The Search for Benchmarks: When Do Crowds Provide Wisdom?

HARVARD | BUSINESS | SCHOOL

The Search for Benchmarks: When Do Crowds Provide Wisdom?

Charles M.C. Lee
Paul Ma
Charles C.Y. Wang

Working Paper

15-032

November 10, 2014

The Search for Benchmarks: When Do Crowds Provide Wisdom?*

Charles M.C. Lee
Stanford University
Graduate School of Business

Paul Ma
University of Minnesota
Carlson School of Management

Charles C.Y. Wang
Harvard University
Graduate School of Business Administration

November 10^{th}, 2014

Abstract

We compare the performance of a comprehensive set of alternative peer identification schemes used in economic benchmarking. Our results show the peer firms identified from aggregation of informed agents' revealed choices in Lee, Ma, and Wang (2014) perform best, followed by peers with the highest overlap in analyst coverage, in explaining cross-sectional variations in base firms' out-of-sample: (a) stock returns, (b) valuation multiples, (c) growth rates, (d) R&D expenditures, (e) leverage, and (f) profitability ratios. Conversely, peers firms identified by Google and Yahoo Finance, as well as product market competitors gleaned from 10-K disclosures, turned in consistently worse performances. We contextualize these results in a simple model that predicts when information aggregation across heterogeneously informed individuals is likely to lead to improvements in dealing with the problem of economic benchmarking.

JEL: D83, G11
Keywords: peer firm, benchmarking, EDGAR search traffic, co-search, analyst coverage, industry classification, wisdom of crowds

*The authors can be contacted at clee8@stanford.edu, paulma@umn.edu, and charles.cy.wang@hbs.edu. We thank Boris Groysberg, Paul Healy, Ryan Buell, Kai Du, Akash Chattopadhyay, Andrew Jing Liu, Daniel Malter, Tatiana Sandino, Pian Shu, Martin Szydlowski, Akhmed Umyarov, and Aaron Yoon for helpful comments and suggestions. We are very grateful to Scott Bauguess at the Securities and Exchange Commission for assistance with the EDGAR search traffic data. We also thank Kyle Thomas for excellent research assistance. All errors remain our own.

1. Introduction

The need to find appropriate economic benchmarks for individual firms is a fundamental issue for both researchers and practitioners. In many common applications, both in research and in practice, we aim to secure a relatively objective point of reference, or a benchmark, by which we can distinguish results (e.g., stock or accounting returns) attributable to firm-specific factors (such as managerial skill, or an idiosyncratic price differential) from results due to common factors (such as those associated with macro conditions that affect the entire set of economically-related firms). The need to identify fundamentally similar benchmarks for these purposes arise in applications ranging from performance evaluation and executive compensation to fundamental analysis, equity valuation, statistical arbitrage, and portfolio construction.

Recent research in the area of peer identification has uncovered a number of interesting findings that point to new ways to think about the age-old problem of economic benchmarking. Whereas traditional benchmarking methods rely primarily on industry classification schemes (such as the SIC codes, the Fama and French, 1997 industry groups, or the MSCI GICS groupings),[2] more recent approaches introduce new dimensions by utilizing novel data sources (such as self-reported competitors in 10-Ks), or new data analytic techniques (such as textual analysis of business descriptions culled from regulatory filings). Some of these approaches suggest we may need to rethink the relatively rigid classification schemes associated with industry groupings.

In this study, we conduct a comprehensive analysis of the state-of-the-art representatives of four broad categories of peer identification schemes nominated by either recent academic studies or financial practitioners as potential solutions to economic benchmarking. First, we consider the Global Industry Classification System (GICS), which has been

[2] Bhojraj et al. (2003) compares the efficacy of alternative industry classification schemes. Sometimes these schemes are augmented on other firm attributes, such as size, profitability, or expected growth (e.g., Bhojraj and Lee, 2002).

shown to be at the frontier of the standard industry classification schemes that group firms into mutually exclusive and collectively exhaustive groupings on the basis of similarities in inputs or product lines. Second, we consider two candidates that represent the frontier of benchmarking schemes that aim to identify product market competitors. One of these candidates, the "Text Network Industry Classification" ("TNIC"), uses a novel textual analytical technique to identify firms belonging to similar competitive product spaces (Hoberg and Phillips, 2010, 2014). Specifically, the more similar are two firms' 10-K business descriptions, the more likely they are product market competitors. Another candidate comes from Capital IQ ("CIQ"), who collects information on firms' self-reported product market competitors in regulatory filings (Rauh and Sufi, 2012; Lewellen, 2013).

We examine a third category of revealed-choice-based solutions, that is, schemes which aggregate individual agents' revealed choices to extract some latent intelligence or reveal the collective wisdom of investors with respect to the set of economically-related firms. We consider three candidates of this type: the search-based peers (SBPs) identified by Lee, Ma, and Wang (2014), hereafter LMW, based on investors' information co-search patterns on SEC's EDGAR website; the co-searched ticker symbols on Yahoo Finance (YHOO) (Leung et al., 2013); and the economically-related peers identified on the basis of analysts' co-coverage patterns of firms (analyst co-coverage peers, or "ACPs"; e.g., Ramnath, 2002, Liang et al., 2008, Israelsen, 2014, Kaustia and Rantala, 2013, and Muslu et al., 2014). Finally, we also consider peers from Google Finance ("GOOG") as a leading candidate from the class of benchmarking schemes based on a hybrid approach.[3]

Our results suggest that, under appropriate conditions, benchmarks extracted from the revealed choices of investors exhibit greater and more nuanced fundamental similarity with base firms. Our findings confirm and extend the results from LMW. Using a much longer time-series of EDGAR traffic data spanning 2003-2011, we show that their original

[3]Google Finance provides a list of "related companies" based on a proprietary algorithm. These peer firms are not identical to co-search-based peers identified by Google Knowledge Graph, and appear to be at least partially based on Factset industry classification schemes. We therefore refer to these Google "related companies" as the representatives from a hybrid approach.

findings hold over a ten-year period from 2004 to 2013, through both up and down markets. Specifically, in either the S&P500 or S&P1500 sample of base firms, SBPs — peer firms whose fundamentals are most commonly co-searched with the base firms' fundamentals on EDGAR — explain a much larger proportion of the cross-sectional variation in base firms' out-of-sample returns, valuation multiples, growth rates, and financial ratios than any of the alternative approaches. Among the other contenders, we find that ACPs — peer firms that are most commonly co-covered by analysts who cover the base firm — perform best. YHOO peers — those whose information are most commonly co-searched with the base firms' on Yahoo Finance — perform relatively poorly. Similarly, the peer firms identified by Google Finance and peers identified as product market competitors (TNIC and CIQ) turned in consistently worse performances.

Given SBPs' and ACPs' strong out-performance relative to YHOO, a natural question emerges: when does information aggregation across the revealed choices of a population of investors lead to better peer-firm selection? To provide intuition to this question, in the Appendix, we develop a simple model of aggregated co-search (co-coverage) decisions. The model features a population of agents, each of whom receives a private signal on the similarity between the base firm and the candidate peer firm. In this context, we show that the amount of information that can be gleaned through aggregation will depend on: (a) the inherent sophistication of the set of individuals involved, and (b) the size of the sampling population. When the population is "sufficiently sophisticated" (i.e., when the bias of the individual signals is low and the precision is high), the information environment surrounding a firm is of sufficiently high quality, and when there is a sufficiently large number of such agents in our sample making independent choices, their collective wisdom will lead to superior benchmarks. The model suggests that YHOO peers' relatively poor performance can be explained by the sophistication of Yahoo Finance users relative to that of either EDGAR users or sell-side analysts.

We also examine the differences between SBPs and ACPs. At a certain level of

abstraction, SBPs are the result of decisions by buy-side participants (investors) while ACPs are the result of decisions by sell-side analysts. Although both groups are likely to be more sophisticated than users of Yahoo Finance, there are differences in their incentive structures that could color the peer identification process. In particular, prior studies show sell-side analysts' stock recommendations tend to exhibit a bias in favor of larger growth firms with glamor characteristics (Jegadeesh et al., 2004). At the same time, due to informational constraints (Peng and Xiong, 2006; Van Nieuwerburgh and Veldkamp, 2010), sell-side analysts tend to specialize in a particular set of industries or sectors, and are less likely to cover stocks over widely divergent industries (Liang et al., 2008; Groysberg and Healy, 2013).

These priors are broadly confirmed in our tests. Specifically, we find that ACPs tend to be more anchored towards GICS industry classification, and exhibit a bias towards high growth firms. SBPs, on the other hand, are more likely to contain supply chain partners. We conjecture that these tendencies play an important role in explaining SBPs' out-performance relative to ACPs in aggregate.

Finally, we show that it is possible to combine SBPs and ACPs results to create a composite solution for identifying economically related firms. These are the best performing set of economic benchmarks examined in this paper, and outperform the standalone SBPs in explaining cross-sectional variations in returns. The improvement in performance from composite peers over that of SBPs is concentrated among the set of smaller base firms. Consistent with the predictions of our model, these results suggest that there is a greater value to information aggregation for firms operating in poorer information environments, where individual investors' signals are less precise.

Taken together, our results point to the aggregation of informed agents' revealed choices as a particularly promising venue through which to identify economically similar peer firms. The efficacy of this approach will depend on the intrinsic sophistication of the individuals in the population (i.e., the inherent level of collective wisdom attainable

through sampling), the quality of the information environment surrounding the firm, as well as the size of the sample itself. For the moment, it would appear the state-of-the-art benchmarking methodology is one that combines firms identified as SBPs and ACPs.

The remainder of the paper is organized as follows. Section 2 provides more explicit evidence of benchmarking behavior in the co-search patterns of EDGAR users and examines the performance of SBPs relative to six-digit GICS over a ten-year period, from 2004 to 2013. Section 3 compares SBPs' performance to those from alternative state-of-the-art benchmarking schemes suggested by industry and academic literature. Section 4 investigates the differences between SBPs and ACPs, and provides evidence on the performance for a composite revealed-choice-based benchmarking solution. Section 5 concludes.

2. Extended evidence on Lee, Ma, and Wang (2014)

Lee, Ma, and Wang (2014) develop a method for identifying economically-related peers based on investors' co-search traffic patterns at the SEC's EDGAR website. They find that firms appearing in chronologically adjacent searches by the same individual (what they refer to as "Search-Based Peers" or SBPs) are fundamentally similar on multiple dimensions. Specifically, they show SBPs dominate GICS6 industry peers in explaining cross-sectional variations in base firms' out-of-sample: (a) stock returns, (b) valuation multiples, (c) growth rates, (d) R&D expenditures, (e) leverage, and (f) profitability ratios. In addition, they show that "co-search intensity" (the fraction of a base firm's total co-searches owned by a given peer) captures the degree of similarity between firms.

We begin by establishing that the findings of LMW — that SBPs outperform peers from six-digit GICS industries in explaining the cross-sectional performance of firm performance — is not a transient phenomenon attributable to the three-year span from 2008 to 2010. Using an extended dataset and a more generalized co-search and robot detection algorithm, we establish that SBPs' out-performance of GICS6 systematically over a

ten-year period. We also provide evidence of benchmarking behavior on EDGAR.

2.1. Data and Descriptives

Our data comes from the traffic log on the Securities and Exchange Commission's (SEC) Electronic Data-Gathering, Analysis, and Retrieval (EDGAR) website, and is an updated version of the data used in LMW. The main advantage of the update is its greater time coverage; whereas the prior version of the data spans calendar years 2008 to 2011, this updated vintage contains data on visits to the EDGAR website from calendar years 2003 to 2011.[4] The format of the updated data is largely identical to the prior vintage: each observation in the raw data contains information on the visitor's IP address, timestamp, CIK, and accession numbers which uniquely matches to a particular company's specific SEC filing.

An important difference in this update is that the data extraction process the SEC used differs from the one employed for the prior vintage of the data; the process was changed in order to accommodate the longer time series. As a result, the new and the prior vintages of data are not identical in the overlapping period from 2008 to 2011.

The differences between these data vintages do not pose an issue so long as they do not systematically exclude certain types of search traffic. To further investigate the differences between these data vintages, Panel A of Table A1 reports a variety of daily-user level search traffic characteristics embedded in each of the two raw data samples in the overlapping time period from 2008 to 2011. Each observation is defined at the daily-user level, defined as a unique IP address on a given calendar day. We report the average of daily-user level search characteristics in each data vintage and compute the differences in the averages.

An immediate notable difference between these two vintages is the number of total daily users contained in the raw data: over the 2008 to 2011 period, the older vintage

[4]The new data sample extends to March 2012, but we do not use the partial data in 2012 in this paper.

contains 39.9 million unique daily-IP observations, whereas the updated data contains 35.0 million.[5] We also summarize the filing types accessed by users in each dataset, in terms of the percent of daily-users that access any 10-K's or 10-Q's, proxy filings, forms 3, 4, or 5 ("Insider"), S-1's, SEC comment letters, 13-F's, 13-G's, 13-D's, 6-K's, and 20-F's. Finally, we summarize and compare the two datasets, in terms of the average number of unique CIKs (firms) accessed, the number of total clicks (downloads), the number of unique filings types accessed, the number of unique file extension types accessed, and the estimated average number of hours spent on the site for a daily-user on EDGAR.

