Product Market Uniqueness, Organizational Form and Stock Market Valuations

Gerard Hoberg and Gordon Phillips[*]

August 31, 2014

ABSTRACT

We introduce a new framework for forming peer firm portfolios that can account for firm uniqueness and organizational form. Our new vocabulary-based peer firm portfolios explain much cross sectional dispersion in firm valuations and generate a direct measure of firm product market uniqueness. We find that firms have higher stock market valuations than their peers when their products are more unique. This result holds for conglomerate and focused single-segment organizational forms. Increased success in patenting, increased branding, and less venture capital financed entry into the firm's product space all contribute to the long-term maintenance of uniqueness and thus higher valuations.

[*] University of Southern California and University of Southern California and National Bureau of Economic Research, respectively. Hoberg can be reached at hoberg@marshall.usc.edu and Phillips can be reached at gordon.phillips@marshall.usc.edu. This paper entirely replaces a previous version titled "Conglomerate Industry Spanning". We thank Michael Cooper, Dongmei Li, Peter MacKay, Oguz Ozbas, Jiaping Qiu, Michael Roberts, Merih Sevilir, Albert Sheen, Denis Sosyura and seminar participants at the Carnegie Mellon, CFEA conference, Duisenberg School of Finance and Tinbergen Institute, Erasmus University, Hebrew University, Interdisciplinary Center of Herzilia, McMaster University, Rotterdam School of Management, Stanford University, Tel Aviv University, Tilberg University, the University of Chicago, University of Illinois, the University of Mannheim, the University of Miami and the University of Utah for helpful comments. All errors are the authors alone. Copyright ©2012 by Gerard Hoberg and Gordon Phillips. All rights reserved.

A fundamental question both in corporate finance and asset pricing is whether and how the stock market uses information from peer firms. The literature has focused, for example, on the effect of peers on capital structure (MacKay and Phillips (2005), Leary and Roberts (2010) and Rauh and Sufi (2010)).[1] It remains unknown the extent to which stock market valuations incorporate peer firm fundamentals. Conglomerate firms, in particular, are frequently cited as being difficult for analysts to follow and value given the multiple business segments in which they operate.

We introduce new generalized benchmarking methods and examine how the stock market incorporates information about peer firm fundamentals using new text-based peer firm portfolios that best replicate the product offerings of a given firm.[2] The new benchmarks replicate the firm's product market offerings and also account for organizational form and the basic accounting characteristics (such as size and age) of a given focal firm. Using these new peer benchmarks, we examine whether firms with more unique products relative to their peers have higher stock market valuations. We then examine the factors that enable firms to maintain product uniqueness over time, focusing on the importance of product market competition, patenting and branding behavior.

We show that our new methodology of identifying peer firms and which peer firms best match a given focal firm can explain the cross-sectional valuation of conglomerate and single-segment firms. While prior studies have focused on the average or median valuation of these firms,[3] we focus in this paper on understanding the cross-sectional distributions of valuations of both conglomerate and single-segment firms. In particular, we focus on why some firms trade at premia or at discounts for each organizational form, an issue Stein (2003) also identifies as important in his survey paper.

We find that text-based replication peers generate significantly better benchmarks for both conglomerate and single-segment firms (relative to traditional indus-

[1] Additional studies on the effects of peer firm financial policies on real decisions in a strategic context include Chevalier (1995), Khanna and Tice (2000) and Phillips (1995).

[2] These weighted "network" benchmarks represent best matches in the product market analogous to a weighted Facebook circle of friends (both close friends and acquaintances).

[3] For early articles focusing on the average or median discount, see Wernerfelt and Montgomery (1988), Lang and Stulz (1994) and Berger and Ofek (1995).

try peers) in their ability to explain valuations and other characteristics. In particular, our replication peers have characteristics that are more correlated with those of a given focal firm relative to other benchmarks. Also relative to other benchmarks, our new benchmarks also have significantly lower mean-squared error on their ability to predict actual firm valuations.

We further show that both conglomerate *and* single-segment firms have higher stock market valuations than replicating peer firms when they are more unique relative to these peers. These higher valuations are long-lasting. We examine the time series persistence of product uniqueness and find that, in particular, patent citations, R&D, branding and less entry of new rivals through venture capital financing and initial public offerings can explain the persistence of firm uniqueness over time. Overall, our findings show that the stock market values firm uniqueness and recognizes peer groups based on fundamental product characteristics that are not reflected in standard industry groupings.

Our valuation results for firm uniqueness are consistent with the conclusion that it is not easy for a competitor to introduce similar successful products when a firm is unique and has patents and brands. The case of Apple versus its peers is illustrative. Apple's peers, Dell and HP, have both tried to introduce successful tablet computers, while Sony and Microsoft have introduced new digital music players. Apple still has very high margins and market shares in each of these markets multiple years after first introducing its products - despite efforts by peers to replicate Apple's successful product offerings.

Many empirical tests, including event studies in corporate finance and anomaly testing in asset pricing, are based on the use of peer firms or counterfactual groups. Standard approaches in the literature often define counterfactuals as averages of a firm's SIC code peers. Some studies additionally limit the SIC peers to those with similar size or age as the focal firm. This general approach has many limitations: (1) SIC codes are not highly informative (see Hoberg and Phillips (2010a)), (2) limiting the counterfactual to just 1-2 firms having similar size and age can introduce noise, and (3) it is difficult to create counterfactuals based on more than one or two characteristics like size and age given high sort dimensions. We provide generalized counterfactuals that overcome these limitations. We identify a portfolio of peer firms that best replicates the product offerings and characteristics (such as size and age) of each given firm. In particular, we use least squares to determine the set of peer weights having a weighted average product market vocabulary that best matches the vocabulary of the firm being replicated. By applying simple constraints to this optimization, we can additionally ensure that the weighted portfolio also matches the focal firm on additional characteristics such as size and age.

In all, our new framework goes beyond the TNIC classification of Hoberg and Phillips (2010a) (HP 2010) in three ways. First, we use information about the number of operating segments a firm has to fully assign conglomerates to multiple locations in the TNIC product market space. This yields improved power and research flexibility, particularly to examine theories of conglomerate structure and organizational form. Second, we compute generalized counterfactual peer groups that replicate a firm based on product markets and can also match a focal firm based on vectors of firm characteristics. This latter feature improves benchmark quality and allows for more flexible counterfactual analysis. Third, and perhaps most importantly, the data structure we create differs from HP 2010 in both content and purpose. As a result, the papers test different theoretical issues (HP do not consider firm uniqueness or conglomerate valuations in cross section).

The objective of HP (2010) is to identify the set of peer firms that are product market rivals, and to identify pairwise proximity in an intransitive network. The objective of the current study is to establish empirical counterfactuals, which is fundamentally different. For example, our counterfactual peer portfolios assign higher weights not only when the given peer is more proximate as in HP (2010), but also when they offer product attributes that a focal firm also has, and that other rivals do not. Thus, it is frequently the case that some proximate rivals would be assigned zero or even negative weights due to their redundancy in replication, whereas some moderately proximate rivals might have relatively high weights. Finally, our accounting for basic characteristics such as size and age departs fully from the approach in HP (2010), but is central in benchmarking, where the objective is often to construct a counterfactual that is in the same product market, and that also matches the focal firm on a number of exogenous characteristics. In all, we believe that HP 2010 and the current study both have many empirical applications, but they are quite distinct and complementary.

We note that our measure of uniqueness is a direct result of how we construct product market replicas for each firm. Because our approach uses a form of least squares, we can compute a replication R^2 for each firm. Intuitively, firms with high replicating peer R^2 are not unique and can be easily replicated by product market peers. This concept of product market uniqueness (defined as one minus this R^2) has economic content. A firm with unique products should be fundamentally different from other firms, and is likely insulated from competition. A primary focus of the current article is not only to explore the impact of improved counterfactuals, but also to examine the role of product market uniqueness in determining economic outcomes. To our knowledge, our study is the first to show that benchmark quality (through product market uniqueness) is important for understanding economic outcomes. Because we find strong evidence that unique firms are fundamentally different, our results further suggest that future studies of counterfactual analysis should consider benchmark quality. For example, our approach allows the researcher to divide a sample into a subsamples with high versus low quality benchmarks. We note that existing benchmarks based on standard industry classifications do not provide analogous measures of benchmark quality.

We use our new benchmarks and measures of firm uniqueness to examine cross-sectional differences in conglomerate *and* single-segment valuations in the stock market. We show that our weighted benchmarks provide economically large improvements relative to existing methods in their ability to accurately predict actual firm valuations and characteristics. We go beyond examining whether there is an average or median conglomerate discount, which many existing studies document to be related to self-selection.[4] Our main valuation finding, which is unique relative

[4] For additional articles on the average or median discount of conglomerate firms, see also Comment and Jarrell (1995), Servaes (1996), Lins and Servaes (1999), Rajan, Servaes, and Zingales (2000) and Lamont and Polk (2002). Denis, Denis, and Yost (2002) report an analogous discount when segments are internationally diverse. The average conglomerate discount has been shown to be related to self-selection by Campa and Kedia (2002), Graham, Lemmon, and Wolf (2002), and Villalonga (2004b), data reporting by Villalonga (2004a) and merger accounting by Custodio (2010).

to existing studies, is that both conglomerate and single-segment firms have higher valuations relative to replicating portfolios when they are unique. In time series tests, we find that uniqueness is related to innovation and branding activities. We do not find evidence of stock return reversals, indicating that the higher valuations of unique firms are not associated with predictable decays, and hence these higher valuations are long lived.

Our paper makes three main contributions. Our first contribution is to examine the link between product-market variables and stock market valuations in cross section. For both single-segment and conglomerate firms, we find that firm valuations are higher when the firm is more unique. These results fill an empirical gap that Stein (2003) identifies in his survey paper. Our second contribution is to show how firms maintain product uniqueness. We find that three key factors are important: increased success in patenting measured by patent applications and citations, increased branding, and less venture capital funding in a firm's product space.

Our third contribution is methodological: we present new text-based methods that use constrained optimization to generate single-segment benchmarks for both single-segment firms and conglomerate firms. We also document how these replicating peers, and their corresponding measures of uniqueness, can be identified using a closed form solution to a constrained optimization. These benchmarks offer significant gains in accuracy relative to existing methods.[5]

Our paper proceeds as follows. In the next section, we briefly discuss the key hypothesis we examine. Section III describes our data, variables, and methods used to examine product relatedness. We develop new methods to computationally weight peer firms based on their product descriptions and other accounting characteristics. Section IV examines how the stock market values uniqueness for both conglomerate and single-segment firms. Section V examines the long-run maintenance of product uniqueness and Section VI concludes.

[5] These optimized peer counterfactuals should prove useful in other research applications such as event studies, or research estimating the magnitude of peer effects. We intend to distribute the replicating peer data on the web.

II Stock Market Valuations and Peer Firms

Our main hypothesis addresses the link between firm uniqueness and the levels of firm valuations in the stock market for firms of different organizational forms. We examine this link for firms that produce in a single industry, and for firms that produce in multiple industries so that we can better understand the role of organizational form. The literature on product differentiation and uniqueness is extensive. Uniqueness can be viewed as analogous to product differentiation. Chamberlin (1933) originally proposed the importance of product differentiation, with recent articles by Berry, Levinsohn, and Pakes (1997), and Seim (2006) also focusing on product differentiation. We focus on stock market valuations and focus more directly on the matter of whether or not firms in the existing universe have adequate product diversity to pose a material replication threat to a given firm under consideration. We postulate that a firm that can be replicated by others firms holds a weaker competitive position and should have a lower valuation compared to firms that are difficult to replicate. We thus consider the following hypothesis:

H1: Stock Market Valuations and Firm Uniqueness: A firm's stock market valuation will be higher when its combination of products is not easily replicated or "spanned" by peer firms (i.e., when the firm is difficult to replicate using the best replicating portfolio of rival firms), and hence, exhibits product differentiation relative to peer firms.

