artillery,guns,armor,hunting,
1989	0.059	hawk,vtol,helicopter,helicopters,takeoff,ustman,fuselage,iiis,wing,tightness,
1989	0.056	overhaul,aircraft,airframe,redistributors,overhauled,rotable,airbus,overhauls,spare,airframes,
1989	0.046	cement,masonry,concrete,roofing,quarries,siding,quarry,kilns,quicklime,limestone,
1989	0.041	asphalt,shale,limestone,asphalts,pulverized,lime,paving,quicklime,kilns,naphthas,
1989	0.038	flag,cruise,ships,cruises,drydock,voyage,crewed,tankers,aboard,maritime,
1989	0.038	flights,airline,airlines,fares,airways,airtran,flown,flight,passengers,designator,
		Panel B: General Dynamics Scope Allocations in 1998
1999	0.091	defense,military,missiles,missile,warfare,aerospace,navy,radar,subcontracts,spacecraft,
1999	0.086	defense,navy,tracor,weapons,missile,military,warfare,logicon,tactical,nasa,
1999	0.060	overhaul,aircraft,airframe,redistributors,overhauled,rotable,airbus,overhauls,spare,airframes,
1999	0.052	reconnaissance,defense,military,warfare,subcontracts,ruggedized,assemblies,procurements,tracor,radar,
1999	0.050	initiators,pyrotechnic,airbag,initiator,military,ruggedized,defense,pyrotechnics,airbags,missiles,
1999	0.047	hawk,vtol,helicopter,helicopters,takeoff,ustman,fuselage,iiis,wing,tightness,
1999	0.046	flag,cruise,ships,cruises,drydock,voyage,crewed,tankers,aboard,maritime,
1999	0.038	precast,paving,concrete,prestressed,runways,gravel,aggregates,roads,highway,excavation,
1999	0.035	propane,armored,bobtail,tanks,tank,safes,refilling,noncyclical,loomis,npga,
1999	0.031	radar,sensing,sensors,military,weapons,reconnaissance,battlefield,radars,explosive,ruggedized,
1999	0.030	radar,dgps,navigation,radars,jamming,military,avionics,rangefinder,missiles,receivers,
1999	0.029	outboard,inboard,watercraft,boats,boat,powerboat,boating,snowmobiles,yachts,fishermen,
		Panel C: General Dynamics Scope Allocations in 2017
2017	0.211	defense,navy,tracor,weapons,missile,military,warfare,logicon,tactical,nasa,
2017	0.177	defense,military,missiles,missile,warfare,aerospace,navy,radar,subcontracts,spacecraft,
2017	0.161	ammunition,firearms,remington,handgun,pistols,incendiary,artillery,guns,armor,hunting,
2017	0.147	initiators,pyrotechnic,airbag,initiator,military,ruggedized,defense,pyrotechnics,airbags,missiles,
2017	0.118	reconnaissance,defense,military,warfare,subcontracts,ruggedized,assemblies,procurements,tracor,radar,
2017	0.100	radar,sensing,sensors,military,weapons,reconnaissance,battlefield,radars,explosive,ruggedized,
2017	0.068	radar,dgps,navigation,radars,jamming,military,avionics,rangefinder,missiles,receivers,
2017	0.066	encryption,cryptographic,authentication,cryptography,firewall,firewalls,encrypted,trusted,encrypt,untrusted,
2017	0.065	hawk,vtol,helicopter,helicopters,takeoff,ustman,fuselage,iiis,wing,tightness,
2017	0.061	flights,airline,airlines,fares,airways,airtran,flown,flight,passengers,designator,
2017	0.056	overhaul,aircraft,airframe,redistributors,overhauled,rotable,airbus,overhauls,spare,airframes,
2017	0.054	propane,armored,bobtail,tanks,tank,safes,refilling,noncyclical,loomis,npga,
2017	0.042	flag,cruise,ships,cruises,drydock,voyage,crewed,tankers,aboard,maritime,
2017	0.039	propane,heating,npga,bobtail,gallons,noncyclical,liquefied,brooding,fuels,gasoline,
2017	0.034	encryption,authentication,cryptographic,cryptography,encrypted,firewalls,encrypt,tokens,trusted,firewall,
2017	0.029	isps,internet,netcom,hosting,uunet,psinet,compuserve,earthlink,online,prodigy,
2017	0.028	ticket,tickets,ticketing,keyboard,scanning,barcode,scanners,scanner,intermec,printer,
2017	0.028	avis,thrifty,rental,hertz,rentals,alamo,renters,reservations,reservation,travel,

