Scope, Scale and Competition: The 21st Century Firm

Gerard Hoberg and Gordon M. Phillips*

November 18, 2020

Abstract

We provide evidence that over the past 30 years, U.S. firms have expanded their scope of operations. Increases in scope and scale 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. We show that traditional concentration ratios do not capture this expansion of scope and are upward biased. After accounting for scope, we do not find evidence that industry concentration is increasing. 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 seminar participants. All errors are the authors' alone.

1 Introduction

The interplay between scope, scale and competition has been the focus of numerous authors including both 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]

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 (71%) during our sample period from 1989 to 2017. Firms have been able to increase scope without increasing the number of operating segments they report during the period of our sample. This motivates a new perspective on the 21st Century version of a multi-product firm that produces in different 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. Instead of multiple distinct segments, firms 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. Our results are consistent with multi-product firms having synergies across related products, and with unrelated diversification not being a major consideration.

An important question is how did this happen? To shed light on this question, we examine how plausibly exogenous variation in the incentives to increase scope has impacted investment decisions. We develop two instruments. The first measures the firm's scope-expansion opportunity set using the diversity of markets served by each firm's distant peers. When these peers have operations that are distributed across multiple well-defined industries,

[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) and Philippon and Gutierrez (2017).

it suggests that the focal firm likely has multiple scope-enhancing projects in its opportunity set that could be considered. Our second instrument considers the cost-side. We examine the asset portfolios of the firm's closer industry peers as compared to the firm's more distant industry peers. 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 new markets with lower adjustment costs.

Both instruments are measures of a focal firm's incentives to increase scope, and neither is measured using any data about the focal firm itself. Instead, both use the characteristics of more distant peers, whose characteristics are not easily controlled or influenced by the focal firm or vice-a-versa. We use these instruments to control for the endogeneity of scope but do not ascribe causality to the final results given limitations on the ability to find instruments that fully satisfy the exclusion restriction in our context. Although ideal instruments are not available, we follow prior studies in the network econometrics literature and focus on distant peers and not the firm itself to increase our ability to account for endogeneity.[3]

We find that firms expand scope through increased acquisitions, fewer divestitures, and increased spending on research and development. We find no evidence that these firms increase capital expenditures. These results are consistent with an ongoing process of asset redeployment across and within firms, which is reinforced by innovation that improves the flexibility and efficiency of these assets when they are used for multi-industry production. The use of innovation and flexibility can explain why we observe increases in scope but we do not find evidence of increased operating segments in the Compustat database.

We also find that increasing scope creates value as a business strategy, and thus it is unlikely that increases in scope are due to bad governance or private benefit extraction. This evidence of value creation is important given the possible empire building incentives managers would have to increase scope. We also find evidence of ex post sales growth and

[3]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 using the attributes of more distant peers can produce exogenous variation that can be used as instruments.

asset growth. However, we do not find any significant impact for profitability in the form of return on assets. Put together, 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 second. Our valuation results - which show that firm market-to-book equity ratios increase with our scope measures - are in contrast to the historical conglomerate literature (Berger and Ofek (1995) and Lang and Stulz (1994)) that finds that firm market-to-book valuation measures go down with an increased number of reported Compustat segments.

We also examine how firms finance scope expansion and find that they rely significantly on equity. They issue more shares and pay lower dividends. We find no significant link to debt financing. These results are consistent with intangibles and the redeployment of existing assets playing an important role in scope expansion. In particular, the method of expansion through intangibles and existing assets does not increase fixed collateral, and the nature of intangibles favors financing with equity.

We note that these conclusions do not use the traditional measure of scope, the number of Compustat segments. In particular, the average Compustat segment count of firms has not increased over time. A key issue with reported segment data is that, almost by definition, it does not facilitate measurement of the "new conglomerate", which emphasizes related expansion. The new conglomerate can span multiple product spaces while preserving a single segment organizational form. Additionally, SFAS 131, which governs segment reporting does not require reporting segments based on the number of industries served. We further examine the comparative validity of our measures and the Compustat segments by relating both to the frequency of explicit statements in firm 10-Ks that indicate broad product scope. Our new measures are highly significant in predicting these statements, whereas the number of Compustat segments is only borderline significant in these regressions, and becomes mostly insignificant when included in the same regression as our new text-based scope measures.

We conclude with an analysis of whether the increases in scope we document can explain why traditional measures of concentration are increasing, and whether or not this indicates an increase in monopoly power among U.S. firms. We first replicate this trend using traditional industry classifications and illustrate that their primary failure is an inability to dynamically adjust granularity when scope changes. During our 30-year sample, scope has increased so much that the level of market overlap of firms sharing the same 3-digit SIC code early in our sample is roughly the same as that of firms sharing the same 2-digit SIC code late in our sample. Hence scope has shifted the focal level of granularity by one full digit of the SIC code system. We adjust traditional HHIs for this scope-induced shift in granularity and find that concentration is not increasing during our sample.

We also assess the trend in concentration using the multi-industry assignments our scope measures generates as part of their construction. The resulting scope-adjusted HHIs again show that concentration is not increasing during our sample period. Crucially, adjusting concentration ratios using Compustat segments does not analogously explain the rise in concentration. As noted above, this is because firms increase their scope by producing related products without increasing their reported number of business segments.

These results suggest that an alternative narrative should be considered to explain the apparent rise in concentration ratios. In particular, the rise in concentration ratios is potentially driven by a rise in scope across related markets that is not easily observed in reported operating segments or traditional industry codes. Thus not adjusting concentration ratios for greater scope generates an upward bias in concentration ratios. An implication is that observed trends might not be problematic in terms of horizontal market power. For example, when purchasing a product in a given market, consumers likely have a similar number of choices in the early and later parts of our sample. It is also possible that any longer-term efficiency gains associated with scope might be passed onto consumers, improving welfare. However, a rise in scope could still be problematic in terms of market power as these firms might offer product bundles aimed at reducing competition, and increases in scale and scope could also lead to anti-competitive conduct along the supply chain. Overall, our results suggest that regulators should not ignore the changes in the market, but rather, they might look for anti-competitive practices in different aspects of market conduct. While we note that our findings are only suggestive regarding regulatory implications, they should be seen as a call to future researchers to further consider the impact of scope on specific markets or in the broader market at large.

2 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.

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 with 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, consistent with rational expansion. Moreover, value increases and sales grow even though return on accounting assets experiences some initial dilution.

Our paper also has further implications for the older literature documenting a diversification discount including highly visible works by Berger and Ofek (1995) and Lang and Stulz (1994). Although many recent studies call the diversification discount into question (see Custodio (2012) and Hund, Monk, and Tice (2020) for example), our paper takes a different approach and portrays an entirely new perspective on multi-industry firms and how they are growing in prevalence. The overall conclusion is that the modern multiple-industry firm may share scarce valuable resources, is able to serve multiple industries while maintaining an efficient single segment organizational form and has a high market valuation.

Our study also contributes to the broader literature on acquisition motives and innovative investment. Our focus is on new synergies due to related operations and redeployable assets. Other studies examining synergies and related-industry acquisitions include Hoberg and Phillips (2010a), Rhodes-Kropf and Robinson (2008), Bena and Li (2014), and Fresard, Hoberg, and Phillips (2020).

3 Data and Methods

3.1 Sample Selection and Panel Structure

Our sample begins with the universe of Compustat firm-years with available 10-K filings either 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 (see Hoberg and Phillips 2016) is based purely on EDGAR filings, its coverage begins in 1996. An important contribution of the current study is thus that we back-extend the TNIC database to 1988 using the 10-Ks obtained from the two aforementioned libraries. To remain in our sample, a firm must have an available 10-K filing both in the current year of observation, and also in the previous year. We exclude financial firms 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 100,525 firm-year observations from 1989 to 2017.

3.2 Novel Measures of Scope

We develop new measures of firm-specific product market scope, which are updated annually. We do so using the textual network spatial representation of the product market developed by Hoberg and Phillips (2010a) and Hoberg and Phillips (2016) (henceforth HP2016), which uses the text in each firm's business description filed with the U.S. Securities and Exchange Commission in each year. These descriptions are updated every year and are required by Regulation S-K to accurately represent the products sold by the firm in the given fiscal year of the 10-K that is filed. Prior research including Hoberg, Phillips, and Prabhala (2014) illustrates that these filings are updated in a highly dynamic way as firms evolve their product portfolios.

At a high level, measuring scope requires scoring firm i in year t regarding how many industry vocabularies (industries are indexed $j \in J$) are discussed in its specific product description. Following the approach in HP, we specify the product market space as the set of all product market words that are in the TNIC database.[4] If there are N words in the product space, then each firm is represented in this space by a vector $V_{i,t}$, which is an N dimensional vector containing a one for words that firm i uses in year t and a zero for words that firm i does not use. Following HP2016, we also set to zero the element for any stop words, which are those appearing in more than 25% of all Item 1's in year t.

Analogously, we generate vector representations for industries $j \in J$. Crucially, unlike firm vectors, we lock in each industry's vocabulary vector and do not allow industry vectors to vary with time. As we discuss later, this ensures that we can measure changes in scope in its purest form, and without measuring the separate and potentially confounding matter of product market innovation, which can include the creation of entirely new markets.[5] We note that studying the creation of new markets is separately interesting, although it is not

[4]This is the set of all words that appear in the union of all firm 10-K Item 1's.

[5]Although test clarity indicates that it is philosophically important to lock in industry definitions, we note that our results are very similar if we instead re-define industries every year (we run this alternative specification using dynamically recomputed FIC-300 industries).

the focus of the current study. We also focus on a fixed set industry vocabularies because it is conservative, as it favors under-measuring the true increase in firm scope that has occurred due to new industries.

