Competition Links and Stock Returns

Assaf Eisdorfer, Kenneth Froot, Gideon Ozik, Ronnie Sadka*

November 2019

ABSTRACT

We consider a firm's competitiveness based on the manner by which other firms mention it on their 10-K filings. Using all public firm filings simultaneously, we implement a PageRank-type algorithm to produce a dynamic measure of firm competitiveness, denoted C-Rank. A high-minus-low C-Rank portfolio yields 16% alpha annually, where return predictability mainly stems from cross-sector competitiveness. The findings are largely consistent with investor underreaction to firm business opportunities identified by other strong firms. Nevertheless, stock return covariation with the C-Rank portfolio spread suggests that part of the return predictability can be interpreted as compensation for systematic cross-sector disruption risk.

Keywords: Text analysis; Competition; Asset pricing

JEL Classifications: G12, G14

---

* Eisdorfer: University of Connecticut, email: assaf.eisdorfer@uconn.edu. Froot: Harvard Business School, email: kfroot@hbs.edu. Ozik: EDHEC, email: gideon.ozik@edhec.edu. Sadka: Boston College, email: ronnie.sadka@bc.edu. Froot, Ozik, and Sadka are affiliated with MKT MediaStats, LLC (www.mktmediastats.com). The views expressed are solely of the authors. We thank Lauren Cohen, Dan diBartolomeo, Abraham Lioui, Anna Scherbina, Mikhail Simutin, Junbo Wang (discussant), and seminar participants at Barclays Sixth Annual Quantitative Research Conference 2019, Brandeis University, IDC Herzliya College, QWAFAFEW, Northfield's Annual Research Conference 2019, and Annual Conference on Financial Economics and Accounting 2019, for insightful comments. We would like to thank Sharon Hirsch for data science and research assistance.

1. Introduction

The information content of corporate financial statements has been explored for many years. Starting with the early studies on the effect of unexpected net income on stock returns (e.g., Ball and Brown (1968) and Beaver, Clarke, and Wright (1979)), through a deeper look at the accounting numbers such as discretionary accruals (e.g., Sloan (1996)), to the resent more sophisticated learning procedures such as text analysis (e.g., Hoberg and Phillips (2010, 2016), Garcia and Norli (2012), and Cohen, Malloy, and Nguyen (2019)). The common objective of these studies is to assess what investors can learn about a certain company from the information embedded in its financial statements.

In this study, we ask what we can learn about a company from the financial statements of *its competitors*. Such releases typically include a competition section in which the company can list and discuss the firms it views as competitors. We hypothesize that a company mentioned as a competitor by many other companies is more likely to be a strong competitor. However, simply counting the number of such mentions is not sufficient, as not all mentions are created equal: a mention by a major competitor should count more than a mention by a minor competitor. It follows that well-measured competitive "strength" requires the solution of a set of simultaneous equations, whereby the strength of a company is a function of the strength of the companies that mention it.

We use an advanced textual analysis technology to identify competitor lists in the cross-section of financial statements. We then implement a Google PageRank-type algorithm in the manner of Page et al. (1999) using cross-mentions in the most recent annual reports of all firms in each month to produce our preferred measure of firm-level competition rank (referred to as 'C-Rank'). The purpose of our C-Rank system is to emphasize that competitive strength cannot be assessed merely based on own-firm characteristics, such as firm size or product market share. Rather, it requires the collective view of all of a firm's competitors.

We further hypothesize that C-Rank signifies a firm's future business opportunities, identified by other firms, yet due to either investor limited attention or its computational complexity, C-Rank might not be fully understood by investors. This can result in underpricing of high C-Rank stocks, reflected in abnormally high future returns. Our first test documents the ability of the C-Rank competition status to predict future stock returns. Due to the positive correlation between C-Rank and firm size, we first orthogonalize C-Rank to size by running monthly cross-sectional regressions of C-Rank on current firm size, and use the regression residuals as the primary sorting variable. (A non-parametric specification controlling for size yields similar results.) This residual thus captures the firm's level of competitiveness relative to "expectation" vis-a-vis its size. We find that C-Rank positively predicts stock returns; an investment strategy that buys high C-Rank stocks and shorts low C-Rank stocks generates a monthly 6-factor alpha of 1.35% (about 16% annually). This effect is robust to various subsamples and investment horizons, and is confirmed using Fama-MacBeth (1973) regressions with various controls.

The positive effect of C-Rank on stock returns appears monotonic yet non-linear. This is consistent with the skewed distribution of C-Rank, which characterizes a typical competitive environment in which a small number of firms are perceived as the lead competitors while most firms are not considered as major competitors. Yet, we confirm that the cross-sectional return predictability of C-Rank does not solely reflect the difference between non-competitive and competitive firms, as it also explains cross-sectional differences within competitive firms alone. Furthermore, we verify that the C-Rank return predictability is not driven by a small group of giant competitors.

It is important to note that while larger firms are likely to be stronger competitors, our competition measure contains relevant information that is not captured by firm size. First, in all our tests of the information content of C-Rank we control for the effect of firm size. Second, we show that the largest companies in the market are not necessarily the strongest competitors. Moreover, we find that high competition status is associated with high returns, in contrast to the underperformance of large firms, on average, over time.

Given the natural tendency of firms in the same sector to compete with each other, we study the C-Rank across and within sectors. Specifically, we produce a cross-sector C-Rank, which considers only mentioned firms that operate in different sectors than the filing firm, and a within-sector C-Rank, which considers only mentioned firms from the same sector as the filing firm. We find that the main results hold for cross-sector C-Rank, not for within-sector C-Rank. This suggests that a firm's effective competition status stems mostly from competing with companies outside of its sector.

The ability of C-Rank to predict returns supports the hypothesis that investors do not fully understand the firm's business opportunities, recognized by other firms. If say a firm such as

Google mentions a given firm outside the technology sector as a competitor, it might indicate that Google finds the business environment of the given firm attractive, a significant and auspicious signal from the perspective of investors. The outperformance of a high C-Rank firm in this case is consistent with prior mispricing based on inferior information as it is updated slowly by investors to include Google's informationally-superior views. This interpretation is consistent with recent studies suggesting that investors might not efficiently process relevant textual information included in financial statements (see, for example, Cohen and Frazzini (2008) and Cohen, Malloy, and Nguyen (2019)).

To further assess the mispricing explanation, we utilize data on analyst coverage. If a given firm is recognized as a competitor by other firms outside its sector—recognition that indicates its business potential—then it is more likely that this information will be known to investors if the financial analysts that cover the firm also cover multiple sectors. We find that this scenario indeed reduces the return predictability of C-Rank, supporting the conjecture that highly competitive firms are underpriced.

As an alternative explanation to stock mispricing, we hypothesize that C-Rank might identify an element of a firm's risk profile. If many strong companies recognize a certain firm as a competitor, they are likely to adjust their strategies to compete more vigorously with that firm. A firm's high C-Rank may suggest therefore that the firm is "targeted" by strong competitors, which can increase the uncertainty about the firm's future performance and value. If this risk is undiversifiable, say an overall market environment of technological disruption, then the outperformance of high C-Rank firms might manifest compensation for risk.

We perform several tests to explore the risk-based explanation. First, we study *changes* in C-Rank. If C-Rank is associated with systematic risk, stock prices should react negatively to substantial unexpected increases in C-Rank. Indeed, on average, firm stock prices drop by roughly 3% over the months that exhibit a significant increase in cross-sector C-Rank, suggesting that investors assign a higher discount rate to such stocks. Second, we study the systematic pricing of the C-Rank portfolio return spread, and find that firms whose stock returns exhibit a high beta with respect to this spread subsequently outperform low such beta stocks, on average. Nevertheless, C-Rank beta explains only a small fraction of the C-Rank portfolio return spread. Therefore, while there is some evidence in favor of systematic risk pricing related to competitiveness, it seems that the return predictability is largely consistent with mispricing, as investors are slow to adjust for valuable information in financial statements.

Note, the C-Rank measure used here has two key features. First, a given firm's competitiveness is defined not only by its own financial statement but also by what other firms say in their financial statements about the given firm. Second, the C-Rank measure assigns more weight to the stronger firms (i.e. those that more firms mention them as competitors). We find that both these features are important. For example, if instead of C-Rank we use a simple count of the number of reports in which the firm is mentioned as a competitor by other firms, and even value these counts by the size of the mentioning firms, the predictive ability declines substantially. Therefore, not only is C-Rank an economically sensible measure, but also its unique features jointly contribute to the results documented in this paper.

This paper is related to a recent literature that studies firm connectedness and predictable returns. For example, Cohen and Frazzini (2008) identify supplier-customer links from 10-Ks, and show that a shock to the customer predicts stock return of the supplier. Antón and Polk (2014) document that firms being commonly held by mutual funds exhibit pairwise return correlation, and show implications for the cross-section of expected returns. Scherbina and Schlusche (2015) establish links between firms via their co-mentioning by the media, and Lee, Sun, Rogfei, and Zhang (2019) demonstrate return predictability via technological links, measured using shared patents.

The paper also contributes to the literature that uses textual analysis to provide economic insight. For example, Hoberg and Phillips (2010, 2016) study the product description section in 10-Ks to measure the similarity between the products of each pair of companies. Garcia and Norli (2012) extract U.S. state name counts from annual reports and find that less state-diversified firms earn higher stock return. Li, Lundholm, and Minnis (2013) measure the competitiveness of a firm as the percent of competition-related words in its 10-K. Cohen, Malloy, and Nguyen (2019) show that when companies make an active textual change in their reporting practices, this conveys an important signal about future firm operations (see also Froot, Kang, Ozik, and Sadka (2017) analysis of earnings call transcripts).

Overall the results in this paper highlight a distinct source of mispricing stemming from the slow reaction of investors to information about potentially profitable business opportunities of a given firm as pointed out by the competitors of that firm.

2. Generating competition rank

Our procedure to measure the relative competition strength of a firm is based on the entire cross-sectional pool of financial statements. Each month we observe the most recent annual report of each firm (if available). That is, all companies are represented in the competition measurement in each month. Using an advanced technology of text analysis, we record the companies that are listed as competitors by each firm (Appendix A describes the text analysis procedure). This provides a system of links between all firms in each month, where each firm can both mention other firms and be mentioned by other firms.

We observe a total of 119,785 10-Ks filed by 11,304 firms over the period 1995-2017, out of which 68,952 reports (58%) include a competition section. The number of firms that are mentioned as competitors in a single report's competition section ranges between zero (61 percent of the reports) to 35. Also, most firms, about 69 percent, are not mentioned at all in other reports. The company with the most mentions at a given time is IBM Corp. ('IBM') which was recognized as a competitor by 136 companies in the annual reports filed during 1997, followed by Microsoft Corp. ('MSFT') with 113 mentions in the reports filed during 1999. Figure 1 presents the distributions of the number of firms mentioned as competitors in a report's competition section, and the number of reports in which a firm is mentioned as a competitor.

Our measure is based on the notion that the competition status of a firm is determined not only by the number of times that the firm is mentioned by other firms, but also by the competition status of the mentioning firms. That is, the competition status of all firms should be evaluated simultaneously. We instantiate this using the PageRank algorithm developed by the founders of Google, Larry Page and Sergey Brin (Page et al., (1999)).

