Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis

Gerard Hoberg
University of Maryland, Robert H. Smith School of Business

Gordon Phillips
University of Maryland, Robert H. Smith School of Business

We use text-based analysis of 10-K product descriptions to examine whether firms exploit product market synergies through asset complementarities in mergers and acquisitions. Transactions are more likely between firms that use similar product market language. Transaction stock returns, ex post cash flows, and growth in product descriptions all increase for transactions with similar product market language, especially in competitive product markets. These gains are larger when targets are less similar to acquirer rivals and when targets have unique products. Our findings are consistent with firms merging and buying assets to exploit synergies to create new products that increase product differentiation. (*JEL* G14, G34, L22, L25)

It has long been viewed that product market synergies are key drivers of mergers. Mergers are a quick way to potentially increase product offerings if synergies arise from asset complementarities. One important dimension of synergies is the ability of merging firms to create new products and differentiate themselves from rivals when merging firms have complementary assets. Rhodes-Kropf and Robinson (2008) model similarity and asset complementarities as a motive for mergers but do not present direct evidence of their importance.

We are especially grateful to Michael Weisbach (the Editor) and an anonymous referee for very helpful suggestions. We thank Alex Edmans, Michael Faulkender, Denis Gromb, Dirk Hackbarth, Kathleen Hanley, Rich Matthews, Nagpurnanand Prabhala, David Robinson, Matthew Rhodes-Kropf, Paul Tetlock, and seminar participants at Duke University, ESSEC, HEC-Paris, Hong Kong University of Science and Technology, Imperial College, Insead, ISTCE (Lisbon), Lausanne, London Business School, National University of Singapore, NBER, Ohio State University, the Securities Exchange Commission, Stockholm School of Economics, the University of Amsterdam, the University of North Carolina, the University of Texas, the University of Maryland, Vienna Graduate School of Finance, and Washington University for helpful comments. Any errors are the authors' alone.
Send correspondence to Gerard Hoberg or Gordon Phillips, University of Maryland, Robert H. Smith School of Business, College Park, MD 20742; telephone: 301-405-9685 or 301-405-0347; fax: 301-405-0359. E-mail: ghoberg@rhsmith.umd.edu or gphillips@rhsmith.umd.edu.

doi:10.1093/rfs/hhq053 Advance Access publication August 15, 2010

Little is actually known about the extent to which new product synergies and asset complementarities impact mergers.

Our article provides evidence of synergies and new product creations using new methods based on textual analysis. We examine whether firms can improve profit margins by merging with a target that is (i) similar to itself and offers asset complementarities, but is also (ii) different from its *rivals*. Our new measures of product similarity and new product introductions come from text-based analysis of 49,408 10-K product descriptions obtained from the Securities Exchange Commission (SEC) Edgar website. The importance of synergies and product differentiation to merger success relates to whether the target has skills or technologies that can help the acquirer differentiate its product offerings from its rivals and improve profitability.[1]

Using our new text-based methods, we find evidence consistent with merging firms that have high ex ante similarity and high ex ante differentiation from the acquirer's rivals increasing cash flows and new product introductions. This evidence is consistent with product market synergies being important for ex post merger success. Our measures use vector representations of the text in each firm's product description to develop a Hotelling-like product location space for U.S. firms.[2] Distances between firms on this 10-K–based product location space measure the degree to which firms are different from rivals (within and across industries), allowing independent measurement of the similarity between acquirers and targets. Henceforth, we classify firms as "similar" when they are close to each other in distance in this product location space and "different" when a potential target is far away from a potential acquirer.

Our new measures allow us to test how asset complementarities (measured using acquirer and target similarity) interact with product market competition (measured using the relative crowding of rivals around a firm), to give firms the incentive to conduct mergers to differentiate their products from their rivals. Our new measures are unique in that they capture the relatedness of firms in the product market space, and they have the ability to measure both within and across industry similarity. In contrast, neither the Standard Industry Classification (SIC) nor the North American Industry Classification System (NAICS) has a corresponding spatial representation or a continuous representation of the pairwise similarity of any two firms.

We first report two new stylized facts. First, merger pairs are far more similar than SIC- or NAICS-based measures would suggest. Even merger pairs in different two-digit SIC codes (which are common) generally have high

1 Many papers show that related mergers have higher ex post cash flows (see Andrade, Mitchell, and Stafford 2001 and Betton, Eckbo, and Thorburn 2008 for two surveys). Potential explanations also include increased cost-efficiencies, decreased agency costs, increased pricing power, and synergies. The methodology in our article can help distinguish among these explanations.

2 In addition to Hotelling's (1929) linear representation of product differentiation, Lancaster (1966) and Salop (1979) show that product differentiation can be represented in a multidimensional spatial framework, and that the degree of differentiation is fundamental to profitability and theories of industrial organization.

similarities with each other. Second, we detect a high degree of heterogeneity in the degree of similarity across merger pairs, allowing substantial power to test our hypotheses. We emphasize that all of our results are robust to controls for horizontal and vertical relatedness used in the existing literature.[3]

We report three central findings. First, transaction incidence is higher for firms that are more broadly similar to all firms in the economy and lower for firms that are more similar to their local rivals. We interpret the first incidence result as an "asset complementarity effect," since a firm that is more broadly similar has more opportunities for pairings that can generate synergies. We refer to the second incidence result as a "competitive effect," since firms with very near rivals must compete for restructuring opportunities given that a potential partner can view its rivals as substitute partners. These incidence results are significant economically and statistically for both large and small firms.

Second, long-term real outcomes are better (higher profitability, sales, and evidence of new product introductions) when the target and the acquirer are more pairwise similar. Our results, especially those related to new product introductions and sales growth, suggest that merging firms with similar assets use synergies from asset complementarities to introduce new products and improve cash flows.

Third, outcomes are better when acquirers reside in ex ante competitive product markets, and when the transaction increases the acquirer's differentiation relative to its rivals.

We differ from the existing literature in two major ways. First, the literature has not focused on how asset complementarities and product differentiation jointly affect acquisition choices, subsequent performance, and new product development. Second, we present an innovative new methodology based on text analysis that allows us to examine hypotheses that cannot be tested using the SIC-code-based approach taken by existing studies. Our work is related to that of Hoberg and Phillips (2009), who employ the same text-based empirical framework to form dynamic industry classifications, which significantly outperform existing classifications in explaining firm characteristics in cross section. Measures of product market competition based on these industries also dominate alternatives in explaining observed firm profitability (central to theories of industrial organization). These findings are relevant to our own analysis of mergers, which requires that measures of product market competition based on our product market space are both powerful and consistent with theory.

Our research contributes to the literature on mergers, similarity, asset complementarity, and industrial organization. Healy, Palepu, and Ruback (1992) and Andrade, Mitchell, and Stafford (2001) have documented increased industry-adjusted cash flows following mergers. However, the literature has

[3] For example, Fan and Goyal (2006) show that mergers that are vertically related using the input-output matrix have positive announcement wealth effects in the stock market.

not been able to identify whether asset complementarities allowing firms to introduce new products are responsible for gains in cash flows. Rhodes-Kropf and Robinson (2008) model asset complementarity and synergies as a motive for mergers. Our evidence is consistent with their model. We discuss the related literature in greater detail in the next section.

We also add to a new, growing literature that uses text analysis in finance. Hanley and Hoberg (2010) address theories of initial public stock offering (IPO) pricing by examining prospectus disclosures on the SEC Edgar website, and separate text into standard and informative content. Loughran and McDonald (2009) examine the SEC's Plain English rules on corporate finance outcomes and disclosure. Outside of corporate finance, other studies that use text-based analysis to study the role of the media in stock price formation include those by Tetlock (2007), Macskassy, Saar-Tsechansky, and Tetlock (2008), Li (2006), and Boukus and Rosenberg (2006).

The remainder of the article is organized as follows: Section 1 discusses merger strategies and incentives to merge, Section 2 presents the data and methodology used in this article and a summary of statistics, Section 3 introduces our new empirical measures of product market similarity, Section 4 presents determinants of the likelihood of restructuring transactions, Section 5 discusses ex post outcomes upon the announcement of a merger and in the long term, and Section 6 offers our conclusions.

1. Incentives to Merge and Merger Outcomes

This section discusses the existing literature on mergers, and develops our hypotheses of how potential changes to product differentiation and competition may affect both a firm's decision to merge and its post-merger performance.

1.1 Relation to Previous Literature on Mergers

Our article focuses on how high ex ante product market competition creates incentives to merge, and how mergers may create profits through synergies and subsequent product differentiation. The central idea is that firms may wish to merge with partners with complementary assets that expand their range of products through new product introductions (enabling them to differentiate from rival firms), while also picking partners that are related enough so that they can skillfully manage the new assets.

This rationale for mergers is distinct from the existing motives in the finance literature.[4] Hackbarth and Miao (2009) present a new theory examining

[4] Existing reasons for mergers include technological industry shocks and excess industry capacity (Morck, Shleifer, and Vishny 1988; Jensen 1993; Mitchell and Mulherin 1996; Andrade and Stafford 2004; Harford 2005), reduction of agency problems (Jensen 1993), agency and empire building (Hart and Moore 1995), demand shocks and efficiency (Maksimovic and Phillips 2001; Jovanovic and Rousseau 2002; Yang 2008), industry life cycle (Maksimovic and Phillips 2008), and to reallocate corporate liquidity (Almeida, Campello, and Hackbarth 2009).

the role of synergies in an oligopoly setting, and Rhodes-Kropf and Robinson (2008) also consider synergies through asset complementarities as a motive for mergers. Neither explores how competition impacts the motive to merge, and neither empirically examines the role of ex post new product introductions. We use text-based methods to directly measure similarity and potential asset complementarity of merger partners rather than inferring it using market-to-book ratios. Perhaps more importantly, using new techniques, our article can measure potential synergies, how similar merging firms are from rivals, and how product descriptions grow post-merger.

In addition to increases in product differentiation, key to understanding ex post performance after mergers are measures of relatedness of the target and acquirer. As emphasized by many authors, related acquisitions have the potential to perform better, since the acquirer is likely to have existing skill in operating the target firm's assets. Kaplan and Weisbach (1992) show that related mergers are less likely to be divested subsequently by the acquirer, although they do not find any difference in the performance of diversifying compared with non-diversifying mergers. Maksimovic, Phillips, and Prabhala (2008) find that acquirers with more skill in particular industries are more likely to maintain and increase the productivity of the assets they acquire and keep in related industries. Fan and Goyal (2006) show that mergers that are vertically related using the input-output matrix have positive announcement wealth effects in the stock market (although Kedia, Ravid, and Pons 2008 show that this reverses after 1996). While partially informative, SIC codes cannot measure the degree to which firms are similar within and across industries. Our study shows that many merging firms have highly related product market text even though they have different two-digit SIC codes that are not related. Most importantly, discrete measures of relatedness cannot capture how related rivals are to the merging firms both within and across industries.

Research in industrial organization has also studied mergers in industries with existing differentiated products. Baker and Bresnahan (1985) theoretically model how mergers can increase pricing power by enabling post-merger firms to increase prices. The prescribed policy is for acquirers to merge with firms whose products are close substitutes to their own, where nonmerging firms produce only distant substitutes. Pricing power is enhanced by merging because post-merger firms face an increased steepness in their residual (inverse) demand curves. More recently, following Berry, Levinsohn, and Pakes (1997), the approach of the product differentiation literature has been to estimate demand and cost parameters in specific markets, including the ready-to-eat cereal market (Nevo 2000).

In this article, however, we focus on the incentives for firms to merge to *differentiate* themselves and to exploit asset complementarities. Using text-based measures, we examine how firms are related to each other without relying on predefined industries. We can measure product similarity across arbitrary firms and in time series, allowing us to test for new product introductions more broadly, and across multiple industries. Our analysis is related conceptually to the empirical question raised by Berry and Waldfogel (2001), who studied competition in the radio broadcasting market and showed that the number of broadcasting formats increased post merger. Our approach is also related to other recent studies, including those by Mazzeo (2002) and Seim (2006), who, while not examining mergers, also focus on the incentives for firms to differentiate themselves.

1.2 Hypotheses

We now develop specific hypotheses based on the aforementioned theories of merger pair similarity and industrial organization. Our first two hypotheses concern the probability that a given firm will become part of an acquisition. Our third hypothesis formulates predictions regarding ex post outcomes.

Hypothesis 1 (H1): Asset Similarity: *Firms are more likely to merge with other firms whose assets are highly similar or related to their own assets.*

Hypothesis 2a (H2a): Differentiation from Rivals: *Acquirers in competitive product markets are more likely to choose targets that help them to increase product differentiation relative to their nearest ex ante rivals.*

Hypothesis 2b (H2b): Competition for Targets: *Firms with high local product market competition are less likely to be targets of restructuring transactions given the existence of multiple substitute target firms.*

Key to testing these hypotheses is the degree of similarity. The rationale for H1 is that firm management has certain skills or abilities that can be applied to the similar assets without decreased returns to scope. H2a captures the effect that acquirers will wish to merge to differentiate themselves from existing competition. H2b suggests that competition will reduce the likelihood that any given firm will merge, because it must share the likelihood of participating in a given opportunity with its highly similar rivals. Key to testing H1 and H2b is the degree of similarity. The effects of H2b are likely to be stronger where firms are very similar. For example, identical firms would have to compete in order to merge with a third firm offering new synergies requiring their technology. In contrast, the effects of H1 are more likely to hold where firms are moderately similar, because such firms are less viable substitutes.