Table IA3: Relative Segment Counts vs Product Market Distance Over Time

The table reports the average number of D2V segment pairs each firm has in each year versus how far the segment pairs are in product market space. We thus take all permutations of the 300 D2V industries and compute pairwise industry similarities for each industry based on the pairwise similarities of the k-means centroids from which each industry was estimated. We then sort all industry pairs into quintiles based on how spatially distant they are from each other (product market distance). We label those industry-pairs in the most similar quintile as "most similar" segments and those in the second most similar quintile as "weakly similar segments". We group the least similar three quintiles into a single group of "likely unrelated" industry pairs as these last three quintiles uniformly have very low similarity and also have far fewer operating pairs as noted in Table 6. We then tabulate how many pairs are operating in each similarity bin for each firm, where counts are weighted such that each segment gets a total weight of unity. In each year, we then average all segments counts in each quintile bin across all firms in each year to generate an average number of segments firms have in each bin in each year. As our goal is to illustrate how segment counts are growing or declining in each bin, we then normalize each by the first year of this analysis 1990 so that each reported figure indicates total growth relative to the base year. For example, the 1.47 in 2017 in the first quintile indicates that near-segments grew by 47% during our sample period.

Year	Most Similar Segments	Weakly Similar Segments	Likely Unrelated Segments
1990	1.000	1.000	1.000
1991	1.027	1.025	1.015
1992	1.030	1.002	0.979
1993	1.017	0.987	0.980
1994	1.024	1.027	0.998
1995	0.992	0.970	0.954
1996	0.981	0.901	0.878
1997	0.967	0.897	0.915
1998	0.997	0.875	0.836
1999	1.024	0.879	0.835
2000	1.074	0.913	0.870
2001	1.116	0.924	0.888
2002	1.157	0.945	0.891
2003	1.200	0.998	0.905
2004	1.251	1.042	0.936
2005	1.251	1.033	0.910
2006	1.265	1.036	0.916
2007	1.297	1.068	0.941
2008	1.328	1.119	0.987
2009	1.348	1.155	0.986
2010	1.372	1.185	1.025
2011	1.384	1.209	1.040
2012	1.417	1.257	1.075
2013	1.437	1.258	1.062
2014	1.441	1.282	1.070
2015	1.445	1.274	1.077
2016	1.441	1.279	1.079
2017	1.470	1.287	1.112

Table IA4: Investment Regressions (by Fama-French-5 Major Sectors)