We consider two methods for identifying the spatial location of industries. The first draws industry vocabularies from the "Fixed Industry Classification" technology developed in HP2016, which uses a clustering algorithm applied on the sample of single segment firms to identify product market clusters. This approach develops vocabularies for each industry, which are computed as the average of the vectors $V_{i,t}$ after normalization based on the firms assigned to each cluster. We specifically use the FIC-300 classification developed by HP2016, which is based on 300 industries, and we use the FIC-300 classification from 1997, which is the base year used by HP2016. We denote the resulting industry vocabulary vectors by $D_{FIC,j}$, where $j \in 1, ..., 300$. The vocabularies are obtained directly from HP2016 without modification. Each vector is then normalized to sum to unity and resides in the same N-dimensional space as do the firm-vectors $V_{i,t}$.

Our second method is based on NAICS industry definitions, as provided in the 2017 version of the NAICS Manual, which is a 963 page document providing extensive and detailed descriptions of each NAICS industry. We use the four-digit NAICS granularity, and group all vocabulary for all industries having the same four digits of their NAICS code into each NAICS' code's total vocabulary that we use to determine the spatial location of each 4-digit NAICS code in our N-dimensional space. Our approach has natural stop-wording built in, as we will ignore any words that are not part of the stop-word-adjusted TNIC vocabulary in each year, as explained above. As an additional precaution, we reviewed the list of common words that appear in the NAICS manual and manually identified a list of stop words (see Appendix) that we additionally remove from the NAICS dialects. We thus create one vector $D_{NAICS,j}$ for each four digit NAICS industry from the 2017 manual, where $j \in 1, ..., 311$, as there are 311 four digit NAICS that we capture using this approach. Each element of this vector is populated with the number of times each word is used in the NAICS manual for the given industry, and the vector is normalized to sum to unity. Each vector resides in the same N-dimensional space as do the firm-vectors $V_{i,t}$.

Now that we have spatial locations for each firm-year and for each industry, we score each firm-year based on how much of each industry's vocabulary it uses. Crucially, we avoid 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 of such firms would become small). Instead, we compute the fraction of each industry's vocabulary that appears in each firm-year's product description as the following overlap ratio (we develop an analogous ratio and method for the NAICS-based scope measure):

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

Finally, to compute how many FIC industries a given firm might operate in, we identify a fixed threshold $\bar{Q_{FIC}}$ above which we deem a firm having $Q_{i,j,t,FIC} > \bar{Q}$ to be operating in industry j in the given year t (we call this an "operating pair"). Crucially, we hold $\bar{Q_{FIC}}$ fixed and do not allow it to vary with time, as otherwise we might create false inferences in our time series analysis. We thus use our base year of 1997 to compute $\bar{Q_{FIC}}$ as the threshold such that 2% of all firm-industry combinations are deemed to be operating-pairs in 1997. We choose the 2% threshold as that is the threshold used in HP2016 to determine the granularity of three-digit SIC industries. Our results are robust to using 1% or 5%. Our main variable FIC-scope is then simply the number of industries the given firm likely operates in, i.e., the number of industries with a similarity above the fixed threshold $\bar{Q_{FIC}}$ (NAICS-scope is computed analogously):

$$FIC - Scope_{i,t} = \sum_{j=1,...300} Indicator\{Q_{i,j,t,FIC} > \bar{Q_{FIC}}\} \tag{2}$$

$$NAICS - Scope_{i,t} = \sum_{j=1,...300} Indicator\{Q_{i,j,t,NAICS} > \bar{Q_{NAICS}}\} \tag{3}$$

We note that these measures are simply estimates of how many markets a given firm operates in based on its disclosed product market offerings. Hence these measures estimate how wide-ranging a firm's product scope is across well-defined and intuitive market definitions. To mitigate the impact of outliers, we winsorize our scope variables at the 1/99% level.

3.3 Local Asset Redeployability and Outward Scope Expansion

We develop two instrumentals for firm scope, which allow us to examine the extent to which plausibly exogenous incentives for firms to increase scope may lead firms to alter their corporate finance policies and to relate scope expansion to ex post outcomes. We use these instruments to control for endogeneity but do not ascribe causality to the final results. We thus view our results based on these instruments as suggestive, given some limitations remain and given the difficulties associated with further testing the exclusion restriction. Our first instrument is based on the redeployability of each firm's assets specifically to redeploy in product markets that are nearby in the product space. The intuition is that a firm that can easily redeploy assets in spatially proximate product markets likely has strong incentives to increase scope because the cost of doing so is likely to be low. In particular, the firm can use its existing asset base with low adjustment costs to expand in these related markets. We focus on this concept of local redeployability rather than broader measures of redeployability to increase the power of our instrument. This is motivated by empirical work by Hoberg and Phillips (2018) and the theory of Crémer, Garicano, and Prat (2007), which indicate that firms expand scope by operating in groups of highly related industries and not highly diversified (or distant) industries.

Our approach to measuring local redeployability follows Kim and Kung (2017) (KK2017), who use the capital flows tables from the Bureau of Economic Analysis to compute measures of broad asset redeployability. We extend and refine their methodology to focus on localized redeployability. In particular, the BEA tables indicate the extent to which each of 123 BEA industries (which can be mapped to NAICS codes) utilizes a set of 180 types of assets (and in what fraction). Intuitively, if two BEA industries utilize the 180 assets in very similar proportions, we would conclude that a firm operating in one of the two industries faces a high degree of asset redeployability should it choose to operate in the second. We view a lack of asset redeployability as a barrier to scope expansion, or analogously, the presence of redeployability across a pair of industries can be seen as an incentive to expand scope in that direction, as it would indicate that scope expansion can be achieved at a low cost. To the extent that the asset allocation vectors across industries are exogenous from a given firm's perspective, it would follow that a firm facing high levels of asset redeployability between its current industries and its nearby industries faces an exogenously higher incentive to increase scope as part of its forward-looking business strategy.

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 take additional precautions to further improve the extent to which these variables are more plausibly exogenous. 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 focal firm using an approach that examines the product market around the firm, and strictly avoids using any data from the focal firm itself. In particular, 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 directly from the 1997 capital flows table, which has

[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 calibrated to be as granular as SIC-3. hence firms that are in TNIC2 but not TNIC3 are "distant peers" as they are still operating in nearby markets but not directly in the focal firm's current market.

reported dollar amounts for 180 assets tracked by BEA for each industry. 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. These two industry distributions reflect likely paths that firms would take if they are outwardly expanding scope, as the close peers would indeed consider the types of industries that the distant peers are operating in as noted above. 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 joint distribution of industry pairs spanned by the two sets of peers:

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

Importantly, the more the close and distant peers operate in the same NAICS-4 industries, the lower is the outward asset redeployability as there is less scope for outward scope expansion when all local firms are all in the same limited set of markets. The more they operate in different NAICS-4 industries, the more the weighted average of the redeployability scores of these pairs will increase the computed Local Asset Redeployability from the focal firm's perspective. Additionally, we note that the 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 identification in network settings (see (Bramoulle, Djebbari, and Fortin, 2009) for example). Also see Cohen-Cole, Kirilenko, and Patacchini (2014) for a

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

related application in finance. Importantly, when local asset redeployability is high, it indicates that the focal firm likely has many ways to increase scope that would be feasible with relatively low adjustment costs given the focal firm's likely asset composition. It thus is a shifter of the focal firm's scope-expansion strategy.

3.4 The Local Scope-Expansion Opportunity Set

Whereas our above instrument is based on incentives rooted in low adjustment costs to expanding scope, our second instrument is based on a measure of the quality of the outward-scope expansion opportunity set as seen from the focal firm's perspective. As was the case for our first instrument, we construct our second instrument by focusing on more distant peers and we do not use the characteristics of the focal firm itself to construct the instrument.

As above, we first identify distant peers for a given focal firm i in a given year t by identifying the rivals that are in the focal firm's TNIC-2 industry but are not in the focal firm's TNIC-3 industry. As above, we compute the distribution of NAICS-4 industries served by these distant peers ($F_{i,t,j,distant}$). To compute the local scope expansion opportunity set, we simply compute one minus the concentration ratio (HHI) based on this distribution:

$$LocalScopeExpansionOpp.Set_{i,t} = 1 - \sum_{j \in NAICS-4} F^2_{i,t,j,distant} \tag{5}$$

When this variable is high, it indicates that nearby peers serve a wide array of closely related product markets. From the focal firm's perspective, the firm thus faces 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 more likely to be first-order 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.

3.5 R&D, Investment and Acquisitions

We examine four investment policies: R&D/assets, CAPX/assets, the decision to acquire assets, and dis-investment in the form of selling assets as a target. The R&D (XRD) and CAPX variables are from Compustat. We scale each by beginning of period total assets (AT). When R&D is missing, we assume it to be zero.[8] 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 any assets from any seller (public or private) in the given year and is zero otherwise. Analogously, 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. Both variables include transactions involving parts of firms or whole firms.

We also consider four other outcome variables including sales growth and asset growth, which are equal to the log of one plus the ratio of current sales to past-year sales and current assets to past-year assets), respectively. We compute firm a firm valuation ratio as its 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 dividends, with all four being scaled by assets.

All accounting ratio variables are winsorized within each year at the 1%/99% level. Please see the complete variable description in the Appendix for more detail.

3.6 Summary Statistics and Correlations

Table 1 displays summary statistics for our 1989 to 2017 panel of 100,525 firm-year observations. The average value of our key FIC-scope and NAICS-scope variables are 6.9 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 well defined FIC or NAICS-4 industries,

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

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.