Our remaining hypothesis focuses on long-term outcomes.

Hypothesis 3 (H3): Synergies through Asset Complementarities: *Acquirers buying targets similar to themselves are likely to have asset complementarities and experience future higher profitability, sales growth, and new product introductions.*

The reasons why acquirer and target pairwise similarity should result in positive outcomes, as in H3, are numerous, as discussed in Section 1.1. Our primary rationale is that, in addition to these motives, new products developed using technology from similar targets with complementary (similar) assets are also more likely to succeed, as shown by Rhodes-Kropf and Robinson (2008). We also note that key to the success of asset complementarities is the ability to differentiate the acquirer from existing rivals.[5] Overall, key to maximizing gains is finding a target that is both similar to self and different from rivals.

We illustrate these hypotheses using the merger of General Dynamics (GD) and Antheon (the “GDA” merger). We base our discussion on actual similarity data (described in Section 2). Figure 1 displays the GDA merger and the ten nearest rivals surrounding both firms. Firms with a label “G” are General Dynamics’ closest rivals, and firms with an “A” are Antheon’s closest rivals. This merger illustrates a key benefit of our methodology over SIC codes: Although they make highly related products, these firms have different two-digit SIC codes.

Both firms offer products that are indeed related (Antheon is GD’s eleventh closest rival and GD is Antheon’s second closest) but are also somewhat different. GD focuses on large “old economy” military equipment, including land vehicles and ships. Antheon serves the same primary client but focuses on military applications based on computing and intelligence. Figure 1 also shows that GD faced competition from rivals, including Lockheed, United Defense, and Northrup Grumman.

The GDA merger is consistent with GD choosing Antheon because it is similar enough to permit asset complementarities and new product synergies (H3), as well as successful managerial integration. Here, new products might take the form of military vehicles that incorporate real-time applications of Antheon’s technology, given the growing role of information in defense. This merger also might help GD to differentiate itself from its rivals and introduce new products, which in turn will improve profit margins. Importantly, this notion of similar but different is underscored by comparing the chosen target with a hypothetical merger between GD and one of its other near rivals, such as Northrup Grumman, which likely offers fewer ways to differentiate GD’s products from its closest rivals.

In addition, in Figure A1 of the Online Appendix, we show a visual representation of the Disney and Pixar merger. These firms were classified as being in different two-digit SIC codes. However, our similarity measure shows a rich array of relatedness prior to the merger as they both are in each other’s 10 most similar firms and they also share many of the same closest 10 firms. Thus the Disney-Pixar merger is consistent with Disney choosing Pixar because it is similar enough to permit asset complementarities and new product synergies. Following the merger, many Disney computer-animated movies (e.g., *Toy Story*,

[5] Recent (Mazzeo 2002; Seim 2006) and historical studies in industrial organization beginning with Hotelling (1929) suggest that firms that differentiate themselves should be more profitable.

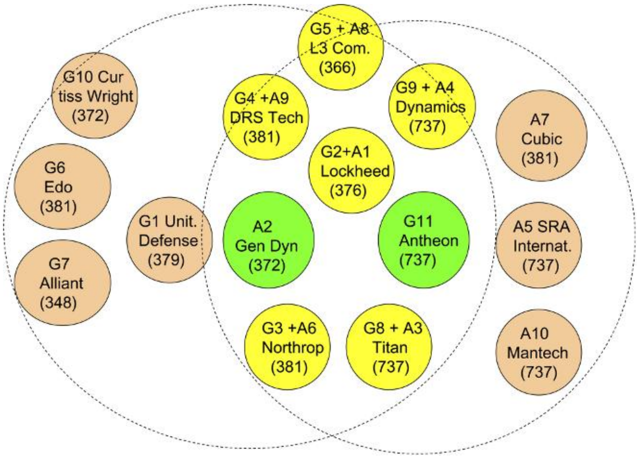

Figure 1
The large dashed circles give a visual depiction of General Dynamics' and Antheon's ten closest rival firms determined using our measure of product similarity described in Section 2.
Antheon is General Dynamics' eleventh-closest rival, and General Dynamics is Antheon's second-closest rival. Each additional firm shown has a header beginning with the letter "G" or "A" followed by a number. This identifies which firm's circle of ten nearest rivals the given firm exists in, and also how close the given firm is to either General Dynamics or Antheon. For example, Alliant has a code "G7" and is thus General Dynamics' seventh-closest rival. The random firm pairings group is based on the pairwise similarity of two randomly drawn firms. For consistency with the other groups, we draw the random firms from the set of firms that participated in acquisitions (results are nearly identical if firms are randomly drawn from the set of firms that did not participate in acquisitions). For each firm, we also report its primary three-digit SIC code in parentheses.

A Bug's Life, Cars) have been produced using Pixar technology and distributed by the merged company.

2. Data and Methodology

2.1 Data Description

A key innovation of our study is that we use firms' 10-K text product descriptions to compute continuous measures of product similarity for every pair of firms in our sample (a pairwise similarity matrix).[6] We construct these text-based measures of product similarity (described later in this section) using firm product descriptions obtained directly from 10-K filings on the SEC Edgar website. Our primary sample includes filings associated with firm fiscal years ending in calendar years 1997 to 2006. We also compute similarities for 1996 and 2007 but use the 1996 data to compute only the starting value

[6] We also considered whether variables based on the properties section of the 10-K might help to separate asset similarities and product similarities. After carefully examining this section for a sample of 100 documents, however, we concluded that the properties section almost exclusively describes real estate holdings and the location of the firm's corporate headquarters, rendering it inapplicable for this purpose.

of lagged variables and the 2007 data to compute only the values of ex post outcomes.[7]

We create our primary firm-year database by merging the 10-K database with the Compustat/Center for Research in Security Prices (CRSP) database using the central index key (CIK), which is the primary key used by the SEC to identify the issuer.[8] We then link this firm-level database to the Securities Data Company (SDC) Platinum database of mergers and acquisitions to create a transaction-level database.

We electronically gather 10-Ks by searching the SEC Edgar database for filings that appear as "10-K," "10-K405," "10KSB," and "10KSB40." The Edgar database for the years prior to 1997 is less complete because electronic filing was not required until 1997. Of the 56,540 firm-year observations with fiscal years ending in 1997–2006 that are present in both CRSP and Compustat, we are able to match (using the CIK number) 55,326 (97.9% of the CRSP/Compustat sample). We can also report that our database is well balanced over time, since we capture 97.6% of the eligible data in 1997, and 97.4% in 2006, and this annual percentage varies only slightly in the range of 97.4% in 2006 to 98.3% in 2001. Because we do not observe much time variation in our data coverage, and because database selection can be determined using ex ante information (i.e., the 10-K itself), we do not believe that our data requirements induce any bias. Our final sample size is 50,104 rather than 55,326 because we additionally require that lagged Compustat data items (assets, sales, and operating cash flow) be available before observations can be included in our analysis.

For each linked 10-K, we extract the product description section. This section appears as Item 1 or Item 1A in most 10-Ks. We use a combination of Perl web-crawling scripts and APL [A Programming Language] to extract and process this section. The web-crawling algorithm scans the Edgar website and collects the entire text of each 10-K annual report, and APL text-reading algorithms then extract its product description and other identifying information, including its CIK number. This process is supported by human intervention when non-standard document formats are encountered. This method is highly reliable and we encountered only a very small number of firms that we were not able to process because they did not contain a valid product description or because the product description had fewer than 1,000 characters.

---

7 Although we use data for fiscal-year endings through 2007, we extract documents filed through December 2008, because many of the filings in 2008 are associated with fiscal years ending in 2007. This is because 10-Ks are generally filed during the three-month window after the fiscal year ends.

8 We thank the Wharton Research Data Service (WRDS) for providing us with an expanded historical mapping of SEC CIK to Compustat gvkey, since the base CIK variable in Compustat contains only current links. Although somewhat rare, a small number of links have changed over time. Our results are robust to using either the WRDS mapping or the Compustat mapping.

2.2 Product Similarity

We consider three measures to evaluate the similarity of the text in product descriptions: basic cosine similarity, local cosine similarity, and broad cosine similarity. Basic cosine similarity is a widely used method for evaluating textual similarity (see, e.g., Kwon and Lee 2003). Local cosine similarity is motivated by the concept of "cliques" in social networks and is analogous to the local clustering measure (see Watts and Strogatz 1998; Granovetter 1973). Broad cosine similarity is the complement of local similarity, and focuses on words that tend to cluster less. The local and broad measures add power to our tests and improve the focus of our interpretations.[9]

Although we provide an overview in this section, we provide a complete technical description of similarity calculations in the Appendix. We start by building a list of all unique words used in all documents in a given year. We then discard all common words (those used in more than 5% of all 10-Ks). The resulting list of N non-common words is the "main dictionary." We then represent each firm as an N-vector summarizing its usage of the N words, and we normalize each vector to unit length. Intuitively, each firm's normalized vector lies on a unit sphere, and hence this method generates an empirical product market space analogous to a Hotelling grid, with all firms having a specific address. The cosine similarity of these vectors is bounded in the range [0,1], and firms having descriptions with more words in common have a higher similarity. The normalization avoids over-scoring larger documents.

The local and broad cosine similarities are computed in an analogous fashion, except that they use different dictionaries. Local similarity is based on words that tend to appear with a common group of related words (i.e., they appear in "word cliques"). Intuitively, product market words are likely to have this property. For example, the words *cola*, *beverage*, and *drink* should often appear together in a 10-K product description if they appear at all. We provide many examples of words in the "local dictionary" in Section 3. We believe that words in the local dictionary are strongly represented by words that are unique to local product markets.

We divide all words in the main dictionary into the local dictionary and the broad dictionary. The broad words are not common words, which have been screened out earlier, as discussed. Rather, they are non-common words that do not generally appear in cliques. Although these words are generally not specific to a local product market, they are often indicative of a firm having assets that can be easily redeployed to other product markets. Examples of words in the broad dictionary consistent with this interpretation include *container, painting, pallet, frames, forecasting, learning, fabric,* and *salespeople*. On the other hand, the broad dictionary is far more likely to contain noisy words with

[9] However, we also note that our study's results are robust to just employing measures based on the basic cosine similarity measure, as was done in an earlier draft. We thank an anonymous referee for motivating us to consider more targeted measures.

less economic content, including *hereby, recognizable, instrumental, reflection,* and *insert*. We construct the following variables for each firm-year in our main database:

Product Similarity (10 Nearest): For a given firm i, this variable is the average pairwise similarity (using the local dictionary) between firm i and its ten most similar rivals j. The closest rivals are the ten firms with the highest local similarity to i.

Broad Similarity (All Firms): For a given firm i, this variable is the average similarity (using the broad dictionary) between firm i and all other firms j in the sample.

% Neighbor Patent Words: For each firm i, this is the percentage of words in its product description having the word roots "patent," "copyright," and "trademark."

Last Year 10 Nearest Fraction Restructured: For firm i, this variable is the fraction of its ten closest rivals (using the local dictionary) that were either targets or acquirers in the previous year according to the SDC Platinum database.

We compute the following variables for each transaction in our transaction database:

Target + Acquirer Product Similarity: For a given merger pair (target and acquirer), this is their pairwise similarity (using the local dictionary).

Gain in Product Differentiation: This variable is the target's average pairwise distance from the acquirer's ten nearest rivals, minus the acquirer's average distance from its ten nearest rivals, all using the local dictionary. This variable measures the degree to which an acquirer gains product differentiation from its rivals by purchasing the given target.

Given our article's focus on local product markets, we use the local dictionary to construct all but one variable. Because this list has fewer noisy words, this approach increases the accuracy of our measures, as well as the clarity of their interpretation. Although the one measure based on the broad dictionary (we interpret broad similarity as asset redeployability) might be subject to greater noise in its measurement, this concern is offset by the fact that this variable is also constructed as a similarity averaged over far more firm pairings, which reduces noise.

In an Online Appendix, we test whether our similarity variables are valid measures of product market competition. This validation is based on a large body of theoretical and empirical literature predicting that higher product similarity to near rivals will be associated with lower profitability. This literature includes recent models that endogenize product choice (Mazzeo 2002; Seim 2006), and historical studies that first considered product differentiation

(Chamberlin 1933; Hotelling 1929). We find strong evidence that our product similarity variable is negatively related to profitability. This test is also related to the study by Hoberg and Phillips (2009), who employ the same empirical product market space to form new industry classifications and test links to profitability. These findings are relevant because our discussion relies on this validated link between 10-K product similarity and product market competition.