The table reports the second stage results of 2-stage instrumental variable regressions where the dependent variable is a firm investment policy such as acquisitions, divestitures (target of an acquisition), R&D/assets or CAPX/assets. These regressions use the same specification as in Table 12 except we now run them using 5 industry subsamples based on the Fama-French-5 industry groupings (see Panel headers). Our instrumented variable of interest is a measure of scope (D2V-Scope). The first-stage regressions are displayed in Table 11 and include two instruments for scope (explained in detail in Table 11). The first is a measure of the extent to which the broader product market surrounding a focal firm is characterized by a high degree of outward-directed asset redeployability indicating a low cost to scope expansion by existing firms. The second is a measure of the size of the focal firm's outward-expansion opportunity set. All regressions include firm and year fixed effects, and *t*-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	Scope Variable	Log Assets	Log Age	# Obs
		Panel A: D2V-Scope (Tech-Industry Subsample)			
(1)	Acquirer Dummy	0.004	0.041	–0.059	26,697
		(0.460)	(4.630)	(–2.870)	
(2)	Target Dummy	–0.005	0.028	0.061	26,697
		(–0.880)	(4.260)	(4.370)	
(3)	R&D/Assets	0.003	–0.023	0.021	26,697
		(2.460)	(–11.900)	(4.970)	
(4)	CAPX/Assets	0.000	–0.001	–0.007	26,697
		(0.330)	(–1.090)	(–3.040)	
		Panel B: D2V-Scope (Manufacturing-Industry Subsample)			
(5)	Acquirer Dummy	0.029	–0.050	–0.076	21,060
		(2.640)	(–2.850)	(–3.740)	
(6)	Target Dummy	–0.014	0.074	0.045	21,060
		(–1.720)	(5.670)	(2.920)	
(7)	R&D/Assets	0.000	–0.003	0.005	21,060
		(–0.260)	(–2.400)	(3.200)	
(8)	CAPX/Assets	–0.002	0.001	–0.009	21,060
		(–1.390)	(0.740)	(–3.070)	
		Panel C: D2V-Scope (Consumer-Industry Subsample)			
(9)	Acquirer Dummy	0.027	–0.018	–0.034	22,227
		(1.740)	(–1.010)	(–1.770)	
(10)	Target Dummy	–0.012	0.039	0.025	22,227
		(–0.870)	(2.800)	(1.880)	
(11)	R&D/Assets	0.002	–0.003	0.000	22,227
		(1.710)	(–2.320)	(0.000)	
(12)	CAPX/Assets	0.001	–0.006	–0.013	22,227
		(0.690)	(–2.950)	(–4.790)	
		Panel D: D2V-Scope (Health-Industry Subsample)			
(13)	Acquirer Dummy	0.013	0.012	–0.030	12,700
		(1.310)	(1.070)	(–1.090)	
(14)	Target Dummy	–0.032	0.045	–0.008	12,700
		(–3.400)	(4.010)	(–0.380)	
(15)	R&D/Assets	0.009	–0.034	0.008	12,700
		(2.390)	(–7.430)	(0.790)	
(16)	CAPX/Assets	0.001	–0.001	–0.007	12,700
		(0.720)	(–0.380)	(–2.020)	
		Panel E: D2V-Scope (Misc-Industry Subsample)			
(17)	Acquirer Dummy	0.048	–0.053	–0.067	15,521
		(2.410)	(–2.560)	(–2.320)	
(18)	Target Dummy	–0.003	0.031	0.065	15,521
		(–0.270)	(2.370)	(4.040)	
(19)	R&D/Assets	0.001	–0.003	0.001	15,521
		(1.150)	(–1.850)	(0.950)	
(20)	CAPX/Assets	–0.004	–0.022	0.00	
		(1.130)	(–1.860)	(–5.380)	

Table IA5: Investment Regressions (One-Stage Regressions Using Scope Incentive Variables)

The table reports one-stage results based on the two-stage regression results for investment variables in Table 12. The key difference in this table is that we include the two scope incentive variables directly as key RHS variables instead of using these two variables as instruments. The first scope incentive variable is a measure of the extent to which the broader product market surrounding a focal firm is characterized by a high degree of outward-directed asset redeployability indicating a low cost to scope expansion by existing firms. The second is a measure of the size of the focal firm's outward-expansion opportunity set. As always, all RHS variables are measurable as of year $t-1$ and all dependent variables are as of year t. We also include controls for size, age, market to book, and the TNIC HHI. All regressions include firm and year fixed effects, and t-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	Asset Redeployability Scope Incentive	Opportunity Set Scope Incentive	Log Assets	Log Age	Mkt/Book	TNIC HHI	# Obs
(1)	Acquirer Dummy	0.002	0.049	0.018	–0.051			99,514
		(0.110)	(3.450)	(5.690)	(–5.770)			
(2)	Target Dummy	–0.027	–0.021	0.029	0.046			99,514
		(–1.630)	(–1.960)	(12.500)	(7.250)			
(3)	R&D/Assets	0.014	0.001	–0.012	0.007			99,514
		(3.590)	(0.310)	(–15.030)	(4.280)			
(4)	CAPX/Assets	–0.001	0.000	–0.001	–0.013			99,514
		(–0.300)	(0.150)	(–2.840)	(–10.530)			
(5)	Vertical Integration	0.004	0.005	0.001	0.001			99,434
		(6.490)	(14.210)	(8.910)	(3.270)			
(6)	Outsourcing	0.005	0.050	0.040	0.021			16,759
		(0.110)	(1.890)	(3.880)	(0.650)			
(7)	Acquirer Dummy	–0.004	0.046	0.024	–0.036	0.019	0.012	98,945
		(–0.200)	(3.250)	(7.450)	(–4.100)	(16.530)	(1.450)	
(8)	Target Dummy	–0.026	–0.019	0.028	0.042	–0.005	0.004	98,945
		(–1.560)	(–1.790)	(11.700)	(6.570)	(–6.910)	(0.670)	
(9)	R&D/Assets	0.014	–0.001	–0.012	0.006	–0.001	–0.007	98,945
		(3.510)	(–0.330)	(–15.350)	(3.940)	(–2.730)	(–6.070)	
(10)	CAPX/Assets	–0.002	–0.001	0.000	–0.010	0.004	–0.002	98,945
		(–0.810)	(–0.470)	(–0.700)	(–8.110)	(20.020)	(–1.650)	
(11)	Vertical Integration	0.004	0.004	0.001	0.001	0.000	–0.002	98,874
		(6.400)	(13.500)	(7.840)	(3.600)	(0.180)	(–8.790)	
(12)	Outsourcing	0.003	0.051	0.041	0.016	0.002	0.005	16,698
		(0.070)	(1.930)	(3.800)	(0.480)	(0.430)	(0.320)	