Examining these user characteristics, we conclude that the daily-user level search characteristics between these two datasets are not systematically and economically different. Because there are more than 30 million observations in each dataset, all differences between the datasets are statistically significant at the 5% level. However, none of these differences are economically significant. Across all search and user characteristics examined, the average absolute percentage point difference between the old and the new vintage of the dataset is 1.7%.

Table A1 is also interesting in offering several new stylized facts about raw EDGAR usage patterns. Specifically, we find that: 53% of daily-users click on either a 10-K or 10-Q, 28% click on an 8-K, 11% click on a proxy statement, and 10% click on an insider filing (forms 3,4 or 5) or a S-1, with diminishing interest for the other filings considered. We also report that the average daily-user spends 36 minutes on the site.[6]

For completeness, Panel B of Table A1 reports means and differences-in-means in these daily-user level search characteristics between the early half of the updated data (2003-2007) and the latter half (2008-2011).[7] An immediate difference to be noted is the

[5]In un-tabulated reports we confirm that within the overlapping 2008-2011 subsample, the new data is a strict subset of the prior vintage used in LMW.

[6]The estimated time on site per day is calculated by adding time spent per user session within a given day. As defined in LMW, a user session ends when there is no activity within a 60 minute window since the last action of the user.

[7]Note that the new data vintage misses several months of data between 2005 and 2006 due to the SEC's system constraints. Specifically, in un-tabulated reports, 93 (111) days in calendar year 2005 (2006) had fewer than 100 daily-users, compared against the sample average of 17,500 daily-users. This

usage of the EDGAR website, which has increased significantly over time in terms of the total number of unique daily-IPs: in the five-year period from 2003 to 2007, there was a total of 21 million unique daily users on EDGAR, a number that increased to almost 35 million over the four-year period from 2008 to 2011. Usage of EDGAR not only increased in the extensive margin but also the intensive margin: there was an increase in the average number of total clicks and unique CIKs and filing types accessed, as well as an increase in the average total session length in the latter half of the sample period. We note that some of these increases are likely driven by the increasing presence of web-crawlers or robots on the Internet, thus highlighting the importance of filtering for automated search traffic in this line of research.[8]

In general, the user-level characteristics of searches has remained relatively stable across the two time periods, though with some notable difference. For example, there has been an increase in the incidence of 10-K or 10-Q, 8-K, S-1, comment letter, 13-F, 6-K, and 20-F downloads. These patterns are consistent with EDGAR users on average downloading more information in the post 2008 period in a given daily session. In contrast, there was heightened demand for proxy statements and insider filings in the pre-2008 period relative to the post-2008 sample, which could be in part explained by the effect of governance failures and Sarbanes-Oxley. Overall, we conclude that the patterns observed in the updated vintage of the EDGAR search data do not exhibit any systematic biases or errors that raise concerns about their integrity.

2.2. Updated robot rule

One of the updates we make here to LMW is the methodology used in identifying and filtering out search traffic generated by automated scripts ("robots"), written to download

form of missing data is not likely to introduce a systematic bias, but would reduce the power of our approach.

[8]For example, http://www.incapsula.com/blog/bot-traffic-report-2013 reports that in 2013, 60% of website visits are robot-generated. This number represented a 21% increase from 2012, when 51% of all website visits are made by robots.

massive numbers of filings. We expect such search traffic to be uninformative and thus devise algorithms to filter them out. LMW used an absolute cutoff that classifies all daily IPs downloading more than 50 unique CIKs as a robot, a cutoff that corresponded to the 95^{th} percentile of user search activity in the 2008 to 2011 sample. Given the longer time series in our updated search data, in lieu of an absolute cutoff, we now use a robot identification strategy that classifies any daily user downloading more than the 95^{th} percentile in the distribution of unique CIK's within the corresponding calendar year as a robot.[9] As reported in Panel A of Table 1, keeping the traffic from daily IP addresses that searched for at least 2 unique CIKs' fundamentals — which we require for the co-search algorithm we describe below — and less than the 95^{th} percentile of the unique CIKs downloaded in that year reduces our sample from 3.72 billion (56.2 million) to 351.45 million daily pageviews (16.63 daily unique visitors).

2.3. Updated co-search algorithm and evidence on benchmarking

We infer investors' perceptions of relevant economic benchmarks by aggregating information from their fundamental information acquisition behavior. Under the assumption that the population of EDGAR users is collectively searching for firm fundamentals to aid their investment decisions, and that an important part of this process involves the comparison of these fundamentals to economically related benchmarks, we expect EDGAR users search sequences to be informative of their perceptions of the most relevant set of benchmark firms.

Empirically we observe evidence consistent with EDGAR users acquiring information on EDGAR for benchmarking purposes. For example, the average daily-user on EDGAR downloads information for two firms. Figure 1 shows additionally that, among those EDGAR users searching for information of more than one firm, a vast majority are

[9] A technicality to note here is that we use the 95^{th} percentile in the raw daily-user population. On the other hand, LMW identified the 50 CIK rule on the basis of the 95^{th} percentile in the distribution of daily-users that looked for information of at least two unique CIKs on EDGAR.

downloading the fundamentals of between two to five firms, consistent with benchmarking.

Benchmarking behavior can also be seen in other search sequence characteristics. Table 2 summarizes search sequence composition characteristics, conditional on having accessed a particular firm's filing type and accessing information for more than one firm. The first row shows that, when an investor has accessed one firm's 10-K or 10-Q, 44.84% of the remainder of her searches in the same session are for other firms' 10-K or 10-Qs. In contrast, a substantially smaller proportion of the remaining searches are for 8-Ks, comment letters, insider filings, proxies, S-1s, and other filings. The remaining rows show that when investors access non-10-K and non-10-Q other forms, they tend to co-search across a variety of forms of other firms. The substantially greater coincidence of 10-K and 10-Q searches across different firms suggests that benchmarking behavior is likely to be most pronounced among search sequences that access 10-Ks and 10-Qs.

Following this observation, we restrict our analysis to searches for 10-Ks and 10-Qs — including their amendments or small business variants — to focus on investors' patterns of acquiring fundamental information that most likely captures benchmarking behavior. The final filtered sample, as reported in row 4 in Panel A of Table 1, contains just over 115 million pageviews from 10.96 million daily users.

Using this filtered data, we extract the set of most relevant economic benchmarks to any particular firm i by defining *Annual search fraction*, f_{ij}^t, between the base firm i and a peer firm j in calendar year t:

$$f_{ij}^t = \frac{\Sigma_{d=1}^{365}(\text{Unique daily-user searches for } i \text{ and } j)_d}{\Sigma_{d=1}^{365}(\text{Unique daily-user searches for } i \text{ and any firm } j \neq i)_d}. \tag{1}$$

In words, f_{ij}^t is the fraction of unique daily-users searching for firm i's and another firm's information in a calendar year that also searched for j's information. This is a more generalized version of the co-search algorithm employed in LMW, which defined co-searches based on chronologically sequential clicks. For example, if a user clicks on

Google and then Yahoo, Yahoo is considered a peer of Google, but not vice versa. The co-search algorithm used in this study relaxes these chronological ordering restrictions, and consider firms i and j to be benchmarks for each other, so long as they are co-searched together by the same user within the same daily calendar EDGAR session window. In other words, we are building a network of firms with weighted undirected edges defined through co-searches, which is reasonable under the assumption that investors do not search for information of benchmarks in any particular systematic order.

Our *Annual search fraction* measure sums to one for each firm in a given year, and is easy to interpret. For example, $f^{2008}_{GOOGLE,YHOO}$=0.0602 means that 6.02% of daily-users searching for Google's fundamental information and at least one other firm in calendar year 2008, also searched for Yahoo's information. By construction, we do not use any information from users who only search for a singular firm's filings before leaving the EDGAR website.

Based on this measure, we define a given base firm's top 10 SBPs in a given calendar year as those peer firms with the ten highest values of *Annual search fraction* in the preceding calendar year. The analyses of this paper focuses on the set of base firms that belong to the S&P1500 index as of January 1 of each calendar year; however, no such restrictions are placed on the set of benchmark firms.[10] Panel B of Table 1 summarizes the coverage of base firms with valid SBPs in our final sample as well as the median number of SBPs per firm by year. Again, all our analyses below focus on the top ten SBPs of base firms.

2.4. Price co-movement

We now turn to investigate the performance of SBPs over the ten-year period from 2004 to 2013. Note that although we have only search traffic data from 2003 to 2011, we extend the 2012 SBPs derived from calendar year 2011 search traffic by one more year to

[10]Previously in LMW, peer firms were restricted to be within the same S&P1500 universe as base firms.

create SBPs for 2013, thus completing a ten-year sample.

Following LMW, our tests compare GICS6 and SBPs in their abilities to explain the cross-sectional variation in base firms' monthly stock returns and firm fundamentals. The intuition for these tests is that peer firms that are more economically related to their base firms should exhibit greater contemporaneous correlation with them in returns and in various accounting fundamentals.[11]

In Table 3, we estimate the following cross-sectional regression, for every month from 2004 to 2013:

$$R_{i,t} = \alpha_t + \beta_t R_{p_{i,t}} + \epsilon_{i,t}, \tag{2}$$

where $R_{i,t}$ is the CRSP monthly cum-dividend return for each base firm i, taken from the CRSP monthly files, and $R_{p_{i,t}}$ is the average monthly returns for a portfolio of benchmark firms specific to base firm i. We assess the relative performance between GICS6 and SBPs by comparing the average R^2 produced by monthly regressions using benchmark portfolios of all firms (excluding the base firm) selected from the base firms' GICS6 industries versus the average R^2 produced by portfolios of the base firms' top 10 SBPs. To avoid contamination from simultaneity of information, our SBPs are always identified using search traffic from the prior calendar year. For example, the *Annual search fraction* measure f_{ij}^t used to identify SBPs in calendar year 2009 are computed using calendar year 2008 data. Thus, we estimate cross-sectional regressions of Eq.(2) for every month from 2004 to 2013 and obtain an average R^2 based on the 120 regressions.

We consider two types of peer portfolios using SBPs. The first type of peer portfolio, denoted "SBP EW," takes the closest 10 peer firms as implied by our *Annual Search Fraction* measure and forms an equally weighted portfolio. The second type of peer portfolio, denoted "SBP TW" (traffic-weighted), takes the closest 10 firms as implied through our *Annual search fraction* measure but forms a weighted average portfolio, where

[11] See LMW as well as Lewellen and Metrick (2010) for detailed discussions on the mapping between higher R^2 and greater fundamental similarity between base and benchmark firms.

a firm's portfolio weight is the *Annual search fraction* measures rescaled to sum to one. To facilitate comparisons, all the regressions are conducted using the same underlying set of base firms, so that our analyses include only base firms with sufficient data from both GICS and SBPs.

Table 3 reports the average R^2 values from monthly regressions of Eq.(2), using base firms from the S&P1500 and the S&P500. Our results show that SBP portfolios significantly outperform GICS6 peer portfolios over the 10-year period.[12] For the group of S&P1500 base firms, their GICS6 peer portfolios explain, on average, 10.2% of the cross-sectional variation in monthly returns, significantly lower than the 12.8% (14.1%) explained by their SBP EW (TW) portfolios. Similarly, for the set of S&P500 base firms, their GICS6 peer portfolios explain, on average, 15.2% of the cross-sectional variation in monthly returns, again significantly lower than the 21.2% (23.6%) explained by their SBP EW (TW) portfolios. Finally, we also observe the out-performance of SBPs among the S&P MidCap 400 and S&P SmallCap 600 firms (collectively labeled as "S&P1000").

2.5. Co-movement in valuation multiples, financial ratios, and other characteristics

We also assess the extent to which SBPs explain the cross-section of base firms' valuation multiples, financial ratios, and key accounting measures. To perform these additional tests, we gather quarterly data from Compustat and Institutional Brokers' Estimate System (IBES) on a range of valuation multiples, financial ratios, and other fundamental characteristics, including the price-to-book multiples (pb), enterprise value-to-sales multiples (evs), price-to-earnings multiples (pe), returns on net operating assets ($rnoa$), returns on equity (roe), asset turnover (at), profit margins (pm), leverage (lev), long-term analyst growth forecasts ($ltgrowth$), one-year-ahead realized sales growth ($salesgrowth$),

[12]Unlike LMW, who formed GICS6 portfolios using 10 random GICS6 peers, we use all available GICS6 firms outside of the base firm. We choose this in part to better reflect GICS6 fixed effects, but also because this substantially improves the performance of GICS6 benchmark portfolios.

and research and development expenses scaled by net sales ($rdpersales$). The exact computation of these variables (as well as all others used in this paper) are detailed in Table A2.