Product similarities are most intuitive when firms have only one segment. Our computer-based algorithms are not able to separate the text associated with each segment of conglomerate firms.[10] However, we believe that similarities measured relative to conglomerates are still informative regarding the competition faced by the firm in all of its segments. To the extent that multiple segments might add noise to our measures, we do not see this as a problem, since it would only bias our results away from finding significant results. However, to ensure that our inferences are robust, we also rerun all of our tests using the subsample of single segment firms. Although not reported to conserve space, our results change little in this subsample.[11]

2.3 Other Control Variables

Past studies seeking to measure product differentiation have been forced to rely on industry definitions based on SIC codes. We thus control for the following standard measures of industry competition and merger similarity based on SIC codes:

Sales Herfindahl-Hirschman Index (HHI) (SIC-3): We use the two-step method described by Hoberg and Phillips (2010) to compute sales-based Herfindahl ratios for each three-digit SIC code. This method uses data based on both public and private firms to compute the best estimate of industry concentration given the limited data available on private firm sales.

Last Year SIC-3 % Restructured: For a given firm i, we compute the percentage of firms in its three-digit SIC code that were involved in restructuring transactions in the previous year according to the SDC Platinum database.

Same SIC-3 Industry Dummy: For a given merger pair, this variable is 1 if the two firms are in the same three-digit SIC code.

Vertical Similarity Dummy: For a given merger pair, this variable is 1 if the two firms are at least 5% vertically related. We use the methodology

[10] Multiple segments appear in product descriptions, but automated text-reading algorithms cannot separate the text attributable to each due to the high degree of heterogeneity in how segment descriptions are organized.

[11] In some cases, significance levels decline from the 1% level to the 5% or 10% level due to reduced power.

described in Fan and Goyal (2006) to construct this variable. In particular, based on four-digit SIC codes of both the target and the acquirer, we use the Use Table of Benchmark Input-Output Accounts of the US Economy to compute the fraction of the inputs that are from the other firm's SIC industry. If this percentage exceeds 5% for either firm, then the dummy is set to 1.

Although SIC codes are informative in many applications, we present evidence that SIC codes only weakly measure competition and merger similarity. We also include controls for the following variables at the transaction level:

Target and Acquirer Profitability: The pre-announcement profitability for targets and acquirers is included when we estimate our predicted target and acquirer regressions. Profitability is defined as operating cash flow divided by sales.

Target Relative Size: The pre-announcement market value of the target divided by the sum of the pre-announcement market values of the target and the acquirer.

Merger Dummy: A dummy equal to 1 if the given transaction is a merger and to zero for an acquisition of assets. We examine only these two transactions.

Merger × *Relative Size*: The cross product of the above two variables.

Log Total Size: The natural logarithm of the sum of the pre-announcement market values of the target and the acquirer.

3. Mergers and Product Market Similarity

We begin by comparing our similarity measures with historical measures. To show their uniqueness relative to SIC codes, we consider the set of acquisitions that are in different two-digit SIC codes. Although past studies would label them as dissimilar, we show that they are in fact highly similar. Table 1 lists the 35 restructuring pairs in 2005 that are most pairwise similar despite the fact that they reside in different two-digit SIC codes. All 35 are in the 99th percentile of similarity relative to the distribution of all firm pairings.[12] The firms on the list, given their business lines, suggest that the high degree of similarity we report is, in fact, due to real product similarities. For example, petroleum and pipeline firms are related. Newspapers and radio are also related and can be viewed as

[12] Although we report only the top 35 due to space constraints, there are actually 57 acquisitions in different two-digit SIC codes in the 99th percentile in 2005 and 98 deals in the 95th percentile or better.

substitute sources of advertising despite their being in different two-digit SIC codes.

To further illustrate how the algorithm rated these firms, Table 2 displays the full list of words that were common for the first ten of these related transactions. In the Online Appendix, we present the word lists for the remaining transactions in Table 1. Table 2 further illustrates the limitations of using SIC codes as an all-or-nothing classification of merger pair similarity. The word lists suggest that the similarity calculations are indeed driven by product market content. Key to this result is our focus on non-common words and on words in the local similarity dictionary. These steps help to eliminate document templates, legal jargon, and other non-product content. The list of similar words for each pair is also substantial, indicating that our identification of similarity is informative.

Table 3 displays summary statistics for our firm- and transaction-level databases. Fifteen percent of the firms in our firm-level database were targets of either a merger or an acquisition of assets, and 28.7% were acquirers. These numbers are somewhat larger than some existing studies because (i) our sample includes more recent years in which transactions were more common; (ii) these figures include both mergers and partial acquisitions; and (iii) transactions are included if the counterparty is public or private. We next report the fraction of targets and acquirers by transaction type, and find that mergers (4.2% targets and 10.6% acquirers) are less common than asset acquisitions (10.8% targets and 18.1% acquirers).

Product similarities are bounded in the range [0,1]. The average broad similarity (across all firms) is 0.022 (or 2.2%). The average local similarity between a firm and its ten closest neighbors is considerably higher at 20.1%. The average sales-based HHI for firms in our sample is 0.048. The average percentage of 10-K words having the word roots "patent," "copyright," and "trademark" is 0.255%.

In Panels B and C, we report summary statistics for the transaction-level database and ex post outcomes. The average acquirer experienced an announcement return very close to zero, and the average target experienced an announcement return of 5.4%. The average merger pair is 11.4% similar. Panel C shows that the average acquiring firm experiences little change in industry-adjusted profitability or industry-adjusted sales growth over one- to three-year horizons.

Table 4 displays Pearson correlation coefficients between our measures of product differentiation and other key variables. Broad similarity relative to all firms is 13.7% correlated with local similarity relative to the ten nearest neighbors.[13] We also find that the sales HHI variable is roughly −10% correlated

[13] Although we focus on the 10 nearest neighbors in most of our analysis, our results change little if we instead measure similarity relative to the 100 nearest neighbors.

Table 1
Merging firms in 2005 with very high percentile similarity (top 35) but different two-digit SIC codes

Acquirer	Target	Acquirer SIC-3	Target SIC-3
Amgen Inc	Abgenix Inc	SIC3=283, Medicinal Chemicals & Botanical Products	SIC3=873, Commercial & Biological Research
Atlas Pipeline Partners LP	Enogex Arkansas Pipeline Corp	SIC3=492, Natural Gas Transmission	SIC3=131, Crude Petroleum & Natural Gas
Belo Corp	WUPL-TV,New Orleans,LA	SIC3=271, Newspapers: Publishing & Printing	SIC3=483, Radio Broadcasting Stations
Chemicon International Inc	Cell & Molecular Technologies-	SIC3=283, Medicinal Chemicals & Botanical Products	SIC3=873, Commercial & Biological Research
Correctional Properties Trust	Geo Group Inc-Lawton	SIC3=679, Miscellaneous Investing	SIC3=874, Services-Management Services
Eagle Hosp Prop Trust Inc	Hilton Glendale,Glandale,CA	SIC3=679, Miscellaneous Investing	SIC3=701, Hotels & Motels
Express Scripts Inc	Priority Healthcare Corp	SIC3=874, Services-Management Services	SIC3=809, Services-Misc Health & Allied Services
First Niagara Finl Grp,NY	Burke Group Inc	SIC3=603, Savings Institution, Federally Chartered	SIC3=874, Services-Management Services
General Dynamics Corp	Anteon International Corp	SIC3=372, Aircraft & Parts	SIC3=737, Computer Programming, Data Processing
GEO Group Inc	Correctional Services Corp	SIC3=874, Services-Management Services	SIC3=922, Miscellaneous
Hampshire Group Ltd	Kellwood Co-David Brooks Bus	SIC3=225, Knitting Mills	SIC3=233, Women's, Misses', and Juniors Outerwear
Hercules Offshore Corp	Superior Energy Svcs-Liftboats	SIC3=138, Drilling Oil & Gas Wells	SIC3=353, Construction, Mining & Materials Handling
Highland Hospitality Corp	Hilton Boston Back Bay Hotel	SIC3=679, Miscellaneous Investing	SIC3=701, Hotels & Motels
Highland Hospitality Corp	Westin Princeton at Forrestal	SIC3=679, Miscellaneous Investing	SIC3=701, Hotels & Motels
Host Marriott Corp	Starwood Hotel-Hotel Portfolio	SIC3=679, Miscellaneous Investing	SIC3=701, Hotels & Motels
Innkeepers USA Trust	Westin Governor Morris,NJ	SIC3=679, Miscellaneous Investing	SIC3=701, Hotels & Motels
Journal Communications Inc	Emmis Comm-Radio Stations(3)	SIC3=271, Newspapers: Publishing & Printing	SIC3=483, Radio Broadcasting Stations
L-3 Communications Hldg Inc	Titan Corp	SIC3=367, Electronic Components & Accessories	SIC3=737, Computer Programming, Data Processing
LaSalle Hotel Properties	Hilton San Diego Resort,CA	SIC3=679, Miscellaneous Investing	SIC3=701, Hotels & Motels
MAF Bancorp,Clarendon Hills,JL	EFC Bancorp Inc,Elgin,Illinois	SIC3=671, Holding Offices	SIC3=602, National Commercial Banks
McDATA Corp	Computer Network Technology	SIC3=737, Computer Programming, Data Processing	SIC3=357, Computer & office Equipment
Medco Health Solutions Inc	Accredo Health Inc	SIC3=632, Accident & Health Insurance	SIC3=809, Services-Misc Health & Allied Services
MetLife Inc	CitiStreet Associates LLC	SIC3=631, Life Insurance	SIC3=641, Insurance Agents, Brokers & Service

(continued)

Table 1
Continued

Acquirer	Target	Acquirer SIC-3	Target SIC-3
Nektar Therapeutics	AeroGen Inc	SIC3=283, Medicinal Chemicals & Botanical Products	SIC3=384, Surgical & Medical Instruments
Nevada Gold & Casinos Inc	Colorado Grande Casino	SIC3=651, Real Estate Operators & Lessors	SIC3=701, Hotels & Motels
Pennsylvania Real Estate Inv	Gadsden Mall,Gadsden,Alabama	SIC3=679, Miscellaneous Investing	SIC3=651, Real Estate Operators & Lessors
Secure Computing Corp	CyberGuard Corp	SIC3=357, Computer & office Equipment	SIC3=737, Computer Programming, Data Processing
SeraCare Life Sciences Inc	Celliance Corp-Product Line	SIC3=809, Services-Misc Health & Allied Services	SIC3=283, Medicinal Chemicals & Botanical Prod
Sun Microsystems Inc	Tarantella Inc	SIC3=357, Computer & office Equipment	SIC3=737, Computer Programming, Data Processing
Sunstone Hotel Investors Inc	Renaissance Hotels-Hotel	SIC3=679, Miscellaneous Investing	SIC3=701, Hotels & Motels
Sunstone Hotel Investors Inc	Renaissance Washington DC	SIC3=679, Miscellaneous Investing	SIC3=701, Hotels & Motels
WellPoint Inc	WellChoice Inc	SIC3=809, Services-Misc Health & Allied Services	SIC3=632, Accident & Health Insurance
Willow Grove Bancorp Inc,PA	Chester Valley Bancorp,PA	SIC3=603, Savings Institution, Federally Chartered	SIC3=671, Holding Offices
YDI Wireless Inc	Proxim Corp	SIC3=506, Wholesale-Electrical Apparatus, Wiring	SIC3=366, Telephone & Telegraph Apparatus
Yellow Roadway Corp	USF Corp	SIC3=473, Arrangement of Transportation of Freight	SIC3=421, Trucking & Courier Services (No Air)

This table presents a list of firms that are both (1) in different two-digit SIC codes; and (2) the first 35 mergers with the highest pair similarities in 2005 (all are in the 99th percentile).