Table IA6: Outcomes Regressions (One-Stage Regressions Using Scope Incentive Variables)

The table reports one-stage results based on the two-stage regression results for investment variables in Table 13. The key difference in this table is that we include the two scope incentive variables directly as key RHS variables instead of using these two variables as instruments. The first scope incentive variable is a measure of the extent to which the broader product market surrounding a focal firm is characterized by a high degree of outward-directed asset redeployability indicating a low cost to scope expansion by existing firms. The second is a measure of the size of the focal firm's outward-expansion opportunity set. As always, all RHS variables are measurable as of year $t-1$ and all dependent variables are as of year t. We also include controls for size, age, market to book, and the TNIC HHI. All regressions include firm and year fixed effects, and t-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	Asset Redeployability Scope Incentive	Opportunity Set Scope Incentive	Log Assets	Log Age	Mkt/Book	TNIC HHI	# Obs
(1)	Valuation (M/B)	0.252	0.157	–0.391	–0.381			98,947
		(3.100)	(3.350)	(–25.530)	(–10.940)			
(2)	Sales Growth	0.015	0.081	–0.070	–0.208			99,133
		(0.690)	(5.810)	(–19.030)	(–25.500)			
(3)	Asset Growth	0.037	0.104	–0.165	–0.076			99,511
		(1.820)	(8.540)	(–44.750)	(–10.090)			
(4)	OI/Assets	–0.007	0.000	0.008	0.005			99,310
		(–0.670)	(0.060)	(4.250)	(1.260)			
(5)	Valuation (M/B)	0.153	0.098	–0.300	–0.124	0.325	–0.007	98,714
		(2.550)	(2.770)	(–26.270)	(–4.720)	(34.970)	(–0.320)	
(6)	Sales Growth	0.007	0.076	–0.055	–0.169	0.050	0.020	98,576
		(0.330)	(5.580)	(–15.230)	(–20.930)	(26.700)	(2.300)	
(7)	Asset Growth	0.013	0.090	–0.144	–0.018	0.072	–0.008	98,945
		(0.680)	(7.830)	(–42.050)	(–2.500)	(39.710)	(–1.240)	
(8)	OI/Assets	–0.009	0.000	0.011	0.012	0.008	0.003	98,749
		(–0.860)	(–0.040)	(5.660)	(2.850)	(10.000)	(0.880)	

Table IA7: Financing Regressions (One-Stage Regressions Using Scope Incentive Variables)

The table reports one-stage results based on the two-stage regression results for investment variables in Table 15. The key difference in this table is that we include the two scope incentive variables directly as key RHS variables instead of using these two variables as instruments. The first scope incentive variable is a measure of the extent to which the broader product market surrounding a focal firm is characterized by a high degree of outward-directed asset redeployability indicating a low cost to scope expansion by existing firms. The second is a measure of the size of the focal firm's outward-expansion opportunity set. We also include controls for size, age, market to book, and the TNIC HHI. As always, all RHS variables are measurable as of year $t-1$ and all dependent variables are as of year t. All regressions include firm and year fixed effects, and t-statistics are clustered by firm and shown in parentheses.