We measure firm uniqueness using the product descriptions and accounting characteristics of firms and calculate a best portfolio of replicating peer firms whose product descriptions and accounting characteristics most closely match the focal firm. The specifics of our methodology are described fully in the next section.

We then use our methods to value both single-segment and conglomerate firms given that the valuation of these firms requires identification of peer benchmarks. Our methods have the key advantage that they produce continuous measures of goodness of fit for how good a set of benchmark firms are at replicating a given firm's product offerings and its accounting performance. We can thus measure the extent to which both conglomerate and single-segment firms produce products that are not produced by their peer firms (uniqueness). Lastly, we assess how long firms are able to maintain product uniqueness and what factors are associated with the maintenance of product uniqueness. We focus especially on innovative activity and product branding as key contributors to the maintenance of product uniqueness.

III Data and Methodology

A The Sample of 10-Ks

The methodology we use to extract 10-K text follows Hoberg and Phillips (2010a). The first step is to use web crawling and text parsing algorithms to construct a database of business descriptions from 10-K annual filings from the SEC Edgar website from 1996 to 2008. We search the Edgar database for filings that appear as "10-K," "10-K405," "10-KSB," or "10-KSB40." The business descriptions appear as Item 1 or Item 1A in most 10-Ks. The document is then processed using APL to extract the business description text and the company identifier, CIK.[6] Business descriptions are legally required to be accurate, as Item 101 of Regulation S-K requires firms to describe the significant products they offer, and these descriptions must be updated and representative of the current fiscal year of the 10-K.

B Product Word Descriptions

Based on the database of business descriptions, we form word vectors for each firm based on the text in product descriptions of each firm. To construct each firm's word vector, we first omit common words that are used by more than 25% of all firms. Following Hoberg and Phillips (2010a), we further restrict our universe in each year to words that are either nouns or proper nouns.[7] Let M_t denote the number of such words. For a firm i in year t, we define its word vector $W_{i,t}$ as a binary M_t-vector,

[6] We thank the Wharton Research Data Service (WRDS) for providing us with an expanded historical mapping of SEC CIK to COMPUSTAT gvkey, as the base CIK variable in COMPUSTAT only contains the most recent link.

[7] We identify nouns using Webster.com as words that can be used in speech as a noun. We identify proper nouns as words that appear with the first letter capitalized at least 90% of the time in the corpus of all 10-K product descriptions. Previous results available from the authors did not impose this restriction to nouns. These previous results were qualitatively similar.

having the value one for a given element when firm i uses the given word in its year t 10-K business description. We then normalize each firm's word vector to unit length, resulting in the normalized word vector $N_{i,t}$. Importantly, each firm is represented by a unique vector of length one in an M_t-dimensional space. Therefore, all firms reside on a M_t-dimensional unit sphere, and each firm has a known location.

C Replicating Peers for Different Organizational Forms: Existing Methods

Throughout our discussion of replicating peers, we will adopt the following terminology. We will refer to the firm being replicated as the "focal firm", and the set of firms used to construct its replica as its "replication peers" or simply "peers". As we aim to build peer replicas of both conglomerate and single-segment focal firms, we will generally use the term "pure plays" to refer to the set of replication peers used in a given replication calculation. We use this term for parsimony, and to emphasize that we only consider single-segment firms as candidate replication peers. Limiting our candidate replication peers to single-segment firms ensures that our measures maintain a transparent interpretation, and that they are not influenced by issues underlying why firms choose to be conglomerates.

Although we depart significantly from the literature, we first consider a slightly modified algorithm based on Lang and Stulz (1994) (LS) and Berger and Ofek (1995) (BO) as a benchmark representing the existing literature.[8] Although these studies focus on conglomerate focal firms, we note that the methods used by LS and BO apply to single-segment focal firms as well, and we consider both types of firms.

We follow LS and BO by defining a universe of candidate single-segment pure plays to replicate each conglomerate focal firm segment. In BO, this universe is initially defined as all pure plays operating in the firm's four digit SIC industry. However, if the number of pure plays in this universe is less than five, then the pure plays in the given segment's three-digit industry are used. Finally, coarseness is increased to the two digit or even the one digit level until a universe of at least five pure plays is identified. Because changing the level of coarseness can alter the

[8] Many studies including Campa and Kedia (2002) and Villalonga (2004b) use this methodology.

economic information contained in the benchmark (due to economies of scope or irrelevant peers), we exclusively use three-digit SIC industries as our starting point following the broader literature on industry analysis in Finance. However, we can report that using varying levels of coarseness, as used in BO, does produce materially similar results.

Our second step also follows BO's framework, and we compute the firm value to sales ratio for each candidate replication peer (we do this for each segment of the focal firm). We then compute the median firm value to sales ratio over all of the replication peers associated with each segment. We compute these medians using the three digit SIC code of each segment as discussed above. Each segment's imputed value is the segment's observed sales multiplied by this median value to sales ratio of the given segment's replication peers. Medians are used in this literature to reduce the impact of outliers, as firm value to sales ratios can become extreme, especially when firms have low sales or high growth options. Finally, the imputed value of a conglomerate focal firm as a whole is the sum of the imputed values of the given conglomerate's segments. For a single-segment focal firm, the calculation is the same and the firm imputed value is simply the imputed value of the single segment. Excess value is the natural logarithm of the focal firm's actual firm value divided by its imputed firm value. This calculation can also be done using assets as an alternative to sales. A negative excess value, intuitively, suggests that the focal firm is valued less than the value of its replication peers. We refer to this method as the "Berger+Ofek Baseline" method.

D Replicating Peers: Unconstrained Text-Based Methods

We note three key limitations of the existing LS and BO method. A first is the equal treatment of all pure plays in computing the segment discount or premium. This assumption can reduce accuracy, as additional information exists on the nature of the products each pure play produces, and their comparability to a given focal firm being replicated. Our methods weight more relevant pure plays more heavily. A second limitation is the use of SIC codes to identify the universe of relevant pure play benchmarks. We expand the universe of potential peer firms to include firms identified as competitors using TNIC text-based industries. A third limitation is the focus on a single accounting characteristic, such as sales or assets. Candidate pure play firms likely vary along many other dimensions that can also explain valuation differences. For example, some pure plays might be very young, and therefore might not be relevant as a benchmark for a mature focal firm.

Henceforth, we refer to these three limitations as the "equal weighting limitation", the "limited universe limitation", and the "single characteristic limitation", respectively. Text-based methods offer a solution to all three limitations. In this section, we first examine vocabulary decompositions that address the first two limitations. We address the third limitation in the next section.

We consider relaxing each of these limitations sequentially so we can also assess the impact of each limitation. Our most basic text-based reconstruction method therefore holds fixed the set of pure-play benchmarks as used in our "Berger+Ofek Baseline" method (those in the same three-digit SIC code). However, we use a textual decomposition to determine weights based on which pure plays use product vocabulary that best matches that of the focal firm. We use these weights to replace the BO equal-weighted median calculation with a weighted median calculation. To determine the weights, we use least squares to decompose the business description of the conglomerate or single-segment focal firm being replicated into a linear combination of vocabularies from each of the replication peer firms.

Using the same notation from Section III, let M_t denote the number of unique words in the corpus, i denote a given focal firm being reconstructed, t denote the year of the given focal firm observation, and $N_{i,t}$ denote the focal firm's (M_t x 1) normalized word vector. Further suppose that the given focal firm-year observation has $N_{it,bench}$ candidate benchmark pure play firms to use in its reconstruction. Each pure play has its own normalized word vector. Let $BENCH_{it}$ denote a M_t x $N_{it,bench}$ matrix in which the normalized word vectors of the benchmark pure plays are appended as columns. We thus identify the set of pure play weights (w_{it}) that best explains the firm's observed product market vocabulary as the solution to the following least squares problem.

$$\underset{w_{it}}{MIN}(N_{it} - BENCH_{it} \cdot w_{it})^2 \tag{1}$$

The solution to this problem (w_{it}) is simply the regression slopes associated with a no-intercept regression of the conglomerate's observed word usage vector N_{it} on the word usage vectors of the $N_{it,bench}$ candidate replication peers. Importantly, unlike the BO method where pure plays are treated equally, this method assigns greater weight to pure plays whose product vocabulary best matches that of the focal firm. Imputed value is therefore computed by first computing the weighted median value to sales ratio for all $N_{it,bench}$ pure plays using the weights w_{it}. We then multiply the resulting value to sales ratio by the focal firm segment's observed sales to get the segment's imputed value. We add over segments to get the firm's total imputed value, and excess value is then equal to the natural logarithm of the firm imputed value divided by the observed focal firm value. We refer to this most basic text reconstruction, which addresses the "equal weighting limitation", as the "SIC Universe: Unconstrained" method.

Because the optimization problem in equation (1) is based on a least squares problem and has a regression analog, we can also compute an R^2 for each focal firm in each year. This R^2 effectively measures the quality of the replicating portfolio w_{it}. When this R^2 is high, the given firm is less unique and can be easily replicated by combining the resources of the replicating peer firms. We thus define "product market uniqueness" as $(1 - R^2)$. Higher uniqueness indicates that the given firm cannot be easily replicated by combining the peer firm resources.

We also consider a further generalized method with a single additional enhancement that also addresses the "limited universe limitation". In this case, we add to the pure play universe by adding pure play firms that are in the focal firm's TNIC industry as defined in Hoberg and Phillips (2010a). These firms have products that are similar to the focal firm's product description, and we also note that the TNIC industry classification is equally as coarse as are SIC-3 industries, so this does not introduce scope inconsistencies. The calculation is then analogous to that described above, except in this case the number of benchmarks $N_{it,bench}$ is as large (if no pure play TNIC peers exist) or larger (if pure play TNIC peers do exist). We refer to this method as the "SIC+TNIC Universe: Unconstrained" method.

E Replicating Peers: Constrained Text-Based Methods

We next consider the third limitation, the "single characteristic limitation". The LS and BO method has an underlying assumption that a single firm characteristic, for example sales or assets, is a sufficient statistic to explain firm valuations. Because asset valuations are forward looking and depend on many characteristics (such as firm sales, assets, and age), this limitation can be quite severe. We consider a constrained least squares approach to construct a more refined imputed value that holds any number of accounting characteristics fixed to those of the conglomerate itself.

Using the same notation, suppose a focal firm has $N_{it,bench}$ candidate pure play firms. Suppose the researcher identifies N_{char} accounting characteristics they wish to hold fixed when computing imputed valuations. In our case, we consider $N_{char} = 3$, and account for the following three accounting characteristics: log age, log sales and log assets. Let C_{it} denote a N_{char} x 1 vector containing the focal firm's actual characteristics for these three variables. Let Z_{it} denote a $N_{it,bench}$ x N_{char} matrix in which one row contains the value of these three characteristics for one of the pure play benchmark candidates. We then consider the set of weights w_{it} that solve the following constrained optimization:

$$\underset{w_{it}}{MIN}(N_{it} - BENCH_{it} \cdot w_{it})^2 \text{ such that } \; Z'_{it}w_{it} = C_{it} \tag{2}$$

The solution to this problem (w_{it}) is simply the set of slopes associated with a no-intercept constrained regression of the conglomerate's observed word usage N_{it} on the word usage vectors of the $N_{it,bench}$ pure plays. The closed form solution for these weights is:

$$w_{it} = (BENCH'_{it}BENCH_{it})^{-1}(BENCH'_{it}N_{it} - Z_{it}\lambda), \text{ where} \tag{3}$$

$$\lambda = [Z'_{it}(BENCH'_{it}BENCH_{it})^{-1}Z_{it}]^{-1}[Z'_{it}(BENCH'_{it}BENCH_{it})^{-1}BENCH'_{it}N_{it} - C_{it}]$$

Intuitively, this set of weights identifies the set of pure plays that use vocabulary that can best reconstruct the focal firm's own vocabulary, and that also exactly match the focal firm on the N_{char} characteristics. We refer to this method as the "SIC+TNIC Universe: Constrained" method.