Table 2
Common words for merging firms from Table I with high similarity

Acquirer (Industry) + Target (Industry): list of common words
Amgen Inc (SIC3=283, Medicinal Chemicals & Botanicals) + Abgenix (SIC3=873, Biological Research): abbott,abgenix,abnormal,alpha,amgen,antibodies,antibody,appeared,autoimmune,beneficiaries,biogen, biologic,biologics,biology,biopharmaceutical,boehringer,bone,brain,breast,bristol,calcium,cancers,carcinoma, chemotherapy,chugai,cinacalcet,circumvent,colorectal,cytotoxic,dialysis,disorder,doses,dosing,egfr,elevated, endothelial,endpoint,epidermal,expression,gene,genentech,genmab,glaxosmithkline,hormone, hyperparathyroidism,idec,immune,immunex,immunology,inducing,infections,inflammation,inflammatory, ingelheim,inhibiting,inhibitor,insulin,interleukin,investigator,kidney,kinase,kirin,malignancies,medically, merck,metabolic,metastases,metastatic,mimpara,mineral,molecule,monoclonal,monotherapy,myers, neurological,novartis,oncology,outpatient,panitumumab,parathyroid,payers,pdgf,pfizer,pharma,pivotal, platelet, progression,prostate,psoriasis,radiation,reactions,receptor,receptors,recombinant,refractory,renal, repligen,roche,sensipar,serum,squibb,statistically,supportive,suppression,symptomatic,systemic,tissues, transkaryotic,tumor,tumors,vascular,wyeth
Atlas Pipeline Partners LP (SIC3=492, Natural Gas Transmission) + Enogex Arkansas Pipeline (SIC3=131, Crude Petroleum & Natural Gas): abandonment,anadarko,basin,basins,belvieu,compressors,condensate,continent,conway,cubic,diameter,discriminatory,drilled,enogex,ferc,hedges,hydrocarbon,indices,intrastate,koch,liquids,mont,oneok,pipe,residue,tulsa, wellhead
Belo Corp (SIC3=271, Newspapers) + WUPL-TV,New Orleans (SIC3=483, Radio Broadcasting): advertiser,advertisers,affirmed,attribution,audience,audiences,broadcaster,broadcasters,broadcasting, broadcasts,carriage,compulsory,digitally,distant,dmas,duopolies,duopoly,fare,flag,frequencies,hispanic, indecency,indecent,morning,multichannel,netratings,nielsen,norfolk,orleans,piracy,primetime,purport, reauthorization,remand,remanded,retransmission,revoke,saturday.subscriptions,supreme, syndicated,tacoma, transmissions,viewer,viewers,voices,warner
Chemicon International Inc (SIC3=283, Medicinal Chemicals & Botanicals) + Cell & Molecular Technologies-(SIC3=873, Biological Research): antibodies,assays,bacteria,biochemical,biology,biomedical,bioreactors,biosciences,chemicon,cloning, enzymes,experimental,experiments,expression,fermentation,fractionation,gene,genes,genetically,gpcr, hormones,infectious,isolate,mammalian,merck,molecule,neuroscience,organisms,purification,purified, reagents,receptor,recombinant,selectivity,sera,serologicals,signaling,stem,viral,vivo,yeast
Correctional Properties (SIC3=679, Misc Investing) + Geo Group (SIC3=874, Management Services): adelanto,appropriations,beds,broward,counseling,detainees,detention,falck,hobbs,inmate,inmates,lawton,male, mental,misconduct,odoc,offenders,prison,prisons,privatized,queens,rehabilitative,wackenhut,zoley
Eagle Hosp Prop (SIC3=679, Misc Investing) + Hilton Glendale (SIC3=701, Hotels & Motels): accumulates,bookings,contrary,embassy,franchisees,franchisor,guest,guests,hilton,illiquidity,insuring,lodging, loyalties,renovation,repositioning,reservation,reservations,revpar,suites,swimming,upscale
Express Scripts (SIC3=874, Management Services) + Priority Healthcare (SIC3=809, Misc Health Services): accreditation,admit,advancepcs,alerts,asthma,beneficiaries,caremark,counseling,dispense,formularies, formulary,harbors,hipaa,hmos,implicate,inspector,kickback,medco,medication,medications,medicines,nurses, nursing,outpatient,pbms,pharmacies,pharmacist,pharmacists,pharmacy,ppos,prescribing,prescriptions, referring,reimbursable,remuneration,scripts,soliciting,stark,subpoena,unclear,unconstitutional,violating, willfully
First Niagara Finl (SIC3=603, Savings Institution) + Burke Group (SIC3=874, Management Services): accruing,annuities,bureaus,conservator,continuance,creditworthy,disbursement,fnma,gnma,improbable, inadequately,insiders,iras,marketability,mention,negotiable,nonaccrual,obligor,pertinent,pledging,qualifies, questionable,reckless,repurchases,revising,student,substandard,tying,uncollectible,unimpaired,whichever, worse
General Dynamics (SIC3=372, Aircraft & Parts) + Anteon International (SIC3=737, Computer Programming): appropriated,ballistic,battle,cargo,civilian,combatant,commanders, corps,deployments,destroyer,fighter, littoral,missile,missions,munitions,naval,reconnaissance,shipbuilding,submarine,submarines,tactical, undersea,unfunded,warfare,weapons
GEO Group (SIC3=874, Management Services) + Correctional Services (SIC3=922, Miscellaneous): accreditation,anger,appropriations,beds,confinement,counseling,custody,detainees,detention,inmate,inmates, jail,male,marshals,mental,misconduct,offenders,prison,prisoners,prisons,privatized,psychiatric,rehabilitative, renovate,renovation,sexual,vocational,wackenhut

This table presents the common words for the first 10 mergers from Table I. Firms are both (1) in different two-digit SIC codes; and (2) the first 35 mergers with the highest pair similarities in 2005 (all are in the 99th percentile). The word lists for the remaining mergers from Table I are presented in Appendix II.

Table 3
Summary statistics

Variable	Mean	Std. Dev.	Minimum	Median	Maximum	Obs.
Panel A: Firm Variables						
Target Dummy	0.150	0.357	0.000	0.000	1.000	50,104
Acquirer Dummy	0.287	0.452	0.000	0.000	1.000	50,104
Target of Merger Dummy	0.042	0.201	0.000	0.000	1.000	50,104
Acquirer in Merger Dummy	0.106	0.308	0.000	0.000	1.000	50,104
Target of Acq. of Assets Dummy	0.108	0.311	0.000	0.000	1.000	50,104
Acquirer of Acq. of Assets Dummy	0.181	0.385	0.000	0.000	1.000	50,104
Broad Similarity (All Firms)	0.022	0.006	0.003	0.022	0.059	50,104
Product Similarity (10 nearest)	0.201	0.088	0.031	0.183	0.740	50,104
Fraction 10 Nearest Restructuring	0.377	0.200	0.000	0.400	1.000	50,104
Fraction SIC-3 Restructuring	0.363	0.144	0.000	0.353	1.000	50,104
Log Assets	5.483	2.108	−2.919	5.405	14.210	50,104
SIC-3 Industry Sales-based HHI	0.048	0.026	0.000	0.044	0.229	50,104
% Patent Words	0.255	0.346	0.000	0.130	5.268	50,104
Panel B: Transaction Level Variables						
Target Ann. Return (event day)	0.054	0.172	−0.834	0.006	4.373	6,629
Acquirer Ann. Return (event day)	0.000	0.054	−0.620	−0.000	1.657	6,629
Combined Firm Ann. Return (event day)	0.004	0.042	−0.337	0.001	0.755	6,629
Gain in Product Differentiation	0.005	0.078	−0.658	0.002	0.726	6,629
Merger Pair Similarity	0.114	0.089	0.000	0.097	0.310	6,629
Panel C: Acquirer Ex Post Real Performance						
1-Year Δ Profitability (scaled by assets)	−0.005	0.083	−0.873	0.001	0.902	4,779
3-Year Δ Profitability (scaled by assets)	−0.014	0.111	−1.145	−0.002	0.932	4,779
1-Year Δ Profitability (scaled by sales)	−0.004	0.123	−1.148	−0.003	1.321	4,779
3-Year Δ Profitability (scaled by sales)	−0.021	0.175	−1.291	−0.010	1.595	4,779
1-Year Sales Growth	0.035	0.555	−1.859	−0.008	10.000	4,779
3-Year Sales Growth	−0.019	1.071	−4.990	−0.128	10.055	4,779

Summary statistics are reported for our sample based on 1997 to 2006. Product similarities are measures that lie in the interval (0,1) based on the degree to which two firms use the same words in their 10-K product descriptions (see Appendix 1). A higher similarity measure implies that the firm has a product description more closely related to those of other firms. We compute local product similarities based on the ten nearest firms and a broad similarity based on all firms excluding the 10 nearest. The fraction of nearest neighbors that were involved in restructuring transactions in the past year is computed based both on the firm's ten nearest neighbors and on three-digit SIC codes. The % patent words variable is the percentage of words in the 10-K product description having the same word root as the words *patent*, *copyright*, and *trademark*. Announcement returns are net of the CRSP value-weighted index and are measured relative to the announcement day. The gain in product differentiation is the product distance from the target to the acquirer's ten nearest neighbors, less the acquirer's distance to its ten nearest neighbors. We compute three measures of ex post acquirer real performance. All are based on the first set of accounting numbers available after the transaction is effective (call this the "effective year"), and we consider one- to three-year changes in industry-adjusted performance thereafter (this method avoids bias from trying to measure the pre-merger performance of two separate firms). We compute industry-adjusted profitability as operating income divided by assets or sales in each year, and we then truncate the distribution at (-1,1) to control for outliers (winsorizing produces similar results). We then compute the change in this variable from the effective year to one to three years thereafter. We compute industry-adjusted sales growth as the ex post sales divided by the level of sales in the effective year.

with the product similarity variables. This suggests that firms in concentrated industries have somewhat lower product similarities, which is consistent with higher product similarity and lower HHI both being associated with product market competition. However, this correlation is modest, and both measures contain distinct information. Overall, we conclude that most correlations are small and that multicollinearity is unlikely to be an issue.

Table 4
Pearson correlation coefficients

Row	Variable	Broad Similarity (All Firms)	Product Similarity (10 Nearest)	Last Year 10 Nearest % Restructured	Last Year SIC-3 % Restructured	Industry HHI
		Panel A: Correlation Coefficients				
(1)	Product Similarity (10 Nearest)	0.137				
(2)	% Restructured Last Year (10 Nearest)	0.022	−0.12			
(3)	% Restructured Last Year (SIC-3)	0.013	−0.093	0.423		
(4)	Industry Sales based Herfindahl Index (SIC-3)	−0.033	−0.093	0.043	0.105	
(5)	% Patent Words	−0.043	−0.265	−0.143	−0.185	0.016

Pearson correlation coefficients are reported for our sample based on 1997 to 2006. Product similarities are measures that lie in the interval (0,1) based on the degree to which two firms use the same words in their 10-K product descriptions (see Appendix 1). A higher similarity measure implies that the firm has a product description more closely related to those of other firms. We compute product similarities based on the ten nearest firms and based on all firms excluding the 10 nearest. We also include the fraction of nearest neighbors that were involved in restructuring transactions in the past year, as well as a similar fraction based on three-digit SIC code groupings. The % patent words variable is the percentage of words in the 10-K product description having the same word root as the words *patent*, *copyright*, and *trademark*.

3.1 The Similarity Measure

Figure 2 displays the distributional properties of our local pairwise similarity measure. The uppermost graph displays the distribution of local similarities for all randomly chosen firm pairs (i.e., we do not condition on restructuring). The vertical axis is the frequency, and the horizontal axis is the pairwise similarity expressed as a percentage. Randomly chosen firms generally have similarity percentages ranging from zero to 3, but a relatively fat tail also stretches to scores of 10%. The second graph displays similarities for firms entering into restructuring transactions. Our broad conclusions are that restructuring pairs are highly similar relative to randomly selected firms, and that merger pair similarities are quite diverse, with considerable mass attached to values ranging from zero all the way to 40.

Existing studies measure merger pair similarity by asking if the target and acquirer reside within the same SIC code. The third and fourth graphs confirm that product similarities are very high for merger pairs in the same two- and three-digit SIC codes. However, the high diversity of similarities within these groups illustrates that SIC codes are too granular to capture all product heterogeneity.

Perhaps more striking are the high levels of similarity for merger pairs residing in different two-digit SIC codes in the bottommost graph. When compared with the topmost graph, this is striking because studies assessing merger pair similarity on the basis of SIC groupings would label these pairs as dissimilar.[14] The first and last graphs of Figure 2 suggest that identifying diversifying mergers using two-digit SIC codes is likely inaccurate, since most of these transactions are actually very similar. Many mergers with different two-digit SIC codes actually have much higher measures of similarity than random firm pairs. Thus the findings of studies related to that by Kaplan and Weisbach (1992) that mergers classified as diversifying perform as well as those classified as related might now be consistent with the proposition that related mergers perform better.

4. Merger and Asset Acquisition Likelihood

In this section, we use logistic models to test whether firms are more likely to merge when they are broadly more similar to other firms (H1) and when they are locally more similar to their nearest rivals (H2b).

4.1 Transaction Incidence

Table 5 displays the results of logistic regressions in which one observation is one firm in one year. All reported figures are marginal effects, and *t*-statistics are reported in parentheses. The dependent variable is a dummy equal to 1 if

[14] These results are also robust to excluding vertically related industries.