Row	Dependent Variable	Asset Redeployability Scope Incentive	Opportunity Set Scope Incentive	Log Assets	Log Age	Mkt/Book	TNIC HHI	# Obs
(1)	Equity Issuance	0.016	0.018	–0.038	–0.018			99,514
		(2.670)	(5.660)	(–30.980)	(–7.660)			
(2)	Debt Issuance	–0.014	0.011	–0.010	0.008			99,514
		(–1.200)	(1.480)	(–5.480)	(1.720)			
(3)	Dividends/Assets	–0.003	–0.001	0.000	0.002			99,414
		(–2.410)	(–0.650)	(–0.410)	(4.440)			
(4)	Repurchases/Assets	–0.006	–0.003	0.004	0.008			92,086
		(–1.580)	(–1.060)	(8.940)	(6.330)			
(5)	Equity Issuance	0.009	0.015	–0.034	–0.006	0.015	–0.005	98,945
		(1.650)	(4.840)	(–29.580)	(–2.450)	(23.890)	(–2.430)	
(6)	Debt Issuance	–0.017	0.010	–0.010	0.009	0.002	–0.006	98,945
		(–1.410)	(1.310)	(–5.300)	(1.980)	(3.220)	(–1.430)	
(7)	Dividends/Assets	–0.004	0.000	0.000	0.003	0.001	0.001	98,848
		(–2.570)	(–0.500)	(0.670)	(5.260)	(7.510)	(2.500)	
(8)	Repurchases/Assets	–0.006	–0.003	0.005	0.009	0.002	0.003	91,555
		(–1.680)	(–0.910)	(10.570)	(7.170)	(6.950)	(2.620)	

Table IA8: Corporate Finance Regressions (Robustness to Adding Industry x Year Fixed Effects)

The table reports the results of second-stage 2-stage instrumental variable regressions where the dependent variable is a firm investment policy, an outcome variable, or a firm financing policy (as noted in the first column). This table is run using the same models as our baseline results in Tables 12, 13, and 15. However, in this table, we add industry x year fixed effects in addition to the firm and year fixed effects already included in the baseline.

Row	Dependent Variable	D2V-Scope	Log Assets	Log Age	# Obs
(1)	Acquirer Dummy	0.021	–0.004	–0.049	96,771
		(3.610)	(–0.560)	(–5.220)	
(2)	Target Dummy	–0.014	0.040	0.044	96,771
		(–3.110)	(7.960)	(6.690)	
(3)	R&D/Assets	0.003	–0.014	0.007	96,771
		(3.360)	(–12.790)	(4.170)	
(4)	CAPX/Assets	0.000	–0.001	–0.014	96,771
		(–0.150)	(–1.430)	(–10.800)	
(5)	Vertical Integration	0.002	–0.002	0.001	96,692
		(12.640)	(–7.760)	(3.070)	
(6)	Outsourcing	0.026	0.012	–0.003	16,173
		(1.890)	(0.600)	(–0.070)	
(7)	Valuation	0.103	–0.481	–0.384	96,204
		(5.140)	(–18.620)	(–10.190)	
(8)	Sales Growth	0.036	–0.106	–0.205	96,401
		(5.890)	(–14.040)	(–21.340)	
(9)	Asset Growth	0.049	–0.214	–0.070	96,768
		(8.540)	(–27.970)	(–6.760)	
(10)	OI/Assets	–0.001	0.010	0.005	96,577
		(–0.560)	(3.340)	(1.170)	
(11)	Equity Issuance	0.010	–0.049	–0.016	96,771
		(7.280)	(–25.010)	(–5.920)	
(12)	Debt Issuance	0.002	–0.012	0.007	96,771
		(0.790)	(–3.430)	(1.530)	
(13)	Dividends/Assets	–0.001	0.001	0.002	96,675
		(–2.060)	(1.640)	(4.150)	
(14)	Repurchases/Assets	–0.002	0.007	0.008	89,305
		(–1.660)	(4.670)	(6.140)	
(15)	VC Funding Similarity	1.820	–1.191	–0.415	96,752
		(15.800)	(–8.230)	(–1.720)	
(16)	Product Market Fluidity	0.819	–0.427	–0.648	95,715
		(14.690)	(–6.500)	(–6.100)	