With each of these variables, we run the analogous cross-sectional regression,

$$Variable_{i,t} = a_t + \beta_t Variable_{p_{i,t}} + \epsilon_{i,t}, \tag{3}$$

where $Variable_{i,t}$ is the variable of interest for each base firm i and the regressor $Variable_{p_{i,t}}$ is the portfolio mean value for i, based on either other firms belonging to the same GICS6 group or one of our two traffic-based measures (SBP EW and SBP TW). We estimate these regressions on a quarterly basis, at the end of March, June, September, and December of each calendar year from 2004 to 2013. The relevant variables are computed using financials that are available at the end of each quarter.[13] Similarly, we obtain the most up-to-date median long-term analyst forecasts from IBES at the end of each calendar quarter.

Following Bhojraj et al. (2003), for the entire firm quarter–year sample, we drop observations that are missing data on total assets, long-term debt, net income before extraordinary items, debt in current liabilities, or operating income after depreciation. We also drop observations with negative common or total equity and keep only observations with net sales exceeding $100 million and a share price greater than $3 at the end of the fiscal quarter. Finally, to mitigate the influence of outliers, we truncate observations at the first and 99th percentiles for each of the variables for each regression equation.[14] In addition, we require net income before extraordinary items to be positive and require non-missing values for current liabilities, current assets, and property, plants, and equipment in computing $rnoa$. To facilitate comparisons, all the regressions are conducted using the same underlying set of base firms.

[13]To ensure that our valuation multiples reflect publicly available accounting information, we use Compustat data for which earnings have been officially announced by the end of each quarter.

[14]This is done on an equation-by-equation basis to avoid losing observations unnecessarily.

Table 4 compares GICS6 and SBP portfolios and shows that SBP portfolios outperform the GICS6 peer portfolios for nearly all of the variables tested. Within the S&P1500 base firm sample, reported in Panel A, the SBP EW portfolios explain a significantly greater proportion of the cross-sectional variations than the GICS6 peer portfolios in all of the variables. We find similar results among the subset of S&P500 and S&P1000 firms. Across both subsamples, SBP EW (SBP TW) portfolios explain a significantly greater proportion of the cross sectional variation for 10 of the 11 (all 11 of the) variables examined, at the 10% or better level.

Overall, our results confirm and support the findings of LMW that SBPs substantially outperform GICS6 in terms of their ability to explain cross-sectional variations in stock prices and key valuation multiples, financial statement ratios, and other fundamental characteristics. More importantly, the prior findings which were established on the basis of the three-year period from 2008 to 2010 were not a transient phenomenon, but represent a systematic pattern over a 10 year period from 2004 to 2013. This evidence supports the powerful idea of extracting latent information from market participants' information acquisition patterns in identifying fundamentally-related firms.

3. Comparisons to alternative peer identification schemes

Having reaffirmed the out-performance to GICS6, in this section we extend the analyses above by comparing the performance of SBPs to a number of alternatives that collectively represent the frontier of peer identification schemes proposed by both industry and academia.

3.1. Google Finance and Capital IQ peers

We begin by comparing SBPs to the set of firm benchmarks on Google Finance and Capital IQ. We assembled "GOOGLE" peers by downloading the "Related Firms" listed on each firm's Google Finance page as of June 2014. Our understanding is that Google generates the list through a proprietary algorithm, with FactSet Research's own proprietary industry classification as one of the inputs. We also download a June 2014 snapshot of product market competitors from Capital IQ. Capital IQ collects the set of companies that a given firm i considers to be its competitors (coded as "Named by Company"), as self-disclosed in the company's SEC filings, the set of companies that considers firm i a competitor (coded as "Named by Competitor"), as disclosed by their SEC filings, and finally the set of firms considered to be firm i's competitors as disclosed in third party firms' SEC filings (coded as "Named by Third Party"). We define a firm's "CIQ" peers to be those competitors who are "Named by Company" or "Named by Competitor," similar to Rauh and Sufi (2012) and Lewellen (2013).

Panel A of Table 5 reports summary statistics of the alternative peers that we collected. We have GOOGLE and CIQ peers for 1,088 and 1,160 base firms, respectively. On average, each base firm has 7.69 GOOGLE peers and 5.13 CIQ peers. Finally, on average 69% of GOOGLE peers belong to the same GICS6 industry as the base firms, whereas 59% of CIQ peers belong to the same GICS6 as the base firms. Panel B shows that there's a substantially higher correspondence between a base firm's top 10 SBPs and its GOOGLE peers compared to the correspondence with CIQ peers. 62% of the top-ranked SBPs are also a GOOGLE peer; in contrast, only 22% of top-ranked SBPs are also a CIQ peer. Panel B also reveals a fast decay in this correspondence for both GOOGLE and CIQ peers: only 17% (7%) of the 10^{th}-ranked SBPs are also GOOGLE (CIQ) peers. These summary statistics suggest that while there is some level of similarity between SBPs and GOOGLE or CIQ peers, there are also substantial differences between

them.

Table 6 compares each of the alternative peer identification schemes to SBPs in terms of its performance in explaining the cross-sectional variation in base firms' returns. The tests follow the same estimation specifications, i.e. Eq.(2), and requirements as in Table 3: for example, comparisons between SBPs and an alternative scheme are performed based on the base-firm-month observations for which we have data on peers through the alternative scheme and also data on SBPs. Unlike our baseline tests in Table 3, however, we do not have 10 years' worth of valid peer data for all the alternative schemes. Since both GOOGLE and CIQ peers represent June 2014 snapshots, we have limited ability to assess their performance over time. Thus, we make a conservative assumption in our tests that the peers from these schemes are valid in the 24 months from January of 2012 to December of 2013.[15]

Panel A1 of Table 6 shows that both equal-weighted and traffic-weighted portfolios of SBPs significantly outperform both equal-weighted portfolios of GOOGLE and CIQ peers in explaining the cross-sectional variation in the monthly returns of S&P1500 base firms. The out-performance is both statistically and economically significant. For example, SBP TW portfolios outperform GOOGLE peers by 52% and CIQ peers by 277%, both of which are significant at the 1% level. Panels B1 and C1 of the same table show that SBPs consistently outperform these alternatives across large base firms that belong to the S&P500 and the smaller base firms that belong to the S&P MidCap 400 and S&P SmallCap 600 (collectively denoted S&P1000 in this paper).[16] Interestingly, while SBPs'

[15]To the extent that these assumptions create biases, we expect them to be in the direction favoring these alternative schemes, since we are using base firms' future benchmarks to capture co-movements in future performance. Our sense, however, is that benchmarks produced from various sources tend to be fairly sticky over time, and we do not expect there to be significant variation in a two-year span.

[16]We also considered value-weighted GOOGLE peer portfolios. Google Finance reports a rank ordering of peers based on some proprietary algorithm; our value-weighted portfolio weights each peer firm based on the order in which it appears in Google Finance's listing of "Related Firms." For example, the firm that is reported first out of ten will receive the weight of $\frac{10}{\sum_{i=1}^{10} i} = \frac{2}{11}$. In unreported results, we find that taking into account the relative rank improves the performance of GOOGLE peers only marginally. Both SBP EW and SBP TW continue to significantly outperform both economically and statistically. We only consider equal-weighted portfolios for CIQ peers since there is no meaningful ranking that we can observe.

out-performance over GOOGLE peers tends to be larger for smaller base firms, its out-performance over CIQ base firms tends to be larger for larger base firms.

These results may reflect potential biases embedded in the alternative classification schemes. For example, the disclosure of competitive peers in SEC filings is a voluntary choice and may be driven by strategic considerations. A large firm that views itself as a stand-alone leader in a market may not view—thus name—any specific companies as a competitor; a newcomer to a market, on the other hand, may name the market leaders as its competitors aspirationally. This can potentially explain why CIQ performs especially poorly relative to SBPs and why this out-performance is greater among the larger base firms. Another possibility explaining GOOGLE and CIQ peers' performance may be their relative paucity. As explained in the preceding section, benchmark portfolios consisting of fewer firms are more exposed to peer firms' idiosyncratic shocks, which reduces the peer portfolios' abilities to explain variations in base firms' returns. Finally, our findings here may reflect SBPs' incorporating other, possibly more nuanced, dimensions of fundamental similarity.

3.2. Text Network Industry Classification peers

We also consider peers belonging to the same "Text Network Industry Classification" (TNIC). This classification scheme, developed by Hoberg and Phillips (2010) and Hoberg and Phillips (2014), infer product market peers and group firms into different "industry" groupings by analyzing and quantifying textual similarities in firms' self-reported business descriptions in their 10-K filings.

Data on TNIC peers are obtained from the Hoberg and Phillips Data Library online.[17] Because TNIC is based on 10-K data, we assume that TNIC peers from fiscal year t are usable for out-of-sample tests from July of $t+1$ to June of $t+2$. Overall, we collected data on TNIC peers for 1,465 unique base firms from January 2004 to June of 2013. Panel

[17] We downloaded the July, 2013 version from http://alex2.umd.edu/industrydata/industryclass.htm

A, Table 5 also reports that, on average, each base firm has 79 TNIC peers and that, on average, 48% of them belong to the same GICS6 as the base firm. Moreover, Panel B shows a substantial correspondence between a base firm's top 10 SBPs and its TNIC peers: whereas 73% of top-ranked SBPs are also a base firm's TNIC peers, this correspondence diminishes to 43% for the 10^{th}-ranked SBPs. Given the substantially larger size in TNIC peers, relative to SBPs as well as GOOGLE and CIQ, it is not surprising that SBPs' correspondence to TNIC is also substantially larger compared to the smaller alternative schemes.

Panel A2, Table 6 shows that both equal-weighted portfolios of SBPs and a traffic-weighted portfolios of SBPs significantly outperform equal-weighted portfolios of TNIC peers in explaining the cross-sectional variation in the monthly returns of S&P1500 base firms in the 114 months from January 2004 to June 2013. Whereas SBP EW outperforms TNIC by 71%, SBP TW outperforms TNIC by 88%, both statistically significant at the 1% level. Panels B2 and C2 of the table confirm that this out-performance is consistent across the large and small base firms. Interestingly, SBPs' out-performance of TNIC peers is stronger among the larger base firms.

While we view TNIC as being an innovative method for classifying firms belonging to similar competitive spheres, we conjecture that its performance in explaining the cross-sectional variations in prices and fundamentals may be a result of the relatively large number of firms belonging to a given TNIC "industry," which hampers the ability of the TNIC peer portfolio to capture economic similarities with the base firm. However, it is possible that the performance of TNIC benchmark portfolios could improve using the closest peers in terms of their textual distance to the base firm.

3.3. Yahoo! Finance and analyst co-coverage peers

Our evidence thus far highlights the possibility that SBPs' performance stems from the ability of our co-search algorithm to aggregate and extract the collective wisdom of

investors in identifying fundamentally similar firms for benchmarking purposes. To the extent that this is a primary driver for SBPs' out-performance, we believe this powerful concept can be illustrated, extended, and exploited in different contexts. We do so by considering two peer identification schemes that also aim to capture the collective wisdom of investors.

We first consider the set of co-search-based peers available on Yahoo! Finance (henceforth YHOO peers). Yahoo Finance makes available to users the set of firms which also viewed the base firm: for example, when searching for Google's information Yahoo Finance reports "People viewing GOOG also viewed PCLN AMZN BIDU AAPL MA NFLX." We collected YHOO peers in June 2014 for a total of 922 unique base firms.[18] As we report in Panel A, Table 5, each base firm has on average 5 YHOO peers, with 28% of them on average sharing the same GICS6 as the base firm. Panel B shows that with the exception of the top-ranked SBP, there is on average fairly low correspondence, thus substantial differences, between SBPs and YHOO peers.

Like GOOGLE and CIQ peers, YHOO peers represents a snapshot from June, 2014, limiting our ability to assess their performance over time. Consistent with CIQ and GOOGLE peers, we make a conservative assumption in our tests that YHOO peers are valid in the 24 months from January of 2012 to December of 2013. Panel A3, Table 6 shows that both equal-weighted portfolios of SBPs and a traffic-weighted portfolios of SBPs significantly outperform equal-weighted portfolios of YHOO peers in explaining the cross-sectional variation in the monthly returns of S&P1500 base firms from 2012 to 2013. Whereas SBP EW outperforms YHOO by 136%, SBP TW outperforms YHOO by 165%, both statistically significant at the 1% level. Panels B1 and C1 of the table confirm that this out-performance is consistent across both large and small base firms. Strikingly, SBPs' out-performance of YHOO peers is much stronger among the smaller

[18]Note that Yahoo displays the co-searched tickers on a randomized basis per page refresh- consistent with the issue highlighted in Kremer et al. (2014) that full transparency is inefficient due to reduced incentives to provide novel information. Our algorithm refreshes the page until Yahoo displays the results. In contrast, EDGAR co-searches and search-based peers are not observable by investors.

base firms. For example, whereas SBP EW (TW) outperforms YHOO by 22% (42%) among the S&P500 base firms, this out-performance expands to 177% (205%) among the S&P1000 base firms.