We also note that an analog to R^2 can be computed for these constrained regressions. As with unconstrained regression, higher R^2 indicates a higher quality counterfactual fit. Hence, as we did for the unconstrained methods, we compute product market uniqueness using the constrained regression in equation (2) as $(1 - R^2)$. A higher product market uniqueness indicates a given firm cannot be easily replicated by combining the product market resources, and the basic accounting characteristics, of its replication peers.

The notion of constrained optimization as a means to hold fixed characteristics is related to the portfolio methods in Hoberg and Welch (2009) (and also relates to innovations in Fama (1976)), who show that constrained methods outperform sort-based methods in examining whether asset pricing anomalies are due to characteristics or risk factors. The goal of the current article is different: to focus on improved product market peers (not addressed in the existing studies), and to understand how product market relatedness interacts with firm valuation in both magnitude and timing. We consider basic accounting characteristics - such as size and age - which are much more exogenous that other accounting characteristics and are commonly used in many corporate finance applications.

F Replicating Peers: Accounting for Segment Sales

The discussion in this section is only relevant for conglomerate firms. The objective is to potentially account for the fact that the conglomerate segment tapes not only contain information about the industries in which a conglomerate operates, but also information about how large each segment is. This data is in the form of sales, which are reported at the segment level.

The LS and BO method computes imputed values segment-by-segment, and therefore utilizes information contained in reported segment-by-segment sales. To the extent that sales explains valuations better than other characteristics, this information might be useful. The basic text-based methods described above do not use segment-by-segment sales, and instead rely on the weights obtained from the textual reconstruction to derive imputed value. In this section, we also extend our approach to account for firm segment sales. We therefore consider a method that is identical to the "SIC+TNIC Universe: Constrained" method described above, except that we add an additional set of constraints based on the segment sales to ensure that the imputed value is weighted by sales across segments as is the case for the BO method.

Consider a conglomerate focal firm having $N_{it,seg}$ segments, and let S_{it} denote the $N_{it,seg}$ x 1 vector of sales weights (one element being a given segment's sales divided by the total sales of the conglomerate). To compute imputed values that impose segment sales-based weights, we make two modifications to the constrained optimization. First, we append the vector S_{it} to the vector C_{it}. Second, we create a $N_{it,bench}$ x $N_{it,seg}$ matrix of ones and zeros. A given element is one if the pure play associated with the given row is in the industry space corresponding to the given segment of the conglomerate focal firm associated with the given column. This matrix is populated based on how the pure-play benchmarks are selected. If the benchmark is selected due to its residing in a three digit SIC industry of a given segment, then the given pure play firm is allocated to that segment. If the benchmark is selected due to its residing in the TNIC industry of the conglomerate focal firm itself, then it is allocated to the segment to whose SIC-benchmarks it is most similar (as measured using the cosine similarity method of Hoberg and Phillips (2010a)). We then append this $N_{it,bench}$ x $N_{it,seg}$ matrix of ones and zeros to the matrix Z_{it} and re-solve the constrained optimization problem in equation (2).

The solution to the resulting constrained optimization is a set of new weights w_{it} that has the property that the sum of weights allocated to each segment equals the given segment's sales divided by the total sales of the conglomerate as a whole. Therefore, imputed values can be computed segment by segment for the focal conglomerate firm. As this is also a constrained regression, we can also compute an analogous measure of product market uniqueness as $(1 - R^2)$. We refer to this method as the "SIC+TNIC Universe: Constrained, Segment-by-Segment" method.

IV Results: Firm Valuation

In this section, we first assess the similarity of replicating peers using the reconstruction methods discussed in the previous section. Because the literature on replicating peers has a long history of focusing on conglomerates, and because the reconstruction methods materially differ for conglomerate and single-segment firms, we separately present results for conglomerate and single-segment firms. This separate reporting also ensures the comparability of our results to past studies. In particular, many past studies in the conglomerate literature consider benchmarking in the analysis of conglomerate excess valuations, and hence the results in our study can thus be more directly compared to the methods used in those studies.

Our main tests in this section examine our first hypothesis on whether the existence of high-similarity replicating peers explains firm valuations in cross section for firms of different organizational forms. In particular, we examine if firms that have high-similarity replicating peers, and thus are less unique, trade at stock market discounts while those that have low-similarity replicating peers and thus are more unique trade at stock-market premia. We examine this hypothesis for *both* conglomerate and single-segment firms.

A Methodological Validation

Following the methodology discussion in Section III, we examine excess valuations using six different replication peer identification methods. In particular, we consider six methods discussed in the previous section for identifying replicating peers: two variations of the Berger and Ofek (1995) benchmark, and four text-based methods aimed at addressing potential limitations in the BO method. Table I (conglomerates) and Table II (single-segment firms) display average excess valuations, and mean squared error statistics based on these methods.

We examine average excess valuations for comparability with existing studies, and we also consider mean squared error (MSE) statistics to compare relative valuation accuracy across valuation methods. A method with a lower MSE generates predicted valuations that are closer to actual valuations, and is therefore more accurate.

Panel A of each table presents summary statistics based on raw data. Following convention in the literature in Panel B of each table, we discard an excess value calculation if it is outside the range $\{-1.386, +1.386\}$ (in actual levels instead of natural logs this range is $\{\frac{1}{4}, 4\}$), to reduce the effect of outliers. Therefore, the observation counts available for each valuation method vary slightly as more accurate valuation methods generate excess valuations outside this range less often, and thus have higher observation counts. In Panel C, we omit a firm-year for all calculations if its estimated excess value is outside this range using any calculation method we consider (as this allows a comparison that holds the sample size fixed across all methods). Both tables report mean excess value, MSE statistics, and observation counts for excess value calculations based on sales (first three columns) and assets (last three columns).

Following conventions in the literature, we apply additional screens to the sample included in this part of our study. In particular, we require lagged COMPUSTAT data, we drop firms with sales less than $20 million, firms with zero assets, and we require that a sufficient number of pure play firms exist in segment industries to compute excess valuations. We also require that 10-K text data is available (this screen affects less than one-percent of observations). For conglomerates, we also apply one additional screen following existing studies, and discard conglomerates for which summed segment sales disagrees with the overall firm's sales by more than 1%.

[Insert Table I Here]

Table I displays results for conglomerate firms. Panel A shows that as more refined text-based valuation methods are used, the conglomerate discount almost disappears. For excess valuations based on sales, the 8.2% discount for the Berger and Ofek benchmark in row one declines to just 2.2% using the text-based method that addresses all three limitations. The discount using the BO baseline method that considers four digit SIC codes when available decreases to 6.6%. The most basic text-based benchmark, which holds fixed the same SIC-universe of pure play candidates, results in a decline in the excess value discount to 5.8%. Therefore, changing the weighting of single-segment firms from equal weighting as in Berger and Ofek to textual importance weights is partially, but not fully responsible for our ability to reduce the discount. Row 4 of Panel A expands the universe of firms eligible to receive positive weights to include the TNIC pure play rivals of the conglomerate. This expansion reduces the discount to 4.6%. Matching jointly on both the textual vocabulary dimension, and the three key accounting characteristics, row 5 shows that the discount remains at 4.6%. In row 6 of Panel A, when we further constraining the weights to match segment-specific sales ratios the discount declines rather sharply to 2.2%.

When excess valuation is based on assets in the fourth column, we see that the discount of 2.7% using the Berger and Ofek benchmark declines analogously to nearly zero (1.0%) using the unconstrained text-based benchmark in row four. We conclude that our ability to explain the benchmark is due to three factors: (1) Using weights based on textual decompositions, (2) improving the benchmark candidates to include both SIC and TNIC peers, and for the sales-based benchmark (3) constraining the benchmark to have similar accounting characteristics relative to the conglomerate's segments being reconstructed.

Columns two and four, which report mean squared error statistics, strongly support the conclusion that the constrained model based on the enlarged SIC+TNIC universe offers the most accurate set of conglomerate replicating peers. When based on sales, the mean squared error in row 5 of 0.275 reaches a minimum and is 42% smaller than the mean squared error of 0.474 associated with the Berger and Ofek benchmark. These results suggest that the improvements in accuracy are very large in economic terms.

The most relevant comparisons are in Panels B and C. In these panels, we omit excess valuations outside the interval $\{-1.386, +1.386\}$. Panel C omits the firm-year observation if any of the six valuation method places the value outside this range. In Panel C, we see the excess value discount disappearing using our text-based methods when we move to accounting characteristics matching, and we also observe mean squared error decreasing. In Panel C, the excess value discount entirely disappears for the asset based methods.

We conclude that using higher similarity replicating peers can explain the previously reported conglomerate discount, and also dramatically improve valuation accuracy. The intuition behind this result squares well with the original intent: a portfolio of pure plays that matches the conglomerate in product offerings and accounting characteristics should be a valid benchmark for the conglomerate itself. It represents a more accurate benchmark of how the conglomerate would be valued if it instead operated its segments as a portfolio of single-segment firms. Our results therefore support recent studies and do not support the conclusion that conglomerate firms trade at discounts. Other recent studies that draw the same conclusion using other methods include Campa and Kedia (2002), Villalonga (2004b), and Graham, Lemmon, and Wolf (2002). We view this result as important to illustrate that our methods are well constructed and consistent with existing studies.

We now turn to our more substantive and unique contributions where we consider (1.) the valuations of single-segment firms, (2.) the firm characteristics that influence which firms are the best matched peers for different organizational forms, (3.) whether the discounts and premia of both single-segment and conglomerate firms can be explained by firm uniqueness relative to peers, and (4.) what factors are related to the persistence of firm uniqueness.

[**Insert Table II Here**]

For single-segment firms, interestingly, Panel A of Table II shows an initial discount of nearly 5% for single-segment firms. This discount is related to the fundamental fact, shown in the next table, that firms are different with respect to characteristics such as size and profitability, even within the single-segment category. As it does for conglomerates, this discount disappears when text-based valuation methods are used. Furthermore, Panel B and Panel C show that the discount also disappears when outliers are dropped, both for the BO method and for text-based methods. That is, regardless of how outliers are handled, the excess values remain close to zero for text-based replicating peer methods.

More importantly, the MSE calculations in Table II show that text-based methods also offer improvements in valuation accuracy for single-segment firms. Moreover, the more sophisticated text-based methods offer the highest improvements. For example, MSE declines from 60.2% for the BO method using raw data in Panel A, to 50.7% using expanded TNIC and SIC peers, and then declines further to 34.2% when we further construct replicating peers based on both text and accounting data. These improvements are remarkably similar to the improvements noted above for conglomerates. For example, both single-segment firms and conglomerate firms experience a large decrease in MSE when comparing the BO benchmark to the "SIC+TNIC universe: Constrained" model. Our conclusion at this point is that text-based peer firm identification methods offer advantages in benchmarking that apply to both conglomerate and single-segment firms. These results show that even a complex firm's product offerings can be reconstructed using replicating peer firms.