The basic aim of the PageRank algorithm is to assess the importance of a website page by counting the number and quality of links to the page, assuming that more important websites are likely to receive more links from other websites. In the same way we assume that more competitive companies are likely to be mentioned as competitors by other companies. Each month we run a

PageRank-type algorithm that iteratively solves a system of n (where n is the number of unique-firm reports) simultaneous equations to produce a firm-level competition status, which we refer to as 'C-Rank'.[1] (Appendix B provides a simple example to illustrate this algorithm.) As stated above, to make sure that all firms participate in calculating monthly C-Ranks, for each month we use the most recent annual report of every firm over the past twelve months. For example, the C-Ranks for April 2005 are based on all annual reports from May 2004 to April 2005.

The primary C-Rank measure is based on the cross-section of all firms in the market. We also produce two alternative sector-related measures, based on the eleven GIC sector classifications. The first measure considers only competitors outside the sector of the filing firm (cross-sector C-Rank), and the second measure considers only competitors from inside the sector of the filing firm (within-sector C-Rank).

3. C-Rank distribution and correlations

Our procedure produces the three C-Rank measures for each firm in each month over the period 1995-2017. Our sample includes 1,664,271 firm-month C-Rank values. Table 1 presents key statistics of the C-Rank distribution. By construction the C-Rank values summed to 1 (see Appendix B), which explains the relatively high C-Ranks within-sector, for which much fewer firms are used compared to the full market and cross-sector. All C-Rank means are higher than the medians, indicating a positively skewed distribution. This makes sense in a competitive environment as a small number of firms are perceived as the lead competitors while most firms are not considered as major competitors. On a monthly average, about 60 percent of the firms are not mentioned by any other firm, and thus receive the lowest C-Rank in a given month. Not surprisingly, the three C-Rank measures are positively correlated; the full market and cross-sector C-Ranks exhibit a correlation coefficient of 0.59, whereas the correlations between these measures and the within-sector C-Rank are 0.20 and 0.27, respectively. These gaps in correlations are expected as the sample of firms used to generate the full market and cross-sector C-Ranks are more overlapped than with the sample used for the within-sector C-Rank. These correlation values suggest that each form of C-Rank may capture different aspects of competition.

---

[1] We use a damping factor of 0.7 in applying the algorithm. The results remain similar when using different damping factors between 0.5 to 0.9. See Page et al. (1999) for details on the PageRank algorithm.

A reasonable question regarding our C-Rank measure is whether it simply captures the market capitalization of the firm, as larger firms are typically more competitive. We confirm that C-Rank captures information that is incremental to firm size in several ways. We show that the largest companies in the market are not necessarily the strongest competitors. We control for the effect of firm size in all of our tests, using several specifications for size to demonstrate that lack of meaningful effect on the conclusions. Moreover, we find that high competition status is associated with high returns, which goes against the negative effect of firm size on stock returns.

Table 2 presents the top five firms by the full market C-Rank every year over the period 1995-2017, as well as the largest companies over the same period. The C-Rank leaders exhibit an interesting pattern where from 1995 to 2012 IBM and Microsoft were the only companies with top C-Ranks, and since 2013 Google LLC ('GOOGL', now Alphabet Inc.) gets the top C-Rank every year. Comparing the top competitors to the list of the largest companies indicate indeed that very large firms are likely to be also very strong competitors, as several firms appear in both lists. Yet this association does not seem extremely strong as some of the largest firms in the market do not have the equivalent competition status. For example, General Electric Co. ('GE') had the largest market capitalization in eight years between 1995 and 2005, yet it was in the top five competitors only once during these years, in 1995. Similarly, Exxon Mobil Corp. ('XOM') had been constantly the largest company between 2006 and 2011, but was never in the group of the top five competitors. This provides a first indication that our competition measure contains information that is not captured by firm size.

We further calculate the correlation between C-Rank and firm size and other common risk factors: market-to-book ratio, past stock return, profitability, investment intensity, market beta, and idiosyncratic volatility. All market and accounting data are obtained from CRSP and COMPUSTAT. Because most firms are not mentioned as competitors in any report, we show the correlations both for the full sample and for the sample of competitive firms only (those with at least one mention in other reports). To eliminate time effects, we compute the cross-sectional correlations for each month over the sample period and report the time-series averages in Table 3. As expected, there is a positive correlation between C-Rank and firm size: 0.24 to 0.58. This is consistent with the results in Table 2, indicating that competition represents a firm characteristic that is not entirely captured by the size of the firm. High C-Rank firms are also more profitable and with lower idiosyncratic volatility, yet all average correlations for these and the other characteristics are fairly low. This suggests that C-Rank is not likely representing any of these risk factors.

4. C-Rank and stock returns

An important feature of our measure of the competitiveness of a firm, the C-Rank, is that it is not an independent assessment based on observed firm-specific characteristics, such as firm size or product market share, or even the competitive nature of the text in the firm's 10-K (Li, Lundholm, and Minnis (2013)). Rather, C-Rank reflects the collective view, across all companies, regarding the strong competitors in the market. This feature therefore raises the question of whether investors fully understand the competitive strength of a firm, as recognized by its competitors. We address this question by studying the ability of C-Rank to predict stock returns.

4.1 Portfolio sorts

We first examine the association between C-Rank and future stock returns using portfolio sort analysis. Due to the positive correlation between C-Rank and firm size, we first eliminate the size effect on stock returns. We run monthly cross-sectional regressions of C-Rank as of three months earlier (assuming it takes three months to release financial reports) on current firm size, and use the regression residuals as our sorting variable. Each month over the period 1995-2017 we divide all stocks into five equal-sized portfolios according to their C-Rank-residual. The portfolios are equal-weighted and held for one month. (Value-weighted portfolios also yield statistically significant results.)

Table 4 reports the monthly returns on each portfolio as well as the returns to the hedge portfolio that is long the highest C-Rank quintile and short the lowest C-Rank quintile. In addition to reporting the average return in excess of the risk-free rate, we also report the alphas from factor models. All factor returns are downloaded from Ken French's website. All returns and alphas are in percent per month and numbers in parentheses denote the corresponding *t*-statistics. Panels A, B, and C report the results for the full market, cross-sector, and within-sector C-Ranks, respectively.

The full market C-Rank's results in Panel A show that excess returns and factor-model alphas are generally monotonically increasing as one moves from quintile 1 (least competitive stocks) to quintile 5 (most competitive stocks). The long-short hedge portfolio has an excess return of 0.93% per month (*t*-statistic=3.76). Factor-model alphas are consistent with the excess return, ranging between 0.77% (CAPM) to 1.35% (6-factor model), all are statistically significant (*t*-statistics between 3.17 and 7.00). We further note that the C-Rank return predictability is mostly driven by the top quintile firms, as the difference between quintiles 5 and 4 is typically much larger than the differences across quintiles 1 to 4. This result is consistent with the skewed distribution of C-Rank shown in Table 1. Yet, we confirm in unreported results that the C-Rank effect is not driven by a small group of top competitors such as Google or Microsoft. For example, removing the highest one-hundred C-Rank firms each month from the sample, as well as all firms that they mention, has no material impact on our results. The results in Panel A therefore uncover a clear strong relation between the firm's competitiveness and subsequent returns.

Panels B and C address the role of sectors in the context of firm competitiveness. The results in Panel B based on cross-sector C-Rank are similar to those based on the full market C-Rank; the monthly returns/alphas range between 0.78% to 1.30% (*t*-statistics between 2.72 and 5.22). However, the effect of within-sector competitiveness on stock returns (Panel C) is insignificant and even negative under some models. This may suggest that the firm's real competition status is generated mostly by competing with companies outside its sector.

We note that the C-Rank return predictability is not driven by firms in a particular sector. Results, unreported for brevity, show that removing from the sample all firms from one sector at a time, as well as all firms that they mention, yields a significant alpha for each case; monthly alpha point estimates range from 1.20% (when excluding information technology) to 1.38% (when excluding industrials and financials).

Figure 2 shows the performance of the C-Rank hedge portfolio over the period 1995-2017. While the effect seems somewhat stronger in the early years, it is consistently upward sloping over the sample period, yielding a cumulative excess return of 253% and 6-factor alpha of 369%.

To verify that the positive effect of C-Rank on stock return does not represent other well-documented stock characteristics that are associated with firm risk, we perform a double-sort analysis. We first sort all stocks into equal-sized quintiles based on a stock characteristic. The stocks are then further sorted into quintiles according to their C-Rank/size regression residual, yielding 25 characteristic/C-Rank portfolios. For each of the portfolios we calculate the equal-weighted monthly stock return, and then for each C-Rank quintile we average across the characteristic quintiles, yielding five quintile-mean C-Rank returns. The stock characteristics we consider include firm size, market-to-book ratio, past stock return, profitability, investment intensity, market beta, and idiosyncratic volatility.

The double-sort results reported in Table 5 are consistent with the single-sort results. The 6-factor alpha of the average returns of the hedge C-Rank portfolios across all stock characteristics quintiles is positive and significant for the full market and cross-sector competitors, but not for the within-sector competitors. The results in Table 5 thus confirm that the high stock returns to firms with high C-Rank, especially cross-sector, are not captured by common firm risk characteristics.

We further examine the robustness of the results to different subsamples and return horizons in Table 6. To reduce the clutter in the table, we report only the 6-factor alphas for each portfolio. To facilitate comparison with the main results, we also report the full-sample results in the first row of the table. We consider three different kinds of subsamples. The first simply tabulates results when excluding the month of January. The second subsample excludes recession periods. We use NBER recession dummy as an indicator of the health of the economy for this exercise. Third, we tabulate the results separately for the early years (1995-2006) and the late years (2007-2017).

Panel A of the table shows that the hedge portfolio alpha is somewhat lower when excluding January, but is still significant; the 6-factor alphas is 0.80% with a *t*-statistic of 5.01. The results seem insensitive to the state of the economy, as excluding recessions shows a significant 6-factor alphas of 1.39%. Consistent with Figure 2, the effect of C-Rank is stronger in the early years, yielding an alpha of 1.96% per month (*t*-statistic 5.76), yet is still significant in recent years with an alpha of 0.77% and a *t*-statistic of 4.48.

We look at the horizon effect in Panel B of Table 6. We consider holding periods of 3, 6, 12, and 18 months. This means that we have overlapping portfolios. We take the equal-weighted average of these overlapping portfolios similar to the approach of Jegadeesh and Titman (1993). The 6-factor alphas of the hedge portfolio are positive and statistically significant for horizons up to 18 months, although they decline monotonically as we increase the horizon, from 1.35% for one-month horizon to 0.91% for 18-month horizon. All portfolio sort results are therefore robust to different subsamples and horizons.

4.2 Fama-MacBeth regressions

We further examine the association between C-Rank and subsequent returns using Fama and MacBeth (1973) regressions. Beyond serving as an additional diagnostic check, these regressions offer the advantage of controlling directly for well-known determinants of the cross-sectional patterns in returns and thus check for the marginal influence of C-Rank on our results. Accordingly, we run these cross-sectional regressions and report the results in Table 7. The dependent variable is the excess stock return and the main independent variable is C-Rank, orthogonalized to size as in the portfolio sort analysis. The control variables are log market capitalization, log market-to-book, past six-month return, profitability, investment intensity, market beta, and idiosyncratic volatility. We winsorize all independent variables at the 1% and 99% levels to reduce the impact of outliers. All reported coefficients are multiplied by 100 and we report Newey-West (1987) corrected (with twelve lags) *t*-statistics in parentheses.