These results are revealing of the potential conditions under which the "collective wisdom" of investors are likely to be useful in producing fundamentally similar benchmarks. In the absence of total visibility into the underlying data and Yahoo's algorithms, our conjecture is that search traffic on Yahoo! Finance is driven more by retail investors. If so, we expect their traffic to be more concentrated around large and salient firms and we expect them to be less sophisticated than EDGAR users. These factors could explain our joint findings that 1) SBPs outperform YHOO peers and 2) the out-performance is greater among smaller base firms. Thus, the usefulness and power stemming from the "collective wisdom" critically depends on the level of sophistication of investors underlying the search traffic.[19]

Whereas SBPs implicitly harness the collective wisdom of investors on EDGAR, we further illustrate the power of this idea by examining the collective wisdom of sell-side analysts. Theoretically, analysts have an incentive to cover economically similar firms because of the reduced cost of information acquisition (e.g., Peng and Xiong, 2006). Empirically, research has shown that sell side analysts tend to specialize in industries and cover multiple firms belonging to her primary industry of expertise (e.g., Liang et al., 2008). On the other hand, there can be various other factors — for example, relating to the analysts' incentives or brokerage house characteristics — that drive analysts' coverage decisions. Liang et al. (2008) documents that analysts are more likely to cover a firm based on reasons idiosyncratic to the brokerage house: when the brokerage house has had

[19]The role of user sophistication may also help rationalize literature findings in the aggregated wisdom of stock opinions. Antweiler and Frank (2004) aggregate Yahoo and Raging Bull's message board chats to study dispersion of beliefs and stock volatility. They also find an economically small effect in the aggregated messages' tone's ability to predict future returns. In contrast, Chen et al. (2014) find that aggregated tone from articles on Seeking Alpha helps predict earning surprises and subsequent stock returns, with economically significant magnitudes. A factor which reconciles their findings may be the underlying sophistication of the users within their respective samples.

a recent investment banking relationship with the firm;[20] or when the firm was previously followed by another analyst employed in but who is no longer forecasting for the same brokerage house. Liang et al. (2008) also document the possibility that there may be systematic biases in analysts' coverage decisions: for example, analysts are more likely to cover high growth firms. Thus, while analysts' coverage decisions are in part driven by fundamental similarities between firms and in part due to non-fundamentals-related factors, our thesis is that, like patterns of co-search for firm fundamentals, aggregate patterns of analysts' co-coverage decisions can be informative of fundamental similarities between firms.[21]

To construct analyst co-coverage of firms, we obtain IBES forecast file covering the universe of analyst EPS forecasts for the period 2003-2013. To qualify as an analyst covering firm i, the analyst must have made at least one forecast for the firm in the calendar year. We then apply the same logic of our co-search algorithm and generate an analyst co-coverage fraction between firms i and j in year t:

$$\textit{Analyst co-coverage fraction}_{ijt} = \frac{\#\text{ of analysts who co-cover } i \text{ and } j}{\#\text{ of analysts who cover } i}. \tag{4}$$

This fraction is defined as the percentage of analysts covering i who also cover j. Note that an analyst is defined as the unique combination of the broker and analyst ID from IBES such that an analyst who moves from one broker to another would be treated as a different analyst in our sample.

Based on *Analyst co-coverage fraction*, we consider two types of analyst co-coverage peers (ACPs). First, we consider all peers that are co-covered, and second, we consider only the top 10 ACPs, the same as our treatment of SBPs. Table 5 reports that there are valid ACPs for 1,291 unique base firms in 2013. On average, each base firm has 94 ACPs,

[20] However, this effect diminishes after Regulation Fair Disclosure in 2000.

[21] The idea of identifying related firms based on analysts' coverage choices have been explored in the works of Ramnath (2002), Israelsen (2014), Kaustia and Rantala (2013), and Muslu et al. (2014).

and 39% of these ACPs share the same GICS6 as the base firm. Moreover, Panel B of the table suggests that a fairly high percentage of SBPs are also ACPs. As with TNIC peers, this reflects the relatively numerous ACPs. When we consider the top 10 ACPs, the correspondence with SBPs declines substantially. While 63% of the top-ranked SBPs are also top 10 ACPs, only 19% of the 10^{th}-ranked SBPs belong to the top 10 ACPs. Thus there is also substantial disagreement between SBPs and ACPs in terms of which peer firms are the most related to the base firm.

The last 4 rows in Panel A3 of Table 6 compare the performance of ACPs in explaining the cross-sectional variation in base firms' monthly returns from January 2004 to December 2013. The second and fourth rows of Panel A3 consider equal-weighted [EW] ACP portfolios consisting of all ACPs and top 10 ACPs, respectively; the third and fifth rows consider value-weighted [VW] ACP portfolios consisting of all ACPs and top 10 ACPs, respectively, weighting each ACP by the relative magnitude of its *Analyst co-coverage fraction*.

Our results suggest that the collective wisdom gleaned from analysts' co-coverage patterns produce peers that perform substantially better relative to the other alternatives. The EW and VW portfolios of all ACPs explain 10.2% and 11.6%, respectively, of base firms' monthly returns. Restricting these portfolios to the top 10 ACPs improves the performance substantially, explaining 12.1% and 12.6% of base firms' monthly returns, respectively. Moreover, it's interesting to observe that, similar to our *Annual search fraction* measure, weighting by the relative magnitudes of *Analyst co-coverage fraction* improves the performance of peer portfolios, consistent with co-coverage patterns containing information about underlying firm relations.

While the performance statistics of ACPs approach those of SBPs, SBPs continue to outperform each of the four ACP portfolios considered. In particular, our SBP TW portfolios explain 13% of base firms' monthly returns, an out-performance of 41% and 24% over the EW and VW portfolios of all ACPs, respectively, and an out-performance

of 19% and 14% over the EW and VW portfolios of top 10 ACPs.

3.4. Accounting fundamentals tests

We extend the above comparisons to alternative peer identification schemes by examining their performance in explaining the cross-sectional variation in firm fundamentals, following the analyses and tests of Table 4. Table 7 shows that SBPs also outperform the alternative peer schemes in explaining the cross-sectional variation in a variety of firm fundamentals.

For 9 of the 11 of the fundamental measures considered, SBP EW portfolios outperform equal-weighted portfolios of GOOGLE peers significantly at the 1% level, with a median out-performance of 44%. SBP EW portfolios also outperform equal-weighted portfolios of CIQ and TNIC peers significantly, at the 1% level, for 10 of 11 and 11 of 11 of the measures considered, and a median out-performance of 127% and 39% respectively.

SBPs also compare favorably against YHOO peers as well as ACPs. For all 11 measures considered, SBP EW portfolios outperform YHOO peer portfolios significantly at the 1% level, with a median out-performance of 81%. Similar to the price co-movement tests, the equal- and value-weighted portfolios consisting of top 10 ACPs perform the best. Out of the 11 fundamental variables considered, SBP EW (TW) peer portfolios outperform Top 10 Co-Coverage EW (VW) in 7 (9) at the 5% level, with SBPs' median out-performance of 13% (14%) in the cross sectional variation in base firms' fundamentals.

Together, these findings show that EDGAR users in aggregate are able identify a set of economically-related firms, and this collective wisdom of investors is incremental to existing alternative peer identification schemes. Across our cross sectional tests involving returns, valuation multiples, financial statement ratios, and other fundamental characteristics, SBPs in the aggregate do a better job explaining their cross sectional variation relative to the alternative peer schemes.

3.5. Discussion of the relative performance of revealed-choice-based methods

Though ACPs do not overall perform as well as SBPs in these tests, we view their good performance relative to the alternative peer identification schemes as further evidence of the idea that harnessing the "collective wisdom" of market participants is a powerful way to identify economically-related firms. In the case of ACPs, while an individual analyst's choice of firms to cover may have idiosyncrasies, the collective co-coverage patterns across analysts reflect underlying structure in the fundamental relations between firms.

In the Appendix, we propose a simple model to provide further intuition for why, and under what circumstances, aggregate co-search decisions by investors (the model also applies to co-coverage decisions by analysts) can be expected to uncover the underlying fundamental similarities between firms. This model is anchored on the assumptions that investors, who intend to make an investment decision for some base firm, are performing benchmarking analyses to put the base firm's fundamentals in context. For simplicity, and consistent with the empirical evidence that on average EDGAR users search for the fundamentals of 2 firms, we assume that investors co-search for the information of one benchmark based on her private signals of candidate firms' fundamental similarities with the base firm.

This model makes three intuitive predictions. First, all else equal, search fractions are more "informative" — i.e., of the rank ordering of benchmarking firms based on fundamental similarities with the base firm — when there are more investors independently searching. Second, search fractions are less informative when investors have noisier signals for the fundamental similarities between firms, or when there is a relatively small number of investors independently searching. Three, investors' systematic biases can corrupt the informativeness of co-search fractions, even when there are a large number of investors searching for information. However, when such biases are well-behaved, i.e., preserve the

true order of the fundamental similarities between firms, co-search fractions will continue to be informative. These predictions can be easily extended to the analyst co-coverage context.

The relative performance of SBPs, ACPs, and YHOO peers are consistent with the predictions of this simple model. For example, the finding that YHOO peers perform especially poorly among small firms is consistent with the model if retail investors do indeed perform fewer searches of small firms' fundamental information on Yahoo Finance, thus reducing the informativeness of YHOO co-search fractions. Moreover, the fact that ACPs perform so well, and YHOO peers do not, is also consistent with the model's predictions that the properties of investors' collective biases influences the informativeness of co-search (or co-coverage) fractions. This evidence supports our conjecture that the usefulness and power stemming from the "collective wisdom" critically depends on the level of sophistication of market participants underlying the relevant decision context. Loosely speaking, the level of sophistication determines the collective size and the properties of investors' biases in our model; the less sophisticated the investors, the more likely that systematic biases are order-preserving and less likely that co-search (or co-coverage) fractions are informative of fundamental similarities between firms. Whereas YHOO peers are likely generated or influenced by a disproportionate number of retail investors, analyst co-coverage patterns reflect revealed decisions of relatively more sophisticated (though perhaps still biased) sell-side stock analysts.

Finally, though conceptually similar, the set of investors that generate EDGAR search traffic and their information sets may be quite different from the set of analysts that generate ACPs. Our results above also suggest that SBPs and ACPs produce significantly different results, perhaps resulting from the different size, information sets, incentives, and biases represented by these two pools of market participants.

4. Exploring differences between SBPs and ACPs

In this section we explore the similarities and the differences in the two best-performing peer identification approaches from the preceding section: SBPs and ACPs, representing the collective wisdoms of EDGAR users and sell-side analysts respectively.

4.1. Characteristics of base firms and SBP-ACP disagreement

We begin by exploring the extent to which agreements between SBPs and ACPs are associated with the characteristics of the underlying base firm. To explore if agreement between SBP and ACP is a function of the underlying base firm's characteristics, we estimate the following specification:

$$Agree(SBP, ACP)_{i,t} = \pi' \Psi_{i,t} + \epsilon_{i,t} \tag{5}$$

where the outcome variable is the degree of agreement between the top ten SBP and ACP firms for a given base firm i in year t. $Agree(SBP, ACP)$ ranges from 0 to 1, where 0 denotes no overlap between a firm's top 10 SBPs and top 10 ACPs and 1 denotes 100% overlap.[22] $\Psi_{i,t}$ represents a vector of base firm characteristics including *log size*, *pb*, *evs*, *pe*, *rnoa*, *at*, *lev*, *salesgrowth*, *rdpersales*, and *complexity*, where *complexity* is defined as the number of base firm's operating segments with different SIC2 codes, following Cohen and Lou (2012).

Columns 1 and 2 of Panel A, Table 8 estimate Eq. (5) using OLS. Whereas column 1 includes all base firms in our sample from the S&P 1500 that have ACPs and SBPs, column 2 focuses on the subsample of firms with non-missing values for *ltgrowth*, *eps spread*, and *ltgrowth spread*.

The estimates in column 1 indicate that there is more disagreement between SBPs

[22]However, the rank ordering of the peers need not be identical across the two peer identification schemes.

and ACPs among firms with base firms that are smaller (lower *log size*), more glamorous (higher *pe*, higher *evs*, and lower *rnoa*), and more complex (higher *complexity*). We perform a similar exercise in columns 3 and 5, but use log of $Agree(SBP, ACP)_{it}$ as the dependent variable in an OLS specification and the number of agreed upon peers between a firm's ACPs and SBPs in an ordered logistic model, respectively. These alternative specifications yield qualitatively identical results compared to the baseline OLS specification of column 1.

The finding that there is a greater disagreement between SBPs and ACPs among smaller firms is intuitive, as the information environment around smaller firms is less certain. In the even columns of Panel A, Table 8, we include additional controls for base firms' characteristics relating to analysts' forecasts: *coverage*, *ltgrowth*, *eps spread*, and *ltgrowth spread*. In these specifications, *log size* loses significance and is replaced by a significant coefficient on *coverage*, again consistent with poorer information environment— i.e., when there are fewer analysts covering the base firm— resulting in greater disagreement between a base firm's SBPs and ACPs.