Table IA9: Corporate Finance Regressions (Robustness to Dropping 50 Largest Firms in Each Year)

The table reports the results of second-stage 2-stage instrumental variable regressions where the dependent variable is a firm investment policy, an outcome variable, or a firm financing policy (as noted in the first column). This table is run using the same models as our baseline results in Tables 12, 13, and 15. However, in this table, we drop the 50 largest firms in each year to illustrate that our results are not driven by mega-firms. Our results also remain robust if we drop the 100 largest, 500 largest, or even 1000 largest firms in each year.

Row	Dependent Variable	D2V-Scope	Log Assets	Log Age	# Obs
(1)	Acquirer Dummy	0.021	–0.004	–0.049	96,771
		(3.610)	(–0.560)	(–5.220)	
(2)	Target Dummy	–0.014	0.040	0.044	96,771
		(–3.110)	(7.960)	(6.690)	
(3)	R&D/Assets	0.003	–0.014	0.007	96,771
		(3.360)	(–12.790)	(4.170)	
(4)	CAPX/Assets	0.000	–0.001	–0.014	96,771
		(–0.150)	(–1.430)	(–10.800)	
(5)	Vertical Integration	0.002	–0.002	0.001	96,692
		(12.640)	(–7.760)	(3.070)	
(6)	Outsourcing	0.026	0.012	–0.003	16,173
		(1.890)	(0.600)	(–0.070)	
(7)	Valuation	0.103	–0.481	–0.384	96,204
		(5.140)	(–18.620)	(–10.190)	
(8)	Sales Growth	0.036	–0.106	–0.205	96,401
		(5.890)	(–14.040)	(–21.340)	
(9)	Asset Growth	0.049	–0.214	–0.070	96,768
		(8.540)	(–27.970)	(–6.760)	
(10)	OI/Assets	–0.001	0.010	0.005	96,577
		(–0.560)	(3.340)	(1.170)	
(11)	Equity Issuance	0.010	–0.049	–0.016	96,771
		(7.280)	(–25.010)	(–5.920)	
(12)	Debt Issuance	0.002	–0.012	0.007	96,771
		(0.790)	(–3.430)	(1.530)	
(13)	Dividends/Assets	–0.001	0.001	0.002	96,675
		(–2.060)	(1.640)	(4.150)	
(14)	Repurchases/Assets	–0.002	0.007	0.008	89,305
		(–1.660)	(4.670)	(6.140)	
(15)	VC Funding Similarity	1.820	–1.191	–0.415	96,752
		(15.800)	(–8.230)	(–1.720)	
(16)	Product Market Fluidity	0.819	–0.427	–0.648	95,715
		(14.690)	(–6.500)	(–6.100)	

Table IA10: Corporate Finance Regressions (Robustness to using near-peers opportunity set instrument)

The table reports the results of second-stage 2-stage instrumental variable regressions where the dependent variable is a firm investment policy, an outcome variable, or a firm financing policy (as noted in the first column). This table is run using the same models as our baseline results in Tables 12, 13, and 15. However, in this table, we replace our baseline opportunity set instrument, which is based on the distribution of distant peers serving different NAICS codes, with a version of the same variable computed using close peers instead of distant peers. This version is potentially more endogenous but we report results for robustness purposes.