Because most firms are essentially not competitive (firms that are not mentioned in other reports and thus get the same lowest C-Rank value), we examine the effect of C-Rank also among competitive firms only (those with at least one mention in other reports). The result show that the full market C-Rank has a positive and significant effect on stock return for the full sample (*t*-statistic=5.24). When removing the non-competitive firms the results are weaker, but still significant (*t*-statistics=2.91). This suggests that the effect of C-Rank on stock returns is not coming only from the difference between non-competitive and competitive firms, but is also important within the competitive firms. These results therefore corroborate the portfolio sort analysis, indicating that a higher competition status is associated with higher stock returns.

The cross-sector C-Rank also exhibits predictive ability over stock returns (*t*-statistics of 3.02 and 1.99 for the full and competitive firms samples), where the within-sector C-Rank does not show any significant effect in both samples. This is consistent with the portfolio sort results and may suggest that a firm's competition status is reflected more by the pool of firms that operate in different sectors than by those operating in the same sector.

5. Mispricing and analyst coverage

A relation between firm characteristic and future returns, not captured by documented risk factors (size, value, etc.), can signify temporary mispricing. Specifically, if a group of strong companies point to a given firm as a competitor, it might indicate that they find the business environment of that firm attractive, more than currently valued by investors. The outperformance of high C-Rank firms in this case is consistent with a mispricing explanation—these firms gradually grow in value as investors slowly digest the information.

Presumably, a significant amount of information across firms and industries flows through analyst reports. Therefore, we can further test the mispricing hypothesis by tracing the analyst links along the competition links. If a given firm is recognized as a competitor by many other firms outside its sector—a recognition that indicates potentially profitable business opportunities—then it is more likely that this information will be known to investors if the financial analysts that cover the given firm also cover many sectors.

We utilize data on analyst coverage to test for mispricing due to slow diffusion of information. We generate a measure of the concentration of a firm's analysts across industries. First, for each analyst appearing in the IBES dataset, we calculate the proportions of firms in each two-digit SIC industry that the analyst covers during a given year. From these industry proportions we calculate the Herfindahl-Hirschman Index (HHI) as a measure of the analyst's industry concentration. For each firm in a given year, we calculate the mean industry concentrations of all analysts that cover the firm during the year. To reduces the staleness of this analyst-based measure, we use the previous year measure for portfolios sorted during the first half of a given year, and the current year measure for portfolios sorted during the second half of a given year. (In this context, the analyst-based results are helpful in understanding the economic source of the C-Rank return outperformance, but they cannot be interpreted as tradable portfolios.)

We use the firm's mean analyst industry concentration to divide all firms each month to three equal-sized groups, and calculate the 6-factor alpha of the C-Rank hedge portfolio for each group. The results reported in Table 8 show a clear relation between the C-Rank return predictability and the analyst industry concentration. For the full market C-Rank, the alphas of the high- and low-concentration firms are 0.73% and 0.41%, respectively, although the difference in not statistically significant. More relevantly, for the cross-sector C-Rank, firms covered by highly industry-concentrated analysts show an alpha of 0.91%, compared to 0.21% for the low-analyst-concentration firms, where the difference is statistically significant (*t*-statistic=2.19). This result is consistent with the mispricing hypothesis, as analysts that cover multiple industries are more likely to capture out-of-sector recognitions of business opportunities, and thus reduce the extent of underpricing for the mentioned firms.

6. Testing for risk

As discussed above, the C-Rank return predictability may reflect underpricing of highly competitive firms driven by investors not fully aware of their attractive business opportunities as recognized by other companies. Yet, the high stock returns gained by companies with high competition status can also be consistent with a risk-based explanation. That is, being "targeted" by strong companies as a competitor imposes uncertainty as to the firm's future performance and value. To the extent that this form of disruption risk is systematic and recognized by the market, it should be compensated by high expected stock returns.

We perform a set of tests to explore this risk-based explanation. First, we study *changes* in C-Rank. If a large increase in a firm's C-Rank indicates that the firm is under a bigger threat because more and stronger companies are pointing at it now, then the firm's market value should react negatively, reflecting an elevated discount rate. To address this effect, each month we divide all companies that are recognized as competitors by other firms into five quintiles according to the change in C-Rank from the prior month. We then look at the difference between the average cumulative excess returns of the top and the bottom quintiles (i.e., the hedge portfolio) around the month of change.

Figure 3 shows the cumulative returns. For cross-sector competitors, the hedge portfolio's value drops sharply by more than 3% over the months that exhibit a significant increase in cross-sector C-Rank. The pools of all competitors and within-sector competitors also show reductions in stock prices, although at a slower phase than that of the cross-sector competitors. The negative price responses to large changes in C-Rank are consistent with the risk associated with high C-Rank values.

In the second test, we study the systematic pricing of C-Rank. We examine whether stocks that are more sensitive to a 'C-Rank factor' gain higher returns than stock that are less sensitive to the factor. We estimate the monthly C-Rank factor as the excess return of the C-Rank hedge portfolio (the difference between the returns of the top and bottom C-Rank quintiles). For each stock every month, we compute a 'C-Rank beta' using rolling regressions over the past 36 months of the firm's excess return on the C-Rank factor. The regressions control for the Fama and French (2015) five factors and the momentum factor. Every month we sort all stocks into five equal-sized portfolios based on their C-Rank beta. The portfolios are equal-weighted and held for one month.

The results reported in Table 9 suggest that high C-Rank beta firms outperform low C-Rank beta firms. The 6-factor alpha of the full market C-Rank beta hedge portfolio is 0.52% per month with a *t*-statistic of 2.47. Consistent with the effect of the C-Rank itself on stock return, the effect of C-Rank beta is also derived by cross-sector competitors. The positive relation between C-Rank beta and future stock returns is consistent with the argument that C-Rank captures some element of systematic risk.

To further assess the significance of this possible risk, we re-examine the pricing of C-Rank beta while controlling for the C-Rank level. We construct 5x5 double-sorted portfolios, first by C-Rank level and then by C-Rank beta. We calculate the beta return spread in each C-Rank level group and then average these return spreads each month. The average time-series return of this series can be interpreted as C-Rank beta spread neutralized to C-Rank level. The results reported in the upper panel of Table 10 show that the average risk-adjusted returns for the full market and cross-sector C-Ranks are 0.28% and 0.23%, respectively (*t*-statistics of 1.71 and 1.44). This means that the C-Rank level can explain roughly 50% of the C-Rank beta return spread. Performing the opposite sorting, first by C-Rank beta and then by C-Rank level (reported in the lower panel), indicates that C-Rank beta explains only a small part of C-Rank level return (controlling for C-Rank beta, the 6-factor alpha drops from 1.35% to 0.99%). Additionally, untabulated results show that when including both C-Rank beta and C-Rank level in cross-sectional regressions, only C-Rank level remains statistically significant. We conclude that while there is some evidence in favor of systematic risk pricing related to competitiveness, it seems that the return predictability is largely consistent with mispricing, as investors are slow to adjust for valuable information in financial statements.

7. The importance of the C-Rank features

As described in Section 2, the PageRank-type algorithm we employ to produce C-Rank gauges the competition-importance of any individual firm from the simultaneous competition-link system across all firms. This means that the C-Rank measure is based on two key and unique features.

The first feature is that a given firm's competition is determined not only by its own financial statement but also by what other firms say about the given firm in their reports. The second feature is that C-Rank gives more weight to the stronger firms (i.e. those that more firms mention them as competitors). We demonstrate that both these features are important in capturing firm competitiveness.

To address the importance of the first feature we posit that the market value of a company is likely negatively affected by the success of its real competitors. We therefore study the sensitivity of the firm's market value to the performance of two groups of competitors: its mentioning companies, and the companies it mentions. A stronger effect of the mentioning firms will support the importance of C-Rank, i.e., that the competition status of a firm cannot be fully assessed by only looking at the firm's own statement.

Inspired by Cohen and Frazzini (2008), we perform an event-time analysis. At the beginning of each month we divide all firms into five equal-sized portfolios according to the average past 12-month return of (i) their mentioning firms (the companies that mention the firm in their recent annual financial statement), and (ii) their mentioned firms (the companies that the firm mentions in its recent annual financial statement). Information from annual statements is taken with a three-month lag. The average return of each competitor group is value-weighted by the firm C-Rank. Figure 4 shows the average buy-and-hold abnormal return of companies with under- and over-performing mentioning firms, and Figure 5 shows the abnormal return of companies with under- and over-performing mentioned firms. Abnormal stock returns are given by comparing raw returns to size/book-to-market/industry benchmarks (the equal-weighted average return of firms in the industry-specific 5x5 size/book-to-market portfolio that includes the firm).

The abnormal returns in Figure 4 show clearly that when the mentioning competitors from outside the sector perform well in a given year, the mentioned firms underperform in the next two years, by 2% against their benchmarks. And if the mentioning competitors perform poorly, the mentioned firms overperform against their benchmarks, by up to 5% in the following two years. The performance of mentioning competitors from inside the sector do not show a clear effect on the performance of the mentioned firms. These results demonstrate that a firm's real competition is captured by its cross-sector C-Rank: if the mentioning firms do well, they might be able to adversely affect the mentioned firm.

The return patterns displayed in Figure 5 suggest that the past performance of the own-firm-mentioned competition group positively predicts the firm's return, especially cross-sector. This result contrasts the negative effect of the mentioning firms, suggesting that the competition captured by the firm's C-Rank cannot be uncovered by looking only at the firm's own statement.

Given the important role that the mentioning firms play in determining the competition status of a firm, we turn to addressing the second key feature of C-Rank, which is assigning more weight to stronger firms based on the cross-sectional competition links. As discussed above, the C-Rank provides a more accurate assessment of firm competitiveness than a simple mention count, as the C-Rank gives the appropriate weight to each mention. Yet because C-Rank and simple mention count are highly correlated (85-90% over the sample period), and because obtaining the C-Rank requires high computer processing power (solving simultaneously a dynamic system of thousands of equations), a valid question is how substantial the benefits from using C-Rank over a simple mention count are.

To address this question, we replicate the portfolio sort analysis of Table 4 when using the simple mention count (number of mentioning firms) as the sorting criterion. To incorporate the expected relevancy of the size of the mentioning firms, we consider two additional measures: the mean and the sum of the market capitalizations of the mentioning firms. As with C-Rank, we run monthly cross-sectional regressions of the three measures as of three months earlier on current firm size, and use the regression residuals as the sorting variables.

Figure 6 shows the mean excess return and 6-factor alpha of the hedge portfolios. All three alternative measures have a positive effect on future stock returns. Among the three measures, the simple mention count shows the strongest effect with mean excess return of 0.67% and 6-facor alpha of 1.01% per month. Yet these effects are still much weaker than that of the C-Rank, with return and alpha of 0.93% and 1.35%, respectively. These results indicate that C-Rank contains information relevant to a firm's competition strength that is not entirely captured by the alternative simple measures. This further emphasizes the importance of the C-Rank feature of giving an appropriate weight to each competition mention.

8. Conclusions

We produce a dynamic measure of firm competitiveness by analyzing the cross-references of firms to their competitors in annual financial statements. Our procedure is based on an advanced text analysis technology that allows identifying competitors in financial reports, and on a PageRank-type algorithm that simultaneously assesses the value of each firm's reference in its competitors' reports.