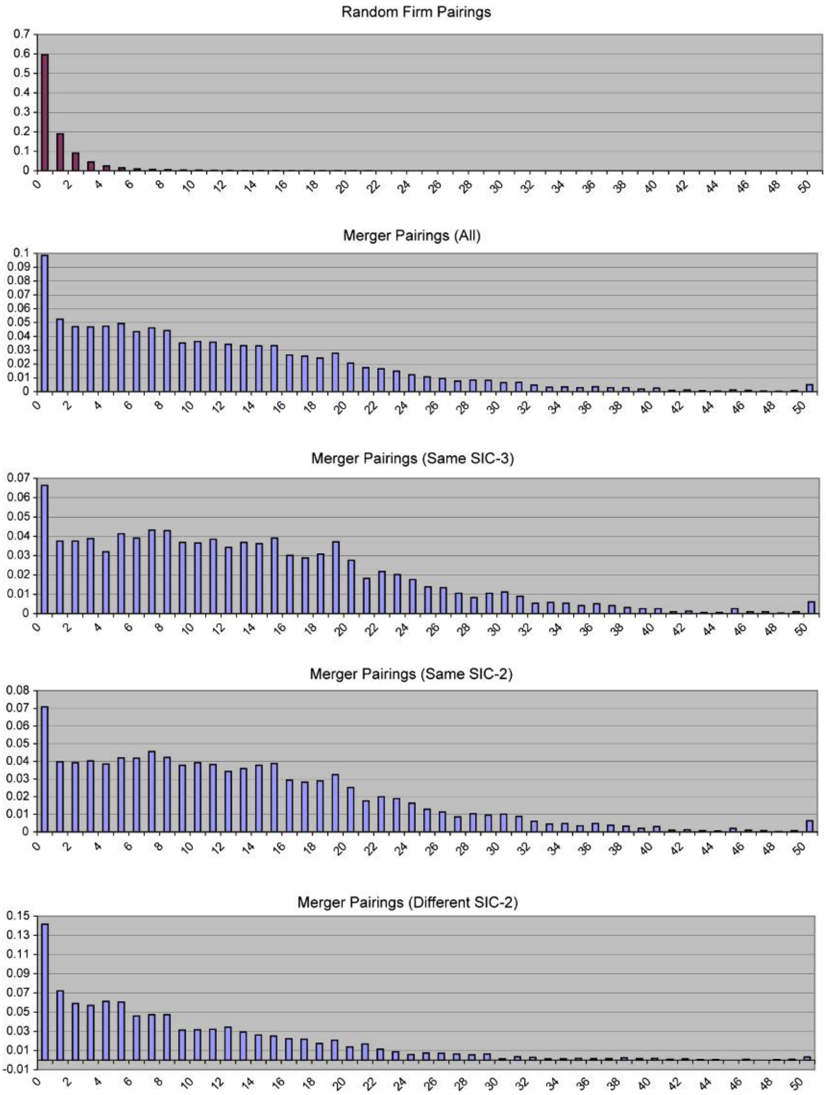

Figure 2
Distribution of product similarity for random firm pairings and merger pairings.
Each plot is an empirical density function, and total probability mass sums to one. The lower axis reflects similarities between zero and 100 (similarities are displayed as percentages for convenience). We truncate displayed results at 50%. The small number of outliers with values higher than 50% are represented by the probability mass assigned to the last bin. The random firm pairings group is based on the subsample of firms that merged, but the differences are taken with respect to a randomly chosen pair of firms in this subsample (results nearly identical in the set of firms that did not merge). The lower four plots are based on actual merger pairs.

Table 5
Mergers and acquisitions and product similarity

Row	Dependent Variable	Broad Similarity (All Firms)	Product Similarity (10 Nearest)	Industry HHI (SIC-3)	Last Year 10 Nearest % Restruct.	Last Year SIC-3 % Restruct.	Last Year % Patent Words	Last Year Log Total Assets	Last Year Profit-ability	Downstream Demand Shock	Obs
					Panel A: Acquirer Likelihood						
(1)	Acquirer?	2.138	−3.782		4.644		−0.448	9.801	1.756		50,104
		(9.84)	(−11.38)		(20.48)		(−1.72)	(32.01)	(3.80)		
(2)	Acquirer?	1.649			5.311		0.508	9.224	1.567		50,104
		(7.58)			(20.92)		(1.86)	(30.14)	(3.31)		
(3)	Acquirer?		−3.423		4.743		−0.592	9.974	1.212		50,104
			(−10.42)		(20.62)		(−2.24)	(33.54)	(2.62)		
(4)	Acquirer?			−1.627		5.461	0.765	9.673	1.237		50,104
				(−3.62)		(16.86)	(2.76)	(31.64)	(2.45)		
(5)	Acquirer?	1.955	−3.617	−2.129	3.320	3.674	0.107	9.798	1.761	1.617	50,104
		(9.19)	(−11.46)	(−5.88)	(14.87)	(13.88)	(0.41)	(30.38)	(4.13)	(3.28)	
					Panel B: Target Likelihood						
(6)	Target?	0.699	−1.917		2.307		0.263	9.157	−1.584		50,104
		(3.37)	(−8.07)		(12.21)		(1.38)	(45.31)	(−5.89)		
(7)	Target?	0.442			2.615		0.736	8.914	−1.643		50,104
		(2.07)			(14.02)		(3.72)	(44.80)	(−5.75)		
(8)	Target?		−1.787		2.344		0.201	9.202	−1.766		50,104
			(−7.83)		(12.18)		(1.08)	(45.65)	(−6.19)		
(9)	Target?			0.268		2.561	0.850	9.037	−1.679		50,104
				(1.26)		(14.00)	(4.55)	(44.49)	(−5.49)		
(10)	Target?	0.654	−1.672	0.054	1.673	1.711	0.550	8.964	−1.551	0.195	50,104
		(3.22)	(−6.87)	(0.27)	(8.88)	(9.52)	(2.81)	(43.91)	(−5.82)	(0.79)	

The table displays marginal effects of logistic regressions where the dependent variable is a dummy indicating whether the given firm is an acquirer of a merger or an acquisition of assets (Panel A) or a target (Panel B). Product similarities are measures that lie in the interval (0,1) based on the degree to which two firms use the same words in their 10-K product descriptions (see Appendix 1). A higher similarity measure implies that the firm has a product description more closely related to those of other firms. The independent variables include the fraction of nearest neighbors that were involved in restructuring transactions in the past year, the average product similarity of each firm relative to its ten nearest neighbors, and broader similarity relative to all firms excluding its ten nearest neighbors. We also include the fraction of past-year restructurings based on three-digit SIC code groupings. The % patent words variable is the percentage of words in the 10-K product description having the same word root as the words *patent*, *copyright*, and *trademark*. The sample is from 1997 to 2006. t-statistics are adjusted for clustering at the year and industry level. Marginal effects are displayed as percentages.

the given firm is an acquirer of a restructuring transaction in a given year (Panel A) and a dummy equal to 1 if the given firm is a target (Panel B). In all specifications, we report t-statistics that account for clustering at the year and industry level.

Table 5 shows that a firm is more likely to be an acquirer or a target (especially an acquirer) if its broad product similarity to all firms is high. These results are consistent with H1. This result is highly significant for acquirers at the 1% level regardless of specification. It is significant at the 1% level or the 5% level for targets.[15] Our later tests support the interpretation of these results through the lens of asset complementarity, and not with cost reductions.

Second, consistent with H2b, firms in highly competitive local product markets are less likely to restructure either as targets or as acquirers. The local product similarity relative to a firm's ten nearest rivals significantly negatively predicts transaction likelihood at the 1% level for both targets and acquirers. We interpret this as a "competitive effect," since firms having very similar rivals must compete for restructuring opportunities. Later in this section, we confirm that both the asset complementarity effect (H1) and the competitive effect (H2b) are economically significant.[16]

Table 5 also shows that firms using more patent words in their product descriptions are more likely to be targets. A likely explanation is that patents, copyrights, and trademarks serve as measures of the potential for unique products. Hence a firm that needs a technology protected by patents has few options to acquire it outside of merging with the patent holder.

The table also shows that the SIC-based *Sales HHI* variable is negatively related to restructuring for acquirers, and generally unrelated for targets. A possible explanation for the strong negative link for acquirers is that firms might anticipate federal regulations that block acquiring firms in concentrated industries. It is also possible that HHIs load on the asset complementarity effect more than the competitive effect. We conclude that concentration measures and product similarity measures should be jointly considered in studies of product market competition because they contain distinct information. We also find that recent restructuring predicts future restructuring for both SIC-based and product-similarity–based measures. Moreover, consistent with Maksimovic and Phillips (2001), we find that more profitable firms are more likely to be acquirers, and less profitable firms are more likely to be targets.

15 The results are also not driven by functional form, since they change little in an ordinary least squares linear probability model.

16 In the Online Appendix, we separately reproduce the tests in Table 5 for small and large firms and by transaction type for mergers and acquisitions of asset transactions. Results are broadly robust in most subgroups. Perhaps the only exception is that most variables do not predict which firms are likely to be the target of a merger, even though they do explain acquisition of asset transactions. This might be due, at least in part, to the fact that acquisition of assets is more common, resulting in more power.

Although industry shocks likely cannot explain our similarity results due to the fact that firm similarities appear to be stable over time, we examine the role of shocks to ensure robustness. Our proxy for industry shocks is based on industrial product shipment data from the Bureau of Labor Statistics (BLS). We define an industry's lagged demand shock at the three-digit SIC level as the logarithmic growth in its shipments from year $t-2$ to year $t-1$.[17] We consider two variations: own-industry shock, and downstream industry shock (downstream industries are identified by the input/output tables previously discussed), but we report results for only the latter to conserve space. Not surprisingly, both are less than 10% correlated with our similarity variables, and our results are virtually unchanged when we control for industry shocks. Although industry shocks do not explain our current findings, we do find that firms in industries with positive demand shocks are more likely to be acquirers.

4.2 Economic Magnitudes

In this section, we summarize the economic magnitude of our findings regarding transaction likelihood. We examine the effect of changing one of three key variables on the probability of a given transaction. In later sections, we also examine economic magnitudes related to announcement returns and real outcomes. Because some of our models are logistic, and others are based on ordinary least squares (OLS), we adopt a general framework based on predicted values. We first compute a model's predicted value when all of the independent variables are set to their mean values. We set one of our key variables to its 10th percentile or 90th percentile values and recompute the predicted value holding all other variables fixed. We then report how a given dependent variable changes when a key independent variable moves from its 10th percentile value, to its mean value, and to its 90th percentile value. Key benefits of this approach include its generality and its ability to show a variable's mean and variation around it.

Table 6 confirms that our findings regarding transaction incidence are economically relevant. The effect of changing local similarity (ten nearest) from the 10th percentile to the 90th percentile changes acquirer incidence from 33.4% to 24.0%. The economic magnitude of our "competitive effect" is thus substantial, especially for big firms in row 4. This effect is somewhat smaller, but still large, for targets at 17.2% to 12.9%. The *Broad Similarity* variable is also large, with a range from 26.2% to 31.2% for acquirer incidence. This similarity effect is considerably smaller for target incidence with a 1.7% spread. The *% Neighbor Patent Words* variable also has a relevant economic impact on target incidence (14.3% to 15.7%) but not on acquirer incidence.

[17] We also include a dummy variable for industries where BLS shipment data are not available.

Table 6
Economic magnitudes of predicting transaction incidence

Row	Description	*Product Similarity (10 Nearest)*			*Broad Similarity (All Firms)*			*% Patent Words*		
		10 %ile	Mean	90 %ile	10 %ile	Mean	90 %ile	10 %ile	Mean	90 %ile
	Panel A: Target and Acquirer Logit Models (Based on models in Table 5)									
1	All Firms: Target Incidence	17.2%	15.0%	12.9%	14.2%	15.0%	15.9%	14.3%	15.0%	15.7%
2	All Firms: Acquirer Incidence	33.4%	28.7%	24.0%	26.2%	28.7%	31.2%	28.6%	28.7%	28.8%
	Panel B: Large Firms									
3	Big Firms: Target Incidence	24.3%	21.2%	18.2%	18.8%	21.2%	23.7%	19.2%	21.2%	23.3%
4	Big Firms: Acquirer Incidence	43.0%	37.6%	32.1%	35.0%	37.6%	40.1%	35.4%	37.6%	39.7%
	Panel C: Small Firms									
5	Small Firms: Target Incidence	9.5%	8.8%	8.2%	9.1%	8.8%	8.5%	9.0%	8.8%	8.7%
6	Small Firms: Acquirer Incidence	23.5%	19.9%	16.2%	17.4%	19.9%	22.3%	21.8%	19.9%	17.9%
	Panel D: Mergers Only									
7	Mergers Only: Target Incidence	3.9%	4.2%	4.6%	4.0%	4.2%	4.4%	4.5%	4.2%	3.9%
8	Mergers Only: Acquirer Incidence	11.4%	10.6%	9.8%	10.0%	10.6%	11.2%	10.2%	10.6%	11.0%
	Panel E: Acquisition of Assets Only									
9	Acq Assets Only: Target Incidence	13.6%	10.8%	8.0%	9.8%	10.8%	11.8%	9.6%	10.8%	12.0%
10	Acq Assets Only: Acquirer Incidence	22.4%	18.1%	13.8%	16.4%	18.1%	19.8%	19.0%	18.1%	17.2%

The table displays economic magnitudes associated with merger incidence. All magnitudes are predicted values, and all magnitudes are conditional and thus account for the effects of industry, year, and all control variables (based on models in earlier tables as noted in panel headers). For each dependent variable being considered (noted in the description column), we first set all control variables to their mean values and compute the model's predicted value. The result of this calculation is the value displayed in the "mean" column in each category. For each independent variable whose economic magnitude we are measuring (product similarity 10 nearest, product similarity overall, and patent words), which is noted in the column headers, we also compute the model's predicted value assuming the given independent variable is expected to be in the 10th and 90th percentile of its distribution, while still holding all control variables fixed at their mean. Product similarities are measures that lie in the interval (0,1) based on the degree to which two firms use the same words in their 10-K product descriptions (see Appendix 1). We consider similarity based on a firm's ten closest rivals, and broader similarity based on all firms excluding these ten nearest firms. A higher similarity implies that the firm has a product description more closely related to those of other firms. The % patent words variable is the percentage of words in the 10-K product description having the same word root as the words *patent*, *copyright*, and *trademark*. The sample is from 1997 to 2006.