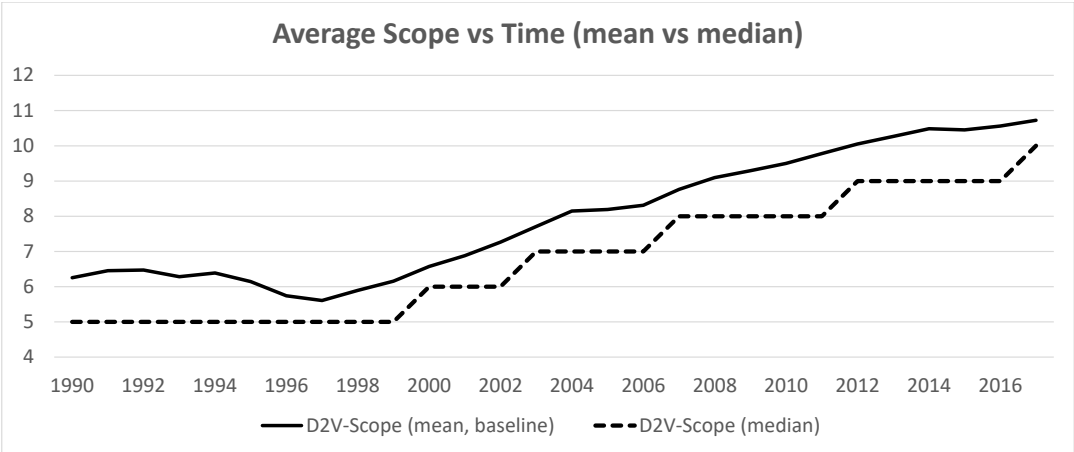
Row	Dependent Variable	D2V-Scope	Log Assets	Log Age	# Obs
(1)	Acquirer Dummy	0.016	0.002	–0.047	98,205
		(4.620)	(0.530)	(–5.260)	
(2)	Target Dummy	–0.007	0.036	0.045	98,205
		(–2.550)	(10.060)	(6.980)	
(3)	R&D/Assets	0.002	–0.013	0.007	98,205
		(3.700)	(–14.300)	(4.340)	
(4)	CAPX/Assets	0.002	–0.003	–0.013	98,205
		(4.220)	(–4.900)	(–9.890)	
(5)	Vertical Integration	0.002	–0.001	0.001	98,125
		(14.670)	(–6.550)	(3.610)	
(6)	Outsourcing	0.019	0.019	0.002	16,478
		(1.750)	(1.150)	(0.060)	
(7)	Valuation	0.049	–0.438	–0.376	97,634
		(4.140)	(–21.010)	(–10.620)	
(8)	Sales Growth	0.026	–0.096	–0.203	97,834
		(7.940)	(–18.030)	(–23.080)	
(9)	Asset Growth	0.036	–0.199	–0.069	98,202
		(11.020)	(–35.570)	(–7.730)	
(10)	OI/Assets	0.000	0.008	0.005	98,004
		(0.080)	(3.290)	(1.300)	
(11)	Equity Issuance	0.007	–0.045	–0.017	98,205
		(8.770)	(–28.690)	(–6.680)	
(12)	Debt Issuance	0.003	–0.013	0.009	98,205
		(1.710)	(–5.250)	(1.920)	
(13)	Dividends/Assets	–0.001	0.000	0.002	98,106
		(–2.610)	(1.770)	(4.310)	
(14)	Repurchases/Assets	–0.001	0.006	0.007	90,689
		(–2.690)	(7.670)	(6.290)	
(15)	VC Funding Similarity	1.252	–0.651	–0.491	98,186
		(22.530)	(–7.800)	(–2.900)	
(16)	Product Market Fluidity	0.936	–0.560	–0.572	97,100
		(24.170)	(–9.840)	(–4.760)	

Table IA11: Competition Complaints vs HHIs and Granularity

The table reports annual cross sectional OLS descriptive regressions where the dependent variable is the intensity of the firm's competition complaints in its 10-K, which is computed as the number of 10-K paragraphs that mention competition divided by the total number of paragraphs in the 10-K. The two RHS variables are measures of concentration (with the sign reversed so they can be interpreted as positive measures of competition) at different granularities. In particular include the Compustat SIC-3 HHI and the Compustat SIC-2 HHI. Each is computed as the sales-based concentration among firms in the given SIC code defined based on three digit and two digit SIC codes, respectively. Finally, we report the fraction of 2-digit granularity as the (1 - SIC-2 HHI) coefficient divided by the sum of the coefficients for both HHIs (truncated at zero in the first three years). This indicates the fraction of total HHI weights that are attached to the more coarse granularity. A high fraction indicates that, in the given year, the economy is such that competition takes place mostly at the 2-digit granularity rather than at the 3-digit granularity. A low value for this fraction indicates the converse.