Our primary results indicate that firms with higher competition ranking (C-Rank) gain higher subsequent stock returns. This effect is significant after controlling for firm size and other common risk factors. The long-short investment strategy that buys high C-Rank stocks and shorts low C-Rank stocks generates an annualized 6-factor alpha of about 16%. Various robustness tests as well as Fama-MacBeth regressions corroborate this effect. The result is largely consistent with investor underreaction to firm business opportunities identified by other strong firms. Further tests utilizing data on analyst coverage support this conjecture. Nevertheless, stock return covariation with the C-Rank portfolio spread suggests that part of the return predictability can be interpreted as compensation for systematic disruption risk.

The results throughout the paper show consistently that the high return associated with high C-Rank firms mainly stems from cross-sector mentioning, suggesting that a firm's competitiveness is coming primarily from its ability to compete across different business environments.

References

Antón, Miguel, and Christopher Polk, 2014, Connected stocks, *Journal of Finance* 69, 1099-1127.

Ball Ray, and Philip Brown, 1968, An empirical evaluation of accounting income numbers, *Journal of Accounting Research* 6, 159-178.

Beaver, William H., Roger Clarke, and William F. Wright, 1979, The association between unsystematic security returns and the magnitude of earnings forecast errors, *Journal of Accounting Research* 17, 316-340.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen, 2019, Lazy prices, Harvard Business School working paper.

Cohen, Lauren, and Andrea Frazzini, 2008, Economic links and predictable returns, *Journal of Finance* 63, 1977-2011.

Garcia, Diego, and Øyvind Norli, 2012, Geographic dispersion and stock returns, *Journal of Financial Economics* 106, 547-565.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1-22.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

Frazzini, Andrea, and Lasse H. Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1-25.

Froot, Kenneth, Namho Kang, Gideon Ozik, and Ronnie Sadka, 2017, What do measures of real-time corporate sales tell us about earnings management, surprises and post-announcement drift? *Journal of Financial Economics* 125, 143-162.

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

Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124, 1423-1465.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-91.

Lee, Charles M.C., Stephen Teng Sun, Rongfei Wang, and Ran Zhang, 2019, Technological links and predictable returns, *Journal of Financial Economics* 132, 76-96.

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

Newey, Whitney K., and Kenneth D. West, 1987, A simple positive semidefinite heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd, 1999, The PageRank citation ranking: Bringing order to the web, *Technical Report*. Stanford InfoLab. URL http://ilpubs.stanford.edu:8090/422/.

Scherbina, Anna, and Bernd Schlusche, 2015, Economic linkages inferred from news stories and the predictability of stock returns, working paper.

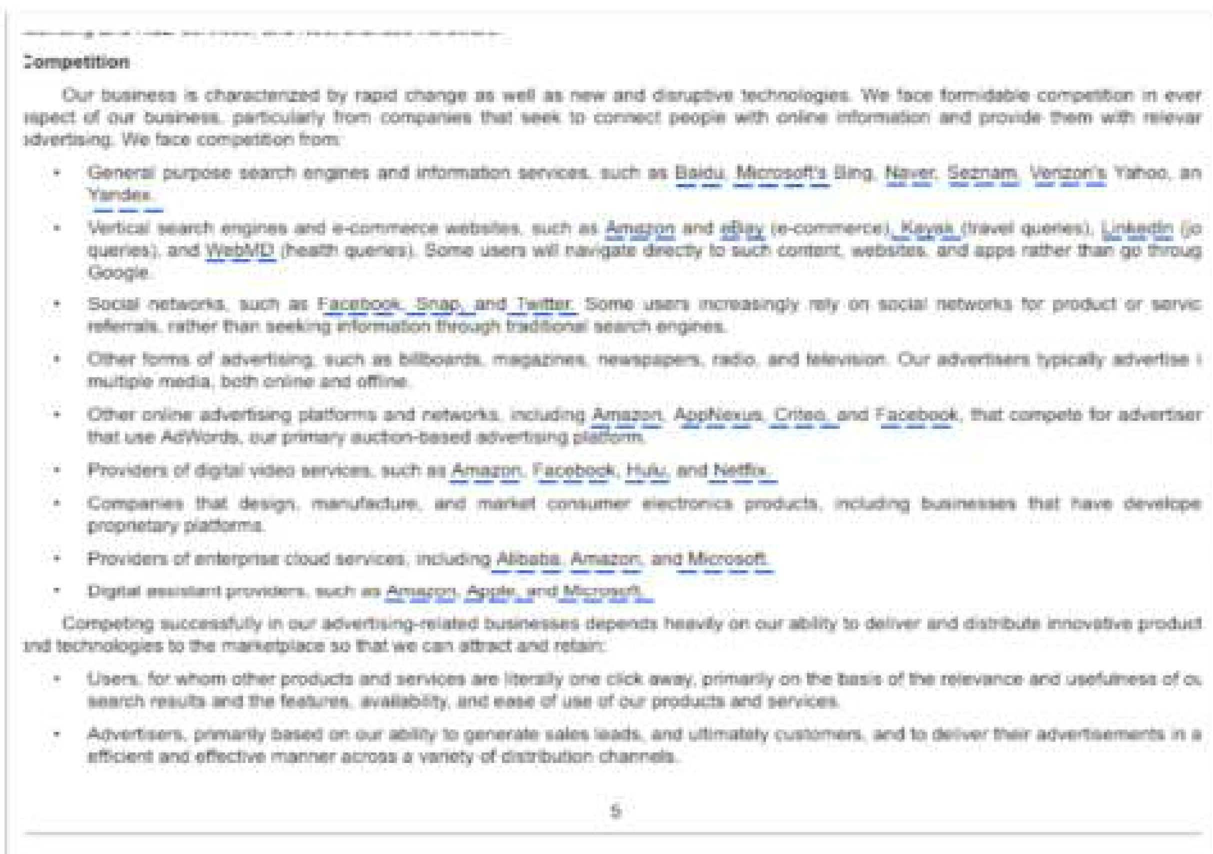
Sloan, Richard G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review* 71, 289-315.

Appendix A. Text analysis of competition sections in 10-Ks

Dataset

We match company tickers to CIKs, identifiers used by SEC-Edgar, and download from SEC-Edgar the 10-K filings. We observe a total of 119,785 10-Ks filed by 11,304 firms over the period 1995-2017. The focus of this paper is Part I / Item 1 – Business of the 10-K form. Although reporting firms are not required to designate a competition section in Item 1, we find that 68,952 of the forms used in this study (58%) include a designated section for competition. And about 39% of these competition sections include names of the company's competitors.

The example below is an extract from the 2017 10-K form filed by Alphabet Inc., parent company of Google. In Part I / Item 1 – Business, Alphabet designates a section to discuss its competitive environment. In this section it lists both the areas in which it faces competition (e.g., general search engines, vertical search engines, social networks, etc.) and the companies it considers as competitors in each of the areas.



Competition

Our business is characterized by rapid change as well as new and disruptive technologies. We face formidable competition in ever respect of our business, particularly from companies that seek to connect people with online information and provide them with relevar advertising. We face competition from:

- General purpose search engines and information services, such as Baidu, Microsoft's Bing, Naver, Seznam, Verizon's Yahoo, an Yandex.
- Vertical search engines and e-commerce websites, such as Amazon and eBay (e-commerce), Kayak (travel queries), LinkedIn (jo queries), and WebMD (health queries). Some users will navigate directly to such content, websites, and apps rather than go throug Google.
- Social networks, such as Facebook, Snap, and Twitter. Some users increasingly rely on social networks for product or servic referrals, rather than seeking information through traditional search engines.
- Other forms of advertising, such as billboards, magazines, newspapers, radio, and television. Our advertisers typically advertise i multiple media, both online and offline.
- Other online advertising platforms and networks, including Amazon, AppNexus, Criteo, and Facebook, that compete for advertiser that use AdWords, our primary auction-based advertising platform.
- Providers of digital video services, such as Amazon, Facebook, Hulu, and Netflix.
- Companies that design, manufacture, and market consumer electronics products, including businesses that have develope proprietary platforms.
- Providers of enterprise cloud services, including Alibaba, Amazon, and Microsoft.
- Digital assistant providers, such as Amazon, Apple, and Microsoft.

Competing successfully in our advertising-related businesses depends heavily on our ability to deliver and distribute innovative product and technologies to the marketplace so that we can attract and retain:

- Users, for whom other products and services are literally one click away, primarily on the basis of the relevance and usefulness of ou search results and the features, availability, and ease of use of our products and services.
- Advertisers, primarily based on our ability to generate sales leads, and ultimately customers, and to deliver their advertisements in a efficient and effective manner across a variety of distribution channels.

5

In total Alphabet lists twenty individual companies as competitors. These include domestic US firms such as Verizon and Microsoft, foreign firms (e.g., Baidu), and also private companies and private subsidiaries of public companies such as Hulu and Yahoo respectively. Some of the listed competitors appear multiple times as Alphabet considered them as competitors in multiple areas. Amazon which is mentioned five times is considered by Alphabet as a competitor in e-commerce search, online advertising, digital video, enterprise cloud, and digital assistance services.

Identifying firms in competition section

Once a designated competition section is found on a 10-K filing, our process attempts to identify which specific companies it lists. Since competitors are referred to by names using natural language, matching listed firms to security identifiers requires some additional text and language processing. We use an open-source natural language processing (NLP) tool, StanfordNER,[2] which is designed to label names of "things" in sequences of words. Each of the 68,952 designated competition sections is passed to the StanfordNER tool which is required to provide a list of text parts that are likely names of organizations. We consider each name of organization as a potential public company by matching against databases of public companies.

We apply a matching process that first searches for organization name on Edgar-SEC database, then on company name column of the CRSP master file, and finally we search Wikipedia using suspected organization names and in the cases of public companies parse the ticker following a "traded as" tag.[3] On average, we find 1,940 unique firms mentioned on 10-K filings of other companies each year.

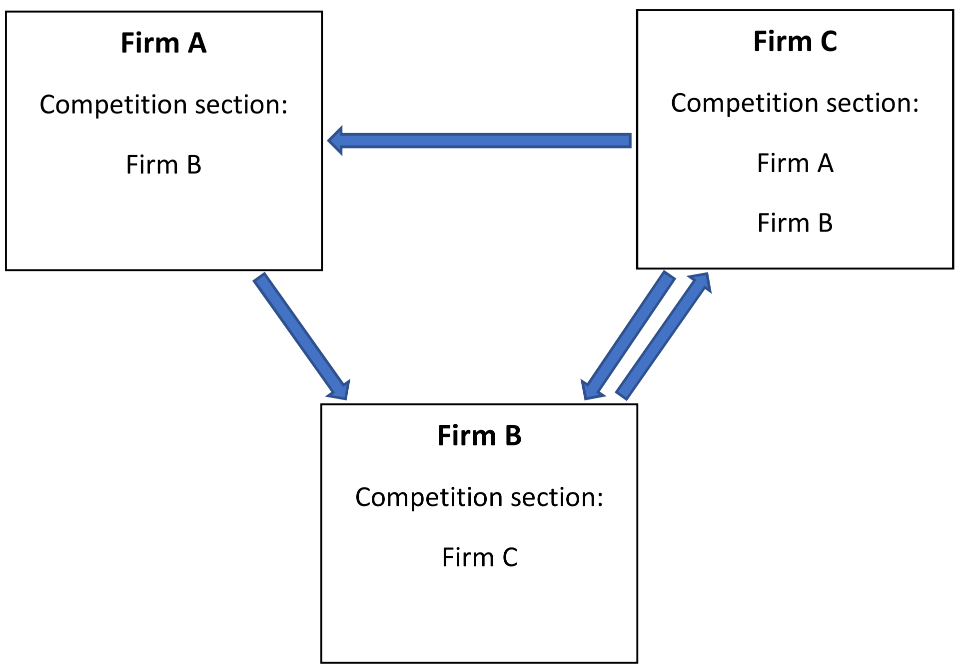
---

[2] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pp. 363-370. http://nlp.stanford.edu/~manning/papers/gibbscrf3.pdf https://nlp.stanford.edu/software/CRF-NER.shtml

[3] To increase the probability of matching suspected names of organizations to public companies we remove generic strings and suffixes such as Corp., LTD, LLC, etc. which are often used prior to processing the matching algorithm. We then use the standard text matching algorithms Sequence Matcher and Levenshtein Distance.