B Replicating Peers and Accounting Characteristics

In Table III, we assess whether replicating peers have similar average accounting characteristics as the conglomerate (Panel A) or single-segment firm (Panel B) they intend to replicate. As the objective of these methods is to rebuild an identical replica of any focal firm using primitives, better peers should match the focal firm along many dimensions beyond valuation (discussed above). For example, they should have similar sales growth, be equally as mature, be as profitable, and have similar investment intensities as the focal firm.

To assess this prediction, we first compute the implied characteristics for each accounting variable using the same methods used to compute imputed valuations in the excess valuation calculation. For example, the implied sales growth of a Berger and Ofek (baseline) benchmark is computed as the equal weighted median of the given characteristic for pure play firms in the same three digit SIC code as the given segment. For textual methods, we simply use the weighted median sales growth using the same set of textual weights as before. This calculation is analogous for single-segment and conglomerate firms, as each simply implies a different set of weights as discussed in the methods section.

[**Insert Table III Here**]

Table III reports correlations between the actual firm characteristics and the implied replicating peer characteristics for each characteristic noted in the first column using each replication method noted in the remaining columns. Higher correlations indicate that the replication was more successful in matching the true firm for the given characteristic. For conglomerate firms, Panel A reveals that the text-based benchmarks yield higher correlations between the focal firm and its peers than the Berger and Ofek benchmark for every single characteristic assessed. Even the simplest text-based methods (that do not constrain accounting characteristics) in columns two and three have significant improvements in correlations compared to the BO correlations in the first column. For example, the 28.9% correlation between the OI/Assets for the BO benchmark increases dramatically to (35.7% to 42.1%) using these simple, unconstrained text-based weights.

As indicated in the methodology section, the unconstrained text-based weights are purely a function of the vocabulary used by the pure plays and the focal firm, and are not mechanistically related to the accounting numbers that these methods better match in these tests. In the last two columns, not surprisingly, we observe that Pearson correlations rise dramatically when we use the text-based constrained optimization. As these weights constrain the replicating peers to match the focal firm on three key accounting characteristics, it is thus not surprising that these characteristic correlations are higher. We conclude that text-based measures offer substantial improvements over existing methods.

Panel B of Table III shows that improvements in these correlations also exist for single-segment firms, but also that the improvements are less dramatic for the simplest text-based methodology. Some correlations decline slightly from the BO benchmark to the SIC-only unconstrained textual method in the second column. For example, the sales correlation dips from 20.4% to 18.1%, indicating that equally weighting peers can match somewhat better in terms of size (although the text-based methods uniformly outperform on key fundamentals, including profitability, investment style, Tobin's Q, and leverage). Despite this rather modest result for the simplest text-based replication peers, the later columns illustrate that more elaborate text-based peers outperform BO benchmarks on all characteristics, and by a large margin. We also note here that even stronger replicating peers can be constructed if we additionally match using variables such as ex ante profitability and sales growth. However we restrict attention to size and age matching characteristics in this article to ensure that our conclusions are conservative.

It is also natural to ask which type of replicating peers are weighted more than others when reconstructing conglomerates and single-segment firms. Panel A of Table IV explores this question for conglomerates, and Panel B for single-segment firms. Both panels display average accounting characteristics for the replicating peers that are assigned high weights (those in the highest quartile using the text-based conglomerate benchmarks) versus those assigned low weights (those in the lowest quartile). In particular, we construct a large database of high weighted replicating peers based on sorting the peers from each focal firm-year replication into quartiles, and extracting those in the high quartile. We build a similar database for low weighted peers, and we formally compare characteristics across the two databases to examine systematic differences in the firms receiving high versus low weights. This test is not possible using the historical Berger and Ofek method, as that method assigns equal weights to all firms. Our framework generates a strong test of peer attributes, and sheds new light on issues underlying peer selection for conglomerate firms versus single-segment firms.

The first three columns of Panel A are based on the "SIC+TNIC universe (unconstrained)" method. This method is text-based and uses an enhanced set of eligible replicating peers (SIC and TNIC peers) to reconstruct a given conglomerate. In the second three columns in Table IV, we repeat the same exercise using the "SIC+TNIC universe (constrained)" method, which also holds fixed key accounting variables when identifying replicating peers as discussed earlier.

[Insert Table IV Here]

Panel A shows that pure play firms receiving higher weights when matched to conglomerate firms using text decompositions tend to be older, more mature firms with lower sales growth. These firms also have less research and development, are more profitable, and have higher leverage relative to those pure plays matched to conglomerate firms that are assigned lower weights. Because mature firms have lower valuation ratios, this helps to explain why conglomerates appear undervalued using earlier methods.

The results in the latter three columns are similar to those in the first three columns, but are notably sharper. For example, the average difference in age is nearly 7.4 years using the constrained text method, compared to just 4.4 years using the unconstrained text method. We conclude that equally weighting all pure plays, as in the Berger and Ofek benchmark, will overweight high growth firms and thus generate the unwarranted conclusion that conglomerates are undervalued. Our results in the next section formally confirm this conjecture.

Panel B shows that similar results do not obtain for single-segment replicating peers. This result should not be surprising given that all decompositions are based only on single-segment firms to maintain a clear interpretation and to maintain consistency with earlier literature. More succinctly, replicating peers, which are limited to single-segment firms, are unlikely to be systematically different from the single-segment firms they aim to replicate. We thus do not observe material differences in the firms receiving high versus low weights for their size and profitability, and more generally significance levels and difference magnitudes are substantially smaller in Panel B for single-segment firms when compared to the conglomerate firms in Panel A.

C Stock Market Valuations and Firm Uniqueness

In this section, we examine our first central hypothesis: does firm uniqueness explain firm valuations in cross section? For the remainder of the paper, we measure firm uniqueness as one minus the R^2 from the textual decomposition used to compute replicating peers based on both product text descriptions and accounting characteristics (our constrained regression model). We also replicate all of the results of this section, and the subsequent sections, measuring firm uniqueness just based on product text descriptions without matching any accounting characteristics (our unconstrained regression model) and report these results in an online appendix.

We hypothesize that focal firms that are harder to replicate using replication peers, and are thus more unique, will have higher valuations relative to firms that are more easily replicated. In particular, firms that are harder to replicate using their product text descriptions are likely to have more unique and differentiated products, and likely face less direct product market competition as well as less severe competitive threats.

To explore this question, we regress both conglomerate and single-segment firm excess valuations on the measure of firm uniqueness generated by the text-based replication peers replication. Because the existing literature focuses extensively on the excess valuation of conglomerate firms, we examine our hypothesis within the sample of conglomerate firms and single-segment firms separately. We also include controls for document length, and accounting variables used in the existing literature. Finally, we also examine the role of accounting variables such as size and age in benchmarking and uniqueness. Thus we report results for both the unconstrained and the constrained methods discussed in the previous section.

Relevant to interpreting the tests in this section, we show in the next section that our new measure of firm uniqueness is highly persistent, even over long horizons such as ten years. Examining firm uniqueness over time, we document that patenting, patent citations, R&D, branding, and a low rate of entry (as measured by IPO or venture capital activity in the local product market) are also associated with higher levels of future firm uniqueness. These findings are consistent with firm uniqueness capturing product market protection from rivals that is also long-lived.

[**Insert Table V Here**]

Table V displays the results of OLS panel data regressions in which one observation is one conglomerate focal firm in one year (Panels A and B), or one single-segment focal firm in one year (Panels C and D). In Panels A and C, the dependent variable is the focal firm's excess valuation using the unconstrained text-based valuation method (HP: SIC+TNIC Universe (wf): Whole Firm, Unconstrained) of Table I. In Panels B and D the dependent variable is excess valuation using the Berger and Ofek (1995) valuation method. *t*-statistics are shown in parentheses, and standard errors are adjusted for clustering by firm. We also standardize all independent variables to have a standard deviation of one for ease of interpretation and comparison of coefficients. To assess the stability of firm uniqueness, we also lag the right hand side variables in three ways as noted in the first column: no lag, a one-year lag, and a three-year lag.

Our first key finding is that the firm uniqueness variable - both for single-segment firms and for conglomerate firms - is positive and highly statistically significant in all four panels. Because the independent variables are standardized, we can also interpret the coefficients to mean that a one standard deviation increase in firm uniqueness generates a 3% to 6% increase in valuation. Both conglomerates and single-segment firms that are harder to replicate have higher valuations relative to their replication peers. As the uniqueness variable captures the uniqueness of the conglomerate's products relative to its best replicating peers, one would not expect its effect on valuation to be negated out in the difference used to compute excess valuations. Unlike many variables, which have industry and firm level components, this variable is a unique property of any given focal firm that is not necessarily a property of its its industry peers.

[**Insert Table VI Here**]

Table VI displays analogous tests to those in Table V, except we consider firm uniqueness constructed using the text-based replicating peers that control for size and age (HP: SIC+TNIC Universe (wf): Whole Firm, Constrained). Overall, the results in Table VI are very similar to those in Table V. Perhaps the only notable difference is that the constrained method produces more consistent coefficients across conglomerates and single segment firms in Panels A and C, as the coefficients are very similar in magnitude. The unconstrained method in Table V, in contrast, generates stronger results for conglomerates relative to those for single segment firms. These results suggest that accounting characteristics such as size and age are likely differ across conglomerates and single segment firms, a result we confirmed earlier in Table IV. Using size and age to help form peer portfolios helps in that older firms are more likely to become conglomerates as they plausible have exercised more of their growth options than younger, smaller firms. These characteristics also have been used in many prior corporate finance studies to help find peer firm matches.

Our findings on firm uniqueness, which are robust at the 1% level of significance in all specifications, are consistent with unique firms earning higher rents due to the inability of other firms to replicate their unique products. The results for uniqueness are also highly stable, as the firm uniqueness variable experiences only modest degradation (and retains its very strong significance) when comparing the no-lag case to the three-year max lag case. These conclusions hold both in Table V and Table Table VI.

The consistent results of similar magnitude for both conglomerates and single-segment firms also indicate that uniqueness has value in many firm organizational forms. For example, barriers to entry can create a scenario in which a single-segment firm can achieve a high degree of firm uniqueness. Conglomerate firms can generate gains through this same channel, or the conglomerate structure itself can be used to assemble divisions that, when combined, are difficult to replicate by virtue of product market synergies that require multiple technologies from the multiple segments. It is also relevant to note that our results change very little if we additionally include a control for product market concentration.[9] Our control variables indicate that firms are also valued more when they have more investment (R&D and Capital Expenditures), when they are more profitable, and when they are larger.

We also find that the reported R^2s are higher in Panels B and D compared to those in Panels A and C. This result arises because our text-based valuation methods produce benchmarks that are more comparable to the given conglomerate (as shown previously). Hence, spurious differences in valuation relating to mismatched characteristics are less likely using text-based methods, as Panels A and C illustrate. Put differently, excess valuations are more difficult to predict or explain when sys-

[9] In unreported tests, we examine robustness to including product market concentration (as measured the text-based TNIC HHI variable from Hoberg and Phillips (2010a). When HHI is included, the uniqueness variable's coefficient changes little and the HHI variable itself is not statistically significant. If the uniqueness variable is removed and the HHI variable is included, the HHI variable becomes positive and significant with a t-statistic in the interval (2.0,3.0) depending on the specification. We conclude that HHI contains some common information as firm uniqueness, as one might expect, but firm uniqueness subsumes this information and is more relevant in explaining firm valuation.

tematic biases in measurement are removed. The table also shows that the level of significance of our key variable, firm uniqueness, is quite similar in all panels, and thus it is robust to changes in the replicating peer methodology, as well across both conglomerate and single-segment firms.