[**Insert Table 1 Here**]

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 29% of our sample firm-years are involved in an acquisition. The average firm's valuation ratio (market to book) is roughly 1.76, and the average firm spends roughly 1.6% of its assets on repurchases and 0.8% on dividend payments.

Table 2 displays Pearson correlation coefficients for our key variables and illustrates sensible relationships to key existing variables. For example, our two scope variables (FIC-scope and NAICS-scope) are about 20% correlated with the number of Compustat operating segments. This correlation is significant and positive as expected, but it is also far from unity illustrating why evaluating scope using segment counts alone is likely to miss much important variation. Both measures are also roughly 27% correlated with firm size as measured using log assets. This is intuitive and illustrates that larger firms serve a wider array of product markets. This finding also illustrates why controlling for variables such as size is important for our later inferences regarding investment policies. For example, both scope measures are also positively correlated with both the acquisition dummy and the 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. Hence, it is no surprising that when we run formal regression analysis, we find that scope is associated with more acquisitions but less divestitures (targets), which conforms to the intuition that firms with high incentives to increase scope are indeed net acquirers once size is held fixed.

[Insert Table 2 Here]

We also find that our control variables are only modestly correlated with our variables of interest and with each other. Multi-collinearity is thus unlikely to be a concern.

4 Relation to Traditional Scope Variables

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.

4.1 Compustat Segments

The existing literature generally focuses on the Compustat segment tapes when exploring issues relating to scope and often takes the perspective that Conglomerate firms (those with more than one operating segment) have high levels of product scope but they are also diversified and tend to operate in relatively distant product markets. We take the perspective that the segment tapes are problematic for measuring scope, not only due to basic mismeasurement (see Villalonga 2004), but also because we expect modern firms to increase scope without increasing the actual number of rigid operating segments. In particular, modern firms are able to use innovation to increase product scope through more flexible production and by redeploying existing assets. A consequence is that scope is increasing but it would not be observed by plotting Compustat segments over time, which would be stable and not necessarily increasing.

We begin our analysis by computing basic summary statistics for subsamples of firms with different numbers of Compustat segments.

[Insert Table 3 Here]

Table 3 displays the results and shows that moving from one segment to two segments increases FIC-scope and NAICS-cope by roughly one unit. For example, FIC-scope increases from 6.42 to 7.53. The table also shows that adding segments beyond two roughly adds one more unit to our scope measures. These results conform to intuition, and suggest that each additional operating segment adds one additional product market to our measures of scope. However, these expected positive relationships greatly under-state the true variation in our measures of scope, which have standard deviations ranging from five to seven, and hence most of the variation in our measures cannot be explained by the number of Compustat segments a given firm operates in.

4.2 Scope Trends

Figure 1 plots the average number of Compustat segments over time in the upper panel, and the average FIC-Scope and NAICS-Scope in the lower panel. The number of segments initially declines steadily from roughly 1.5 in 1989 to roughly 1.3 by 1997. This trend 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. The jump around 1997 can be explained by SFAS 131, in which FASB changed segment disclosure requirements for filings associated with fiscal years ending after December 15, 1997. In particular, this rule required that managers must report segments based on how managers themselves internally evaluate operating performance. Prior to this rule change, segment reporting was instead based on an industry approach. Yet we note that the rule change itself was precipitated by concerns by market participants that segments were being under-reported, perhaps for strategic reasons. We refer readers to Song (2020) for a more detailed summary of these important events.

[Insert Figure 1 Here]

The events leading up to SFAS 131, and the consequences of this rule change itself, suggest that readers should interpret any trend for Compustat segments in Figure 1 with significant caution. It is unlikely to provide a clear indication of whether the scope of U.S. firms is increasing or not. For example, the alleged practice of under-reporting prior to the rule change calls the declining trend from 1989 to 1997 itself into question. After 1997, we observe a flat trend, but this too is questionable because segment counts are based on how managers internally evaluate performance and not how many product markets the firm actually operates in. For example, a major result we report later is that firms increasing scope choose to operate in industries that are closely related, not industries that are diversified. Intuitively, many managers might prefer to internally assess the performance of such related industry product lines together, and not separately. If so, the number of Compustat segments could strongly understate the increase in scope, as it is likely to only capture instances where firms operate in relatively unrelated product markets, where separate evaluation is more warranted.

The lower panel of Figure 1 displays the average FIC-scope and NAICS-scope over our sample. The coverage dating back to 1989 was made possible by the backward extension of the TNIC database to the late 1980s, which is also a contribution of the current study. The figure illustrates that scope was increasing during our entire sample, with the most rapid rate of increase appearing between 1997 and 2013. During this time, the average scope of firms in our sample increased by a full 50%. Notably, the upper panel shows that the number of Compustat segments did not change during this period. These results are consistent with the view that product market scope did increase materially, but the increases were mainly driven by firms serving multiple industries that are related (for example selling computers and cell phones) rather than diversified and unrelated (such as selling oil and cat food). Put together, our results are consistent with managers evaluating these highly related industry markets together, and hence we observe very few Compustat segments. Our evidence throughout the paper supports this view.

A final note on scope trends is that the increased scale and scope we report can also help to explain why the length of 10-Ks has also increased over time. The upper panel of Figure 2 plots the average number of words in the 10-K Item 1 over time, which increased in the first half of our sample rather steadily until 2003. After this year, the size of the Item 1 has become relatively stable. This suggests that document size is related scope, but also different, as Figure 1 shows that scope has been increasing throughout our sample, even after 2003.

[**Insert Figure 2 Here**]

To further understand the evolution of Item 1 over time, we note that there are 3 sources of variation that might drive its length: (1) increased scope across product markets, (2) increased product variety within markets, and (3) increased boilerplate content. Regarding boilerplate content, our removal of stop words, or any word that appears in more than 25% of all filings in a given year, should greatly reduce the impact of boilerplate content on our measures. Regarding increased scope, as noted above, its time series is different from the trend in document size overall, which suggests that a shift in product variety might also have occurred in parts of our sample. The lower panel of Figure 2 supports this intuition, and reports the average time trend of the number of 10-K Item 1 words per product market the firm likely operates in. This is computed as the number of words in the given firm's Item 1 divided by FIC-scope (number of likely product markets). The figure suggests that within-market product variety has likely increased during the earlier part of our sample and thus can explain some of the increase in 10-K size. These findings motivate future research on this topic, as our study focuses on changes in scope across product markets and we do not examine within-market product variety further.

4.3 Scale and Scope

We now examine the role of increasing firm scale and its relation to firm scope. Figure 3 plots average firm size over time (based on book assets) both in nominal terms and in inflation-adjusted terms. The figure shows that average firm size has increased substantially over time by both metrics. Using the conservative inflation adjusted metric, firm size has roughly tripled during our sample. This increase likely reflects the increases in firm scope we document above, but also increases in firm scale. We thus assess the relationship between scale and scope using sorts.

[Insert Figure 3 Here]

Table 4 reports average scale and scope statistics for size quintiles. Quintiles are formed by sorting on Compustat assets separately in each year. We report these statistics separately for the full sample and for firms that report just one Compustat operating segment.

[Insert Table 4 Here]

The table confirms that all measures of scope sort strongly with firm size. Even for Compustat segments, the smallest quintile firms have an average of 1.22 segments, which grows to 1.94 for the largest firms. Regarding FIC-scope, the interquartile range is from 5.65 product markets to 8.62 product markets. The range is larger for NAICS-scope at 4.02 to 9.17 markets. Yet the growth in scope by any measure across these quintiles pales in comparison to the range of firm size itself. Small quintile firms have about $23 million in assets, whereas the largest quintile firms average $11.7 billion. We conclude that some of the variation we see in firm size and its increase over time, likely is related to corresponding increases in firm scope. However, the sheer magnitude of the scale increase also suggests that the increases in firm size likely have other drivers beyond firm scope. In particular, it is likely that U.S. firms have achieved not only economies of scope, but also economies of scale.

We conclude this discussion with a note about single segment firms. The rightmost columns in Table 4 report the same statistics for the subset of single segment firms. Importantly, we see only modest reductions in the differences across the quintiles for this subsample. This reinforces further our conclusion that segments are not a reliable source of information about firm scope. As we show at the end of this section, basic validation tests using orthogonal queries further reinforce that segment counts contain only a weak signal regarding scope, and our new measures of scope are much more strongly validated.

5 Scope but not Diversification

In this section, we 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. Although older studies generally portray conglomerates as the former, recent studies suggest that modern firms are more like the latter (see theory by Crémer, Garicano, and Prat 2007 and empirical support in Hoberg and Phillips 2018). To examine this issue, we first compute the average product market distance between every permutation of pairs of FIC-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. We thus observe which pairs of industries are proximate and which are distant in the product space.

Next, we consider the firm-to-industry mapping created when computing the FIC-score. As we discussed in Section 3, this calculation first requires us to identify the set of FIC-300 industries that each firm likely operates in based on the given firm's coverage of the given industry's vocabulary in its 10-K item 1. Finally, we use firm configurations across industries to create a database of observed "operating pairs". A firm that maps to industry i and industry j is thus an observation of the operating pair ij. A firm that maps to three industries {i,j,k} is an observation of three operating pairs: {ij,ik,jk}. A firm that maps to just one industry does not have any observed operating pairs. We then tabulate the total number of operating pairs for each pair of industries in each year to compute the distribution of operating pairs for each year.