5. Ex post Outcomes

We now examine ex post outcomes, including combined-firm announcement returns and long-term real cash flow growth and product description growth.

5.1 Announcement Returns

This section examines the returns of the combined acquirer and target firms preceding and surrounding transaction announcements. We focus on H3, the hypothesis that says that total value creation will be larger the more merging firms are similar. Although this hypothesis has predictions regarding the combined firm's returns, it is silent on how the gains would be split between the target and the acquirer. Hence, we focus our analysis on the combined firm.

Table 7 reports OLS regressions with the acquirer's and target's combined abnormal announcement return as the dependent variable. We consider one- to eleven-day event windows ending on the announcement date (from day $t = -10$ to day $t = 0$), and we adjust standard errors to reflect possible clustering at the industry and year levels. The combined firm's raw return is the total market capitalization of both firms (in dollars) at the end of the event window minus the original market capitalization (in dollars), all divided by the original market capitalization. Hence, this is a simple value-weighted return for the combined firm. The abnormal return is this raw return less the return of the CRSP value-weighted market index. Given the possibility of deal anticipation, we examine event windows starting at ten days before announcement.[18]

We find strong support for the conclusion that more value is created upon announcement when the acquirer is in a more competitive product market (the acquirer similarity to its rivals is positive and significant), and the target is in a less competitive market (the target similarity to its rivals is negative and significant). Rows 1, 3, and 5 support this conclusion at the 5% level for longer horizons but are also weaker for the event day itself. This suggests that the market rewards acquirers buying assets in less competitive markets, because this permits new product introductions in more profitable markets.

In rows 2, 4, and 6, we replace the product market competition variables with two other variables to test whether this finding is linked to target and acquirer pairwise similarity (H3), and to potential gains in product differentiation.[19] We find some support for H3 over the longest event window, since the pairwise similarity variable is positive and significant at the 5% level in row 6 for the $t = -10$ to $t = 0$ event window. The results also show that the *Gain in Product Differentiation* variable is positive but not significant in all rows. These

[18] To measure deal anticipation further, we also considered the fraction of a firm's ten nearest rivals that were involved in restructuring in the past year. Consistent with anticipation, this variable negatively predicts announcement returns, but only at the 10% level or less. We omit this variable to conserve space and because it does not alter any inferences.

[19] The four key variables in the alternating rows cannot be included in the same regression because they are highly correlated.

Table 7
Announcement returns

	Event Window	Acquirer Product Simil. to Rivals	Target Product Simil. to Rivals	Gain in Prod. Diff. vs. Rivals	Target + Acquirer Pair Simil.	Target % Patent Words	Same SIC-3 Industry Dummy	Vertical Simil. Dummy	Acquirer Industry HHI (SIC-3)	Target Relative Size	Full Merger Dummy	Merger x Relative Size	Log Total $ Size	R^2	Obs
						Combined Firm Announcement Returns									
(1)	t=0 only	0.010	−0.012			0.004	−0.000	−0.006	0.007	−0.000	−0.003	0.019	−0.002	0.019	6,629
		(1.56)	(−2.08)			(2.33)	(−0.33)	(−2.24)	(1.17)	(−0.14)	(−1.72)	(3.30)	(−7.17)		
(2)	t=0 only			0.007	0.006	0.005	−0.000	−0.006	0.005	−0.000	−0.003	0.019	−0.002	0.018	6,629
				(1.03)	(0.85)	(2.73)	(−0.33)	(−2.21)	(0.74)	(−0.08)	(−1.86)	(3.33)	(−6.77)		
(3)	t=−5	0.022	−0.040			0.003	−0.001	0.005	0.010	−0.001	0.003	0.022	−0.003	0.020	6,629
	to t=0	(2.00)	(−3.99)			(1.31)	(−0.59)	(0.83)	(1.09)	(−0.44)	(1.35)	(2.99)	(−6.39)		
(4)	t=−5			0.015	0.014	0.006	−0.001	0.005	0.002	−0.001	0.003	0.023	−0.003	0.019	6,629
	to t=0			(1.32)	(1.13)	(2.20)	(−0.67)	(0.86)	(0.18)	(−0.38)	(1.09)	(3.04)	(−6.07)		
(5)	t=−10	0.030	−0.036			0.003	−0.000	0.006	0.012	−0.005	0.004	0.031	−0.003	0.014	6,629
	to t=0	(2.13)	(−2.90)			(1.04)	(−0.01)	(0.94)	(0.92)	(−1.27)	(1.30)	(3.58)	(−4.00)		
(6)	t=−10			0.016	0.032	0.006	−0.001	0.006	0.006	−0.004	0.003	0.031	−0.002	0.014	6,629
	to t=0			(1.10)	(2.14)	(1.76)	(−0.25)	(0.95)	(0.47)	(−1.18)	(0.98)	(3.63)	(−3.63)		

The table displays panel data regressions in which the dependent variable is the abnormal announcement return of combined target and acquirer. Announcement returns are computed over various windows including day $t=-10$ to day $t=0$ ($t=0$ is the announcement date) as indicated in the event window column. The combined firm's raw return is the total market capitalization of both firms at the end of the event window minus the original market capitalization, divided by the original market capitalization. The abnormal return results after subtracting the return of the CRSP value-weighted market index over each event window. Product similarities are measures that lie in the interval (0,1) based on the degree to which two firms use the same words in their 10-K product descriptions (see Appendix 1). A higher similarity measure implies that the firm has a product description more closely related to those of other firms. We compute product similarities based on the ten nearest firms (for both the acquirer and the target). We also compute the pairwise similarity of the target and the acquirer. The % patent words variable is the percentage of words in the 10-K product description having the same word root as the words *patent*, *copyright*, and *trademark*. The same SIC-3 industry dummy is one if the target and acquirer reside in the same three-digit SIC code. The vertical similarity dummy is one if the target and acquirer are more than 5% vertically related (based on Fan and Goyal 2006). Target relative size is the ex ante market value of the target divided by that of the acquirer. The merger dummy is one if the transaction is a merger and zero if it is an acquisition of assets. Log total size is the natural logarithm of the summed ex ante market values of the two firms. The sample is from 1997 to 2006. t-statistics are adjusted for clustering at the year and industry level.

results suggest that significant gains accrue as a function of similarity-based traits, especially variables linked to product market competition, despite the low power of this test. We also find that the *% Neighbor Patent Words* variable is positive in many specifications, consistent with larger gains when potential new products are likely to be unique.

We find that announcement returns are larger when the transaction is a merger involving a target firm that is large relative to the acquirer, and smaller when the firms are larger unconditionally. These results likely reflect the leveraged gains that should be observed in the combined firm's returns when transactions involve a larger fraction of the combined firm. The *Vertical Similarity* variable is negative and significant for short horizons (consistent with Kedia, Ravid, and Pons 2008), but this result becomes insignificant for longer horizons.

5.2 Real Performance

Although we observe evidence of financial value creation in Section 5.1, it is important to examine whether value increases are accompanied by real post-transaction gains in sales and profitability and by new product introductions, especially in light of the hypothesized role for asset complementarities and synergies.

An important challenge faced by researchers studying ex post restructuring performance is that two separate firms exist ex ante, and one or two firms might exist ex post depending on the transaction type. Measuring ex ante profitability or sales is especially confounding for partial asset purchases. We avoid this issue entirely by considering only the acquirer's post-effective change in performance measured relative to the first set of numbers available after the transaction's effective date. Our hypotheses thus assume that profitability and sales growth accrue over time, as should be the case because new products require time to build. We examine changes from year $t+1$ to year $t+2$ or $t+4$ (one- and three-year horizons). As a result, the sample of transactions used in this section is somewhat smaller than our sample in Section 5.1, because we must further require that the acquirer have valid Compustat data at least two years after a given transaction closes. By examining post-effective changes only, we bias our analysis toward not finding results due to lost power, but we avoid complications associated with attempts to measure year $t = -1$ performance. The need to focus on post-effective changes is further underscored by Maksimovic, Phillips, and Prabhala (2008), who document that many transactions also involve selling off divisions at the time of the transaction. Our results are conservative and likely understate the true relations between our key variables.

We consider three measures of ex post performance: (i) the change in industry-adjusted operating income divided by assets; (ii) the change in industry-adjusted operating income divided by sales; and (iii) industry-adjusted sales growth. To mitigate the effect of outliers, we truncate both profitability variables to lie in the interval $[-1, 1]$. For sales growth, we truncate the distribution to lie in the interval $[-1, 10]$.[20] Changes are computed from year $t+1$ to year $t+4$. To reduce survivorship issues, we assign any missing values for a given horizon the value of the last known horizon (e.g., if three-year sales growth is missing, we populate the given observation with two-year sales growth or one-year sales growth).[21]

Table 8 reports the results of OLS regressions where the ex post change in performance (horizons noted in column 2) is the dependent variable. In each panel, we first consider acquirer local product similarity. We then consider two other variables that more directly test H3 (merger pair similarity), and the role of increasing product differentiation.

We find that acquirers residing in highly competitive product markets experience positive changes in industry-adjusted profitability and higher industry-adjusted sales growth. The gains in profitability in Panel A and the sales growth in Panel C generally appear as significant at the 5% level after one year, and continue to accrue over three years (generally 1% significant). The weaker results in Panel B for profitability normalized by sales rather than assets are likely due to the high sales growth documented in Panel C, which coincides with the profitability growth in Panel A (sales is in the denominator in Panel B).

Table 8, rows 3 and 4 in Panel A, and rows 11 and 12 in Panel C, show strong support for the conclusion that the gains in profitability and sales growth are related to merger pairwise similarity and, to some extent, the potential for gains in product differentiation. The positive and highly significant *Target + Acquirer Pairwise Similarity* coefficient supports the conclusion that similarity is a key driver of real merger gains.

Because gains in profitability are possible for acquirers in ex ante competitive product markets, our results suggest that firms can significantly influence the degree of competitive pressure they face by restructuring. The mechanism is linked to the potential introduction of new products through asset complementarities. The strong results for sales growth are especially consistent with new product development playing a key role (we provide more supporting evidence in the next section). This result helps to separate our findings from the hypothesis based on only product market power of Baker and Bresnahan (1985). We now test the role of new products more directly.

5.3 Product Descriptions

In this section, we consider the prediction that new product development will accompany positive real outcomes. New product development is especially likely when the target firm is similar to the acquirer due to complementary assets (H3), and when additional gains in expected profitability are possible.

20 These truncations are similar to winsorizing at the 1% level.

21 Our results are robust to simply discarding these observations rather than using the last value.

Table 8
Long-term performance of acquirers

Row	Horizon	Acquirer Product Simil. (10 Near.)	Gain in Prod. Diff. vs. Rivals	Target + Acquirer Pair Simil.	% Patent Words	Same SIC-3 Industry Dummy	Vertical Simil. Dummy	Acquirer Industry HHI (SIC-3)	Target Relative Size	Merger Dummy	Merger x Relative Size	Log Total $ Size	R^2	Obs
					Panel A: Industry-Adjusted Operating Income/Assets									
(1)	1 Year	0.033			−0.001	−0.001	−0.018	0.026	0.008	0.007	−0.009	0.000	0.006	4,779
		(1.66)			(−0.13)	(−0.20)	(−1.79)	(1.40)	(1.90)	(2.07)	(−0.67)	(0.44)		
(2)	3 Year	0.079			0.002	−0.006	−0.031	0.066	0.009	0.003	0.006	0.001	0.012	4,779
		(2.64)			(0.25)	(−1.55)	(−3.12)	(1.75)	(1.44)	(0.69)	(0.43)	(0.68)		
(3)	1 Year		0.021	0.043	−0.000	−0.001	−0.018	0.030	0.008	0.006	−0.010	0.001	0.007	4,779
			(1.28)	(2.52)	(−0.07)	(−0.41)	(−1.80)	(1.66)	(1.93)	(1.85)	(−0.70)	(0.81)		
(4)	3 Year		0.021	0.074	0.001	−0.007	−0.031	0.079	0.009	0.002	0.005	0.001	0.013	4,779
			(0.83)	(2.88)	(0.16)	(−1.71)	(−3.13)	(2.05)	(1.57)	(0.44)	(0.37)	(1.12)		
					Panel B: Industry-Adjusted Operating Income/sales									
(5)	1 Year	0.050			0.007	−0.004	−0.012	0.033	0.014	0.005	−0.000	−0.002	0.006	4,779
		(1.46)			(0.96)	(−0.99)	(−1.20)	(1.14)	(2.08)	(0.85)	(−0.01)	(−1.60)		
(6)	3 Year	0.040			0.011	−0.014	−0.039	0.110	0.016	0.000	0.023	−0.004	0.011	4,779
		(0.73)			(1.11)	(−2.26)	(−2.73)	(1.96)	(1.62)	(0.02)	(0.91)	(−1.89)		
(7)	1 Year		0.026	0.032	0.006	−0.004	−0.012	0.039	0.014	0.005	−0.001	−0.002	0.006	4,779
			(0.85)	(1.16)	(0.85)	(−0.93)	(−1.18)	(1.45)	(2.17)	(0.84)	(−0.04)	(−1.51)		
(8)	3 Year		0.005	0.026	0.010	−0.014	−0.039	0.118	0.017	−0.000	0.023	−0.004	0.011	4,779
			(0.11)	(0.59)	(1.03)	(−2.15)	(−2.72)	(2.12)	(1.67)	(−0.01)	(0.89)	(−1.84)		