We next assess in Table VII the economic magnitudes of our findings for firm uniqueness captured by the difficulty of pure plays to replicate the firm. In each year, we sort firms into quintiles based on firm uniqueness, and we then compute the average excess valuation for each group. In Panels A and B, firm uniqueness is defined as 1-R^2 from the regressions finding the best set of peer firms using just product text from Table I (HP: SIC+TNIC Universe (wf): Whole Firm, unconstrained). In Panels C and D, we additionally account for the accounting variables size and age (HP: SIC+TNIC Universe (wf): Whole Firm, constrained). We also compute the average residual excess valuation, where residuals are from a regression of excess valuation on all of the variables in Table V (Panels A and B) and Table VI (Panels C and D), with the exception of the firm uniqueness variable. We compute results separately for conglomerate and single-segment firms.

[**Insert Table VII Here**]

Table VII displays the results for conglomerate firms in Panels A and C and single-segment firms in Panels B and D. Using the text-based model, the fourth column of Panel A shows that residual conglomerate excess valuations are higher for the highest uniqueness quintile (+6.3%) relative to the lowest quintile (-5.1%). This large 11.4% valuation spread increases to 13.5% in Panel C when we additionally account for accounting variables in the text-based constrained replication. This effect also persists but is weaker when the Berger and Ofek method is used to compute excess valuations as in column 5 (6.7% to 8.1% spread in the two Panels).[10] Panels B and D show analogous results for single-segment firms. The inter-quintile range is 7.9% for residual excess valuations in Panel B, and 12.8% for the constrained text-based method in Panel D. These results suggest that our results are economically

[10] We note that although the BO method can be used to compute excess valuations, it cannot be used to compute uniqueness itself, and hence this comparison only illustrates the weaker valuation identification potential associated with the BO method.

large, and also that there are some additional gains to controlling for accounting variables.

Overall, our results are consistent with firms having higher valuations when they are more unique with product configurations that are more difficult to replicate. This suggests that such firms extract more value through differentiated product offerings that cannot be easily raided by potential rival peer firms. Going further, in unreported tests, we find that firm uniqueness does not predict abnormal stock returns. Hence, the high valuations associated with firms which are more difficult to replicate are likely more permanent, and reflect the stock market recognizing the value of firm uniqueness that is based on firm fundamentals such as protection from rivals.

V Persistence of Product Market Uniqueness

Given we show that firms that are more unique have higher valuations, we now explore two questions on the time-series persistence of uniqueness. First, we examine how long uniqueness lasts. Second, we consider what factors drive changes in firm uniqueness. To explore the time-series persistence of uniqueness, we examine simple correlation statistics over time.

To examine time-series persistence, we consider Pearson correlation coefficients between current uniqueness and uniqueness measured one to ten years later for each firm. Firm Uniqueness is one minus the R^2 from the regression that constructs peer firm portfolios based on best matches of the focal firm's product description and accounting characteristics (HP: SIC+TNIC Universe (wf): Whole Firm, Constrained) of Table I. The results are reported in Table VIII. The appendix presents the same results using firm uniqueness constructed without matching accounting characteristics (HP: SIC+TNIC Universe (wf): Whole Firm, Unconstrained) of Table I. We find that uniqueness is highly persistent over one year, as next-year uniqueness is 84.3% correlated with current uniqueness. In addition, uniqueness appears to decay at a slow rate over time, as uniqueness today remains 66.1% correlated with uniqueness ten years later.

[Insert Table VIII Here]

Given this persistence in uniqueness, we next examine the factors that are related to uniqueness in time series. We consider three factors: brand names, innovative activity as measured by R&D and patenting activity, and entry by new firms. This question is important because it allows us to examine whether some forms of expenditures are longer-lasting than others, and hence, more influential in explaining our valuation results. As Sutton (1991) emphasizes, firms spend money on advertising and R&D to differentiate themselves and to build endogenous barriers to entry.

We consider a firm's brands as measured by each firm's use of brand-specific vocabulary in its 10-K. Specifically, we define brand vocabulary as the words indicated by the brand list website: http://www.namedevelopment.com/brand-names.html.

To assess R&D and innovative activity, we use the amount of money firms spend on R&D and also two measures of its patenting activity. Using the NBER patent database, we include the number of patents for which a firm has applied, and the number of forward looking citations to its patents. We control for document length, firm size, and time and firm fixed effects. The firm fixed effect controls are important to highlight, as they ensure that our identification is within-firm and hence our results are not driven by unobserved firm characteristics. We also cluster standard errors by firm.

In Table IX, we examine whether these branding and innovation variables drive changes in firm uniqueness over time. We further consider whether these variables continue to affect uniqueness in a long-lasting fashion one to three years into the future.

[Insert Table IX Here]

Table IX shows that, after one year, the primary drivers of future uniqueness are a firm's branding activity, the number of patents, and especially the degree of successful innovation as measured by the number of patent citations. In particular, the results suggest that R&D spending only influences longer-term uniqueness if it generates successful innovation. The other columns indicate that the single best factor explaining future uniqueness is successful innovation, as only patent citations remain significant over two and three year horizons. These intuitive results shed new light on how firms maintain their market position over longer periods of time. Thus, the key to a firm's maintenance of product uniqueness is not only to spend on R&D, but also to generate successful high quality patents.

We also examine if venture capital and IPO activity in a given firm's product market relate to own-firm product uniqueness. We follow Hoberg, Phillips, and Prabhala (2014) and consider text-based measures of VC and IPO activity. These variables measure and give a score to the level of IPO and VC activity in the local product market based on how similar a firm's product market text is to the text of firms undertaking IPOs, or those receiving venture capital financing (this private firm product text is obtained from the SDC Platinum database). We report the results for just venture capital activity in Table IX, but we obtain similar results if we instead consider the corresponding IPO-score variable based on IPO activity in the local product market.

We find that firms in product markets where less VC activity has occurred in the past (and in the present) have higher ex-post uniqueness. The results are strong, and are still highly significant if we predict two-year-ahead product market uniqueness. These results are intuitive given IPO and VC activity are forms of financed entry of rival firms, and any barriers to such entry that result in fewer new firms obtaining financing should improve ex-post uniqueness.

In the on-line appendix in Table A2 we present the results where we only match based on product text and not on accounting characteristics. Patent cites and venture capital financed entry remain significant in explaining uniqueness in this auxiliary test. One change in the results is that the brand vocabulary results and patents applied are not significant in explaining uniqueness in this auxiliary text. This difference in results suggests that controlling for size and age is relevant, as there is variation in the necessary time to build brand value, and it also takes time and larger firm resources to develop successful brands. This comparison also highlights the value of the research flexibility inherent to the constrained text-based method, which is flexible enough to match based on any set of accounting characteristics.

A Organizational Forms and the Persistence of Uniqueness

Table X examines if the factors that drive product market uniqueness differ across single-segment and conglomerate firms.

[Insert Table X Here]

Table X shows that our previous innovation and branding determinants of uniqueness matter in similar ways for both single-segment and conglomerate firms. However, the table also reveals two important differences. First, the brand vocabulary coefficient is significantly lower for conglomerate firms than for single-segment firms. This result suggests that brands of single-segment firms are more sensitive to branding investments. Second, we find that the effect of patent citations on uniqueness is significantly higher for conglomerate firms than for single-segment firms. Citations thus contribute to longer lasting uniqueness especially for conglomerate firms where such technologies might be more able to be used across multiple markets.

The differential importance of patent citations and branding across single-segment and conglomerate firms confirms the intuition from earlier parts of our paper, which suggests that conglomerate firms are generally not comparable to single-segment firms. The results suggest that conglomerates rely more on successful past innovation and precluding entry by innovative rivals to maintain their uniqueness, whereas single-segment firm uniqueness is sensitive to investment in branding. Interestingly, in Table A3 of our on-line appendix we present results that show these differences across organizational form remain when constructing uniqueness based on product text alone and not matching based on accounting characteristics. Thus, when comparing organizational forms, it is less important to control for size and age, as the organizational forms themselves have firms of more similar size and age.

Taken together with our earlier findings from Table IV, a more comprehensive view of organizational form emerges. Table IV shows that, relative to peers, conglomerates are more mature, have lower Tobin's qs, less overall R&D, and are older and larger. Because informational asymmetry is typically reduced for these more mature firms, and product innovation rates are also typically low, our results suggest that conglomerates aim to use technological synergies that span multiple product markets (see Bena and Li (2014) for supporting evidence) to differentiate their products. This strategy might also allow conglomerates to create sustainable barriers to entry that single-segment firms cannot easily overcome, thus maintaining product uniqueness and longer-term profitability. The creation of technological and scope-based monopolies also makes it clear why successful entry by newly financed venture-backed firms or IPOs is among the biggest threats to these firms as they may offer less mature rivals new ways to innovate.

In contrast, our results from Table IV show that single-segment firms are younger, invest more, have higher Tobin's qs, and are smaller. Their product offerings are more focused, and hence, they cannot as readily benefit from technologies that improve the cost structure or the product features of more complex baskets of products like those offered by a conglomerate. Hence, these firms compete in more focused markets with more focused rivals. In this environment, given that cross-market technology is not relevant, high product branding may offer a more viable alternative for these firms to help them establish higher valuations.

VI Conclusions

We examine how the stock market uses information about firm and peer firm product offerings using text-based computational methods. We examine this question using a new method that identifies portfolios of peer firms using text-based analysis that assigns different *weights* based on how each peer contributes to creating a near replica on the basis of both product offerings and accounting characteristics for a given firm under consideration. This method provide a closed form solution that identifies the best set of replicating peers for any conglomerate or single-segment firm. This calculation generates a new measure of firm uniqueness based on the extent to which the replica's representative product market vocabulary matches the firm being replicated.

We find that firms whose products and characteristics are more unique (difficult for industry peer firms to replicate) have higher stock market valuations. These higher valuations are thus based on firm fundamentals. These results hold for both conglomerate and single-segment firms, consistent with both types of firms being more highly valued by investors when they have more unique products.

Examining the time-series persistence of uniqueness, we find that firm uniqueness is highly persistent over long ten-year horizons, consistent with these firms being able to maintain a competitive advantage. We find that the primary drivers of a firm's ability to maintain a high level of uniqueness over long periods of time are patent citations and the lack of entry into a firm's product market. Although R&D spending is highly correlated with simultaneously measured uniqueness, it does not correlate strongly with future uniqueness in a setting where we control for within-firm variation and patent citations. Only realized successful innovation (high patent counts and especially patent citations) strongly correlate with future uniqueness. Intuitively, R&D spending alone cannot ensure successful maintenance of uniqueness. Rather, R&D spending must produce successful innovation as measured by patents with citations. R&D can thus be viewed as a risky form of investment whose long-term impact on the firm's uniqueness depends on realized success in developing new products or new product features that rivals cannot replicate.

Our results show that the stock market values firm uniqueness and shows what factors contribute to firms maintaining uniqueness and product differentiation. Overall, our methodology and new peer groupings allow more accurate benchmarking of competitor firms. These new benchmark peer firms should be useful in many other corporate finance applications or asset pricing research where performance benchmarking or counterfactual analysis is important.

References

Bena, Jan, and Kai Li, 2014, Corporate innovations and mergers and acquisitions, *forthcoming Journal of Finance.*

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

Berry, Steven, James Levinsohn, and Ariel Pakes, 1997, Automobile prices in market equilibrium, *Econometrica* 63, 841–890.

Campa, Jose, and Simi Kedia, 2002, Explaining the diversification discount, *Journal of Finance* 57, 1731–1762.

Chamberlin, EH, 1933, *The Theory of Monopolistic Competition* (Harvard University Press: Cambridge).