[**Insert Table 5 Here**]

In each year, we sort industry pairs into deciles based on the TNIC similarity of the pair. Table 5 then reports the fraction of all observed operating pairs that are in each 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. For the full sample and for the single and multi-segment firms individually, we find that almost 40% of all operating pairs are in the highest decile of TNIC industry pairwise similarity. An additional 13% to 15% are in the next decile. These results indicate that modern multi-industry firms are not the diversified conglomerates portrayed in the early corporate finance literature. The modern multi-industry firm operates in highly related industries, which are likely feasible to span without a complex conglomerate governance structure. Following results in Hoberg and Phillips (2010a) and Bena and Li (2014) for example, these related industries likely generate both product and operational synergies for these multi-industry firms, suggesting that this practice is likely value creating. Indeed we report later that scope-enhancement appears to be rewarded with higher valuations and sales growth.

5.1 Validation of Scope Measures

The results presented above provide support for the proposition that text-based measures of scope, which are backed by Regulation S-K which requires disclosure of the products sold by the firm in the given fiscal year, should have many advantages over using Compustat segments. Segment reporting, in contrast, is governed by SFAS 131, which is linked to internal performance evaluation rather than industry organization. In this section, we augment this evidence with a formal validation test that uses direct firm statements to evaluate candidate measures of scope.

We consider four queries of firm 10-Ks to identify direct statements indicating the firm offers a selection of products that has a high degree of scope. These queries are based on the following 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 program to compute four variables of interest to validation. Our first "Product Breadth" is the number of paragraphs in each firm's overall 10-K that mentions a phrase in List A, scaled by the total number of paragraphs in the firm's 10-K. The variable "Product/Svc Breath" is analogously defined based on List B. Our second two variables are more specific. Our third variable "Product Breadth Detail" is the number of paragraphs that contain a phrase in List A and also a word from List C. Our final variable "Product/Svc Breadth Detail" is analogously defined using List B and List C. Intuitively, when these scores are higher, it is likely that the firm offers a wide-scope array of products and services. We also note that the latter two variables are quite stringent and are based on proximity searches. This anchor-phrase approach is more rigorous than are basic unconditional word counts as it would be less informative if words from List C appeared in different paragraphs than those where Lists A and B appear.

To validate our measures of scope and compare them to Compustat segments, we regress the four variables above on our scope measures. We additionally include controls for size, age, market to book, and the TNIC HHI, and we also include firm and year fixed effects in all specifications. For validation, a measure of scope should have a strong positive and significant coefficient.

[Insert Table 5 Here]

Table 5 displays the results. The first four rows only include the controls as a baseline, and illustrate for example that size (log assets) is not surprisingly related to our direct statements of scope, and the t-statistic is 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 for the first two broader validation variables (t-statistic of roughly 2.7) but is only significant at the 10% level for the more stringent variables. Rows (9) to (12) additionally add our new measure FIC-scope to the regression. We find that FIC-scope is much more positive and significant than both firm size and Compustat segments. Its t-statistic is roughly 8.5 for the broad validation measures and 6.5 for the strict validation measures. Additionally, including FIC scope reduces the significance of the Compustat segment variable by roughly one third. Rows (13) to (16) reproduce this test for the NAICS-scope variable, and we find similar but slightly weaker validation (t-statistics range from 5.3 to 6.2). We conclude that FIC-scope has the strongest support in this validation exercise, but NAICS-scope also performs quite well. However, Compustat segment counts are only very weakly validated.

6 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 high importance for understanding corporate finance in general, and also issues of relevance to regulators. For example, is the increase in scope we report related to the high level of acquisition activity reported in the popular press over the past couple decades? Additionally, is innovation investment in R&D needed to support increases in scope, or is it achieved instead through acquisitions and capital expenditures. We examine these issues in this section using plausibly exogenous shifters of the incentives firms have to increase scope. We also assess the link between scope and firm performance, and we explore how increases in scope are financed.

6.1 Instruments and First-Stage Analysis

We first examine the relationship between our proposed instruments and our novel measures of scope. This analysis constitutes the first stage that we will ultimately use in our two-stage least squares analysis of the impact of scope incentives on ex-post investment, performance and financing in the next section. We use these instruments to control for endogeneity but do not ascribe causality to the final results.

Our first instrument is Sectoral Redeployment Potential, which we explain in detail in Section 3.3. This variable measures the extent to which the assets owned by a focal 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, it indicates that the focal firm likely has the ability to increase its scope outward at low cost, as its assets are likely redeployable to assist in production in these nearby product markets. Our second instrument is Sectoral Opportunity Set Potential, which is based on the supply of scope-expansion opportunities rather than the cost of executing 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, it indicates that these peers span many related product markets. As a consequence, the focal firm likely sees a favorable distribution of industries to which it can increase its scope (a "thick" opportunity set). Importantly, both measures are computed without using the characteristics of the focal firm itself, and are weighted heavily on the more distant peers. The use of distant peers, as explained earlier, is supported by econometric research as being moor plausibly exogenous due to the the second-degree (rather than first-degree) network linkages of these peers.

In our first stage analysis, we regress our measures of scope on both instruments and we include all control variables that are also included in our two stage models. Additionally, we include firm and year fixed effects in both stages. The results are displayed in Table 7.

[Insert Table 7 Here]

Row (1) shows that both instruments are strongly positively related to FIC-scope. Sectoral Redeployment Potential is positive with a t-statistic of 4.3, and Sectoral Opportunity set Potential has a positive t-statistic of 11.9. Results are similar for NAICS scope. For comparison, we run the same regression with the # of Compustat segments as the dependent variable and we find the results are much weaker. The first instrument is not significant, and the second is much more weakly significant with a t-statistic of 2.9. Overall these first stage results indicate that our instruments are powerful regarding variation in FIC-scope and NAICS-scope, but not for the number of segments. For this reason, we only consider FIC-scope and NAICS-scope in our second stage to avoid weak instruments.[9]

6.2 Corporate Finance Policies and Scope Expansion

We now consider the second stage regressions where we assess the impact of ex-ante scope increase incentives on ex-post investments, performance, and financing policies. We start with investments and consider ex post acquisitions, divestitures (target of acquisition), R&D, and CAPX. In particular, we consider two-stage least squares regressions where we instrument either FIC-scope or NAICS-scope using our two instruments discussed in the previous section. We then consider the investment policies as the key dependent variables. We also control for size, age, and firm and year fixed effects. Finally, we also consider a specification where we additionally control for valuation (market to book) and the TNIC HHI.

[Insert Table 8 Here]

The results for investment policies are displayed in Table 8. In Panel A, we find that firms with high ex-ante incentives to increase scope are more likely to do an acquisition

[9] As expected in unreported tests, our results are weaker and sporadically significant in the second stage if we do use the number of segments instead of FIC-scope or NAICS-scope.

($t - statistic$ of 4.7) and they are less likely to divest ($t - statistic$ of -2.3). These results are as we predicted because acquisitions are a natural way to increase scope, and avoiding divestitures is also necessary to avoid losing any previous gains in scope. These results are important as they confirm that increases in product scope are an important acquisition motive, and this in turn allows for the possibility that scope increases might be one reason why acquisitions became so prevalent over the past two decades. Documenting this link is also important from a regulatory standpoint given the more controversial link between acquisitions and market power increases suggested in the popular press.

A second important finding is that firms with high incentives to increase scope also increase their R&D expenditures, but they do not increase capital expenditures. The increase in R&D is significant with a t-statistic of 3.9. This finding indicates that innovation likely facilitates the scope expansion in tandem with acquisitions. For example, when assets can be redeployed across product markets, some innovation spending is likely synergistic as it can serve to improve productive efficiency and flexibility. In turn, this model likely facilitates the creation of multi-industry firms that do not need multiple operating segments. For example, all products can be produced in more universal and flexible production sites. Hence the benefits of increased scope are feasible in the 21st century without having to accept the dark side of negative governance externalities that might associate with a complex conglomerate structure. Indeed our earlier results suggest that increases in scope were achieved during our sample with almost no change in the average number of Compustat operating segments.

[**Insert Table 9 Here**]

Table 9 reports the results of analogous regressions for ex-post performance metrics. We consider ex-post valuations (market to book), sales growth, asset growth and return on assets. The table shows that firms with high scope expansion incentives experience higher ex-post valuations, higher sales growth and higher asset growth. However, we do not observe a significant coefficient for profitability measured as return on assets (ROA). Overall, the higher valuations suggest that scope expansion is a positive net present value investment and that investors expect higher profits in the future. We also note that the non-result for ROA is consistent with the view that firms add new industries to their portfolio in an assortive way, and focus on the most profitable markets first. This interpretation is consistent with all four of our findings on performance. Our valuation results - which show that firm market-to-book equity ratios increase with our scope measures - are in contrast to the historical conglomerate literature (Berger and Ofek (1995) and Lang and Stulz (1994)) that finds that firm market-to-book valuation measures go down with an increased number of reported Compustat segments.

[**Insert Table 10 Here**]

Table 10 reports the results of analogous regressions for ex-post financing policies. The question of interest is how do firms likely finance scope expansions? Our results suggest that equity is more commonly used than debt. Increased equity in the capital structure appears to accrue both through the issuance of new shares and through lower overall dividend payments. These results are consistent with innovation and asset redeployment being used to facilitate scope expansion, as neither creates a significant amount of new fixed collateral that is traditionally associated with debt financing.