The finding that there is more disagreement between SBPs and ACPs among complicated base firms is particularly interesting in conjunction with the finding that firms' top 10 ACPs have a tighter correspondence with GICS6 than SBPs. If an analyst's coverage portfolio decision is more anchored upon traditional industry classification schemes (compared to EDGAR users' co-search decisions), the finding that SBPs disagree more with ACPs among the set of more complicated base firms — for whom industry classification schemes are least appropriate (e.g., Cohen and Lou, 2012) — could be a possible explanation for SBPs' out-performance of ACPs.

The finding that there is more disagreement between SBPs and ACPs among base firms that are higher growth or more glamorous could reflect systematic biases in analysts' preferences for covering such types of firms, which have been documented in prior literature (e.g., Bhushan, 1989; Cowen et al., 2006; Irvine, 2000; Liang et al., 2008; Mc-

Nichols and O'Brien, 1997; Daniel et al., 2002). In other words, if analysts are biased towards covering higher growth firms on the margin, we would capture these preferences in the aggregate co-coverage patterns too. In untabulated results, we find that ACPs on average indeed command higher price multiples and receive higher long-term growth forecasts from analysts in comparison to SBPs. Such systematic biases in analysts' coverage decisions could further explain SBPs' out-performance of ACPs.

4.2. Determinants of top SBPs and ACPs

To further understand the differences between SBPs and ACPs, and potential biases that may drive each peer identification scheme, we complement the above analyses by examining the determinants of top SBPs and ACPs in a multivariate setting. In particular, we examine the relative importance of similarities in base-to-peer-firm fundamental characteristics in determining the likelihood of a candidate peer firm being a top SBP or ACP.

Our empirical approach begins with matching each base firm in the S&P1500 sample to its top 10 SBPs and the set of alternative peer firms coming from each base firm's GICS2 sector that are not already included in its top 10 SBPs. In an analogous sample, we match each base firm to its top 10 ACPs and the set of alternative peer firms from each base firm's GICS2 sector that are not already included in its top 10 ACPs. In each of these samples, we compute, for each base-to-peer-firm pair, the absolute percentage difference ($|\frac{peer}{base} - 1|$) in each of the following fourteen characteristics: *size*, *pb*, *pe*, *rnoa*, *roe*, *at*, *evs*, *lev*, *salesgrowth*, *rdpersales*, *coverage*, *eps spread*, *ltgrowth*, and *ltgrowth spread*. Each of these differences is then decile ranked within year in order to reduce the influence of outliers in estimation: higher decile values correspond to greater absolute percentage differences in the relevant characteristic between the base and peer firm relative to other base-peer pairs in the cross section. We also define a dummy variable, *supply chain*, that equals 1 if the base-peer pair are supply chain partners, following the procedure in Cohen

and Frazzini (2008).[23] Finally, we define a dummy variable, *Different GICS4*, that equals 1 if the base firm and the peer firm belong to different GICS4 industry groups.

Using pooled probit models, we estimate the likelihood that a candidate peer firm is a top 10 SBP or ACP as a function of differences in fundamental characteristics and year fixed effects. Table 9 reports the estimated marginal effects for the S&P1500 base firm sample: with the exception of *Supply Chain*, *Different GICS4*, and the year fixed effects, which are all evaluated at 0, all other marginal effects are evaluated at 1. Following the format of Table 8, even columns include only base and peer firms that have valid data for *coverage*, *ltgrowth*, *eps spread*, and *ltgrowth spread*, a restriction that substantially attenuates the set of candidate peers.

Overall the various specifications in Table 9 paint a consistent picture illustrating key differences between the determinants of SBPs and ACPs. Relative to potential peer firms in the same GICS2 sector, a firm is more likely to be a top 10 SBP or ACP when it is more similar to the base firm in fundamental characteristics. With the exception of *size* and *pb*, all decile ranked percent differences in fundamental characteristics are negative and significant at the 1% level in all specifications, consistent with SBPs and ACPs capturing a set of peer firms that are fundamentally more similar to the base firm of interest.[24]

These results also capture potential biases in how each scheme aggregates the "collective wisdom" of market participants. First, greater differences in size increase the likelihood of a peer firm being a top 10 SBP or ACP. In un-tabulated summary statistics, we find that indeed both ACPs and SBPs are on average larger than their base firms in market capitalization. These findings are consistent with the possibility that: 1) EDGAR users have a systematic tendency to benchmark to larger firms, which is not surprising if larger and salient firms influence investors' search behaviors; and 2) analysts have a

[23] In particular we use the Compustat customer-supplier database identify customer-supplier links at the yearly level.

[24] Though the coefficients on decile ranks of differences in *pb* are statistically significant, they are economically negligible. For example, column 2 (4) suggests that a peer firm in the 10th decile in the percent difference in price-to-book multiple has a 0.6% (0.9%) greater likelihood of being a base firm's top 10 SBP (ACP) compared to a peer firm in the 1st decile of price-to-book difference, all else equal.

tendency to cover large blue chip firms within a certain industry, all else equal, which is not surprising given that such firms are expected to command the greatest institutional ownership and interest (e.g., Bhushan, 1989). Thus, while both SBPs and ACPs exhibit a similar bias in size, we conjecture that they are a result of different factors. Interestingly, this size effect is stronger for SBPs, implying a greater propensity among EDGAR users to co-search for or benchmark to larger firms: a peer firm in the 10th decile in the percent difference in market capitalization has a 6 percentage point greater likelihood of being a base firm's top 10 SBP compared to a peer firm in the 1st decile of size difference, all else equal; in contrast, a peer firm in the 10th decile in the percent difference in market capitalization has a 2 percentage point greater likelihood of being a base firm's top 10 ACP compared to a peer firm in the 1st decile of size difference, all else equal.

The results of Table 9 are also consistent with the view that there is a greater bias towards high growth or glamour firms among ACPs. For example, the negative marginal effects on the decile ranked differences in *pe*, *evs*, and *salesgrowth* are smaller in magnitude for explaining determinants of top ACPs relative to SBPs. For each of these variables, going from the 1st decile to the 10th decile in percent difference approximately reduces the likelihood of being in a base firm's top 10 SBPs by approximately twice as much as the reduction in the likelihood of being in a base firm's top 10 ACPs.

The most economically meaningful effects in Table 9 are captured by the dummy variables *Supply Chain* and *Different GICS4*. In particular, SBPs are more likely to capture supply chain partners relative to ACPs. Column 2 suggests that, all else equal, being a supply chain partner to the base firm increases a firm's likelihood of being a top 10 SBP by 30 percentage points, all else equal; in contrast, being a supply chain partner only increases the likelihood of being a top 10 ACP by 19 percentage points, all else equal.

Finally, we find ACPs are much more anchored on GICS classifications compared to SBPs. Whereas being in a different GICS4 industry grouping reduces a candidate peer firm's likelihood of being a top 10 SBP by about 8 percentage points, the effect on the

likelihood of being a top 10 ACP is a reduction of 16 percentage points, all else equal.

Collectively the evidence presented here highlights that while both SBPs and ACPs in general capture peer firms that are fundamentally similar to the base firm, each exhibits its own unique biases. Whereas SBPs exhibits a slightly greater bias towards large firms and are more likely to contain supply chain partners, ACPs tend to anchor more to GICS industry groupings and have a greater bias towards high growth and glamour firms. These observations can be reconciled with our findings that SBPs tend to outperform ACPs along multiple dimensions. The fact that SBPs are more likely to capture base firms' supply chain partners can explain the greater ability of SBPs to explain cross-sectional variations in returns and fundamentals; after all, many economic shocks can stem from (and be captured by) the supply chain. Moreover, as explained in Cohen and Lou (2012) and LMW, picking economically related firms for complex or conglomerate firms can be particularly difficult using traditional classification schemes that, by default, organize firms into mutually exclusive and collectively exhaustive groupings. Thus the fact that ACPs anchor more on GICS, combined with the finding that ACPs disagree with SBPs more when the base firms are complex, can also explain SBPs' superior performance.

4.3. Performance of composite peers

We now turn to investigate the incremental information captured by the disagreements between SBPs and ACPs. For example, despite the general superior performance of SBPs, there may still be incremental information in ACPs missing in SBPs. In Table 10, we investigate whether a hybrid approach that combines both set of revealed-choice-based benchmarks is incremental to the stand-alone performance of SBPs. For brevity, we focus on the price co-movement test from Eq 2.

Column 3 of Panel A reports the R^2 of the regression using the union of the set of top 10 SBPs and ACPs ("SBP∪ACP") while column 1 replicates the baseline SBP results. Across both the S&P1500 base firm sample, the S&P500 subsample of larger base firms,

and the S&P1000 subsample of smaller base firms, we find that the union of the peer sets modestly outperform the standalone SBP grouping, ranging from a 3.6% to 10.3% improvement in column 5. In particular, the incremental improvement is greater among the smaller firms.

In lieu of the union, we also investigate the alternative composite group formed using the intersection of the two groupings ("SBP∩ACP"). To perform the test, we restrict the base firm sample to those firms with at least one top 10 SBP that is also in its top 10 ACPs.[25] These results are reported in Panel B of the same table in column 2. Column 4 shows that SBPs that do not belong to this intersection provide significant incremental information, as the performance of SBP exceeds that of SBP∩ACP by 27% (31%) for the set of S&P1500 (S&P1000) base firms, again concentrated around smaller base firms. Moreover, a comparison of "SBP∪ACP" to SBP here again reveals a small but significant improvement in performance, ranging from 3% among S&P500 base firms to 8% among the smaller S&P1000 base firms.

The findings in this section provides a best-performing set of revealed-choice-based benchmarks that combines the collective wisdom of EDGAR users and sell-side analysts. Moreover, the finding that the improvements in the performance of composite peers over SBPs are concentrated among the set of smaller base firms suggests that in poorer information environments, there is a greater value in aggregating and combining the collective wisdom gleaned from the behavior of different types of sophisticated market participants.

5. Conclusion

In an increasingly service- and knowledge-based economy characterized by quickly-changing competitive landscapes, traditional industry classifications are unlikely to cap-

[25] This reduces the average number of firms in each cross section from 1,311 to 1,199. Smaller base firms disproportionately affected here, going from 899 to 785 firms on average, reflecting greater disagreements between ACPs and SBPs among smaller base firms.

ture nuanced or changing economics in firms. This paper argues that the class of benchmarking solutions that harness the collective wisdom of investors is a promising path for the future.

We provide evidence that SBPs, which aggregate EDGAR users' perceptions of fundamentally related benchmarks, not only systematically outperform peers derived from standard industry classification schemes like the GICS over a ten-year period, but they also outperform alternative state-of-the-art benchmarking solutions proposed by the academic literature and industry.

Strikingly, among the alternative schemes we considered, the next-best alternative also represents a revealed-choice-based solution that embodies the collective wisdom of sell-side analysts, gleaned from aggregate patterns of their co-coverage decisions. While there is substantial overlap in the set of peer firms identified as SBPs and ACPs, we find greater disagreement between the two groups amongst both growth firms and complicated firms. We also find that ACPs are more anchored on GICS than SBPs, and exhibit a bias towards higher growth firms. SBPs, on the other hand, are more likely to capture supply chain partners of the base firm.

Collectively, the evidence put forth in this paper suggests that while these aggregated revealed-choice-based approaches have great potential in resolving long-standing benchmarking problems in accounting and finance. Future research that seeks to add to this class of benchmarks should focus on areas where there is a critical mass of sophisticated market participants, and where the market participants' collective actions are not likely to exhibit collective biases unrelated to the fundamental relatedness between firms. Finally, in poorer information environments, there are greater benefits from combining the collective wisdom of different types of sophisticated market participants in identifying economic benchmarks. We hope that the findings of this paper stimulates further research in this area.

References

Antweiler, W., Frank, M. Z., 2004. Is all that talk just noise? the information content of internet stock message boards. The Journal of Finance 59, 1259–1294.

Bhojraj, S., Lee, C. M. C., 2002. Who is my peer? A valuation-based approach to the selection of comparable firms. Journal of Accounting Research 40, 407–439.

Bhojraj, S., Lee, C. M. C., Oler, D. K., 2003. What's my line? A comparison of industry classification schemes for capital market research. Journal of Accounting Research 41, 745–774.

Bhushan, R., 1989. Firm characteristics and analyst following. Journal of Accounting and Economics 11, 255–274.

Chen, H., De, P., Hu, Y. J., Hwang, B.-H., 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. Review of Financial Studies 27, 1367–1403.

Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. The Journal of Finance 63, 1977–2011.

Cohen, L., Lou, D., 2012. Complicated firms. Journal of Financial Economics 104, 383–400.

Cowen, A., Groysberg, B., Healy, P., 2006. Which types of analyst firms are more optimistic? Journal of Accounting and Economics 41, 119–146.

Daniel, K., Hirshleifer, D., Teoh, S. H., 2002. Investor psychology in capital markets: Evidence and policy implications. Journal of Monetary Economics 49, 139–209.

Fama, E. F., French, K. R., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.

Groysberg, B., Healy, P., 2013. Wall Street Research: Past, Present, and Future. Stanford University Press.