Year	One minus SIC-2 HHI	One minus SIC-3 HHI	Adj R^2	Fraction 2-digit Granularity	# Obs.
1997	-0.001	0.011	0.014	0.000	5,521
	(-0.46)	(8.15)			
1998	-0.003	0.011	0.012	0.000	5,297
	(-1.34)	(7.63)			
1999	-0.003	0.012	0.017	0.000	5,076
	(-1.24)	(8.66)			
2000	0.006	0.010	0.022	0.369	4,827
	(2.46)	(7.90)			
2001	0.007	0.010	0.019	0.401	4,359
	(2.37)	(6.74)			
2002	0.009	0.005	0.010	0.642	3,954
	(3.17)	(3.30)			
2003	0.009	0.006	0.007	0.600	3,631
	(2.34)	(2.90)			
2004	0.011	0.008	0.028	0.577	3,540
	(4.44)	(6.02)			
2005	0.011	0.007	0.023	0.604	3,465
	(4.25)	(5.11)			
2006	0.007	0.006	0.019	0.565	3,378
	(3.34)	(4.91)			
2007	0.008	0.004	0.016	0.661	3,305
	(3.85)	(3.80)			
2008	0.009	0.004	0.023	0.674	3,120
	(4.54)	(4.29)			
2009	0.008	0.005	0.024	0.636	3,006
	(4.21)	(4.78)			
2010	0.010	0.004	0.024	0.744	2,899
	(5.20)	(3.72)			
2011	0.013	0.003	0.029	0.825	2,755
	(6.47)	(2.91)			
2012	0.005	0.003	0.009	0.655	2,665
	(2.51)	(2.92)			
2013	0.004	0.004	0.014	0.516	2,654
	(2.10)	(4.23)			
2014	0.004	0.003	0.011	0.538	2,695
	(1.98)	(3.43)			
2015	0.003	0.004	0.016	0.445	2,643
	(1.82)	(4.57)			
2016	0.003	0.004	0.015	0.484	2,546
	(1.93)	(4.21)			
2017	0.005	0.003	0.017	0.619	2,453
	(2.96)	(3.67)			

Figure IA1: The figures report robustness regarding how scope changes over time. Our baseline results are based on the average (mean) D2V-Scope over time. The first panel below compares our baseline based on the mean to the same calculation done using median D2V-scope in each year. The bottom figure on the bottom reports the average D2V-Scope for firms with above versus below median assets in each year.



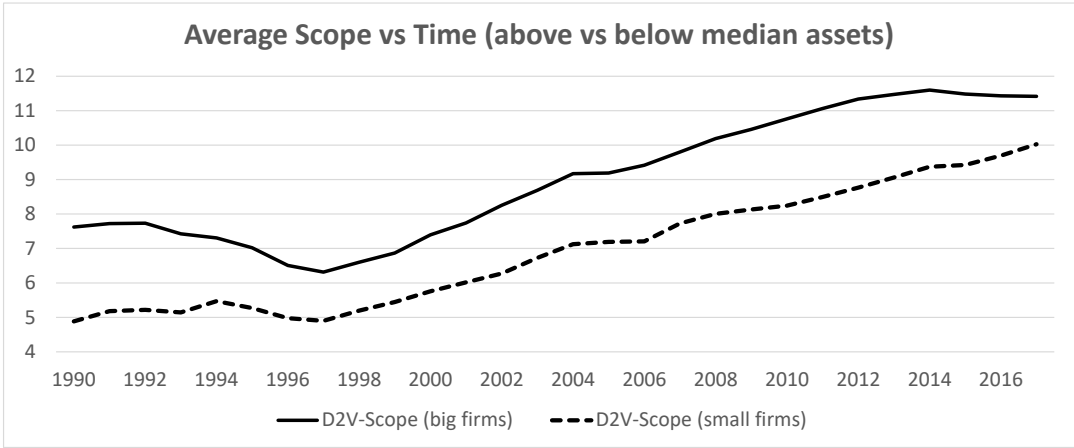


Figure IA2: The upper figure reports the average number of words in the firm's 10-K Item 1 business description divided by the number of industries (D2V-based or NAICS-based scope) the firm likely operates in. The goal is to measure the average degree of product variety within industries over time. The lower figure displays the average size of the 10-K Item 1 over time.