7 Increasing Concentration or Increasing Scope

Although our paper's primary objective is to document the rise in scope and its roots in corporate finance strategies and performance, we also examine the trend of increasing industry concentration documented in the literature (see Grullon, Larkin, and Michaely (2019) for example). Intuitively, if the number of firms in the economy were held fixed, and every firm expanded its scope to serve twice the number of product markets, it follows that competition would increase economy-wide as more and 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 would not reflect the higher competition. This intuition is particularly clear in the extreme case, where all firms sell products in all markets. Here, all firms are direct competitors in all markets, and competition would be strong economy wide. This section examines if the magnitude of the increase in scope reported earlier is large enough to explain all or part of the apparent rise in concentration. The results could have important implications for the regulatory and academic debate surrounding this issue.

Existing industry classifications do not allow researchers to account for such increases in scope. Many earlier studies measure competition by computing HHIs based on industry codes from Compustat. The key limitation is that each firm, regardless of its actual scope, is assigned to one and only one industry code. If scope was increasing, it would not be observable in these data structures. 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 note 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 the increases in scope that occurred later in our sample. We consider two approaches to adjust concentration ratios for increasing scope.

One limitation of our study is that we only have scope data for U.S. publicly traded firms. Hence, we do not account for the competitive impact of private or foreign firms. The goal of this section is thus to illustrate that an alternative narrative should be considered, and to provide suggestive evidence on its economic magnitude. However, we also note that stylized facts suggest that accounting for the influence of foreign competitors and private firms would likely reinforce our finding that concentration is not increasing as sharply as suggested by prior studies. For example, globalization is increasing over time (see Hoberg and Moon (2017) for example) and accounting for foreign competition would likely further reduce the growth rate of concentration over time. Analogously, studies including Ewens and Farre-Mensa (forthcoming) suggest that larger firms are staying private longer, and accounting for the larger private firms (which are most material) might further reduce the measured growth rate of concentration. Yet we advocate caution in drawing strong conclusions regarding these limitations, and suggest that future research accounting for either should be fruitful.

7.1 Broadening Scope of Businesses

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 11.

The table illustrates an economically large trend toward increasing importance of production across coarser SIC-2 codes in understanding firm production and competition thus arises from multiple markets. 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 concentration is not rising materially in our sample.

7.2 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 FIC-300 classification and the firm-specific market definitions implied by equation (2). 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.

[Insert Figure 4 Here]

To illustrate the impact of scope on industry boundaries and the role of granularity, the upper graph in Figure 4 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 actually higher than the level of market overlap for SIC-3 pairs early in our sample. This indicates that industry boundaries are as strong today at the two-digit SIC level as they were at the three-digit SIC level 25 years ago.

Existing studies note 1997 as a pivotal year in which the rise in concentration began to accelerate. At the start of this year, SIC-2 market overlap was 4.04% and SIC-3 market overlap was 5.72%. By the end of our sample, SIC-2 market overlap rose to 4.88%, and this 0.84% rise is enough to close 50% of the ex-ante gap of 1.68% between SIC-2 and SIC-3. It follows that between 1997 and present day, the granularity at which competition takes place moved roughly 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 4. These include the benchmark SIC-2 HHI and the SIC-3 HHI as computed in the existing literature,[10] 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 concentration is not rising materially during our sample 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

[10] 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 industries if indicated in the Compustat segment tapes. The annualized average HHIs are also weighted by sales.

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.

7.3 Scope Adjustment via Implied Multi-Industry Assignments

The construction of the FIC-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 FIC-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 $Q_{i,j,t,FIC}$ in equation (1)) that were used to construct the classification itself. HHIs are then computed at the FIC 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 $Q_{i,j,t,FIC}$). 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.[11] We then plot both estimates of the HHI faced by average firm in each year in the Figure 5. The figure illustrates, as was the case with the scope adjustment used in the previous section, that concentration levels are not rising materially after 1997.

The results in this section imply that a different narrative might be relevant to understand

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

the rise in industry concentration reported in the literature. This alternative narrative is that scope has been rising rapidly, and as a consequence, traditional HHIs are increasing but this fact might be an artifact of using overly rigid industry classifications that do not account for scope and assign firms to single or historical industry categories. We show that scope-adjusted concentration ratios do not rise materially. This result implies that the number of options consumers have to purchase a particular product is not declining on average. Hence, concerns about within-market horizontal market power rising might not be fully warranted. Notwithstanding that, it is still important to note that increased scope can also generate anti-trust concerns through other channels. For example, market power can arise through anti-competitive product bundling strategies or through market power increases within supply chains. Firms achieving high levels of both scale and scope can accumulate bargaining power, which may be detrimental to healthy competition.

8 Conclusions

We use textual analysis of firm 10-Ks to compute novel measures of scope at the firm-year level. Using our new measures, we find that the scope of U.S. firms has increased dramatically during our sample period from 1989 to 2017. The 21st century high-scope firm operates in multiple product markets, and can do so without needing an inefficient conglomerate structure. Our results suggest that using innovation and acquisitions, these firms are capable of servicing many related product markets using a simple but flexible structure. Analogous analysis of the Compustat segment tapes indicates that historical Compsutat segment data is generally uninformative regarding the scope of U.S. public firms. We reach this conclusion both on the grounds of empirical validation and also on theoretical grounds, as our thesis is that modern firms can achieve multiple-industry product portfolios while maintaining a single segment organizational form.

We find that firms increase scope by acquiring more, divesting less, and increasing innovation spending in R&D, 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. They finance scope expansion using equity rather than debt, consistent with intangibles and asset redeployment not creating material amounts of new collateral. We document that scope expansion is highly valued by the market unlike the discount for conglomerates previously documented in the earlier conglomerate literature for unrelated business lines.

We conclude our analysis with evidence that the increase in scope we report might explain why other studies have documented a trend of 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 using these new text-based measures of firm scope, concentration is essentially unchanging since 1997. These results suggest that a new narrative that considers scope-based growth is important to understand the previous reports of increasing HHIs and the high levels of M&A activity. In particular, much M&A has been targeted at increasing scope, a business strategy that can produce positive net present value as our findings indicate. Yet these results must be interpreted with care, as increased scope can still lead to antitrust concerns in the form of product bundling and increased bargaining power within supply chains. We believe that future research examining scope and market power within specific product-based areas could be particularly impactful given the importance of these issues to regulators and society at large.

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.

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*.

Ewens, Michael, and Joan Farre-Mensa, forthcoming, The deregulation of the private equity markets and the decline in ipos, *Review of Financial Studies*.

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*.

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 S Katie Moon, 2017, Offshore activities and financial vs operational hedging, *Journal of Financial Economics* 125, 217–244.

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.

———, 2016, Text-based network industry classifications and endogenous product differentiation, *Journal of Political Economy* 124, 1423–1465.

———, 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, .

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

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.

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.*

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

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
FIC-Scope	The number of TNIC (see Hoberg and Phillips 2016) industries (using the FIC-300 classification based on 1997 industry clusters) 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 FIC industries.	
NAICS-Scope	This is computed in a similar way to the FIC-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 FIC 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).	

Variable	Definition
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.
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.

Additional stop words dropped from the NAICS manual vocabularies: 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 1: Summary Statistics

Summary statistics are reported for our sample of 100,525 observations based on annual firm observations from 1988 to 2017. Our main variables of interest, TNIC-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 FIC-scope, fixed industries are based on the TNIC FIC-300 industries (see Hoberg and Phillips 2016), 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 3 of the paper.

Variable	Mean	Std. Dev.	Minimum	Median	Maximum	# Obs
		Panel A: Scope and Segment Variables				
TNIC-Scope	6.923	5.482	0.000	6.000	30.000	100,525
NAICS-Scope	6.269	7.519	0.000	4.000	47.000	100,525
# Compsutat Segments	1.452	0.862	1.000	1.000	11.000	100,525
		Panel B: Accounting Variables				
R&D/Assets	0.056	0.127	0.000	0.000	2.944	100,525
CAPX/Assets	0.058	0.070	-0.000	0.037	0.725	100,525
Acquisition Dummy	0.287	0.452	0.000	0.000	1.000	100,525
Target Dummy	0.126	0.332	0.000	0.000	1.000	100,525
Valuation (M/B)	1.758	2.156	0.064	1.204	203.313	99,938
Sales Growth	0.110	0.437	-6.177	0.076	9.383	100,112
Asset Growth	0.077	0.356	-4.294	0.050	5.529	100,494
Equity Issuance	0.051	0.152	-0.002	0.004	2.895	100,525
Debt Issuance	0.103	0.207	0.000	0.002	1.999	100,525
Dividends/Assets	0.008	0.028	0.000	0.000	1.248	100,424
Equity Repurchase	0.016	0.042	0.000	0.000	0.518	93,049
Log Assets	5.446	2.122	0.694	5.327	13.590	100,525
Log Age	2.622	0.766	0.693	2.565	4.220	100,525

Table 2: Pearson Correlation Coefficients

Pearson Correlation Coefficients are reported for our sample of 100,525 observations based on annual firm observations from 1988 to 2017. Our main variables of interest, TNIC-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 FIC-scope, fixed industries are based on the TNIC FIC-300 industries (see Hoberg and Phillips 2016), 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 3 of the paper.