Chevalier, Judith A., 1995, Capital structure and product-market competition: Empirical evidence from the supermarket industry, *American Economic Review* 85, 415–435.

Comment, Robert, and Gregg Jarrell, 1995, Corporate focus and stock returns, *Journal of Financial Economics* 37, 61–87.

Custodio, Claudia, 2010, Mergers and acquisitions accounting can explain the diversification discount, Arizona State University Working Paper.

Denis, David, Diane Denis, and Keven Yost, 2002, Global diversification, industrial diversification, and firm value, *Journal of Finance* 57, 1951–1979.

Fama, Eugene, 1976, *Foundations of Finance* (Basic Books).

Graham, John, Michael Lemmon, and Jack Wolf, 2002, Does corporate diversification destroy value?, *Journal of Finance* 57, 695–720.

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

——— , and N.R. Prabhala, 2014, Product market threats, payouts, and financial flexibility, *forthcoming Journal of Finance.*

Hoberg, Gerard, and Ivo Welch, 2009, Optimized vs. sort based portfolios, working paper, University of Maryland and UCLA.

Khanna, Naveen, and Sheri Tice, 2000, Strategic responses of incumbents to new entry: The effect of ownership structure, capital structure and focus, *Review of Financial Studies* 13, 749–779.

Lamont, Owen, and Christopher Polk, 2002, Does diversification destroy value? evidence from the industry shocks, *Journal of Financial Economics* 63, 51–77.

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

Leary, Mark, and Michael Roberts, 2010, Do peer firms affect corporate capital structure?, Working Paper, University of Pennsylvania.

Lins, Karl, and Henri Servaes, 1999, International evidence on the value of corporate diversification, *Journal of Finance* 54, 2215–2240.

MacKay, Peter, and Gordon M. Phillips, 2005, How does industry affect firm financial structure?, *Review of Financial Studies* 18, 1433–66.

Phillips, Gordon M., 1995, Increased debt and industry product markets: An empirical analysis, *Journal of Financial Economics* 37, 189–238.

Rajan, Raghuram G., Henri Servaes, and Luigi Zingales, 2000, The cost of diversity: the diversification discount and inefficient investment, *Journal of Finance* 55, 35–80.

Rauh, Joshua, and Amir Sufi, 2010, Explaining corporate capital structure: Product markets, leases, and asset similarity, Northwestern University Working Paper.

Seim, Katja, 2006, An empirical model of firm entry with endogenous product choices, *Rand Journal of Economics* 37, 619–40.

Servaes, Henri, 1996, The value of diversification during the conglomerate merger wave, *Journal of Finance* 51, 1201–1225.

Stein, Jeremy, 2003, Agency, information and corporate investment, *Handbook of the Economics of Finance* pp. 110–63.

Sutton, John, 1991, *Sunk Costs and Market Structure* (MIT Press: Cambridge, Mass).

Villalonga, Belen, 2004a, Diversification discount or premium? new evidence from business information tracking series, *Journal of Finance* 59, 479–506.

——— , 2004b, Does diversification cause the diversification discount, *Financial Management* 33, 5–27.

Wernerfelt, Birger, and Cynthia Montgomery, 1988, Diversification, ricardian rents, and tobin's q, *Rand Journal of Economics* 19, 623–632.

Table I: Quality of Excess Valuation Calculations Across Methods (Conglomerates)

This table displays summary statistics for conglomerate benchmark valuations. Panel A is based on all conglomerates, Panel B restricts attention to conglomerate firms with excess valuations within the interval $\{ln(-4), ln(4)\}$, and Panel C restricts attention to observations for which all methods generate excess valuations within this range (this holds the sample size fixed). The **Berger+Ofek Baseline** benchmarks are based on Berger and Ofek (1995). The specification with a "(SIC-3)" label only considers firms in the same SIC-3 as benchmarks. The specification with a "(Variable SIC)" label uses SIC-4 as benchmarks, but substitutes for SIC-3 when the number of peers is fewer than five. The **SIC Universe: Whole Firm, Unconstrained** benchmarks use text-based weights to construct the benchmarks. The **HP: SIC+TNIC Universe: Whole Firm, Unconstrained** benchmarks extend this method by expanding the set of available pure plays to include TNIC peers. The **HP: SIC+TNIC Universe (wf): Whole Firm, Constrained** benchmarks extend this method further using constrained regression to match the conglomerate on log age, log sales, and log assets. The **HP: SIC+TNIC Universe: Constrained, Segment-by-Segment** benchmarks additionally account for segment-by-segment sales.

		Excess	MSE		Excess	MSE		
		Value	Excess Val.	# Obs.	Value	Excess Val.	# Obs.	Std. Dev.
Row	Benchmark	(Sales	(Sales	(Sales	(Assets	(Assets	(Assets	Weights
		Based)	based)	based)	Based)	based)	based)	
	Panel A: Raw Data							
1	Berger+Ofek Baseline (SIC-3)	-0.082	0.474	12714	-0.027	0.288	10916	
2	Berger+Ofek Baseline (Variable SIC)	-0.066	0.459	12714	-0.031	0.292	10916	
3	HP: SIC Universe (wf): Unconstrained	-0.058	0.463	12714	-0.037	0.256	10916	0.041
4	HP: SIC+TNIC Universe (wf): Unconstrained	-0.046	0.402	12733	-0.010	0.229	10928	0.031
5	HP: SIC+TNIC Universe (wf): Constrained	-0.046	0.275	12693	-0.011	0.224	10874	0.042
6	HP: SIC+TNIC Universe (ss): Constrained: Segment-by-Segment	-0.022	0.416	12641	-0.014	0.246	10844	0.049
	Panel B: Restrict to Excess Valuations to interval [-1.386,+1.386] (Berger and Ofek)							
7	Berger+Ofek Baseline (SIC-3)	-0.070	0.335	11786	-0.067	0.212	8695	
8	Berger+Ofek Baseline (Variable SIC)	-0.063	0.338	11880	-0.071	0.213	8686	
9	HP: SIC Universe (wf): Unconstrained	-0.048	0.339	11755	-0.034	0.215	8730	0.041
10	HP: SIC+TNIC Universe (wf): Unconstrained	-0.038	0.310	11913	-0.015	0.192	8746	0.030
11	HP: SIC+TNIC Universe (wf): Constrained	-0.041	0.235	12132	-0.013	0.189	8749	0.042
12	HP: SIC+TNIC Universe (ss): Constrained: Segment-by-Segment	-0.021	0.307	11843	-0.015	0.204	8725	0.050
	Panel C: Uniformly Restrict to interval [-1.386,+1.386]							
13	Berger+Ofek Baseline (SIC-3)	-0.061	0.298	10911	-0.047	0.182	7568	
14	Berger+Ofek Baseline (Variable SIC)	-0.054	0.303	11078	-0.055	0.187	7678	
15	HP: SIC Universe (wf): Unconstrained	-0.036	0.304	10911	-0.016	0.188	7593	0.040
16	HP: SIC+TNIC Universe (wf): Unconstrained	-0.023	0.268	10911	0.002	0.169	7610	0.030
17	HP: SIC+TNIC Universe (wf): Constrained	-0.025	0.200	10911	0.001	0.167	7617	0.041
18	HP: SIC+TNIC Universe (ss): Constrained: Segment-by-Segment	-0.008	0.277	10911	0.001	0.184	7603	0.049

Table II: Quality of Excess Valuation Calculations Across Methods (single-segment Firms)

This table displays summary statistics for single-segment benchmark valuations. Panel A is based on all single-segment firms, Panel B restricts attention to those single-segment firms with excess valuations within the interval $\{ln(-4), ln(4)\}$, and Panel C restricts attention to observations for which all methods generate excess valuations within this range (this holds the sample size fixed). The **Berger+Ofek Baseline** benchmarks are based on Berger and Ofek (1995). The specification with a "(SIC-3)" label only considers firms in the same SIC-3 as benchmarks. The specification with a "(Variable SIC)" label uses SIC-4 as benchmarks, but substitutes for SIC-3 when the number of peers is fewer than five. The **SIC Universe: Whole Firm, Unconstrained** benchmarks use text-based weights to construct the benchmarks. The **HP: SIC+TNIC Universe: Whole Firm, Unconstrained** benchmarks extend this method by expanding the set of available pure plays to include TNIC peers. The **HP: SIC+TNIC Universe (wf): Whole Firm, Constrained** benchmarks extend this method further using constrained regression to match the conglomerate on log age, log sales, and log assets.

Row	Benchmark	Excess Value (Sales Based)	MSE Excess Val. (Sales based)	# Obs. (Sales based)	Excess Value (Assets Based)	MSE Excess Val. (Assets based)	# Obs. (Assets based)	Std. Dev. Weights
	Panel A: Raw Data							
1	Berger+Ofek Baseline (SIC-3)	-0.047	0.602	37579	0.045	0.339	37578	
2	Berger+Ofek Baseline (Variable SIC)	-0.030	0.551	37584	0.023	0.321	37583	
3	HP: SIC Universe (wf): Unconstrained	0.012	0.579	37575	0.051	0.354	37574	0.053
4	HP: SIC+TNIC Universe (wf): Unconstrained	-0.019	0.507	37583	0.044	0.320	37582	0.028
5	HP: SIC+TNIC Universe (wf): Constrained	0.010	0.342	37709	0.045	0.309	37708	0.056
	Panel B: Restrict to Excess Valuations to interval [-1.386,+1.386] (Berger and Ofek)							
6	Berger+Ofek Baseline (SIC-3)	-0.019	0.354	34638	0.022	0.262	36635	
7	Berger+Ofek Baseline (Variable SIC)	-0.013	0.365	35533	0.009	0.264	36735	
8	HP: SIC Universe (wf): Unconstrained	0.008	0.343	34744	0.030	0.268	36407	0.053
9	HP: SIC+TNIC Universe (wf): Unconstrained	-0.007	0.330	35254	0.024	0.250	36667	0.029
10	HP: SIC+TNIC Universe (wf): Constrained	0.005	0.269	36526	0.030	0.245	36687	0.051
	Panel C: Uniformly Restrict to interval [-1.386,+1.386]							
11	Berger+Ofek Baseline (SIC-3)	-0.026	0.306	31958	0.034	0.210	31620	
12	Berger+Ofek Baseline (Variable SIC)	-0.017	0.315	32538	0.016	0.221	32224	
13	HP: SIC Universe (wf): Unconstrained	0.011	0.297	31958	0.040	0.223	31603	0.051
14	HP: SIC+TNIC Universe (wf): Unconstrained	-0.008	0.278	31958	0.033	0.208	31651	0.028
15	HP: SIC+TNIC Universe (wf): Constrained	0.012	0.220	31958	0.037	0.207	31700	0.049

Table III: Characteristic Correlations (SIC-3 Peer Firms vs. Text-Based Peer Firms)

The table displays Pearson Correlation coefficients between actual focal firm characteristics and the characteristics of different sets of peer firms. We consider several different replicating peer methods as noted in the column headers. Panel A reports results for conglomerate focal firms and Panel B reports results for single-segment focal firms.