(continued)

Table 8
Continued

Row	Horizon	Acquirer Product Simil. (10 Near.)	Gain in Prod. Diff. vs. Rivals	Target + Acquirer Pair Simil.	% Patent Words	Same SIC-3 Industry Dummy	Vertical Simil. Dummy	Acquirer Industry HHI (SIC-3)	Target Relative Size	Merger Dummy	Merger x Relative Size	Log Total $ Size	R^2	Obs
						Panel C: Industry-Adjusted Sales Growth								
(9)	1 Year	0.244	.	.	−0.013	−0.042	−0.039	0.042	0.129	−0.006	0.208	−0.027	0.027	4,779
		(2.16)			(−0.45)	(−2.34)	(−1.36)	(0.34)	(4.79)	(−0.29)	(2.15)	(−4.51)		
(10)	3 Year	0.619	.	.	−0.083	−0.061	−0.175	0.057	0.256	−0.039	0.168	−0.049	0.024	4,779
		(3.10)			(−1.67)	(−1.91)	(−2.70)	(0.23)	(4.38)	(−0.93)	(1.12)	(−4.74)		
(11)	1 Year	.	0.257	0.172	−0.016	−0.039	−0.039	0.057	0.130	−0.005	0.205	−0.027	0.027	4,779
			(2.05)	(1.94)	(−0.55)	(−2.25)	(−1.36)	(0.47)	(4.83)	(−0.21)	(2.13)	(−4.31)		
(12)	3 Year	.	0.657	0.452	−0.089	−0.055	−0.175	0.094	0.258	−0.036	0.160	−0.048	0.024	4,779
			(2.92)	(2.49)	(−1.83)	(−1.72)	(−2.72)	(0.39)	(4.43)	(−0.85)	(1.07)	(−4.48)		

The table displays panel data regressions in which ex post changes in real performance measures are the dependent variable. For a transaction that becomes effective in year t, ex post profitability change or sales growth is a one- to three-year change from year $t+1$ until year $t+2$ (one-year) or $t+4$ (three-year) as noted in the horizon column. The dependent variable in Panel A is industry-adjusted operating income divided by assets; in Panel B, it is industry-adjusted operating income divided by sales; and in Panel C, it is industry-adjusted sales growth. Product similarities are measures that lie in the interval [0,1) based on the degree to which two firms use the same words in their 10-K product descriptions (see Appendix 1). A higher similarity measure implies that the firm has a product description more closely related to those of other firms. The acquirer product similarity (10 nearest) is the average similarity between the acquirer and its ten closest rivals. The target and acquirer product similarity is the pairwise similarity between the acquirer and target firms' products. The gain in product differentiation is the product distance from the target to the acquirer's ten nearest neighbors, less the acquirer's distance to its ten nearest neighbors. The % patent words variable is the percentage of words in the 10-K product description having the same word root as the words *patent*, *copyright*, and *trademark*. The same SIC-3 industry dummy is one if the target and acquirer reside in the same three-digit SIC code. The vertical similarity dummy is one if the target and acquirer are more than 5% vertically related (based on Fan and Goyal 2006). Target relative size is the ex ante market value of the target divided by that of the acquirer. The merger dummy is one if the transaction is a merger and zero if it is an acquisition of assets. Log total size is the natural logarithm of the summed ex ante market values of the two firms. The sample is from 1997 to 2006. t-statistics are adjusted for clustering at the year and industry level.

We proxy for new product development by examining whether firms experience growth in the size of their product descriptions in the years following the merger's effective date. The first post-merge description reflects the initial state of the post-merger firm. We define "product description growth" as the logarithmic growth in the number of words used in the product market description from year $t+1$ to either year $t+2$ or $t+4$. We then explore whether the same set of variables used to predict ex post real performance also predicts product description growth. We use an OLS specification in which all standard errors are adjusted for clustering at the year and SIC-3 industry level. The sample in this section is slightly smaller than that used in Section 5.2 because Compustat data (needed for the dependent variables in Section 5.2) are slightly more available than our collected 10-K data (needed for the dependent variable in this section).

Table 9 presents the results. We find that new product development is especially likely when acquiring firms reside in competitive product markets (rows 1 to 3), and when the target firm is similar to the acquirer (H3) (rows 4 to 6). We also find gains when there is the potential for higher profitability of new products (*Gain in Product Differentiation*). These results are significant at the 1% level.

Table 9 shows that vertical mergers, as described by Fan and Goyal (2006), independently of our similarity measures, experience some ex post growth in the size of product descriptions for longer horizons. This evidence linking vertical mergers to the introduction of new products supports the conclusion of Fan and Goyal (2006) that vertical mergers create value.

Table 9 also shows that a same-industry SIC dummy variable is insignificant and thus that pairwise similarity measured using SIC codes is too granular to produce similar inferences. Hence, understanding the relation between pairwise similarity and product differentiation relies on the researcher's ability to measure the degree of product similarity. The table also documents that product descriptions have a tendency to mean revert over time, a feature that is likely due to writing style. We control for this feature in addition to the other variables discussed above.

5.4 Economic Magnitudes for Ex Post Outcomes

Table 10 displays the economic magnitude of two key variables (*Product Similarity (10 Nearest)*, and the *Target + Acquirer Pairwise Similarity*) on announcement returns and real outcomes using the same generalized method as in Section 4. The competitive effect (local similarity) increases event-day-announcement returns by 0.2%, and longer-horizon (eleven-day) event returns by 0.7%. This spread is modest but not trivial, given the mean event-day announcement return of 0.4% (SD 4.2%), and the mean eleven-day announcement return of 0.5% (SD 7.8%). The modest size of these returns is consistent

Table 9
Ex post product descriptions of acquirers

Row	Horizon	Acquirer Product Simil. (10 Near.)	Gain in Prod. Diff. vs. Rivals	Target + Acquirer Pair Simil.	% Patent Words	Same SIC-3 Industry Dummy	Vertical Simil. Dummy	Acquirer Industry HHI (SIC-3)	Target Relative Size	Merger Dummy	Merger x Relative Size	Log Total $ Size	Initial Prod. Desc. Size	R^2	Obs
						Panel A: Ex post growth in product description									
(1)	1 Year	0.614			−0.005	−0.018	0.060	−0.181	−0.061	−0.014	0.057	0.019	−0.205	0.117	4,518
		(4.06)			(−0.28)	(−1.31)	(1.12)	(−1.25)	(−2.14)	(−0.62)	(1.30)	(3.68)	(−6.30)		
(2)	2 Year	0.941			0.003	−0.017	0.066	−0.105	−0.083	−0.032	0.083	0.020	−0.328	0.201	4,518
		(5.78)			(0.13)	(−1.14)	(1.25)	(−0.64)	(−2.94)	(−1.30)	(1.51)	(3.79)	(−10.26)		
(3)	3 Year	0.985			0.008	−0.017	0.084	−0.197	−0.059	0.001	−0.058	0.019	−0.380	0.215	4,518
		(5.77)			(0.36)	(−1.06)	(2.11)	(−1.22)	(−1.77)	(0.05)	(−0.94)	(3.53)	(−12.23)		
(4)	1 Year		0.454	0.602	−0.009	−0.020	0.059	−0.112	−0.056	−0.016	0.051	0.021	−0.205	0.117	4,518
			(3.27)	(3.99)	(−0.55)	(−1.40)	(1.10)	(−0.75)	(−2.01)	(−0.71)	(1.18)	(3.81)	(−6.27)		
(5)	2 Year		0.690	0.907	−0.005	−0.020	0.065	0.002	−0.075	−0.035	0.073	0.023	−0.327	0.201	4,518
			(5.17)	(5.90)	(−0.25)	(−1.27)	(1.22)	(0.01)	(−2.72)	(−1.38)	(1.34)	(4.10)	(−10.21)		
(6)	3 Year		0.782	0.915	−0.001	−0.018	0.082	−0.091	−0.052	−0.000	−0.068	0.021	−0.378	0.214	4,518
			(5.58)	(5.81)	(−0.02)	(−1.07)	(2.09)	(−0.56)	(−1.55)	(−0.00)	(−1.11)	(3.80)	(−12.21)		

The table displays panel data regressions in which three-year ex post (from year $t+1$ to $t+4$) logarithmic growth in the size of the firm's product description is the dependent variable. Size of the product description is measured as the number of words. Firms with larger increases in the size of their product description are interpreted as having introduced more products relative to other firms. For a transaction that becomes effective in year t, ex post product line growth is the one- to three-year growth in the size of the product description size from year t + 1 until year t + 2 (one-year), t + 3, and t + 4 (three-year) as noted in the horizon column. Product similarities are measures that lie in the interval (0,1) based on the degree to which two firms use the same words in their 10-K product descriptions (see Appendix 1). A higher similarity measure implies that the firm has a product description more closely related to those of other firms. The acquirer product similarity (10 nearest) is the average similarity between the acquirer and its ten closest rivals. The target and acquirer product similarity is the pairwise similarity between the acquirer and target firms' products. The gain in product differentiation is the product distance from the target to the acquirer's ten nearest neighbors, less the acquirer's distance to its ten nearest neighbors. The % patent words variable is the percentage of words in the 10-K product description having the same word root as the words *patent*, *copyright*, and *trademark*. The same SIC-3 industry dummy is one if the target and acquirer reside in the same three-digit SIC code. The vertical similarity dummy is one if the target and acquirer are more than 5% vertically related (based on Fan and Goyal 2006). Target relative size is the ex ante market value of the target divided by that of the acquirer. The merger dummy is one if the transaction is a merger and zero if it is an acquisition of assets. Log total size is the natural logarithm of the summed ex ante market values of the two firms. The initial product description size is the natural logarithm of the total number of words in the firm's initial (year t + 1) product description. The sample is from 1997 to 2006. t-statistics are adjusted for clustering at the year and industry level.

Table 10
Economic magnitudes of returns and real outcomes

		Acquirer Local Product Similarity (10 Nearest)			*Local Acquirer+Target Pairwise Similarity*		
Row	Description	10 %ile	Mean	90 %ile	10 %ile	Mean	90 %ile
	Panel A: Announcement Returns (Based on models in Table 7)						
1	Combined Firm Ann Returns (t=0)	0.3%	0.4%	0.5%	0.3%	0.4%	0.5%
2	Combined Firm Ann Returns (t=−10 to t=0)	0.6%	1.0%	1.3%	0.6%	1.0%	1.3%
	Panel B: Profitability and Sales Growth (Based on models in Table 8)						
3	Δ OI/Assets: 1 Year	−0.8%	−0.5%	−0.2%	−1.1%	−0.5%	0.1%
4	Δ OI/Assets: 3 Year	−2.3%	−1.4%	−0.6%	−2.4%	−1.4%	−0.4%
5	Δ OI/Sales: 1 Year	−0.9%	−0.4%	0.2%	−0.8%	−0.4%	0.1%
6	Δ OI/Sales: 3 Year	−2.5%	−2.1%	−1.7%	−2.4%	−2.1%	−1.7%
7	Sales Growth: 1 Year	0.9%	3.5%	6.0%	1.2%	3.5%	5.8%
8	Sales Growth: 3 Year	−8.4%	−1.9%	4.6%	−7.9%	−1.9%	4.1%
	Panel C: Growth in Product Descriptions (Based on models in Table 9)						
9	Prod Desc. Growth: 1 Year (A)	−4.4%	2.0%	8.4%	−4.9%	2.0%	8.8%
10	Prod Desc. Growth: 3 Year (A)	−5.9%	4.4%	14.6%	−6.0%	4.4%	14.7%

The table displays economic magnitudes associated with various findings reported earlier in this study. All magnitudes are predicted values, and all magnitudes are conditional and thus account for the effects of industry, year, and all control variables. For each dependent variable being considered (noted in the panel headers and the description column), we first set all control variables to their mean values and compute the model's predicted value. The result of this calculation is the value displayed in the "mean" column in each category. For each independent variable whose economic magnitude we are measuring (product similarity 10 nearest, product similarity overall, and patent words), which is noted in the column headers, we also compute the model's predicted value assuming the given independent variable is expected to be in the 10th and 90th percentile of its distribution, while still holding all control variables fixed at their mean. Product similarities are measures that lie in the interval (0,1) based on the degree to which two firms use the same words in their 10-K product descriptions (see Appendix 1). We consider similarity based on a firm's ten closest rivals, and similarity based on all firms in the universe excluding these ten firms. A higher similarity implies that the firm has a product description more closely related to those of other firms. The sample is from 1997 to 2006.

with most acquired assets being small relative to acquiring firms. The target and acquirer pairwise similarity has a similar economic magnitude.