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

Hoberg, G., Phillips, G. M., 2014. Text-based network industries and endogenous product differentiation. NBER Working Paper No. 15991.

Irvine, P. J., 2000. Do analysts generate trade for their firms? Evidence from the toronto stock exchange. Journal of Accounting and Economics 30, 209–226.

Israelsen, R. D., 2014. Does common analyst coverage explain excess comovement? Journal of Financial and Quantitative Analysis Forthcoming.

Jegadeesh, N., Kim, J., Krische, S. D., Lee, C., 2004. Analyzing the analysts: When do recommendations add value? The Journal of Finance 59, 1083–1124.

Kaustia, M., Rantala, V., 2013. Common analyst–based method for defining peer firms, Working Paper.

Kremer, I., Mansour, Y., Perry, M., 2014. Implementing the "wisdom of the crowd". Journal of Political Economy Forthcoming.

Lee, C., Ma, P., Wang, C., 2014. Search-based peer firms: Aggregating investor perceptions through internet co-searches. Journal of Financial Economics Forthcoming.

Leung, A., Agarwal, A., Konana, P., Kumar, A., 2013. Online search and return comovement. Working paper.

Lewellen, S., 2013. Executive compensation and peer effects. Working paper.

Lewellen, S., Metrick, A., 2010. Corporate governance and equity prices: Are results robust to industry adjustments. Working paper, Yale School of Management.

Liang, L., Riedl, E. J., Venkataraman, R., 2008. The determinants of analyst-firm pairings. Journal of Accounting and Public Policy 27, 277–294.

McNichols, M., O'Brien, P. C., 1997. Self-selection and analyst coverage. Journal of Accounting Research 35, 167–199.

Muslu, V., Rebello, M., Xu, Y., 2014. Sell-side analyst research and stock comovement. Journal of Accounting Research 52, 911–954.

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics 80, 563–602.

Ramnath, S., 2002. Investor and analyst reactions to earnings announcements of related firms: An empirical analysis. Journal of Accounting Research 40, 1351–1376.

Rauh, J. D., Sufi, A., 2012. Explaining corporate capital structure: Product markets, leases, and asset similarity. Review of Finance 16, 115–155.

Van Nieuwerburgh, S., Veldkamp, L., 2010. Information acquisition and underdiversification. The Review of Economic Studies 77, 779–805.

A Appendix: A simple model of aggregate co-search

A.1. Set up

A population of N investors are interested in making investor decisions for some base firm 0 and individually searching for comparative firms to benchmark against firm 0's performance. We assume that there are two potential candidate firms, 1 and 2, whose fundamental similarity to firm 0 are characterized by d_1 and d_2, unobservable to investors. Without loss of generality, $d_1 < d_2$ where a lower d implies greater similarity to the base firm 0.

Each individual investor i receives private signals on the similarity between the base firm and the candidate peer firms:

$$\hat{d}_1 = d_1 + \epsilon_{i,1} \tag{A1}$$

$$\hat{d}_2 = d_2 + \epsilon_{i,2} \tag{A2}$$

where $(\epsilon_{i,1}, \epsilon_{i,2})' \sim_{iid} N(\mu, \Sigma)$. Here, $\mu = (c_1, c_2)$ capture the collective biases that investors may have, and Σ captures the variance-covariance matrix of investor's idiosyncrasies, whose elements are assumed to be finite. Based on the private signals, investor i makes one choice of a benchmarking firm to co-search.[26]

A.2. Co-search fraction and comparative statics

Under this model, investor i will pick firm 1 iff $\hat{d}_1 < \hat{d}_2$, or equivalently $\epsilon_{i,1} - \epsilon_{i,2} < d_2 - d_1$. Thus the probability of selecting firm 1 is $\Phi\left(\frac{(d_2-d_1)-(c_1-c_2)}{\sqrt{\sigma_1^2+\sigma_2^2-2\sigma_{12}}}\right)$, where Φ is the CDF of a standard normal distribution, and $(\sigma_1^2, \sigma_2^2, \sigma_{12})$ represent the variances of the errors and their covariance, respectively.

As the sample of investors $N \to \infty$, the fraction of investors that co-search fundamentals for firm 1 and 2 will be equal to the following co-search fractions:

$$f_{0,1} = \Phi\left(\frac{(d_2 - d_1) - (c_1 - c_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}}\right), \text{ and} \tag{A3}$$

$$f_{0,2} = 1 - \Phi\left(\frac{(d_2 - d_1) - (c_1 - c_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}}\right), \text{ respectively.} \tag{A4}$$

The following comparative statics are implied from the above: $\frac{\partial f_{0,1}}{\partial d_2} > 0$, $\frac{\partial f_{0,1}}{\partial c_2} > 0$, $\frac{\partial f_{0,1}}{\partial d_1} < 0$, $\frac{\partial f_{0,1}}{\partial c_1} < 0$, and $sign(\frac{\partial f_{0,1}}{\partial \sigma_j^2}) = -sign(d_2 - d_1) - (c_1 - c_2))$

[26] We limit this choice to simplify the model, but the assumption can also be thought to reflect search costs and consistent with the observed empirical fact that the modal number of firms for co-searching users is 2.

A.3. Empirical Evidence

We provide some suggestive evidence that the model's comparative statics on $f_{0,1}$ are consistent with the observed empirical search fraction. The model predicts that (assuming order preserving biases) as the investor's signal precision worsens, the search fraction should decrease. We can interpret the signal precision to be worse for smaller firms (with poorer information environments) and for more complicated firms due to increased investor processing costs. Appendix Table 1 illustrates that the average search fraction of a base firm's top 10 search-based peers (SBPs) is increasing in the size quantile of the base firm and decreasing in the number of operating segments within the base firm, a measure of complexity used in Cohen and Lou (2012). The number of base firm-year observations per cell are reported in brackets underneath the average search fraction.

Appendix Table 1: Average search fraction by size and complexity groupings

Size Quantile	Single Segment	2-3 Segments	4 Segments+
1 (smaller)	0.0167	0.0160	0.0146
	[1,632]	[950]	[132]
2	0.0193	0.0171	0.0155
	[1,463]	[1,008]	[165]
3	0.0205	0.0184	0.0161
	[1,333]	[1,110]	[324]
4	0.0209	0.0193	0.0177
	[1,304]	[1,137]	[324]
5 (bigger)	0.0219	0.0206	0.0177
	[1,024]	[1,135]	[562]

A.4. Implications

The basic model generates three key implications.

Implication 1

The collective wisdom of investors reflected in the aggregated co-search fraction $f_{0,1}$ will capture the correct rank ordering of the most fundamentally similar benchmarks $d_2 > d_1$ if and only if

$$\frac{(d_2 - d_1) - (c_1 - c_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}} > 0 \text{ or } d_2 + c_2 > d_1 + c_1. \tag{A5}$$

In other words, as long as investors' biases — e.g., from non-benchmarking behavior, informational errors, or behavioral biases — are order preserving, in large sample, investors' aggregated search fractions reveal the rank ordering of fundamental similarities between firms.

Implication 2

In finite samples with N investors, however, the number of investors that choose firm 1, the correct benchmark, is distributed Binomial(N,Φ), and the observed finite sample search fraction $f_{0,1}$ has a sampling distribution with a mean of Φ and variance of $\frac{\Phi(1-\Phi)}{N}$. This implies that, under the assumption that biases are order-preserving, search fractions are more informative when there are more investors searching.

This follows from the observation that at the limit, the sample search fraction $\hat{\Phi}$ converges to Φ. Under this condition, having more investors independently searching (higher N) reduces sampling variation in the search fraction and increases its ability to correctly reflect the rank ordering of fundamental firm similarities.

Implication 3

Under the assumption that investors' biases are order preserving, the noisier are investors' signals — e.g., due to poor information environments or lower investor sophistication — the less informative are the sample co-search fractions. This follows from the observation that the maximum sampling variation is obtained for $\Phi(0) = \frac{1}{2}$. Thus increasing the noisiness in investors' signals (i.e., increasing σ_1^2 or σ_2^2) tends to increase the sampling variation in the sample co-search fraction $\hat{\Phi}$.

Table A1.
Comparison of SEC Traffic Data Version Vintages

This table reports the means and difference-in-means in observable user search behavior between the data used in Lee, Ma, and Wang (2014) and the more recent data extract made available covering calendar years 2003-2011. Each observation represents a user defined at the daily-IP level. Panel A describes the comparison of the new data versus the LMW dataset for the data overlapping time period which spans 2008 to 2011. Panel B describes the comparison of the new data for the 2003-2007 and 2008-2011 sample periods.

The variables: Any "10-K or 10-Q", "Proxy", "Insider", etc are dummies which equals one if a daily-IP user searched for the particular filing type on a given day. Number of Unique CIKs is the number of unique firms (CIK-based) a user accessed on a daily level. Total Clicks is the raw number of clicks a user generated on a daily level. Number of Unique Filing Types is a count of the total number of filing types (10-K, 10-Q, etc) a user accessed on a daily level. Number of Unique File Extensions is a count of the total number of unique file extensions (.txt, .pdf, .html) a user accessed on a daily level. Total Session Length is the estimated average time (in hours) a user actively accessed the EDGAR website.

Panel A: Daily-User Comparison Between New Data and LMW

	New Data	Lee, Ma, Wang (2014)	Δ	Δ t-stat
Any 10-K or 10-Q	0.533	0.530	-0.002	-18.859
Any Proxy	0.116	0.118	0.002	27.530
Any Insider (3,4,5)	0.101	0.107	0.006	83.885
Any 8-K	0.281	0.282	0.001	10.485
Any S-1	0.091	0.091	0.001	8.004
Any Comment Letter	0.021	0.021	0.000	3.593
Any 13-F	0.034	0.035	0.001	12.743
Any 13-G	0.062	0.063	0.001	16.398
Any 13-D	0.049	0.050	0.001	15.655
Any 6-K	0.050	0.050	0.000	8.503
Any 20-F	0.045	0.045	0.000	3.991
Number of Unique CIKs	13.283	13.322	0.039	0.419
Total Clicks	86.960	88.512	1.552	1.379
Number of Unique Filing Types	2.565	2.589	0.024	18.364
Number of Unique File Extensions Types	1.391	1.473	0.082	482.507
Total Session Length (Hours)	0.598	0.615	0.017	36.679
Observations	34,976,165	39,864,724		

Panel B: Daily-User Comparison Between Pre and Post 2008 of New Data

	2003-2007	2008-2011	Δ	Δ t-stat
Any 10-K or 10-Q	0.527	0.533	-0.005	-39.541
Any Proxy	0.130	0.116	0.014	155.416
Any Insider (3,4,5)	0.164	0.101	0.063	698.741
Any 8-K	0.267	0.281	-0.014	-116.292
Any S-1	0.078	0.091	-0.013	-162.146
Any Comment Letter	0.007	0.021	-0.014	-405.692
Any 13-F	0.023	0.034	-0.011	-242.345
Any 13-G	0.069	0.062	0.007	107.824
Any 13-D	0.052	0.049	0.003	48.357
Any 6-K	0.047	0.050	-0.002	-42.056
Any 20-F	0.042	0.045	-0.003	-50.918
Number of Unique CIKs	7.832	13.283	-5.451	-58.879
Total Clicks	31.806	86.960	-55.154	-53.637
Number of Unique Filing Types	2.559	2.565	-0.006	-4.045
Number of Unique File Extensions Types	1.551	1.391	0.161	841.529
Total Session Length (Hours)	0.487	0.598	-0.111	-219.793
Observations	21,144,241	34,976,165		

Table A2.
Variable Description

Table A2 reports definitions of variables used in our regressions. We use CRSP monthly stock returns and Compustat quarterly data for the sample period 2004–2013. CRSP variable names are in parentheses and Compustat variable names are in square brackets. After collecting the raw Compustat data, in accordance with Bhojraj et al. (2003), we drop all firm–quarter observations missing data on total assets [atq], total long term debt [dlttq], net income before extraordinary items [ibq], debt in current liabilities [lctq], or operating income after depreciation [oiadpq]. Further, we require the raw share price on the last day of each fiscal quarter to be greater than $3, both total common equity [ceqq] and total shareholder equity [seqq] to be positive, and net sales [saleq] to be more than $100 million.