Appendix B. Applying the PageRank algorithm to competition links

We present a simple example to illustrate the use of the PageRank algorithm developed by the founders of Google, Larry Page and Sergey Brin (Page et al., (1999)) to measure firm competitiveness. Consider three firms, named A, B, and C, where each firm includes a competition section in its 10-K. Firm A mentions only Firm B as a competitor, Firm B mentions only Firm C as a competitor, and Firm C mentions both Firms A and B as competitors. The following figure shows the competition links across the three firms.



Applying the PageRank algorithm solves a system of linear equations for each firm C-Rank (CR):

$$CR(A) = \frac{1-d}{N} + d \times \frac{CR(C)}{2}$$

$$CR(B) = \frac{1-d}{N} + d \times \left[CR(A) + \frac{CR(C)}{2}\right]$$

$$CR(C) = \frac{1-d}{N} + d \times CR(B)$$

Where N denotes the number of firms, which is 3 in this example, d is a damping factor that assures that firms that are not mentioned at all will not converge all C-Rank values to zeros, and each firm's C-Rank on the right-hand-side is scaled by the number of firms it mentions (i.e., CR(A) and CR(B) are scaled by 1 and CR(C) is scaled by 2), such that all C-Rank values are summed to 1. Assuming a damping factor of 0.7 yields the following C-Rank values: $CR(A) = 0.2314$, $CR(B) = 0.3933$, and $CR(C) = 0.3753$. That is, Firm B gets the highest C-Rank as it is mentioned by both Firms A and C, and Firm C gets a higher C-Rank than Firm A as it is mentioned by a stronger firm (B and C, respectively).[4]

[4] When the system includes entities that do not point at all to other entities and/or entities that are not pointed at by other entities (as in our 10-K sample), the algorithm is a little more complex, requiring an iterative process of equation solving.

Table 1. C-Rank distribution

The table shows descriptive statistics of the C-Rank measures as described in Section 2, where all statistics are multiplied by 100. The sample includes 1,664,271 firm-month observations over the period 1995-2017.

	Mean	Stdev	min	p25	p50	p75	max
C-Rank full market	0.0172	0.0093	0.0123	0.0136	0.0149	0.0167	0.3422
C-Rank cross-sector	0.0193	0.0097	0.0143	0.0153	0.0174	0.0191	0.3673
C-Rank within-sector	0.2075	0.2317	0.0519	0.0942	0.1154	0.2324	10.0000

Table 2. Top competitors against largest companies over the sample period

The left panel shows the five companies (by ticker symbol) with the highest full market C-Rank competition status each year over the sample period. The right panel shows the largest companies for the same period.

	Top competitors					Largest firms				
Year	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
1995	IBM	HPQ	GE	NIPNY	ITC	GE	T	XOM	KO	MRK
1996	IBM	MSFT	HPQ	WMT	MSI	GE	KO	XOM	INTC	MSFT
1997	IBM	MSFT	HPQ	LU	JNJ	GE	KO	MSFT	XOM	MRK
1998	IBM	MSFT	HPQ	LU	MSI	MSFT	GE	INTC	WMT	XOM
1999	MSFT	IBM	LU	HPQ	MSI	MSFT	GE	CSCO	WMT	XOM
2000	MSFT	IBM	LU	HPQ	A	GE	XOM	PFE	CSCO	C
2001	IBM	MSFT	MSI	SIEGY	HPQ	GE	MSFT	XOM	C	WMT
2002	IBM	MSFT	HPQ	CSCO	GOOGL	MSFT	GE	XOM	WMT	PFE
2003	IBM	MSFT	CSCO	WMT	JNJ	GE	MSFT	XOM	PFE	C
2004	IBM	MSFT	WMT	CSCO	NVS	GE	XOM	MSFT	C	WMT
2005	IBM	WMT	MSFT	A	PFE	GE	XOM	MSFT	C	PG
2006	MSFT	IBM	WMT	ELMG	ABT	XOM	GE	MSFT	C	BAC
2007	IBM	MSFT	WMT	GE	GSK	XOM	GE	MSFT	T	PG
2008	MSFT	WMT	IBM	GE	A	XOM	WMT	PG	MSFT	GE
2009	IBM	MSFT	GE	ELMG	WMT	XOM	MSFT	WMT	AAPL	JNJ
2010	MSFT	WMT	GE	IBM	CSCO	XOM	AAPL	MSFT	GE	WMT
2011	MSFT	IBM	GE	BAC	ELMG	XOM	AAPL	MSFT	IBM	CVX
2012	MSFT	GOOGL	GE	WMT	IBM	AAPL	XOM	WMT	MSFT	GE
2013	GOOGL	MSFT	AAPL	WMT	IBM	AAPL	XOM	GOOGL	MSFT	GE
2014	GOOGL	MSFT	IBM	FB	WMT	AAPL	XOM	MSFT	JNJ	WFC
2015	GOOGL	FB	IBM	MSFT	MDT	AAPL	MSFT	XOM	AMZN	GE
2016	GOOGL	FB	PFE	NVS	MDT	AAPL	MSFT	XOM	AMZN	JNJ
2017	GOOGL	NVS	MDT	FB	PFE	AAPL	MSFT	AMZN	FB	JNJ

Table 3. Correlation between C-Rank and firm characteristics

The table shows the time-series averages of monthly cross-sectional correlations between the three C-Rank measures and firm characteristics. Firm size is computed as stock price multiplied by the number of shares outstanding (in logs). Market-to-book ratio is the market value of equity divided by the book value of equity (in logs). Past return is based on monthly stock returns over the last six months skipping the most recent month (see Jegadeesh and Titman (1993)). We estimate profitability by return on equity (ROE), computed by the annual income before extraordinary items divided by the previous year's book equity value. We estimate investment by the annual change in gross property, plant, and equipment, plus the change in inventories, scaled by lagged book value of assets. Market beta is estimated using a regression of a firm overlapping 3-day log return on the equivalent market return over the past year (see Frazzini and Pedersen (2014) for a similar procedure). We calculate idiosyncratic volatility for each month by the standard deviation of the residuals of regression of daily stock returns on the daily Fama and French (1993) three factors. Panel A shows the correlations for the full sample and Panel B for a subsample of competitive firms, which includes only firms that are recognized as competitors by other firms at least once over the past year. The sample period is 1995-2017.

Panel A. Full sample			
	C-Rank full market	C-Rank cross-sec	C-Rank within-sec
Log(size)	0.568	0.410	0.239
Log(market-to-book)	0.019	0.008	-0.003
Past return	0.008	0.003	0.012
Profitability	0.080	0.051	0.072
Investment	-0.018	-0.001	0.083
Beta	0.018	0.001	-0.109
Idiosyncratic volatility	-0.107	-0.069	-0.137

Panel B. Competitive firms			
	C-Rank full market	C-Rank cross-sec	C-Rank within-sec
Log(size)	0.583	0.398	0.361
Log(market-to-book)	0.025	0.006	0.000
Past return	0.003	0.002	0.017
Profitability	0.121	0.063	0.127
Investment	-0.039	-0.001	0.053
Beta	-0.074	-0.081	-0.147
Idiosyncratic volatility	-0.170	-0.112	-0.182

Table 4. Returns of portfolios sorted on C-Rank

We run monthly cross-sectional regressions of C-Rank as of three months earlier on current firm size, and use the regression residuals as our sorting variable. Each month we divide all stocks into five equal-sized portfolios according to their C-Rank-residual. The portfolios are equal-weighted and held for one month. The table shows the portfolios' mean excess monthly stock returns (in excess of the risk-free rate) and alphas from factor models. The CAPM uses the market factor. The factors in the 3-factor model are the Fama and French (1993) factors. The factors in the 4-factor model are the Fama-French factors augmented with a momentum factor. The factors in the 5-factor model are the Fama and French (2015) factors. The factors in the 6-factor model are the Fama-French factors augmented with a momentum factor. Panels A, B, and C show the results for the full market, cross-sector, and within-sector C-Rank measures. All returns and alphas are in percent per month and the corresponding *t*-statistics are in parentheses. The sample period is 1995-2017.

Panel A. Full market C-Rank	1-low C	2	3	4	5-high C	high-low
Mean excess return	0.77	0.87	0.80	0.97	1.70	0.93
	(2.60)	(2.40)	(2.25)	(2.55)	(4.03)	(3.76)
CAPM alpha	0.03	0.04	0.02	0.23	0.80	0.77
	(0.28)	(0.20)	(0.09)	(0.86)	(3.13)	(3.17)
3-factor alpha	-0.06	-0.10	-0.10	0.17	0.76	0.83
	(-0.98)	(-1.33)	(-0.84)	(0.77)	(3.68)	(3.73)
4-factor alpha	-0.03	0.00	0.06	0.46	1.07	1.11
	(-0.53)	(0.02)	(0.56)	(2.47)	(6.25)	(5.65)
5-factor alpha	-0.13	-0.04	0.00	0.41	1.03	1.16
	(-1.96)	(-0.48)	(0.03)	(1.85)	(4.97)	(5.32)
6-factor alpha	-0.10	0.03	0.12	0.61	1.24	1.35
	(-1.62)	(0.47)	(1.04)	(3.18)	(7.23)	(7.00)

Panel B. Cross-sector C-Rank	1-low C	2	3	4	5-high C	high-low
Mean excess return	0.80	0.87	0.93	0.90	1.61	0.81
	(2.71)	(2.39)	(2.39)	(2.42)	(3.87)	(2.87)
CAPM alpha	0.05	0.02	0.07	0.16	0.82	0.78
	(0.55)	(0.09)	(0.33)	(0.63)	(2.76)	(2.72)
3-factor alpha	-0.01	-0.10	-0.08	0.06	0.78	0.79
	(-0.16)	(-1.46)	(-0.65)	(0.33)	(3.00)	(2.95)
4-factor alpha	0.02	0.00	0.09	0.34	1.12	1.09
	(0.38)	(-0.07)	(0.86)	(2.02)	(4.88)	(4.46)
5-factor alpha	-0.03	-0.06	0.05	0.25	1.08	1.10
	(-0.41)	(-0.94)	(0.38)	(1.27)	(4.06)	(4.04)
6-factor alpha	0.00	0.00	0.15	0.44	1.30	1.30
	(-0.03)	(-0.01)	(1.50)	(2.56)	(5.58)	(5.22)