Row	Variable	Fic-Scope	NAICS-Scope	# CS Segments	Log Assets	Log Age	R&D/ Assets	CAPX/ Assets	Acquisition Dummy	Target Dummy	Sales Growth
	NAICS-Scope	0.579									
	# Compsutat Segments	0.186	0.198								
	Log Assets	0.269	0.273	0.312							
	Log Age	0.038	-0.047	0.309	0.400						
	R&D/Assets	0.010	-0.039	-0.155	-0.250	-0.147					
	CAPX/assets	-0.016	0.045	-0.043	0.026	-0.097	-0.092				
	Acquisition Dummy	0.067	0.088	0.133	0.301	0.075	-0.089	-0.015			
	Target Dummy	0.054	0.057	0.167	0.248	0.168	-0.056	-0.006	0.156		
	Sales Growth	0.030	0.055	-0.032	-0.013	-0.166	0.032	0.099	0.104	-0.065	
	TNIC HHI	-0.133	-0.299	0.091	-0.182	0.171	-0.149	-0.124	-0.042	-0.011	-0.071

Table 3: Scope Statistics vs Compustat Segment Counts

The table reports scale and scope statistics separately for firms based on how many operating segments the firm reports in the Compustat database. Our main variables of interest, TNIC-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 FIC-scope, fixed industries are based on the TNIC FIC-300 industries (see Hoberg and Phillips 2016), and for NAICS-scope, it is based on 4-digit NAICS industries. Assets are from Compustat (variable AT).

# Segments	FIC-SCope	NAICS-Scope	Assets	# Obs.
1 segment	6.41	5.56	1365	71,575
2 segments	7.53	7.02	3255	17,939
3 segments	8.63	8.57	6192	7,447
4 segments	9.90	10.21	10084	2,353
5+ segments	12.21	15.17	30776	1,211

Table 4: Scope Statistics vs Firm Size

The table reports scale and scope statistics separately for firms sorted into size quintiles. Sorts are annual and are based on Compustat assets (variable AT). Our main variables of interest, TNIC-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 FIC-scope, fixed industries are based on the TNIC FIC-300 industries (see Hoberg and Phillips 2016), and for NAICS-scope, it is based on 4-digit NAICS industries. Assets are from Compustat (variable AT). We report statistics for the full sample and separately for single segment firms only (see column headers).

		All Firms					Single Segment Firms Only				
Row	Firm Size Quintile	# Segments	FIC-Scope	NAICS-Scope	Assets	# Obs.	# Segments	FIC-Scope	NAICS-Scope	Assets	# Obs.
	Small Firms	1.22	5.65	4.02	23	20,094	1	5.50	3.88	19	14,304
	Quintile 1	1.26	6.16	5.04	102	20,110	1	6.07	4.72	74	14,321
	Quintile 2	1.35	6.69	5.93	305	20,113	1	6.28	5.39	203	14,320
	Quintile 3	1.50	7.50	7.18	936	20,111	1	6.76	6.18	604	14,321
	Big Firms	1.94	8.62	9.17	11728	20,097	1	7.41	7.64	5926	14,309

Table 5: 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. For each pair of FIC-300 industries in each year, we first tabulate the number of firms that operate in both industries in the pair based on the FIC-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 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 TNIC space. 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	# Obs.
Least Similar	0.036	0.038	0.030	0.001	828,872
Decile 2	0.043	0.046	0.035	0.002	829,542
Decile 3	0.041	0.043	0.036	0.003	829,012
Decile 4	0.052	0.056	0.043	0.004	829,145
Decile 5	0.053	0.053	0.054	0.006	829,064
Decile 6	0.067	0.067	0.068	0.008	829,462
Decile 7	0.086	0.086	0.086	0.010	828,756
Decile 8	0.096	0.091	0.107	0.014	829,414
Decile 9	0.132	0.125	0.147	0.020	828,926
Most Similar	0.393	0.393	0.395	0.047	829,422

Table 6: 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	FIC-Scope	NAICS-Scope	# Segments	Log Assets	Log Age	Log 10K Size	M/B	TNIC HHI	# Obs
(1)	Product Breadth				1.650	0.400	–8.422	–0.209	–2.119	72,280
					(3.640)	(0.260)	(–7.300)	(–2.520)	(–1.760)	
(2)	Prod/Svc Breadth				1.774	0.642	–8.869	–0.209	–2.063	72,280
					(3.840)	(0.410)	(–7.470)	(–2.490)	(–1.680)	
(3)	Prod Breadth Detail				0.763	–1.124	–3.327	0.016	–1.558	72,280
					(3.820)	(–1.700)	(–7.270)	(0.500)	(–3.260)	
(4)	Prod/Svc Breadth Detail				0.791	–0.917	–3.459	0.017	–1.497	72,280
					(3.920)	(–1.370)	(–7.290)	(0.540)	(–3.070)	
(5)	Product Breadth			1.381	1.513	0.140	–8.501	–0.205	–2.124	72,280
				(2.800)	(3.320)	(0.090)	(–7.370)	(–2.470)	(–1.760)	
(6)	Prod/Svc Breadth			1.299	1.646	0.397	–8.943	–0.205	–2.068	72,280
				(2.570)	(3.550)	(0.260)	(–7.530)	(–2.440)	(–1.680)	
(7)	Prod Breadth Detail			0.430	0.720	–1.205	–3.352	0.017	–1.559	72,280
				(1.920)	(3.590)	(–1.820)	(–7.310)	(0.540)	(–3.260)	
(8)	Prod/Svc Breadth Detail			0.415	0.750	–0.996	–3.483	0.019	–1.498	72,280
				(1.820)	(3.700)	(–1.480)	(–7.330)	(0.580)	(–3.070)	
(9)	Product Breadth	0.640		0.959	1.075	0.569	–9.263	–0.224	–0.410	72,280
		(8.430)		(1.960)	(2.390)	(0.380)	(–7.980)	(–2.740)	(–0.340)	
(10)	Prod/Svc Breadth	0.663		0.861	1.192	0.841	–9.732	–0.225	–0.293	72,280
		(8.520)		(1.720)	(2.600)	(0.550)	(–8.150)	(–2.720)	(–0.240)	
(11)	Prod Breadth Detail	0.216		0.288	0.572	–1.060	–3.609	0.011	–0.981	72,280
		(6.360)		(1.300)	(2.910)	(–1.610)	(–7.810)	(0.340)	(–2.010)	
(12)	Prod/Svc Breadth Detail	0.228		0.265	0.594	–0.843	–3.754	0.012	–0.888	72,280
		(6.610)		(1.170)	(2.990)	(–1.260)	(–7.840)	(0.380)	(–1.780)	
(13)	Product Breadth		0.309	1.063	1.253	0.391	–9.075	–0.210	–0.700	72,280
			(6.020)	(2.150)	(2.770)	(0.260)	(–7.780)	(–2.530)	(–0.580)	
(14)	Prod/Svc Breadth		0.327	0.962	1.371	0.662	–9.550	–0.210	–0.562	72,280
			(6.240)	(1.900)	(2.970)	(0.430)	(–7.940)	(–2.500)	(–0.450)	
(15)	Prod Breadth Detail		0.132	0.294	0.609	–1.098	–3.596	0.015	–0.952	72,280
			(5.290)	(1.310)	(3.090)	(–1.660)	(–7.760)	(0.480)	(–1.960)	
(16)	Prod/Svc Breadth Detail		0.137	0.275	0.635	–0.884	–3.737	0.017	–0.867	72,280
			(5.360)	(1.200)	(3.190)	(–1.320)	(–7.770)	(0.520)	(–1.750)	

Table 7: First-Stage Regressions

The table reports the results of first-stage regressions where measures of scope (FIC-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)	FIC-Scope	1.251	6.941	0.882	–0.830	99,506
		(4.330)	(11.910)	(19.290)	(–6.030)	
(2)	NAICS-Scope	1.074	12.772	1.279	–0.976	99,506
		(2.380)	(14.680)	(17.730)	(–5.060)	
(3)	# Segments	0.019	0.293	0.107	0.161	99,506
		(0.300)	(2.900)	(12.340)	(6.260)	

Table 8: 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 (FIC-Scope or NAICS-Scope) as indicated in the panel headers. The first-stage regressions are displayed in Table 7 and include two instruments for scope (explained in detail in Table 7). 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: FIC-Scope is Scope Variable			
(1)	Acquirer Dummy	0.026	–0.006	–0.029			98,196
		(4.660)	(–0.920)	(–2.710)			
(2)	Target Dummy	–0.009	0.038	0.039			98,196
		(–2.270)	(8.570)	(5.390)			
(3)	R&D/Assets	0.003	–0.016	0.009			98,196
		(3.900)	(–12.690)	(4.370)			
(4)	CAPX/Assets	0.000	–0.002	–0.014			98,196
		(0.450)	(–2.630)	(–9.820)			
				Panel B: NAICS-Scope is Scope Variable			
(5)	Acquirer Dummy	0.016	–0.003	–0.035			98,196
		(4.860)	(–0.580)	(–3.540)			
(6)	Target Dummy	–0.005	0.036	0.042			98,196
		(–2.060)	(9.060)	(6.240)			
(7)	R&D/Assets	0.001	–0.015	0.008			98,196
		(3.570)	(–12.860)	(3.980)			
(8)	CAPX/Assets	0.000	–0.002	–0.014			98,196
		(0.570)	(–2.870)	(–10.190)			
				Panel C: FIC-Scope is Scope Variable (extra controls added)			
(9)	Acquirer Dummy	0.036	–0.007	–0.017	0.009	0.107	97,624
		(4.360)	(–0.840)	(–1.500)	(9.180)	(4.560)	
(10)	Target Dummy	–0.012	0.039	0.036	–0.002	–0.028	97,624
		(–2.130)	(7.120)	(4.630)	(–3.400)	(–1.660)	
(11)	R&D/Assets	0.004	–0.017	0.009	–0.001	0.002	97,624
		(3.190)	(–11.910)	(4.190)	(–2.100)	(0.610)	
(12)	CAPX/Assets	0.000	–0.001	–0.013	0.002	–0.003	97,624
		(–0.480)	(–0.790)	(–8.720)	(9.090)	(–1.240)	