Variable	Description	Calculation
returns	Monthly cum-dividend stock returns	(ret)
Valuation Multiples		
pb	Price-to-book ratio	Market cap / total common equity [ceqq]
evs	Enterprise value- to- sales ratio	(Market cap + long-term debt [dlttq]) / net sales [saleq]
pe	Price-to-earnings ratio	Market cap net income before extraordinary items [ibq]
Financial Statement Ratios		
rnoa	Return on net operating assets	Net operating income after depreciation [oiadpq] / (property, plant, and equipment [ppentq] + current assets [actq] - current liabilities [lctq])
roe	Return on equity	Net income before extraordinary items [ibq] / total common equity [ceqq]
at	(Inverse of) Asset turnover	Total assets [atq] / net sales [saleq]
pm	Profit margin	Net operating income after depreciation [oiadp] / net sales [saleq]
lev	Leverage	Long-term debt [dlttq] / total stockholder's equity [seqq]
Other Financial Information		
salesgrowth	One-year-ahead realized sales growth	(Net sale one year ahead in the future - current year net sales) / current year net sales [saleq]
rdpersales	R&D expense- to- sales ratio	R&D expense [xrdq] / net sales [saleq]
ltgrowth	Median analyst long-term growth forecast	
ltgrowth spread	Standard deviation in analyst long-term growth forecast	
eps spread	Standard deviation in analyst one-year-ahead EPS forecast	
coverage	Number of analysts covering firm	
size	Market capitalization	Price (prc) × shares outstanding (shrout)

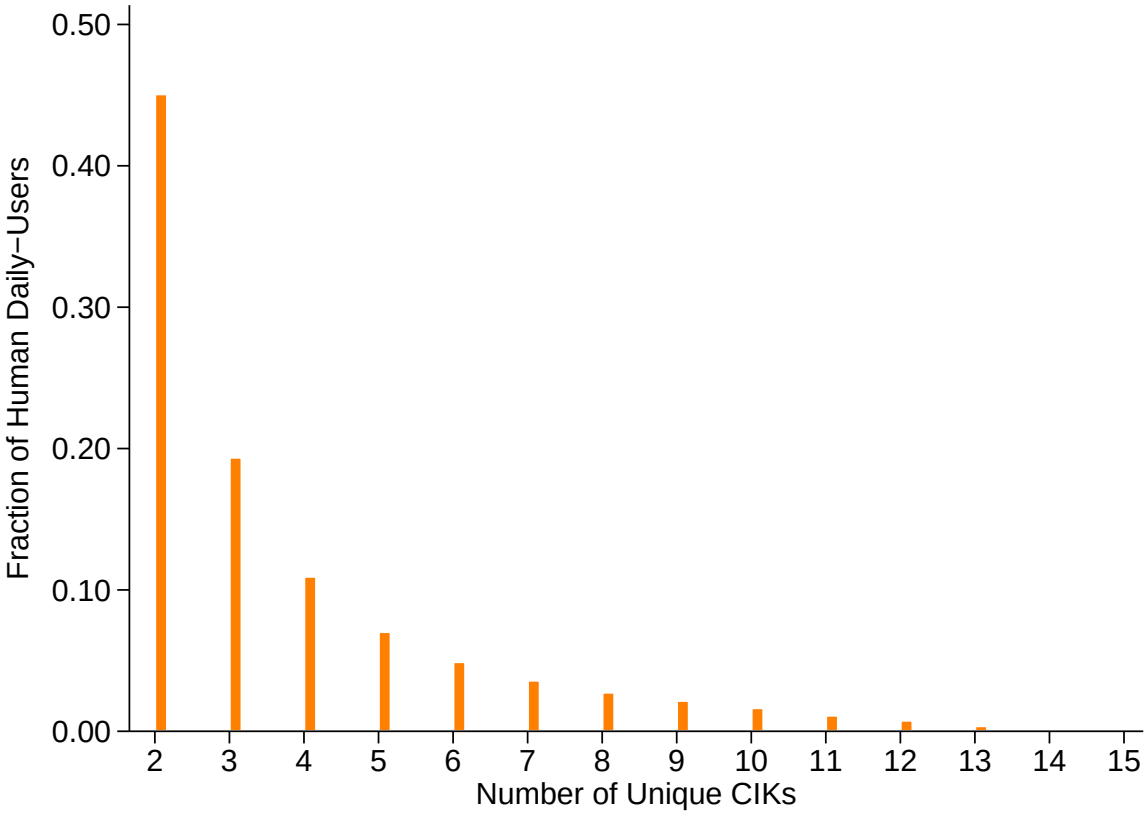

Figure 1. Histogram of Unique Number of CIKs Accessed by Daily Human Users on EDGAR. This figure plots the histogram of number of unique CIKs' fundamental information accessed by daily IP addresses on EDGAR that is classified as "human" by our robot identification rule. We define search traffic from an IP address in a given day as being generated from an automated script or a "robot" if the total number of CIKs whose fundamentals are downloaded by this IP address exceeds the 95^{th} percentile of the aggregate distribution in that calendar year.

Table 1.
Traffic Statistics

This table provides statistics on the sample of SEC EDGAR search traffic occurring between January 2003 and December 2011. Panel A reports our filtering process and the number of observations remaining after each filtering step. Step 1) reports the total number of filing downloads and the total number of daily unique visitors. In Step 2) we restrict traffic to users who search at least two unique firms (CIK-based) in order to apply our co-search algorithm. In Step 3) to reduce the influence of bulk downloaders, we restrict searches to users who do not download more than a cutoff of unique firms in a given day. The cutoff value corresponds to the number of unique CIKs downloaded at the 95% of all users in a given calendar year. In Step 5) we keep only the traffic page views for 10-K and 10-Qs and their variants. Finally in Step 5) we restrict searches to base firms which were in the S&P1500 index as of January 1st of the search traffic year. Panel B reports the number of base firms for which we have valid SBPs. Also reported are the median numbers of total peer firms available by calendar year for the entire base firm sample.

Panel A. Data Filtering Steps

Filter Rule	# of Daily Pageview	# of Daily Unique Visitors
1) Raw Sample	3.72 billion	56.2 million
2) Keep if Unique CIKs> 1	3.55 billion	19.35 million
3) Keep if Unique CIKs≤Annual 95% Cutoff	351.45 million	16.63 million
4) Keep 10-K, 10-Q Searches	115.08 million	10.96 million

Panel B. Coverage of S&P1500 Universe Conditional on 10 Peer Firms

Year	S&P500 Firms	S&P1500 Firms	Median Number of Peer Firms
2004	470	1438	255
2005	471	1444	340
2006	469	1440	248
2007	474	1451	235
2008	485	1465	336
2009	491	1482	451
2010	494	1485	583
2011	491	1481	669
2012	489	1482	718
2013	484	1470	714

Table 2.
Co-Search Characteristics Across Filing Type

This table reports the average composition of the remainder of a search sequence of a daily-user in terms of searches for different document types of other firms conditional on searching for a particular document type of a firm. The filing type 10-K includes 10-Qs and all amendment and small business variants of each. The underlying data is constructed following the data filtering steps in Table 2 with the exception of filtering out 10-Ks and 10-Qs to allow for cosearching across different filing types.

Type	10-K or 10-Q	8-K	Comment	Insider	Other	Proxy	S-1
10-K or 10-Q	**44.84%**	16.50%	0.55%	5.47%	24.64%	5.18%	2.82%
8-K	23.36%	**14.72%**	0.66%	24.86%	21.78%	13.15%	1.45%
Comment	17.47%	21.59%	**4.12%**	17.36%	24.58%	7.71%	7.16%
Insider	26.11%	15.66%	0.58%	**18.16%**	21.56%	16.43%	1.49%
Other	17.21%	21.54%	7.46%	17.18%	**22.08%**	7.51%	7.02%
Proxy	42.64%	16.61%	0.64%	5.96%	21.48%	**10.66%**	2.00%
S-1	21.22%	20.62%	0.79%	20.00%	23.12%	10.51%	**3.74%**

Table 3.
Price Co-Movement Test 2004-2013

This table compares the average R^2 values from monthly cross-sectional regressions of the form

$$R_{i,t} = \alpha_t + \beta_t R_{p_{i,t}} + \epsilon_{i,t}$$

using CRSP returns data from January 2004 to December 2013. Columns 1∼3 report average R^2s from monthly cross-sectional regressions, regressing base firm i's' returns in a given month t on the concurrent returns of a portfolio p_i of peers. Column 1 considers an equal-weighted portfolio of all peer firms from the base firm's GICS6 industry; Column 2 considers an equal-weighted portfolio (SBP EW) of the top 10 SBP firms, ranked by the prior calendar year's *Annual Search Fraction* f_{ij}, defined as the fraction of daily-users searching for both firm i and j's' information on the same day conditional on searching for firm i and any other firm $\neq i$, aggregated over the course of a calendar year; Column 3 considers a portfolio (SBP TW) consisting of the top 10 SBP firms, with each peer firm weighted by the prior calendar year's *Annual Search Fraction* (relative to the top 10 peer firms). With the exception of calendar year 2013, SBPs and portfolio weights are generated based on prior calendar year's EDGAR search traffic (e.g., the regressions in 2012 are generated with weights from calendar year 2011). SBPs and portfolio weights for calendar year 2013 are generated using search traffic from calendar year 2011. Columns 4 and 5 test for the significance of the differences in average R^2s between the two SBP portfolio formulations and the GICS6 peer portfolios.

The results are reported for the sample of base firms that belonged to the S&P1500, S&P500, and S&P1000 at the beginning of a given calendar year. To facilitate comparisons, all the regressions are conducted using the same underlying set of firms. The variable N in parentheses represents the average cross-sectional sample size for each monthly regression and standard errors are reported in square brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

	GICS6 (1)	SBP EW (2)	SBP TW (3)	(2)-(1) (4)	(3)-(1) (5)
SP1500 Base Firms	0.102***	0.128***	0.141***	0.027***	0.040***
(N= 1,461)	[0.005]	[0.006]	[0.006]	[0.002]	[0.003]
SP500 Base Firms	0.152***	0.212***	0.236***	0.060***	0.084***
(N= 480)	[0.007]	[0.007]	[0.008]	[0.004]	[0.005]
SP1000 Base Firms	0.087***	0.103***	0.114***	0.016***	0.027***
(N= 981)	[0.005]	[0.005]	[0.006]	[0.002]	[0.003]
Number of Months	120	120	120	120	120

Table 4.
Fundamentals Co-movement Test Over 2004-2013

This table compares the average R^2 from quarterly cross-sectional regressions of the form

$$Var_{i,t} = \alpha_t + \beta_t Var_{p_{i,t}} + \epsilon_{i,t}$$

using most recently observable quarterly financial statement data from Compustat and market capitalization data from CRSP on March, June, September, and December of each year from 2004 to 2013. Columns 1∼3 report average R^2s from quarterly cross-sectional regressions, regressing base firm i's Var in a given month t on the concurrent Var of a portfolio p_i of peers. Each row considers a different Var, as defined in Table A2. Column 1 considers an equal-weighted portfolio of all peer firms from the base firm's GICS6 industry; Column 2 considers an equal-weighted portfolio (SBP EW) of the top 10 SBP firms, ranked by the prior calendar year's *Annual Search Fraction* f_{ij}, defined as the fraction of daily-users searching for both firm i and j's' information on the same day conditional on searching for firm i and any other firm $\neq i$, aggregated over the course of a calendar year; Column 3 considers a portfolio (SBP TW) consisting of the top 10 SBP firms, with each peer firm weighted by the prior calendar year's *Annual Search Fraction* (relative to the top 10 SBP firms). With the exception of calendar year 2013, SBPs and portfolio weights are generated based on prior calendar year's EDGAR search traffic (e.g., the regressions in 2012 are generated with weights from calendar year 2011). SBPs and portfolio weights for calendar year 2013 are generated using search traffic from calendar year 2011. Columns 4 and 5 test for the significance of the differences in average R^2s between the two SBP portfolio formulations and the GICS6 peer portfolios.