Panel C. Within-sector C-Rank	1-low C	2	3	4	5-high C	high-low
Mean excess return	0.96	0.93	0.97	1.33	0.94	-0.01
	(1.98)	(2.87)	(3.06)	(3.36)	(2.98)	(-0.04)
CAPM alpha	-0.10	0.19	0.27	0.52	0.24	0.34
	(-0.35)	(1.13)	(1.51)	(2.03)	(1.36)	(1.15)
3-factor alpha	-0.03	0.06	0.14	0.40	0.10	0.13
	(-0.16)	(0.57)	(1.23)	(2.21)	(0.75)	(0.58)
4-factor alpha	0.19	0.22	0.30	0.62	0.25	0.06
	(1.04)	(2.26)	(2.91)	(3.77)	(1.94)	(0.26)
5-factor alpha	0.46	0.06	0.14	0.55	0.07	-0.39
	(2.64)	(0.55)	(1.12)	(2.94)	(0.52)	(-1.86)
6-factor alpha	0.59	0.17	0.25	0.70	0.18	-0.41
	(3.71)	(1.75)	(2.34)	(4.15)	(1.36)	(-1.98)

Table 5. Portfolio sorts based on C-Rank controlling for stock characteristics

We run monthly cross-sectional regressions of C-Rank as of three months earlier on current firm size, and use the regression residuals as our sorting variable. Each month, we first sort all stocks into quintiles based on a stock characteristic as described in Table 3. Within each characteristic quintile, the stocks are further sorted into quintiles according to their C-Rank/size regression residual, yielding 25 characteristic/C-Rank portfolios. For each of the portfolios we calculate the equal-weighted monthly stock return, and then for each C-Rank quintile we average across the characteristic quintiles, yielding five quintile-mean C-Rank returns. The table reports the 6-factor alpha of the difference between the top and bottom quintile-mean C-Rank returns. The factors include the Fama-French (2015) factors augmented with a momentum factor. The "Base results" referred to the single sort alpha appearing in Table 4. All alphas are in percent per month and the corresponding *t*-statistics are in parentheses. The sample period is 1995-2017.

	6-factor alpha of C-Rank quintile spreads		
	Full market	Cross-sector	Within-sector
Base results	1.35	1.30	-0.41
	(7.00)	(5.22)	(-1.98)
Sorting characteristic			
Size	1.07	0.77	-0.29
	(8.42)	(6.83)	(-1.50)
Market-to-book	1.18	1.07	-0.45
	(6.90)	(4.92)	(-2.23)
Past return	1.16	1.27	-0.11
	(7.23)	(5.96)	(-0.64)
Profitability	1.31	1.30	-0.17
	(7.74)	(6.04)	(-0.94)
Investment	1.31	1.34	-0.19
	(7.10)	(5.56)	(-1.04)
Beta	1.13	1.19	0.00
	(6.58)	(4.96)	(0.01)
Idiosyncratic volatility	1.19	1.27	0.04
	(8.04)	(6.58)	(0.22)

Table 6. Robustness over time and horizon

We run monthly cross-sectional regressions of the full market C-Rank as of three months earlier on current firm size, and use the regression residuals as our sorting variable. Each month we divide all stocks into five equal-sized portfolios according to their C-Rank-residual. The portfolios are equal-weighted and held for one month. The table reports 6-factor alphas where the factors are the Fama-French (2015) factors augmented with a momentum factor. The full sample period is 1995 to 2017. The full sample period is broken up into subsamples in Panel A. Recession periods are based on NBER recession dummy. The holding period is increased to 3, 6, 12, and 18 months in Panel B. All alphas are in percent per month and the corresponding *t*-statistics are in parentheses.

	1-low C	2	3	4	5-high C	high-low
Full sample	-0.10	0.03	0.12	0.61	1.24	1.35
	(-1.62)	(0.47)	(1.04)	(3.18)	(7.23)	(7.00)
Panel A. Subsamples						
Excluding January	-0.08	0.04	-0.08	0.03	0.72	0.80
	(-1.25)	(0.59)	(-0.77)	(0.21)	(5.03)	(5.01)
Excluding Recessions	-0.11	0.02	0.15	0.69	1.28	1.39
	(-1.67)	(0.29)	(1.38)	(3.58)	(7.03)	(6.74)
1995-2006	-0.18	0.12	0.35	0.98	1.78	1.96
	(-1.75)	(1.03)	(1.86)	(3.24)	(6.04)	(5.76)
2007-2017	-0.04	-0.06	-0.05	0.22	0.73	0.77
	(-0.72)	(-0.83)	(-0.44)	(0.96)	(4.33)	(4.48)
Panel B. Longer investment horizons						
3 months	-0.10	0.05	0.21	0.64	1.13	1.23
	(-1.57)	(0.70)	(1.88)	(3.35)	(6.95)	(6.80)
6 months	-0.09	0.08	0.25	0.68	1.06	1.15
	(-1.35)	(1.12)	(2.19)	(3.49)	(6.83)	(6.69)
12 months	-0.05	0.06	0.29	0.71	0.96	1.01
	(-0.80)	(0.88)	(2.40)	(3.65)	(6.66)	(6.49)
18 months	-0.03	0.08	0.32	0.71	0.88	0.91
	(-0.38)	(1.14)	(2.60)	(3.67)	(6.28)	(6.07)

Table 7. Fama-MacBeth regressions

We run cross-sectional Fama and MacBeth (1973) regressions each month of excess stock returns. The independent variables are C-Rank/size residual as described in Table 4 (full market, cross-sector, and within-sector), log market capitalization, log market-to-book ratio, past six-month return, profitability, investment intensity, market beta, and idiosyncratic volatility, as described in Table 3. We run the regressions on the full sample and on a subsample of competitive firms, which includes only firms that are recognized as competitors by other firms at least once over the past year. All coefficients are multiplied by 100 and Newey-West corrected *t*-statistics (with twelve lags) are in parentheses. The sample period is 1995-2017.

	C-Rank full market		C-Rank cross-sector		C-Rank within-sector	
	All firms	Competitive firms	All firms	Competitive firms	All firms	Competitive firms
Intercept	3.54	4.07	3.38	2.74	3.41	4.07
	(5.59)	(5.13)	(5.32)	(2.46)	(5.45)	(5.19)
C-Rank	0.71	0.35	0.46	0.31	0.05	0.11
	(5.24)	(2.91)	(3.02)	(1.99)	(0.70)	(1.05)
Log(size)	-0.18	-0.21	-0.17	-0.13	-0.17	-0.21
	(-4.54)	(-4.23)	(-4.31)	(-1.94)	(-4.47)	(-4.26)
Log(market-to-book)	-0.11	-0.02	-0.11	-0.09	-0.11	0.04
	(-1.94)	(-0.36)	(-2.05)	(-1.01)	(-2.02)	(0.41)
Past return	0.27	0.22	0.24	0.23	0.24	0.17
	(1.03)	(0.61)	(0.92)	(0.52)	(0.93)	(0.49)
Profitability	0.20	-0.09	0.20	0.32	0.19	0.00
	(0.67)	(-0.27)	(0.65)	(0.71)	(0.65)	(-0.01)
Investment	-2.32	-1.65	-2.40	-1.84	-2.49	-2.00
	(-5.45)	(-3.03)	(-5.52)	(-2.16)	(-5.99)	(-3.63)
Market Beta	0.08	0.10	0.08	0.12	0.09	0.12
	(0.56)	(0.54)	(0.57)	(0.60)	(0.63)	(0.61)
Idiosyncratic volatility	-9.40	-7.69	-8.22	4.70	-8.26	-6.53
	(-1.48)	(-1.09)	(-1.27)	(0.38)	(-1.29)	(-0.95)

Table 8. C-Rank return predictability and analyst coverage

The table reports 6-factor alphas of the C-Rank hedge portfolios, as described in Table 4, for different subsamples. The first column shows the alphas for all firms as appear in Table 4. The second column includes only firms that are covered by at least three analysts in a year. This subsample is further divided into three equal-sized subgroups of stocks classified by their mean analyst industry concentration, which is measured as follows. For each analyst appearing in the IBES dataset, we calculate the proportions of firms in each two-digit SIC industry that the analyst covers during a year. From these industry proportions we calculate the Herfindahl-Hirschman Index (HHI) as a measure of the analyst's industry concentration. For each firm in each year, we calculate the mean industry concentrations of all analysts that cover the firm. All alphas are in percent per month and the corresponding *t*-statistics are in parentheses. The sample period is 1995-2017.

	6-factor alpha of the C-Rank hedge portfolio					
	All firms	Firms with analyst coverage	Mean analyst industry concentration			
			Low	Mid	High	High-Low
Full market	1.35	0.66	0.41	0.70	0.73	0.32
	(7.00)	(4.32)	(2.32)	(3.85)	(2.85)	(0.94)
Cross sector	1.30	0.58	0.21	0.43	0.91	0.69
	(5.22)	(3.54)	(1.21)	(2.07)	(3.04)	(2.19)
Within sector	-0.41	-0.50	-0.42	-0.54	-0.47	-0.04
	(-1.98)	(-2.20)	(-1.80)	(-2.05)	(-1.40)	(-0.13)

Table 9. Portfolios sorted by C-Rank factor beta

For each firm in each month we run a rolling regression over the past 36 months of the firm's excess returns (in excess of the risk-free rate) on the C-Rank factor, which is the mean excess return of the C-Rank hedge portfolio as described in Table 4. The regressions also control for the Fama and French (2015) five factors and the momentum factor. Referred to the coefficient of the C-Rank factor as 'C-Rank beta'. Each month we divide all stocks into five equal-sized portfolios according to their C-Rank beta. The portfolios are equal-weighted and held for one month. The table shows the portfolios' mean excess monthly stock returns and 6-factor alphas as in Table 4. Panels A, B, and C show the results for the full market, cross-sector, and within-sector C-Rank betas. All returns and alphas are in percent per month and the corresponding *t*-statistics are in parentheses. The sample period is 1998-2017.