Table 9: 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 (FIC-Scope or NAICS-Scope) as indicated in the panel headers. The first-stage regressions are displayed in Table 7 and include two instruments for scope (explained in detail in Table 7). 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: FIC-Scope is Scope Variable				
(1)	Valuation (M/B)	0.080	–0.532	–0.344			97,625
		(4.310)	(–17.410)	(–7.360)			
(2)	Sales Growth	0.031	–0.098	–0.182			97,825
		(5.790)	(–14.990)	(–17.860)			
(3)	Asset Growth	0.047	–0.206	–0.038			98,193
		(8.830)	(–30.220)	(–3.550)			
(4)	OI/Assets	–0.001	0.012	0.000			97,995
		(–0.540)	(3.510)	(0.010)			
			Panel B: NAICS-Scope is Scope Variable				
(5)	Valuation (M/B)	0.040	–0.512	–0.373			97,625
		(3.730)	(–17.880)	(–8.230)			
(6)	Sales Growth	0.018	–0.094	–0.191			97,825
		(5.960)	(–15.990)	(–20.450)			
(7)	Asset Growth	0.028	–0.200	–0.050			98,193
		(9.500)	(–33.210)	(–5.180)			
(8)	OI/Assets	–0.001	0.011	0.000			97,995
		(–0.490)	(3.650)	(0.100)			
			Panel C: FIC-Scope is Scope Variable (extra controls added)				
(9)	Valuation (M/B)	0.095	–0.448	–0.156	0.222	0.246	97,391
		(3.770)	(–14.290)	(–3.850)	(10.340)	(2.690)	
(10)	Sales Growth	0.042	–0.091	–0.150	0.032	0.129	97,267
		(5.460)	(–11.480)	(–13.030)	(12.100)	(5.740)	
(11)	Asset Growth	0.056	–0.192	0.002	0.044	0.137	97,624
		(7.140)	(–23.500)	(0.130)	(16.340)	(6.100)	
(12)	OI/Assets	–0.002	0.014	0.004	0.005	0.000	97,430
		(–0.470)	(3.800)	(0.690)	(4.780)	(0.030)	

Table 10: 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 (FIC-Scope or NAICS-Scope) as indicated in the panel headers. The first-stage regressions are displayed in Table 7 and include two instruments for scope (explained in detail in Table 7). 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.

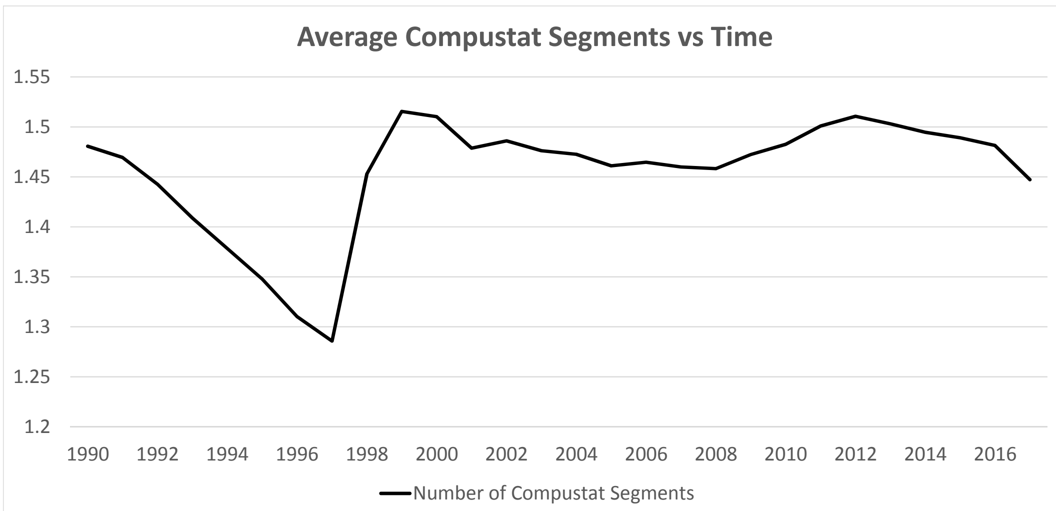
Row	Dependent Variable	Scope Variable	Log Assets	Log Age	Mkt/Book	TNIC HHI	# Obs
		Panel A: FIC-Scope is Scope Variable					
(1)	Equity Issuance	0.009	–0.050	–0.009			98,196
		(7.220)	(–25.060)	(–2.950)			
(2)	Debt Issuance	0.002	–0.011	0.010			98,196
		(0.760)	(–3.960)	(1.990)			
(3)	Dividends/Assets	–0.001	0.001	0.002			98,097
		(–2.870)	(1.900)	(2.440)			
(4)	Repurchases/Assets	–0.001	0.005	0.005			90,680
		(–1.460)	(7.430)	(5.070)			
		Panel B: NAICS-Scope is Scope Variable					
(5)	Equity Issuance	0.005	–0.049	–0.012			98,196
		(7.530)	(–26.340)	(–4.010)			
(6)	Debt Issuance	0.001	–0.011	0.010			98,196
		(0.970)	(–4.390)	(2.050)			
(7)	Dividends/Assets	–0.001	0.001	0.002			98,097
		(–2.710)	(1.520)	(3.300)			
(8)	Repurchases/Assets	0.000	0.004	0.006			90,680
		(–1.290)	(7.860)	(5.590)			
		Panel C: FIC-Scope is Scope Variable (extra controls added)					
(9)	Equity Issuance	0.010	–0.047	0.000	0.010	0.021	97,624
		(5.580)	(–21.180)	(–0.130)	(10.190)	(3.890)	
(10)	Debt Issuance	0.001	–0.010	0.010	0.001	–0.005	97,624
		(0.160)	(–2.930)	(1.840)	(1.990)	(–0.510)	
(11)	Dividends/Assets	–0.002	0.001	0.002	0.000	–0.004	97,528
		(–2.790)	(2.380)	(2.330)	(4.760)	(–1.990)	
(12)	Repurchases/Assets	–0.001	0.005	0.006	0.001	–0.001	90,146
		(–1.350)	(6.960)	(5.240)	(5.780)	(–0.660)	

Table 11: 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 1: Measures of scope versus time. The upper figure plots the average number of Compustat segments per firm over time. The lower figure plots the average values of FIC-scope and NAICS-scope over our sample period. TNIC-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 FIC-scope, fixed industries are based on the TNIC FIC-300 industries (see Hoberg and Phillips 2016), and for NAICS-scope, it is based on 4-digit NAICS industries.



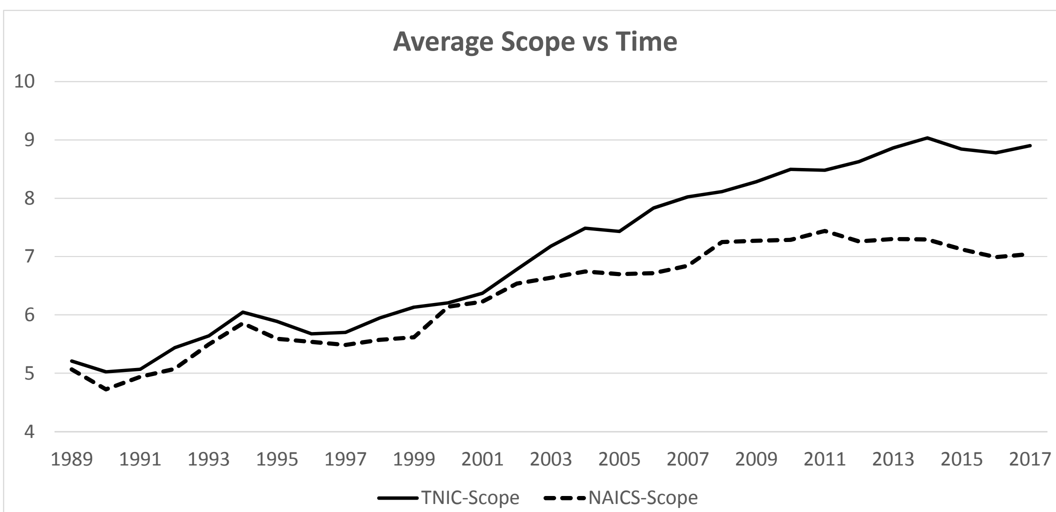


Figure 2: The upper figure reports the average number of words in the firm's 10-K Item 1 business description divided by the number of industries (FIC-based or NAICS-based scope) the firm likely operates in. The goal is to measure the average degree of product variety within industries over time. The lower figure displays the average size of the 10-K Item 1 over time.