The results in Panel B show that our findings regarding real outcomes are economically large. Increasing acquirer local-product similarity from the 10th percentile to the 90th percentile is associated with an increase in profitability growth from −0.8% to −0.2% (one year) and −2.3% to −0.6% (three years). Since profitability does not change much from year to year, this 1.7% shift is economically large, and more than 15% of the standard deviation of the dependent variable. This variable also generates a sales growth spread from 0.9% to 6.0% (one year) and −8.4% to 4.6% (three years). This spread is substantial. The table also shows that the target and acquirer pairwise similarity variable generates similar economic magnitudes.

Panel C displays results for the ex post growth in the size of the product description. The spread for acquirer local product similarity variable is from −4.4% to 8.4% (one year) and −5.9% to 14.6% (three years). This spread is economically large and is similar for the *Target + Acquirer Pairwise Similarity* variable. Overall, our most economically significant findings relate to transaction incidence, sales growth, and product description growth.

5.5 Robustness

In this section, we consider alternative theories and summarize additional tests.

First, one alternative is that our results might be due to expense reductions. In the Online Appendix, we examine whether our variables are related to ex post changes in the cost of goods sold (scaled by sales), selling and administrative expenses (scaled by sales), and capital expenditures (scaled by assets), from year $t+1$ to year $t+2$ or $t+4$. The relationship between our key variables and these expense ratios is not statistically significant, with one exception: acquirer local similarity is associated with reductions in cost of goods sold scaled by sales for the one-year horizon, but not for the three-year horizon. We conclude that cost savings likely cannot explain our product market results.

Second, we examine whether our results are driven by vertical mergers, as in Fan and Goyal (2006). Our similarity measures correlate less than 10% with vertical similarity measured using the input-output tables. We control for vertical similarity in our analysis, and our results are robust to including or excluding vertical controls.

Third, we test whether our results are driven by the 1990s technology boom. Throughout our study, we control for time and industry effects. We also run an unreported test where we exclude all technology firms from our sample (as defined in Loughran and Ritter 2004). Our results are robust in this test.

Fourth, we consider whether our results are related to corporate culture, since firms with similar cultures might have better merger outcomes. We discount this hypothesis for two reasons: First, our word lists (see Table 2) fit a product market interpretation, and second, the corporate culture hypothesis cannot explain why the target's distance from the acquirer's nearest *rivals* matters to ex post outcomes.

Fifth, our results are robust to multiple segment firms. We test this hypothesis by rerunning our tests after excluding multiple segment firms, where multiple segment firms are identified using the Compustat segment tapes. Our results change little in this test, and thus our results are not driven by conglomerates.

Finally, we examine whether our results are driven by repeat-acquiring firms by rerunning our tests after excluding firms that were involved in an acquisition in the past year. Our results are robust to excluding these firms.

6. Conclusions

Using novel text-based measures of product similarity between firms, we analyze how similarity and competition impact the incentives to merge and whether mergers with potential product market synergies through asset complementarities add value. Our conceptual framework is based on creating a Hotelling-style product market space with a location for each firm. This spatial framework allows us to calculate continuous similarity measures between groups of firms, thus replacing zero–one measures of relatedness used in the existing literature. This conceptual framework has advantages over SIC-code measures because it can jointly capture firm similarity relative to other firms within and across product markets, the competitiveness a firm currently faces, and a transaction's potential to increase product differentiation. Traditional SIC codes, even at the two-digit level, or based on using the input-output matrices (vertical relatedness), do not capture the extent to which transactions are related to each other.

We find that firms that are more broadly similar to all firms in the economy are more likely to merge (an "asset complementarity effect"), and firms with more highly similar rivals are less likely to merge (a "competitive effect"). The asset complementarity effect is consistent with these firms having more potential for new product synergies that could be derived from asset complementarities. The competitive effect is consistent with firms having to compete for profitable merger opportunities when they have more rivals that could act as substitute merger partners. We also find that firms with patents, copyrights, and trademarks are more likely to be targets, consistent with merging firms exploiting the potential for unique products.

Examining post-merger outcomes, we find that value creation upon announcement, long-term profitability, sales growth, and, most interestingly, increases in ex post product descriptions are higher when acquirers purchase targets that (i) have high pairwise similarity to the acquirer's own products; and (ii) increase the acquirer's product differentiation relative to its nearest rivals. These gains are larger when there are unique products and patents increasing the potential for new product introductions. Our findings suggest that these economically significant gains are associated with new product introductions.

Overall, our results are consistent with firms merging to use asset complementarities to create value through sales growth and new product introductions. More broadly, our results suggest that firms facing high ex ante competition can actively improve profitability via strategic restructuring transactions that increase ex post product differentiation through product market synergies.

Appendix 1

This appendix explains how we compute the "product similarity" between two firms i and j using the basic cosine similarity method. We first build the main dictionary in each year by taking the list of unique words used in all product descriptions in that year. We then discard words that appear in more than 5% of all product descriptions in the given year, and the resulting list of N words is the main dictionary. Next we take the text in each firm's product description and construct a binary N-vector summarizing its word usage. A given element of this N-vector is 1 if the given dictionary word is used in firm i's product description. For each firm i, we denote this binary N-vector as P_i.

We next define the normalized vector V_i, which normalizes the vector P_i to have unit length:

$$V_i = \frac{P_i}{\sqrt{P_i \cdot P_i}}. \tag{1}$$

To measure how similar the products of firms i and j are, we take the dot product of their normalized vectors, which is then the basic cosine similarity:

$$Product\ Similarity_{i,j} = (V_i \cdot V_j). \tag{2}$$

We define product differentiation as 1 minus similarity:

$$Product\ Differentiation_{i,j} = 1 - (V_i \cdot V_j). \tag{3}$$

Because all normalized vectors V_i have a length of 1, product similarity and product differentiation are bounded in the interval (0,1). This ensures that product descriptions with fewer words are not penalized excessively. This method is known as the "cosine similarity," since it measures the cosine of the angle between two vectors on a unit sphere. The underlying unit sphere also represents an "empirical product market space" on which all firms in the sample have a unique location.

Appendix 2

This appendix describes how we compute local cosine similarity and broad cosine similarity. The calculation is analogous to the local clustering measure used to identify cliques in the social-networking literature (see, e.g., Watts and Strogatz 1998; Granovetter 1973). The first step is to compute the basic cosine similarity matrix for all firm pairs i and j, which is described in Appendix 1.

We then compute a local clustering coefficient for each word w in the main dictionary. Let S denote the set of firms that use word w. Then, let S_{pairs} denote the set of all pairwise permutations of the firms in S. If the set S contains N firms, there are $N_{pairs} = \frac{N^2-N}{2}$ elements in the set S_{pairs}. Where $Similarity_{i,j}$ denotes the basic cosine similarity between firms i and j, the local clustering coefficient for word w is then

$$L_{clus,w} = \sum_{\forall(i,j)\in S_{pairs}} \frac{Similarity_{i,j}}{N_{pairs}}. \tag{4}$$

We assign words in the lowest tercile based on $L_{clus,w}$ to the "broad dictionary," and all other words to the "local dictionary." One technical point is that a word must appear in at least two firm 10-Ks before its corresponding $L_{clus,w}$ can be computed. We assign all words that appear in only one document to the local dictionary because their relevance is clearly local.

We thus have three dictionaries in each year. The basic dictionary includes all words surviving the initial 5% common word screen. The local and broad dictionaries are complementary subsets of the main dictionary. Local similarity is the cosine similarity computed using only words in the local dictionary, and broad similarity is the cosine similarity using only words in the broad dictionary. Because the dictionaries are orthogonal, the correlation between local and broad similarity for a randomly drawn firm is low (13.7%). This ensures the absence of multicollinearity.

Supplementary Data

Supplementary data are available online at http://rfs.oxfordjournals.org.

References

Almeida, H., M. Campello, and D. Hackbarth. 2009. Liquidity mergers. Working Paper, University of Illinois.

Andrade, G., M. Mitchell, and E. Stafford. 2001. New evidence and perspectives on mergers. *Journal of Economic Perspectives* 15:103–20.

Andrade, G., and E. Stafford. 2004. Investigating the economic role of mergers. *Journal of Corporate Finance* 1:1–36.

Baker, J., and T. Bresnahan. 1985. The gains from merger or collusion in product differentiated industries. *Journal of Industrial Economics* 33:427–44.

Berry, S., J. Levinsohn, and A. Pakes. 1997. Automobile prices in market equilibrium. *Econometrica* 63: 841–90.

Berry, S., and J. Waldfogel. 2001. Do mergers increase product variety? Evidence from radio broadcasting. *Quarterly Journal of Economics* 116:1009–25.

Betton, S., E. Eckbo, and K. Thorburn. 2008. Corporate takeovers. In *Handbook of Corporate Finance: Empirical Corporate Finance.* Amsterdam: Elsevier/North-Holland.

Boukus, E., and J. Rosenberg. 2006. The information content of FOMC minutes. Working Paper, Yale University.

Chamberlin, E. 1933. *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Fan, J., and V. Goyal. 2006. On the patterns and wealth effects of vertical mergers. *Journal of Business* 79: 877–902.

Granovetter, M. 1973. The strength of weak ties. *American Journal of Sociology* 78:1360–80.

Hackbarth, D., and J. Miao. 2009. The timing and returns of mergers and acquisitions in oligopolistic industries. Working Paper, Washington University.

Hanley, K., and G. Hoberg. 2010. The information content of IPO prospectuses. *Review of Financial Studies* 23:2821–64.

Harford, J. 2005. What drives merger waves? *Journal of Financial Economics* 77:529–60.

Hart, O., and J. Moore. 1995. Debt and seniority: An analysis of the role of hard claims in constraining management. *American Economic Review* 85:567–85.

Healy, P., K. Palepu, and R. Ruback. 1992. Does corporate performance improve after mergers? *Journal of Financial Economics* 31:135–75.

Hoberg, G., and G. Phillips, 2009. New dynamic product based industry classifications and endogenous product differentiation. Working Paper, University of Maryland.

Hoberg, G., and G. Phillips. 2010. Real and financial industry booms and busts. *Journal of Finance* 65:45–86.

Hotelling, H. 1929. Stability in competition. *Economic Journal* 39:41–57.

Jensen, M. 1993. The modern industrial revolution, exit, and the failure of internal control systems. *Journal of Finance* 48:831–80.

Jovanovic, B., and P. Rousseau. 2002. The *q*-theory of mergers. *American Economic Review* 92:198–204.

Kaplan, S., and M. Weisbach. 1992. The success of acquisitions: Evidence from divestitures. *Journal of Finance* 47:107–38.

Kedia, S., A. Ravid, and V. Pons. 2008. Vertical mergers and the market valuation of the benefits of vertical integration. Working Paper, Rutgers Business School.

Kwon, O., and J. Lee. 2003. Text categorization based on *k*-nearest neighbor approach for website classification. *Information Processing & Management* 39:25–44.

Lancaster, K. 1966. A new approach to consumer theory. *Journal of Political Economy* 74:132–57.

Li, F. 2006. Account report readibility, current earnings, and earnings persistence. Working Paper, University of Michigan.

Loughran, T., and B. McDonald. 2009. Plain English. Working Paper, University of Notre Dame.

Loughran, T., and J. Ritter. 2004. Why has IPO underpricing changed over time? *Financial Management* 33:5–37.

Macskassy, S., M. Saar-Tsechansky, and P. Tetlock. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63:1437–67.

Maksimovic, V., and G. Phillips. 2001. The market for corporate assets: Who engages in mergers and asset sales and are there efficiency gains? *Journal of Finance* 56:2019–65.

Maksimovic, V., and G. Phillips. 2008. The industry life-cycle, acquisitions and investment: Does firm organization matter? *Journal of Finance* 63:673–709.

Maksimovic, V., G. Phillips, and N. Prabhala. 2008. Post-merger restructuring and the boundaries of the firm. Working Paper, University of Maryland.

Mazzeo, M. 2002. An empirical model of firm entry with endogenous product choices. *Rand Journal of Economics* 33:221–42.

Mitchell, M., and H. Mulherin. 1996. The impact of industry shocks on takeover and restructuring activity. *Journal of Financial Economics* 41:193–229.

Morck, R., A. Shleifer, and R. Vishny. 1988. Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics* 20:293–315.

Nevo, A. 2000. Mergers with differentiated products: The case of the ready-to-eat cereal industry. *Rand Journal of Economics* 31:395–421.

Rhodes-Kropf, M., and D. Robinson. 2008. The market for mergers and the boundaries of the firm. *Journal of Finance* 63:1169–211.

Salop, S. 1979. Monopolistic competition with outside goods. *Bell Journal of Economics* 10:141–56.

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

Tetlock, P. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62:1139–68.

Watts, D., and S. Strogatz. 1998. Collective dynamics of small-world networks. *Nature* 393:409–10.

Yang, L. 2008. The real determinants of asset sales. *Journal of Finance* 63:2231–62.