Panel A. Full market C-Rank beta	1-low beta	2	3	4	5-high beta	high-low
Mean excess return	0.88	0.87	0.92	0.98	1.20	0.32
	(2.41)	(2.79)	(2.90)	(2.64)	(2.27)	(1.13)
6-factor alpha	0.20	0.19	0.26	0.37	0.73	0.52
	(2.18)	(2.92)	(3.53)	(3.60)	(3.26)	(2.47)

Panel B. Cross-sector C-Rank beta	1-low beta	2	3	4	5-high beta	high-low
Mean excess return	0.87	0.91	0.92	0.98	1.16	0.29
	(2.39)	(2.95)	(2.88)	(2.64)	(2.20)	(1.00)
6-factor alpha	0.20	0.23	0.25	0.38	0.69	0.49
	(2.12)	(3.30)	(3.60)	(3.56)	(3.03)	(2.15)

Panel C. Within-sector C-Rank beta	1-low beta	2	3	4	5-high beta	high-low
Mean excess return	1.01	0.95	0.92	0.90	1.07	0.06
	(2.06)	(2.77)	(2.98)	(2.74)	(2.45)	(0.20)
6-factor alpha	0.60	0.37	0.30	0.19	0.30	-0.30
	(3.23)	(3.86)	(3.91)	(1.99)	(1.60)	(-1.14)

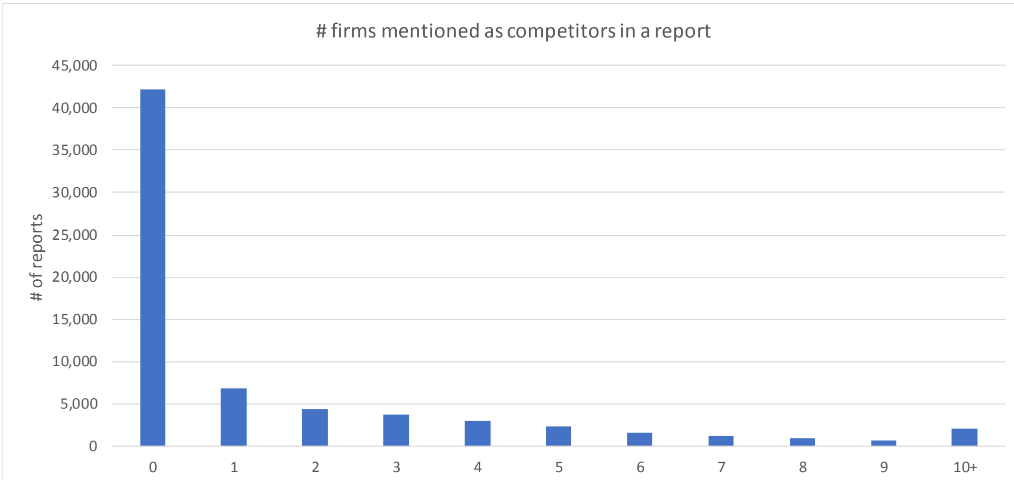
Table 10. 6-factor alphas of portfolios double-sorted on C-Rank and C-Rank factor beta

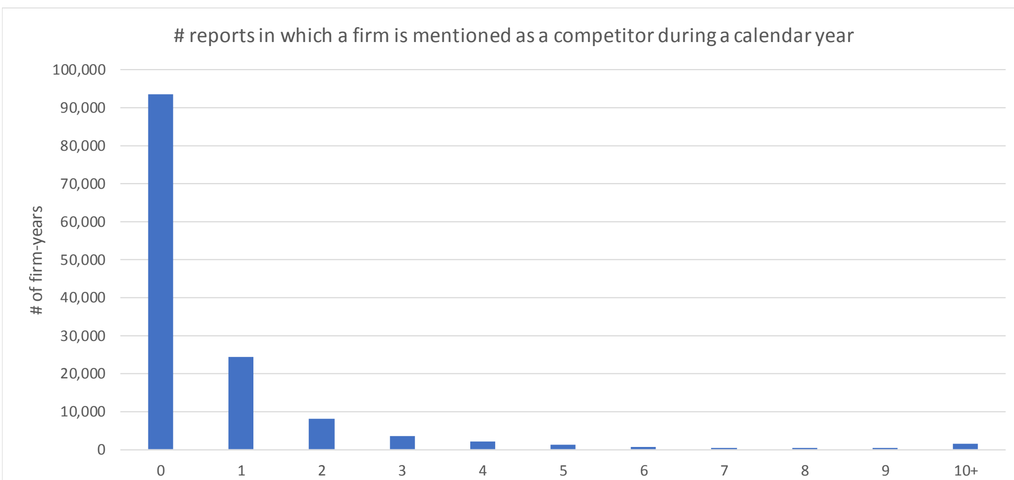
In the upper panel, the 'single sort' column shows the 6-factor alpha on the hedge portfolios sorted on C-Rank factor beta, as appear in Table 9. The 'C-Rank neutral' column shows the 6-factor alphas on the C-Rank-beta hedge portfolio when controlling for the effect of C-Rank, as follows. We first sort each month all stocks into quintiles based on C-Rank as described in Table 4. Within each C-Rank quintile, the stocks are further sorted into quintiles according to their C-Rank factor beta as described in Table 9, yielding 25 C-Rank/C-Rank-beta portfolios. For each of the portfolios we calculate the equal-weighted monthly stock return, and then for each C-Rank-beta quintile we average across the C-Rank quintiles, yielding five quintile-mean C-Rank-beta returns. The panel reports the 6-factor alpha of the difference between the top and bottom quintile-mean C-Rank-beta returns. In the lower panel we reverse the sorting order; the 'single sort' column shows the 6-factor alpha on the hedge portfolios sorted on C-Rank, as appear in Table 4; the 'C-Rank-beta neutral' columns shows the 6-factor alphas on the C-Rank hedge portfolio when controlling for the effect of C-Rank-beta. All alphas are in percent per month and the corresponding *t*-statistics are in parentheses. The sample period is 1998-2017.

C-Rank-beta return spread		
	Single sort	C-Rank neutral
Full market	0.52	0.28
	(2.47)	(1.71)
Cross sector	0.49	0.23
	(2.15)	(1.44)
Within sector	-0.30	-0.25
	(-1.14)	(-1.14)
C-Rank return spread		
	Single sort	C-Rank-beta neutral
Full market	1.35	0.99
	(7.00)	(6.33)
Cross sector	1.30	0.95
	(5.22)	(4.54)
Within sector	-0.41	0.12
	(-1.98)	(0.83)

Figure 1. Distribution of competitor mentions

The upper figure shows the distribution of the number of firms mentioned as competitors in a report for a total of 68,952 10-Ks with competition sections over the period 1995-2017. The middle figure shows the distribution of the number of reports in which a firm is mentioned as a competitor during a calendar year for a total of 135,921 firm-years. The bottom figure shows the joint distribution.





% reports in which a firm is mentioned

% competitors mentioned in a report

	0	1	2	3	4	5	6	7	8	9	10+	Total
0	50.48	5.60	2.12	0.98	0.57	0.36	0.22	0.14	0.09	0.08	0.45	61.09
1	8.37	0.95	0.30	0.12	0.06	0.04	0.02	0.01	0.01	0.01	0.05	9.94
2	5.13	0.76	0.30	0.08	0.07	0.03	0.01	0.01	0.01	0.01	0.02	6.43
3	4.24	0.67	0.29	0.12	0.06	0.05	0.03	0.01	0.01	0.00	0.03	5.51
4	3.15	0.56	0.29	0.12	0.07	0.04	0.02	0.01	0.01	0.01	0.02	4.30
5	2.34	0.46	0.24	0.11	0.06	0.03	0.02	0.02	0.02	0.01	0.02	3.33
6	1.64	0.38	0.16	0.08	0.04	0.02	0.02	0.01	0.01	0.00	0.02	2.38
7	1.20	0.29	0.13	0.06	0.04	0.01	0.01	0.00	0.01	0.00	0.03	1.78
8	0.88	0.20	0.11	0.06	0.03	0.02	0.01	0.01	0.01	0.00	0.01	1.34
9	0.59	0.17	0.08	0.04	0.04	0.02	0.01	0.01	0.00	0.00	0.02	0.98
10+	1.50	0.52	0.33	0.17	0.14	0.09	0.06	0.03	0.02	0.00	0.08	2.94
Total	79.52	10.56	4.35	1.94	1.18	0.71	0.43	0.26	0.20	0.12	0.75	100.00

Figure 2. Cumulative return

The figure plots the cumulative excess return and 6-factor alpha of the zero-investment strategy that buys high C-Rank stocks and shorts low C-Rank stocks, as described in Panel A of Table 4.

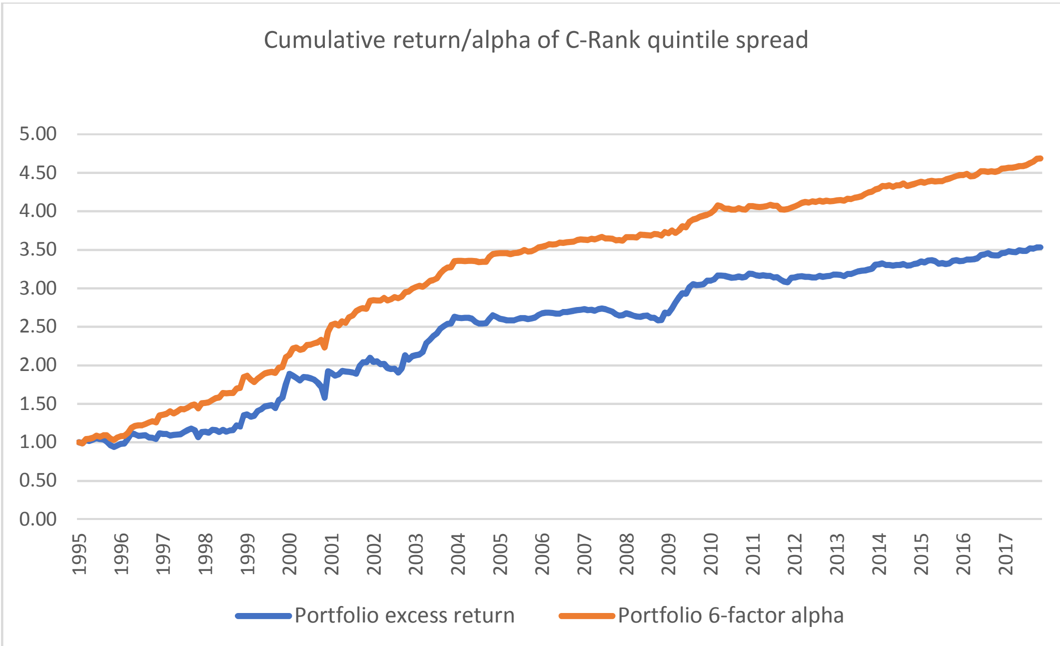


Figure 3. Effect of change in C-Rank

The sample contains all companies that are recognized as competitors by other firms. At the beginning of each month we divide all sample stocks into five equal-sized quintiles according to the difference in C-Rank from the prior month. The figure shows the average cumulative excess return of the hedge portfolio (the difference between the top and the bottom quintiles) from 12 months before the change in C-Rank (month 0) to 24 months after. The results are presented for the full market C-Rank as well as for the cross- and within-sector C-Ranks. The sample period is 1995-2017.