						Text-based
		Berger and Ofek	Text-based	Text-based	Text-based	SIC+TNIC
		SIC Peer firms	SIC only	SIC+TNIC	SIC+TNIC	Constrained
Row	Variable	(Baseline)	No Constr.	No Constr.	Constrained	(Seg by Seg)
			Panel A: Correlation Coefficients: Conglomerates			
1	Assets	0.110	0.194	0.291	0.623	0.619
2	Sales	0.156	0.229	0.385	0.695	0.662
3	Oi/Sales	0.375	0.425	0.479	0.605	0.498
4	Oi/Assets	0.289	0.357	0.421	0.507	0.424
5	R&D/Sales	0.473	0.673	0.705	0.770	0.685
6	Tobin's Q	0.366	0.442	0.469	0.493	0.458
7	Sales Growth	0.241	0.270	0.309	0.332	0.297
8	Log Age	0.268	0.298	0.436	0.932	0.911
9	Book Leverage	0.403	0.418	0.462	0.476	0.435
			Panel B: Correlation Coefficients: single-segment Firms			
10	Assets	0.100	0.063	0.282	0.622	N/A
11	Sales	0.204	0.184	0.368	0.678	N/A
12	OI/Sales	0.402	0.463	0.508	0.638	N/A
13	OI/Assets	0.131	0.190	0.203	0.403	N/A
14	R&D/Sales	0.403	0.594	0.637	0.762	N/A
15	Tobin's Q	0.225	0.330	0.295	0.476	N/A
16	Sales Growth	0.316	0.308	0.360	0.380	N/A
17	Log Age	0.330	0.283	0.392	0.909	N/A
18	Book Leverage	0.472	0.491	0.538	0.545	N/A

Table IV: Which Replicating Peers Match with Conglomerates and single-segment Firms?

The table displays summary statistics for replicating peers assigned above median weights versus below median weights for conglomerate focal firms (Panel A), and single-segment focal firms (Panel B).

		Benchmark Portfolio Weights vs Characteristics					
		SIC+TNIC Universe: Whole Firm, Un-constrained			*SIC+TNIC Universe: Whole Firm, Constrained*		
Row	Variable	Lowest Weights Quartile	Highest Weights Quartile	t-statistic of Difference	Lowest Weights Quartile	Highest Weights Quartile	t-statistic of Difference
			Panel A: Conglomerates				
1	Assets	3418.20	4692.02	6.29	3323.48	5586.47	5.91
2	Sales	1557.47	2091.51	9.57	1542.09	2406.60	9.22
3	oi/sales	0.07	0.08	5.95	0.06	0.08	9.39
4	oi/assets	0.07	0.07	0.63	0.07	0.08	1.06
5	R&D/Sales	0.11	0.09	-13.75	0.11	0.09	-13.33
6	Tobin's Q	2.06	1.92	-5.26	2.07	1.92	-4.33
7	Sales Growth	0.17	0.16	-9.19	0.17	0.16	-11.80
8	Firm Age	25.21	28.96	18.68	24.11	30.51	27.29
9	Book Leverage	0.20	0.21	11.26	0.20	0.21	10.41
			Panel B: single-segment Firms				
10	Assets	7305.50	6584.34	-1.51	7713.76	7046.06	-3.58
11	Sales	1437.45	1392.45	-0.76	1575.01	1423.03	-5.04
12	oi/sales	0.15	0.15	0.35	0.16	0.15	-5.17
13	oi/assets	0.05	0.05	0.78	0.05	0.05	-5.47
14	R&D/Sales	0.08	0.08	-2.38	0.08	0.08	2.19
15	Tobin's Q	1.44	1.42	-2.05	1.43	1.43	-1.19
16	Sales Growth	0.15	0.15	-4.28	0.15	0.15	-1.46
17	Firm Age	23.90	24.29	2.31	24.62	23.99	-6.42
18	Book Leverage	0.18	0.19	4.71	0.18	0.18	1.15

Table V: Uniqueness and Stock-Market Valuations (Unconstrained Text-Based Model)

OLS regressions examining the relation between firm uniqueness and conglomerate and single-segment firm excess valuations. The dependent variable is the focal firm's excess valuation computed using the unconstrained text-based reconstruction based on SIC and TNIC peers ("HP: SIC+TNIC Universe: Unconstrained" method) (Panel A for conglomerate firms and Panel C for single-segment firms) or using the Berger and Ofek reconstruction (Panel B for conglomerate firms and Panel D for single-segment firms). Firm uniqueness is defined as 1-R^2 from the regressions finding the best set of peer firms using just product text from Table I (HP: SIC+TNIC Universe (wf): Whole Firm, unconstrained). The right hand side variables are lagged in three different ways as specified: no lag, one year lag, and 3-year max lag. The 3-year max lag uses the three year lag if available, and otherwise the two year lag if available, and otherwise the one year lag. We include the log of the product description document length, R&D, capital expenditures (CAPX) and operating income (OI) divided by sales and the natural log of firm assets. All independent variables are standardized to have a standard deviation of one for ease of interpretation and comparison of coefficients. All regressions have time fixed effects and standard errors that are adjusted for clustering by firm. Footnotes a, b, c indicate a result is significantly different from zero at the 1%, 5%, 10% level, respectively.

			Log					
	Independent	Firm	Document	R&D/	CAPX/	OI/	Log	# Obs. /
Row	Variable Lag	Uniqueness	Length	Sales	Sales	Sales	Assets	RSQ
		Panel A: Conglomerate Firms Excess Values (Text-based Unconstrained Valuation Model)						
(1)	No Lag	0.053^a	-0.028^a	0.100^a	0.071^a	0.168^a	0.101^a	11,279
		(5.09)	(-2.80)	(10.29)	(8.11)	(15.05)	(9.45)	0.166
(2)	1-year Lag	0.058^a	-0.029^b	0.105^a	0.050^a	0.204^a	0.099^a	8,563
		(4.83)	(-2.51)	(8.29)	(4.61)	(14.51)	(8.30)	0.181
(3)	3-Year Max Lag	0.056^a	-0.022^c	0.099^a	0.054^a	0.188^a	0.101^a	9,062
		(4.14)	(-1.72)	(7.17)	(4.92)	(12.51)	(7.64)	0.162
		Panel B: Conglomerate Firms Excess Values (Berger + Ofek Valuation Model)						
(4)	No Lag	0.046^a	0.007	0.104^a	0.067^a	0.131^a	0.125^a	11,090
		(3.83)	(0.65)	(10.54)	(6.50)	(11.41)	(10.72)	0.147
(5)	1-Year Lag	0.056^a	0.008	0.103^a	0.050^a	0.161^a	0.124^a	8,423
		(4.00)	(0.66)	(8.44)	(4.23)	(11.72)	(9.48)	0.156
(6)	3-Year Max Lag	0.058^a	0.013	0.097^a	0.046^a	0.161^a	0.123^a	8,919
		(3.98)	(0.95)	(7.15)	(4.00)	(11.31)	(9.01)	0.146
		Panel C: Single-segment Firms Excess Values (Text-based Unconstrained Valuation Model)						
(7)	No Lag	0.037^a	-0.005	0.147^a	0.078^a	0.136^a	0.122^a	32,164
		(6.50)	(-0.77)	(19.31)	(13.76)	(16.91)	(18.70)	0.121
(8)	1-year Lag	0.032^a	-0.003	0.142^a	0.062^a	0.140^a	0.114^a	25,112
		(5.06)	(-0.48)	(15.91)	(10.03)	(13.68)	(15.35)	0.109
(9)	3-Year Max Lag	0.026^a	0.001	0.143^a	0.060^a	0.138^a	0.122^a	26,224
		(3.68)	(0.14)	(14.99)	(8.98)	(12.64)	(15.08)	0.113
		Panel D: Single-segment Firms Excess Values (Berger + Ofek Valuation Model)						
(10)	No Lag	0.038^a	0.021^a	0.206^a	0.113^a	0.163^a	0.139^a	31,451
		(6.44)	(3.31)	(26.00)	(21.36)	(18.32)	(19.85)	0.181
(11)	1-Year Lag	0.032^a	0.018^a	0.202^a	0.095^a	0.172^a	0.132^a	24,608
		(4.84)	(2.61)	(21.46)	(16.44)	(15.45)	(16.77)	0.171
(12)	3-Year Max Lag	0.022^a	0.021^a	0.197^a	0.088^a	0.167^a	0.141^a	25,798
		(3.08)	(2.78)	(20.24)	(14.70)	(14.72)	(16.59)	0.171

Table VI: Uniqueness and Stock-Market Valuations (Constrained Text-Based Model)

OLS regressions examining the relation between firm uniqueness and conglomerate and single-segment firm excess valuations. The dependent variable is the focal firm's excess valuation computed using the constrained text-based reconstruction based on SIC and TNIC peers ("HP: SIC+TNIC Universe: Constrained" method) (Panel A for conglomerate firms and Panel C for single-segment firms) or using the Berger and Ofek reconstruction (Panel B for conglomerate firms and Panel D for single-segment firms). Firm uniqueness is defined as 1-R^2 from the regressions finding the best set of peer firms using just product text from Table I (HP: SIC+TNIC Universe (wf): Whole Firm, Constrained). The right hand side variables are lagged in three different ways as specified: no lag, one year lag, and 3-year max lag. The 3-year max lag uses the three year lag if available, and otherwise the two year lag if available, and otherwise the one year lag. We include the log of the product description document length, R&D, capital expenditures (CAPX) and operating income (OI) divided by sales and the natural log of firm assets. All independent variables are standardized to have a standard deviation of one for ease of interpretation and comparison of coefficients. All regressions have time fixed effects and standard errors that are adjusted for clustering by firm. Footnotes a, b, c indicate a result is significantly different from zero at the 1%, 5%, 10% level, respectively.

Row	Independent Variable Lag	Firm Uniqueness	Log Document Length	R&D/ Sales	CAPX/ Sales	OI/ Sales	Log Assets	# Obs. / RSQ
			Panel A: Conglomerate Firms Excess Values (Text-based Constrained Valuation Model)					
(1)	No Lag	0.046^a	-0.003	0.067^a	0.033^a	0.133^a	0.028^a	11,453
		(4.90)	(-0.36)	(7.57)	(4.43)	(13.54)	(2.79)	0.089
(2)	1-year Lag	0.049^a	-0.008	0.072^a	0.023^a	0.140^a	0.033^a	8,657
		(4.53)	(-0.78)	(6.31)	(2.76)	(11.65)	(2.95)	0.089
(3)	3-Year Max Lag	0.041^a	-0.011	0.058^a	0.026^a	0.123^a	0.034^a	9,086
		(3.60)	(-1.01)	(4.74)	(3.15)	(9.69)	(2.91)	0.072
			Panel B: Conglomerate Firms Excess Values (Berger + Ofek Valuation Model)					
(4)	No Lag	0.042^a	0.032^a	0.081^a	0.076^a	0.132^a	0.093^a	10,925
		(3.25)	(2.59)	(7.63)	(7.38)	(10.83)	(7.79)	0.127
(5)	1-Year Lag	0.042^a	0.023	0.078^a	0.062^a	0.162^a	0.089^a	8,266
		(2.91)	(1.66)	(5.80)	(5.17)	(11.15)	(6.55)	0.131
(6)	3-Year Max Lag	0.041^a	0.031^b	0.070^a	0.057^a	0.163^a	0.082^a	8,811
		(2.65)	(2.10)	(4.87)	(4.73)	(10.94)	(5.78)	0.122
			Panel C: single-segment Firms Excess Values (Text-based Constrained Valuation Model)					
(7)	No Lag	0.046^a	0.010^c	0.098^a	0.041^a	0.111^a	0.038^a	33,348
		(8.79)	(1.86)	(14.94)	(10.20)	(15.94)	(6.59)	0.053
(8)	1-year Lag	0.046^a	0.006	0.097^a	0.030^a	0.110^a	0.034^a	26,008
		(7.69)	(1.12)	(12.56)	(6.69)	(13.20)	(5.21)	0.045
(9)	3-Year Max Lag	0.043^a	0.005	0.089^a	0.027^a	0.097^a	0.041^a	26,802
		(6.41)	(0.85)	(10.95)	(5.54)	(10.73)	(5.75)	0.039
			Panel B: single-segment Firms Excess Values (Berger + Ofek Valuation Model)					
(10)	No Lag	0.049^a	0.034^a	0.152^a	0.057^a	0.094^a	0.136^a	31,376
		(7.59)	(5.50)	(20.07)	(10.75)	(11.82)	(19.40)	0.108
(11)	1-Year Lag	0.042^a	0.023^a	0.150^a	0.045^a	0.103^a	0.131^a	24,567
		(5.79)	(3.31)	(17.16)	(7.65)	(10.75)	(16.62)	0.100
(12)	3-Year Max Lag	0.033^a	0.024^a	0.146^a	0.037^a	0.105^a	0.139^a	25,868
		(4.10)	(3.17)	(15.55)	(6.09)	(10.19)	(15.86)	0.101

Table VII: Economic Magnitudes: Firm Uniqueness and Excess Valuation

This table displays average excess valuations for quintiles based on firm uniqueness. In Panels A and B, firm uniqueness is defined as 1-R^2 from the regressions finding the best set of peer firms using just product text from Table I (HP: SIC+TNIC Universe (wf): Whole Firm, unconstrained). In Panels C and D, we additionally account for the accounting variables size and age (HP: SIC+TNIC Universe (wf): Whole Firm, constrained). Panels A and C display statistics for conglomerate focal firms, and Panels B and D display results for single-segment focal firms. For each quintile, we report the average firm uniqueness variable, and average raw excess valuations based on both the respective text based methods and the Berger and Ofek method. Residual excess valuations are residuals from a regression of excess valuation on all of the variables included in Table V (Panels A and B) and Table VI (Panels C and D) excluding the firm uniqueness variable.