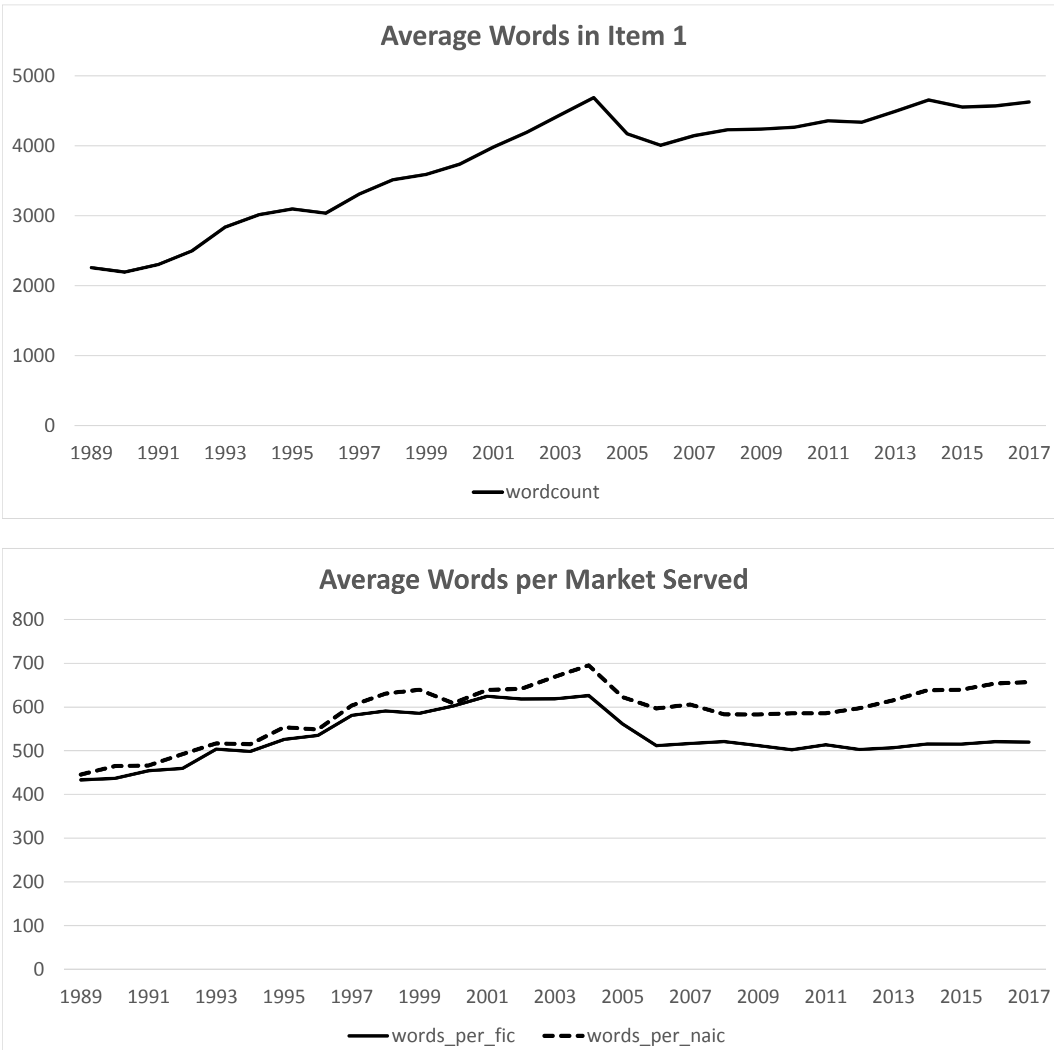


Figure 3: Firm size versus time. The figure displays firm size (measured as Compustat assets, both nominal and inflation adjusted) over time.

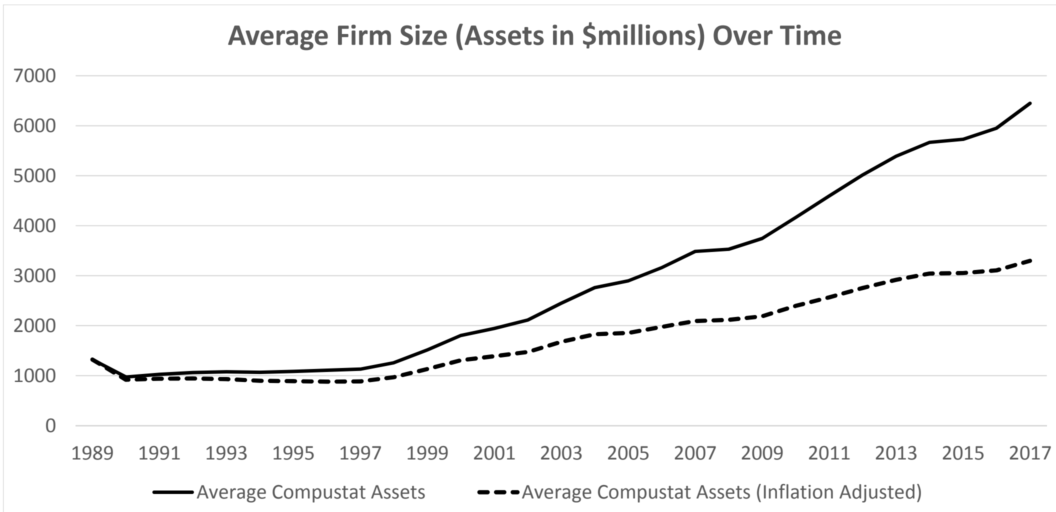
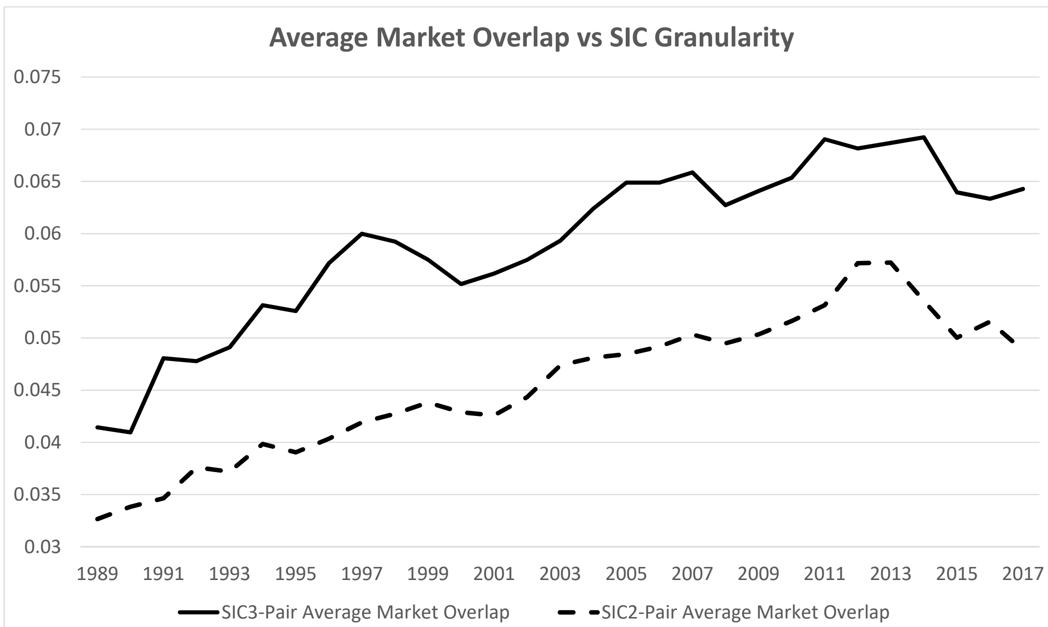


Figure 4: The upper figure reports the average market overlap of pairs of firms that are in the same SIC3 or SIC2 industries. Market overlap at the industry pair level is the average of firm-pair market overlap for all permutations of firms where one firm is in each of the industries being compared at 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 FIC-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.



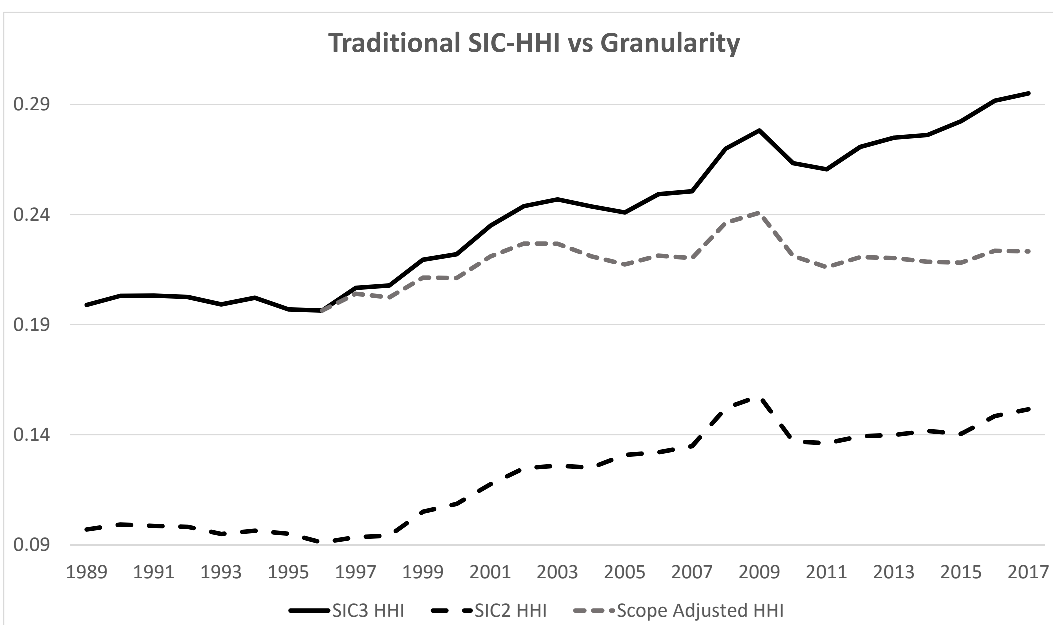


Figure 5: The FIC-Scope Implied HHI computes the HHI after allowing firms to have a presence in multiple industries, as is identified during the derivation of FIC-scope itself. Market shares are based on sales. Each firm's sales are allocated across the multiple sectors each firm is assigned to using similarity weights (similarity weights are defined as $Q_{i,j,t,FIC}$ in equation (1)). HHIs are then computed at the FIC 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 $Q_{i,j,t,FIC}$). 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. The upper figure reports the sales-weighted average HHI over time and the lower figure reports the equal weighted average HHI over time.