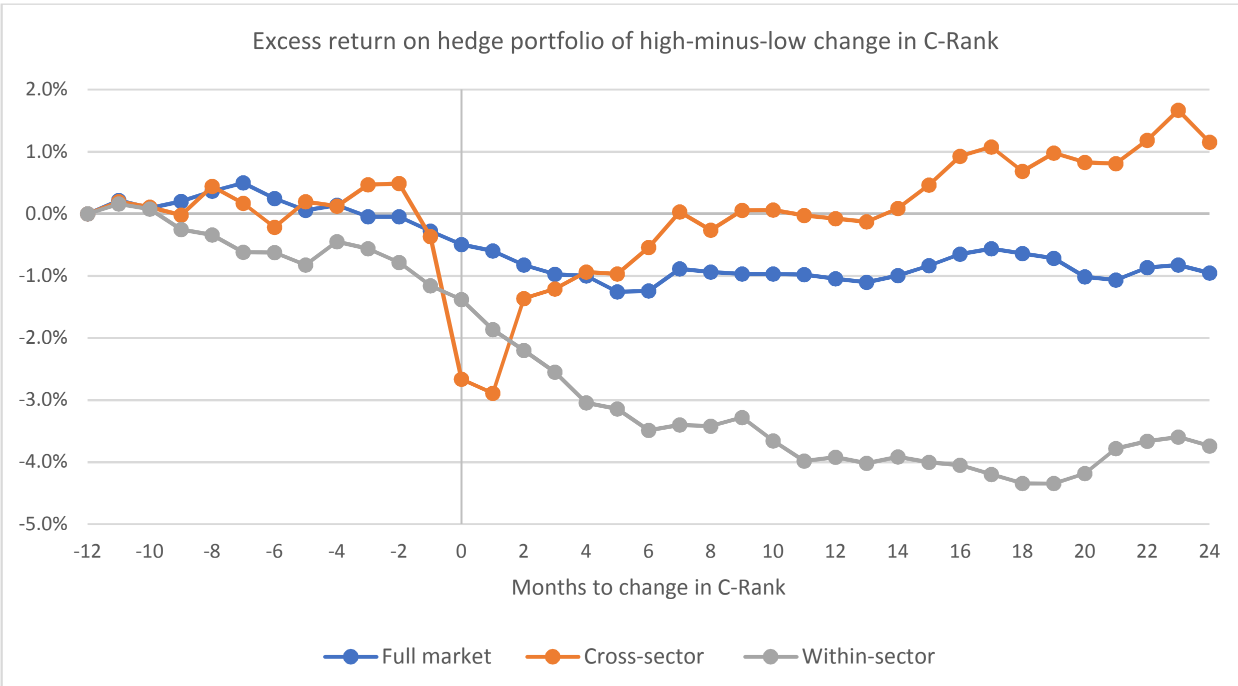
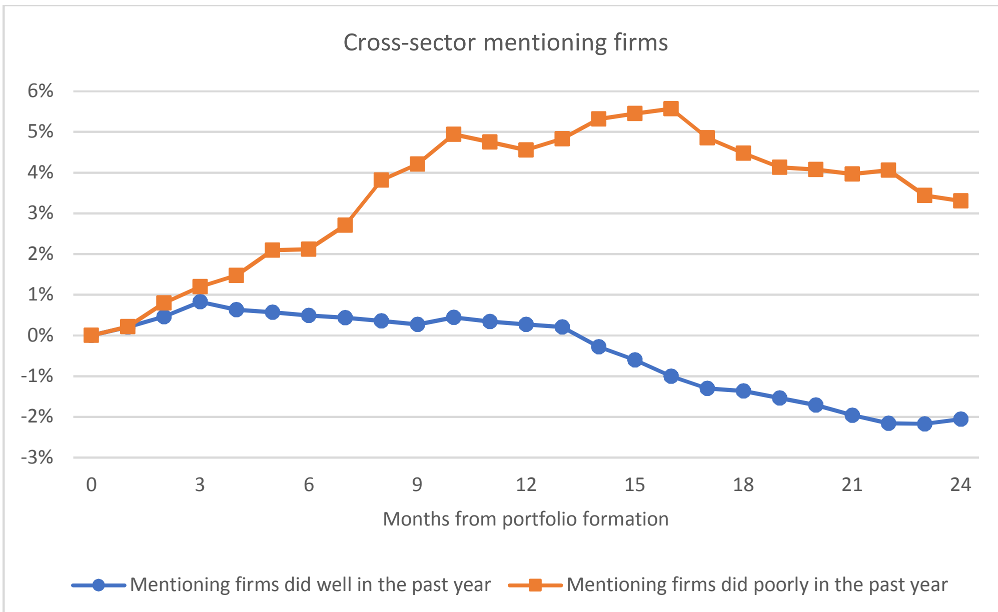


Figure 4. Mentioning firms shock, event-time cumulative return

At the beginning of each month we divide all firms into five equal-sized portfolios according to the average return of their mentioning firms in the past twelve months skipping the most recent month. The mentioning firm group includes the companies that mention the firm in their recent 10-K filings. Information from annual statements is taken with a three-month lag. The average return of each mentioning firm group is value-weighted by the mentioning firm C-Rank. The figure shows the average buy-and-hold abnormal return during the next 24 months for the top and bottom quintiles. Abnormal return is the difference between the firm's stock return and the equal-weighted average return of firms in the two-digit SIC industry-specific 5x5 size/book-to-market portfolio that includes the firm. The sample period is 1995-2017.



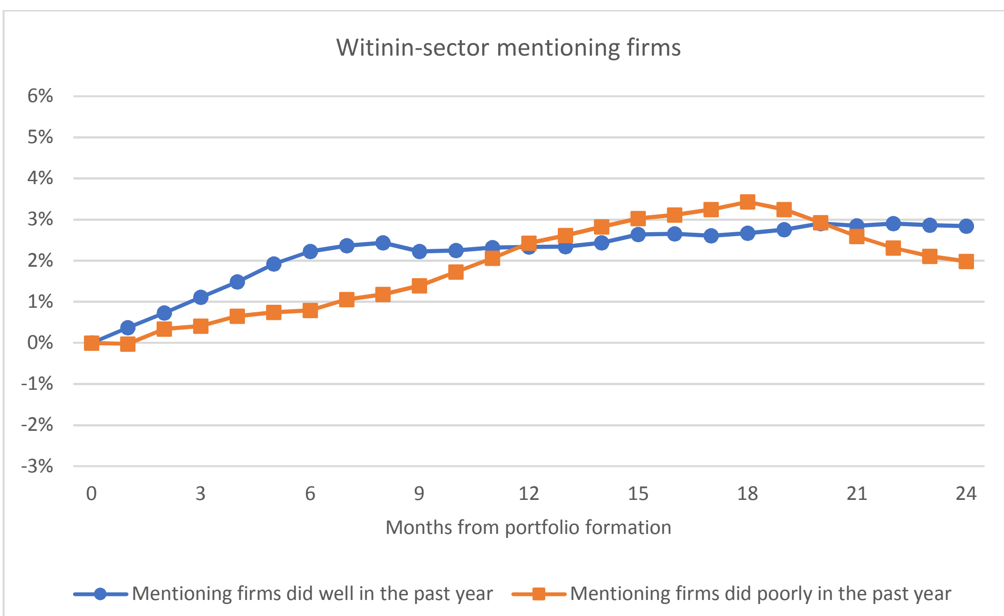
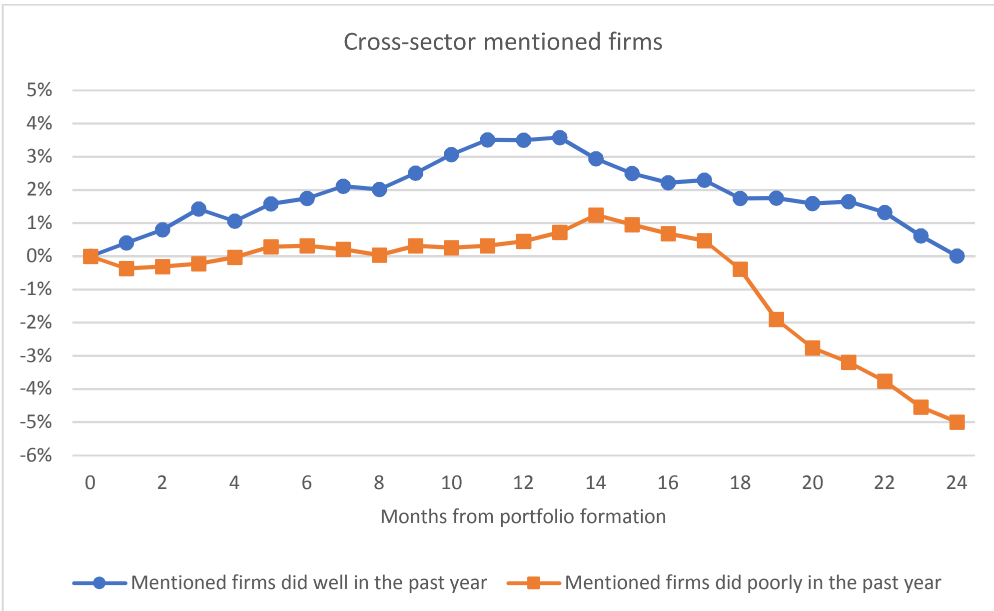


Figure 5. Mentioned firms shock, event-time cumulative return

At the beginning of each month we divide all firms into five equal-sized portfolios according to the average return of the firms it mentions in the competition section of its recent 10-K filing as of three months earlier. The average return of each mentioned firm group is value-weighted by the mentioned firm C-Rank. The figure shows the average buy-and-hold abnormal return during the next 24 months for the top and bottom quintiles. Abnormal return is the difference between the firm's stock return and the equal-weighted average return of firms in the two-digit SIC industry-specific 5x5 size/book-to-market portfolio that includes the firm. The sample period is 1995-2017.



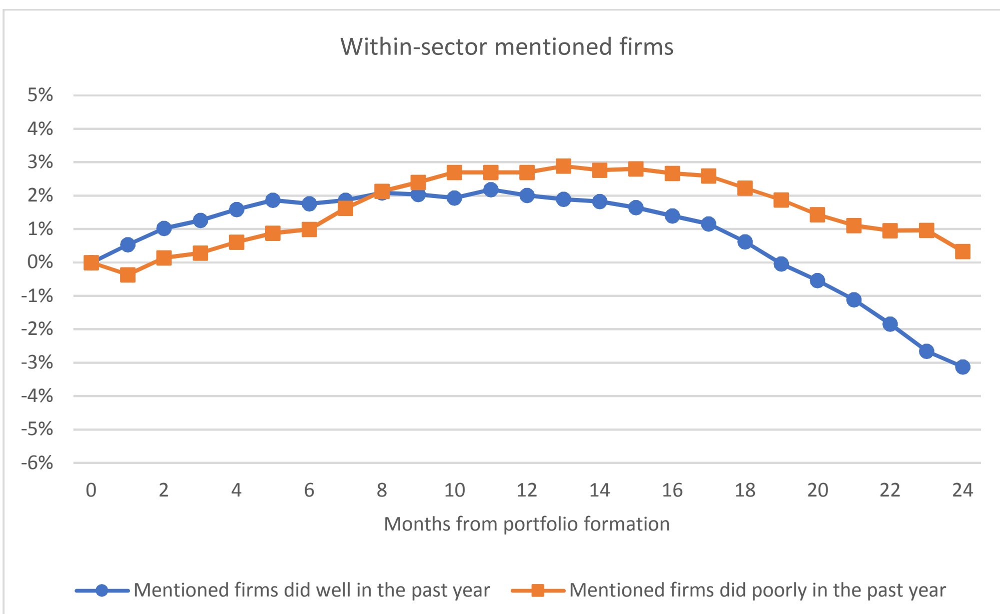


Figure 6. Returns of portfolios sorted on alternative measures of mentioning firms

We replicate the portfolio sort analysis in Table 4 using three alternative measures to C-Rank. The first is the number of 10-Ks in which the firm is mentioned as a competitor over the past twelve months ('number of mentioning firms'), the second is the mean market capitalization of the mentioning firms, and the third is the sum of the market capitalizations of the mentioning firms. As with C-Rank, we run monthly cross-sectional regressions of the three alternative measures as of three months earlier on current firm size, and use the regression residuals as the sorting variables. Each month we divide all stocks into five equal-sized portfolios according to each alternative variable. The portfolios are equal-weighted and held for one month. The table reports the mean excess monthly stock returns (in excess of the risk-free rate) and the 6-factor alpha of the difference between the top and bottom quintiles by each sorting variable. The factors include the Fama-French (2015) factors augmented with a momentum factor. All returns and alphas are in percent per month. The sample period is 1995-2017.