		Raw		Residual		
		Excess	Raw	Excess	Residual	
Firm		Valuation	Excess	Valuation	Excess	
Uniqueness	Firm	(text-	Valuation	(text-	Valuation	
Quintile	Uniqueness	based)	(Berger+Ofek)	based)	(Berger+Ofek)	Obs.
	Summary Statistics by Quintile (Unconstrained Text-Based Method)					
	Panel A: Conglomerates					
Lowest	0.617	-0.055	-0.025	-0.051	-0.018	2,318
Quintile 2	0.713	-0.020	-0.055	0.002	-0.021	2,325
Quintile 3	0.773	-0.031	-0.093	0.001	-0.017	2,327
Quintile 4	0.825	-0.024	-0.083	0.009	-0.007	2,325
Highest	0.895	-0.016	-0.072	0.063	0.063	2,321
	Panel B: Single-segment Firms (Unconstrained Text-Based Method)					
Lowest	0.652	0.020	0.030	-0.051	-0.047	6,792
Quintile 2	0.790	0.036	0.041	-0.001	-0.004	6,801
Quintile 3	0.842	0.022	-0.012	0.026	-0.002	6,803
Quintile 4	0.883	-0.022	-0.068	0.020	0.000	6,801
Highest	0.936	-0.056	-0.083	0.028	0.053	6,794
	Panel C: Conglomerates (Constrained Text-Based Method)					
Lowest	0.610	-0.081	-0.027	-0.069	-0.016	2,310
Quintile 2	0.710	-0.034	-0.058	0.002	-0.016	2,319
Quintile 3	0.775	-0.036	-0.090	0.003	-0.009	2,316
Quintile 4	0.837	-0.017	-0.113	0.024	-0.009	2,319
Highest	0.950	-0.000	-0.087	0.066	0.051	2,312
	Panel D: Single-segment Firms (Constrained Text-Based Method)					
Lowest	0.454	-0.015	0.004	-0.064	-0.070	6804
Quintile 2	0.640	0.023	0.044	-0.012	-0.006	6810
Quintile 3	0.718	0.001	-0.013	-0.003	0.002	6806
Quintile 4	0.796	0.005	-0.026	0.019	0.038	6810
Highest	0.958	0.026	-0.090	0.064	0.035	6805

Table VIII: Product Uniqueness Time Series Correlations (one to ten years)

The table displays Pearson Correlation coefficients between firm uniqueness and various lags (as noted) of firm uniqueness for our panel from 1997 to 2008. Firm Uniqueness is one minus the R^2 from the regression that constructs peer firm portfolios based on best matches of the focal firm's product description and accounting characteristics. Footnotes a, b, c indicate a result is significantly different from zero at the 1%, 5%, 10% level, respectively.

			Product	Product	Product	Product	Product
		Product	Unique.	Unique.	Unique.	Unique.	Unique.
		Unique.	Lagged	Lagged	Lagged	Lagged	Lagged
Row	Variable	(No Lag)	1 Year	2 Years	3 Years	5 Years	9 Years
			Pearson Correlation Coefficients				
(1)	Product Uniqueness (Lagged 1 Year)	0.8435^a					
(2)	Product Uniqueness (Lagged 2 Years)	0.7990^a	0.8511^a				
(3)	Product Uniqueness (Lagged 3 Years)	0.7728^a	0.8057^a	0.8558^a			
(4)	Product Uniqueness (Lagged 5 Years)	0.7289^a	0.7616^a	0.7936^a	0.8249^a		
(5)	Product Uniqueness (Lagged 9 Years)	0.6799^a	0.6924^a	0.7357^a	0.7542^a	0.8142^a	
(6)	Product Uniqueness (Lagged 10 Years)	0.6611^a	0.6814^a	0.6902^a	0.7458^a	0.7859^a	0.9058^a

Table IX: Time Series Determinants of Product Uniqueness (All Firms)

This table uses OLS regressions to examine the effect of R&D, branding, and patenting activity on firm product uniqueness. The dependent variable is current period product uniqueness in the first column. In columns two to four, the dependent variable is one-year, two-year, and three-year out ex-post product uniqueness. The explanatory variables include firm level R&D divided by sales, the fraction of the firm's 10-K business description words that are associated with well-known brands (brand vocabulary), the number of patents a firm applied for in that year, the number of forward looking cites to these patents, and VC Score, which is the similarity between the text of the firm's 10-K business description and that of the business descriptions of firms funded by venture capitalists (from SDC Platinum). We also include the product description length and the natural log of firm assets. All specifications include industry, year and firm fixed effects. Standard errors in parentheses and are adjusted for firm clustering. Footnotes a, b, c indicate a result is significantly different from zero at the 1%, 5%, 10% level, respectively.

	Time t=0	Time t=1	Time t=2	Time t=3
	Product	Product	Product	Product
Dependent Variable	Uniqueness	Uniqueness	Uniqueness	Uniqueness
R&D/Sales	-0.0065	-0.0060	-0.0055	-0.0083^c
	(0.005)	(0.005)	(0.005)	(0.005)
Brand Vocabulary	0.0641^a	0.0295^a	0.0139^c	0.0105
	(0.008)	(0.007)	(0.007)	(0.008)
# Patents Applied	0.0188^a	0.0075^c	-0.0024	-0.0050
	(0.004)	(0.004)	(0.004)	(0.004)
# patent cites	0.0057^c	0.0114^a	0.0165^a	0.0142^a
	(0.003)	(0.003)	(0.003)	(0.003)
VC Score	-0.1281^a	-0.0612^a	-0.0298^a	-0.0069
	(0.006)	(0.006)	(0.006)	(0.006)
Document Length	-0.0645^a	-0.0336^a	-0.0257^a	-0.0136^b
	(0.007)	(0.007)	(0.006)	(0.007)
Log Assets	-0.1835^a	-0.1666^a	-0.1056^a	-0.0584^a
	(0.017)	(0.018)	(0.018)	(0.020)
Log Age	-0.1724^a	-0.0752^a	-0.0347^c	-0.0039
	(0.016)	(0.016)	(0.019)	(0.026)
Constant	-0.1820^a	-0.1384^a	-0.0761	-0.0824
	(0.056)	(0.053)	(0.069)	(0.054)
R^2	0.1286	0.0721	0.0528	0.0391
N	44792	36576	29785	24183

Table X: Time Series Determinants of Product Uniqueness for Conglomerate and Single-Segment Firms)

This table uses OLS regressions to examine the effect of R&D, branding, and patenting activity on firm product uniqueness. Columns one to three include only single-segment firms and columns four to six include only conglomerate firms. The dependent variable is current period product uniqueness in the first and fourth column. In columns (two and five) and (three and six), the dependent variable is one-year and two-year out ex-post product uniqueness, respectively. The explanatory variables include firm level R&D divided by sales, the fraction of the firm's 10-K business description words that are associated with well-known brands (brand vocabulary), the number of patents a firm applied for in that year, and the number of forward looking cites to these patents, and VC Score, which is the similarity between the text of the firm's 10-K business description and that of the business descriptions of firms funded by venture capitalists (from SDC Platinum). We also include the product description length and the natural log of firm assets. All specifications include industry, year and firm fixed effects. Standard errors in parentheses and are adjusted for firm clustering. Footnotes a, b, c indicate a result is significantly different from zero at the 1%, 5%, 10% level, respectively. Footnotes d, e, f indicate a result is significantly different across conglomerates and single-segment firms at the 1%, 5%, 10% level, respectively.

	single-segment Firms			Conglomerate Firms		
	Time t=0	Time t=1	Time t=2	Time t=0	Time t=1	Time t=2
	Product	Product	Product	Product	Product	Product
Dependent Variable	Uniqueness	Uniqueness	Uniqueness	Uniqueness	Uniqueness	Uniqueness
R&D/Sales	-0.0068	-0.0090^{c}	-0.0089^{c}	-0.0195	-0.0056	0.0131
	(0.005)	(0.005)	(0.005)	(0.025)	(0.032)	(0.035)
Brand Vocabulary	0.0928^{a,e}	0.0475^{a,e}	0.0302^{a,d}	0.0087^{e}	-0.0060^{e}	-0.0252^{b,d}
	(0.010)	(0.009)	(0.009)	(0.013)	(0.013)	(0.013)
# Patents Applied	0.0207^{a}	0.0105^{b}	0.0012	0.0154^{b}	0.0042	-0.0055
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
# patent cites	0.0023^{e}	0.0084^{b}	0.0155^{a}	0.0095^{c,e}	0.0144^{b}	0.0166^{a}
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.005)
VC Score	-0.1117^{a,d}	-0.0498^{a,d}	-0.0205^{a}	-0.1395^{a,d}	-0.0691^{a,d}	-0.0465^{a}
	(0.007)	(0.006)	(0.006)	(0.010)	(0.010)	(0.011)
Document Length	-0.0588^{a}	-0.0263^{a}	-0.0178^{b}	-0.0670^{a}	-0.0396^{a}	-0.0358^{a}
	(0.008)	(0.008)	(0.007)	(0.013)	(0.012)	(0.012)
Log Assets	-0.1692^{a}	-0.1588^{a}	-0.1078^{a}	-0.2205^{a}	-0.2159^{a}	-0.1255^{a}
	(0.018)	(0.020)	(0.021)	(0.040)	(0.043)	(0.043)
Log Age	-0.1820^{a}	-0.1062^{a}	-0.0666^{a}	-0.1968^{a}	-0.0401	-0.0036
	(0.021)	(0.020)	(0.022)	(0.029)	(0.026)	(0.038)
Constant	-0.2886^{a}	-0.2276^{a}	-0.1487^{c}	0.1908^{b}	0.1604^{b}	0.2500^{a}
	(0.083)	(0.081)	(0.077)	(0.078)	(0.070)	(0.092)
R^2	0.1283	0.0734	0.0572	0.1533	0.0829	0.0729
N	33347	27040	21852	11445	9536	7933