Regressions are performed for the sample of firms that belong to the S&P1500, S&P500, and S&P1000 base firms in Panels A, B, and C, respectively, as of the beginning of each calendar year. To facilitate comparisons, all the regressions are conducted using the same underlying set of base firms. In addition, for regressions involving *pe*, we also drop observations with negative net income before extraordinary items, and for regressions involving *rnoa*, we drop observations when values are missing for current assets, current liabilities, or property, plant, and equipment. The variable N in parentheses represents the average cross-sectional sample size for each quarterly regression and standard errors are reported in square brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

Table 4.
(Continued)

Panel A: SP1500 Base Firms

	GICS6 (1)	SBP EW (2)	SBP TW (3)	(2)-(1) (4)	(3)-(1) (5)
Valuation Multiples					
pb	0.048***	0.115***	0.112***	0.067***	0.064***
(N= 978)	[0.003]	[0.004]	[0.004]	[0.004]	[0.004]
evs	0.297***	0.425***	0.465***	0.127***	0.168***
(N= 977)	[0.004]	[0.006]	[0.005]	[0.004]	[0.004]
pe	0.031***	0.040***	0.045***	0.009***	0.013***
(N= 878)	[0.003]	[0.004]	[0.004]	[0.003]	[0.003]
Financial Statement Ratios					
rnoa	0.179***	0.223***	0.263***	0.044***	0.083***
(N= 965)	[0.008]	[0.008]	[0.009]	[0.004]	[0.005]
roe	0.031***	0.067***	0.072***	0.035***	0.041***
(N= 973)	[0.003]	[0.004]	[0.004]	[0.003]	[0.003]
at	0.406***	0.559***	0.598***	0.153***	0.192***
(N= 977)	[0.006]	[0.005]	[0.005]	[0.004]	[0.004]
pm	0.199***	0.340***	0.385***	0.141***	0.186***
(N= 972)	[0.006]	[0.008]	[0.008]	[0.005]	[0.005]
lev	0.062***	0.119***	0.110***	0.058***	0.048***
(N= 982)	[0.005]	[0.006]	[0.006]	[0.004]	[0.004]
Other Financial Information					
ltgrowth	0.242***	0.280***	0.312***	0.038***	0.070***
(N= 814)	[0.019]	[0.016]	[0.018]	[0.005]	[0.005]
salesgrowth	0.150***	0.175***	0.197***	0.025***	0.047***
(N= 938)	[0.009]	[0.010]	[0.011]	[0.006]	[0.006]
rdpersales	0.648***	0.688***	0.723***	0.040***	0.075***
(N= 976)	[0.005]	[0.005]	[0.005]	[0.004]	[0.004]
Number of Months	40	40	40	40	40

Table 4.
(Continued)

Panel B: SP500 Base Firms

	GICS6 (1)	SBP EW (2)	SBP TW (3)	(2)-(1) (4)	(3)-(1) (5)
Valuation Multiples					
pb	0.063***	0.117***	0.110***	0.055***	0.048***
(N= 373)	[0.006]	[0.006]	[0.006]	[0.005]	[0.005]
evs	0.341***	0.437***	0.493***	0.096***	0.152***
(N= 372)	[0.008]	[0.011]	[0.010]	[0.007]	[0.006]
pe	0.037***	0.050***	0.060***	0.014***	0.023***
(N= 348)	[0.005]	[0.005]	[0.006]	[0.003]	[0.005]
Financial Statement Ratios					
rnoa	0.221***	0.241***	0.285***	0.020***	0.064***
(N= 368)	[0.011]	[0.009]	[0.012]	[0.007]	[0.007]
roe	0.042***	0.078***	0.077***	0.036***	0.036***
(N= 373)	[0.005]	[0.006]	[0.006]	[0.004]	[0.004]
at	0.377***	0.552***	0.598***	0.175***	0.221***
(N= 374)	[0.010]	[0.008]	[0.009]	[0.006]	[0.007]
pm	0.180***	0.311***	0.369***	0.131***	0.189***
(N= 371)	[0.008]	[0.009]	[0.009]	[0.006]	[0.006]
lev	0.067***	0.151***	0.133***	0.084***	0.066***
(N= 376)	[0.004]	[0.009]	[0.008]	[0.006]	[0.006]
Other Financial Information					
ltgrowth	0.275***	0.323***	0.370***	0.049***	0.095***
(N= 341)	[0.022]	[0.019]	[0.021]	[0.008]	[0.010]
salesgrowth	0.165***	0.209***	0.241***	0.045***	0.076***
(N= 366)	[0.011]	[0.011]	[0.012]	[0.007]	[0.008]
rdpersales	0.722***	0.720***	0.758***	-0.002	0.036***
(N= 374)	[0.006]	[0.006]	[0.006]	[0.004]	[0.004]
Number of Months	40	40	40	40	40

Table 4.
(Continued)

Panel C: SP1000 Base Firms

	GICS6 (1)	SBP EW (2)	SBP TW (3)	(2)-(1) (4)	(3)-(1) (5)
Valuation Multiples					
pb	0.038***	0.089***	0.094***	0.051***	0.056***
(N= 604)	[0.002]	[0.004]	[0.004]	[0.004]	[0.004]
evs	0.254***	0.372***	0.401***	0.118***	0.147***
(N= 605)	[0.004]	[0.006]	[0.006]	[0.005]	[0.005]
pe	0.034***	0.038***	0.041***	0.004	0.007*
(N= 529)	[0.005]	[0.004]	[0.005]	[0.004]	[0.004]
Financial Statement Ratios					
rnoa	0.158***	0.191***	0.226***	0.033***	0.069***
(N= 596)	[0.008]	[0.008]	[0.009]	[0.005]	[0.006]
roe	0.030***	0.048***	0.059***	0.018***	0.029***
(N= 600)	[0.004]	[0.005]	[0.005]	[0.003]	[0.003]
at	0.409***	0.539***	0.570***	0.130***	0.161***
(N= 602)	[0.005]	[0.005]	[0.006]	[0.004]	[0.005]
pm	0.197***	0.299***	0.336***	0.103***	0.139***
(N= 601)	[0.006]	[0.012]	[0.012]	[0.007]	[0.007]
lev	0.063***	0.113***	0.107***	0.050***	0.044***
(N= 605)	[0.007]	[0.007]	[0.006]	[0.005]	[0.005]
Other Financial Information					
ltgrowth	0.221***	0.238***	0.261***	0.018***	0.041***
(N= 473)	[0.019]	[0.016]	[0.018]	[0.006]	[0.006]
salesgrowth	0.144***	0.156***	0.174***	0.012*	0.031***
(N= 572)	[0.010]	[0.010]	[0.011]	[0.006]	[0.007]
rdpersales	0.578***	0.647***	0.679***	0.069***	0.101***
(N= 601)	[0.006]	[0.005]	[0.006]	[0.005]	[0.005]
Number of Months	40	40	40	40	40

Table 5.
Correspondence with Alternative Benchmark Identification Schemes

Row 1 of Panel A reports the number of available base firms within the S&P1500 sample with valid benchmark firms for each identification scheme as of December 2012. Row 2 reports the average number of available benchmark firms for each specific scheme. Row 3 provides the average fraction of peers from each identification scheme which share the same GICS6 grouping as the base firm. The first scheme represents the search-based peers (SBP) of Lee, Ma, and Wang (2014), defined by applying the *Annual search fraction* of Eq 1 to the SEC EDGAR search traffic data. The second, third, and fourth scheme represent peers selected solely based on the GICS2, GICS6, or SIC2 groupings respectively. The fifth scheme represents the list of firms that Google Finance (GOOGLE) reports as a base firm's "Related Firms" as of June 2014. The sixth scheme represents the set of self-reported product market competitors disclosed in SEC filings and collected by CapitalIQ (CIQ). Specifically, the CIQ peer set represents the union of the set of firms j that firm i report as its competitors and also the set of firms j that report i as a competitor. The seventh scheme is the "Text Based Network Industry Classification" (TNIC) of Hoberg and Phillips (2010, 2014), and is derived from the set of peer firms with the most similar self-reported business descriptions in their 10-K filings to the base firm's. The eighth scheme is the list of firms that Yahoo Finance (YHOO) reports as firms that, as of June 2014, are commonly co-searched with the base firm by its users. The ninth scheme represents analyst co-coverage peers (ACP), similar to that of Israelsen (2014), Kaustia and Rantala (2013), and Muslu et al. (2014), defined by applying the *Analyst co-coverage fraction* of Eq 4 to the entire IBES sample. The final scheme ACP (10) restricts the previous ACP scheme to retain only the top ten ACP firms.

Panel B reports the correspondence between Search-Based Peers (SBPs) and major alternative peer identification schemes. The first column SBP Rank denotes the ten closest SBPs as ranked by their search fraction in column 2. GICS2 and GICS6 represents the Global Industry Classification scheme at the 2 and 6 digit level. SIC2 represents the Standard Industry Classification scheme at the 2 digit level. The cells under each of the major classification schemes represent the probability that the alternative classification and the corresponding ith ranked SBP both identify the same peer firm for a given base firm.

Panel A. Summary Statistics of Alternative Peer Classification Schemes

	SBP	GICS2	GICS6	SIC2	GOOGLE	CIQ	TNIC	YHOO	ACP	ACP (10)
N Base Firms	1482	1496	1496	1498	1088	1160	1465	922	1291	1285
N Peers	10	200.7	36.7	61.91	7.69	5.13	79.18	5.04	94.39	10
Correspondence with GICS6	0.61	–	–	0.40	0.69	0.59	0.48	0.28	0.39	0.69

Panel B. Correspondence Between SBPs and Major Alternatives

SBP Rank	Search Fraction	GICS2	GICS6	SIC2	GOOGLE	CIQ	TNIC	YHOO	ACP	ACP (10)
1	0.04	0.88	0.78	0.75	0.62	0.22	0.73	0.43	0.79	0.63
2	0.03	0.85	0.72	0.72	0.54	0.20	0.66	0.31	0.72	0.54
3	0.02	0.84	0.68	0.65	0.45	0.15	0.60	0.21	0.71	0.48
4	0.02	0.80	0.63	0.59	0.35	0.15	0.55	0.16	0.66	0.42
5	0.02	0.77	0.59	0.58	0.32	0.15	0.53	0.16	0.63	0.33
6	0.01	0.75	0.56	0.58	0.26	0.12	0.50	0.12	0.64	0.33
7	0.01	0.74	0.55	0.56	0.23	0.12	0.47	0.10	0.61	0.27
8	0.01	0.74	0.53	0.55	0.20	0.10	0.47	0.10	0.61	0.25
9	0.01	0.73	0.52	0.53	0.19	0.09	0.44	0.09	0.57	0.22
10	0.01	0.73	0.51	0.53	0.17	0.07	0.43	0.09	0.56	0.19
Total	0.02	0.78	0.61	0.61	0.33	0.14	0.54	0.18	0.65	0.37

Table 6.
Price Co-Movement Test: Comparison with Alternatives

This table compares the average R^2 values from monthly cross-sectional regressions of the form

$$R_{i,t} = \alpha_t + \beta_t R_{p_{i,t}} + \epsilon_{i,t}$$

using CRSP returns data. Columns 1∼3 report average R^2s from monthly cross-sectional regressions, regressing base firm i's' returns in a given month t on the concurrent returns of a portfolio p_i of peers. Column 1 considers a portfolio of peers selected from various sources, from Google Finance and Capital IQ in Panel A, to the Text Network Industry Classification (Hoberg and Phillips, 2010, 2014) in Panel B, and finally Yahoo Finance and analysts' co-coverage of firms in Panel C.

Google Finance, Yahoo Finance, and Capital IQ peers comparisons span 24 months, from January 2012 to December 2013, and peer portfolio returns come from equal-weighted returns of peers. The Text Network Industry Classification peers comparisons span 114 months, from January 2004 to June 2013. TNIC peers from fiscal year t are used in returns tests from July of $t + 1$ to June of $t + 2$. Analyst co-covered peer firm comparisons span 120 months, from January 2004 to December 2013. "ACP [EW]" indicates that peer portfolio returns come from equal-weighted returns of peers, and "ACP [VW]" indicates that peer portfolio returns come from a value-weighted returns of peers, weighting by the relative magnitudes of a peer firm's *co-coverage fraction*. The last two rows of Panel C restricts the analysis to the top 10 peer firms ("ACP10") by *co-coverage fraction*.

Column 2 considers an equal-weighted portfolio (SBP EW) of the top 10 SBP firms, ranked by the prior calendar year's *Annual Search Fraction* f_{ij}, defined as the fraction of daily-users searching for both firm i and j's' information on the same day conditional on searching for firm i and any other firm $\neq i$, aggregated over the course of a calendar year; Column 3 considers a portfolio (SBP TW) consisting of the top 10 SBP firms, with each peer firm weighted by the prior calendar year's *Annual Search Fraction*, (relative to the top 10 peer firms). With the exception of calendar year 2013, SBPs and portfolio weights are generated based on prior calendar year's EDGAR search traffic (e.g., the regressions in 2012 are generated with weights from calendar year 2011). SBPs and portfolio weights for calendar year 2013 are generated using search traffic from calendar year 2011. Columns 4 and 5 test for the significance of the differences in average R^2s between the two SBP portfolio formulations and the GICS6 peer portfolios.

The results are reported for the sample of base firms that belonged to the S&P1500, S&P500, and S&P1000 in Panel A, B, and C, respectively, at the beginning of a given calendar year. To facilitate comparisons, all the regressions are conducted using the same underlying set of firms. The variable N in parentheses represents the average cross-sectional sample size for each monthly regression and standard errors are reported in square brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

Table 6.
(Continued)

Panel A: SP1500 Base Firms

	Alternative (1)	SBP EW (2)	SBP TW (3)	(2)-(1) (4)	(3)-(1) (5)
Panel A1: Google Finance (GOOGLE), Capital IQ (CIQ) Peers					
GOOGLE	0.084***	0.114***	0.128***	0.030***	0.044***
(N= 1,084)	[0.007]	[0.010]	[0.010]	[0.005]	[0.005]
CIQ	0.030***	0.098***	0.113***	0.069***	0.083***
(N= 1,121)	[0.004]	[0.009]	[0.009]	[0.007]	[0.007]
Number of Months	24	24	24	24	24
Panel A2: Text Network Industry Classification (TNIC) Peers					
TNIC	0.077***	0.132***	0.145***	0.055***	0.068***
(N= 1,412)	[0.005]	[0.006]	[0.006]	[0.002]	[0.003]
Number of Months	114	114	114	114	114
Panel A3: Yahoo Finance (YHOO), Analyst Co-Coverage (ACP) Peers					
YHOO	0.055***	0.130***	0.146***	0.075***	0.091***
(N= 902)	[0.012]	[0.012]	[0.012]	[0.010]	[0.011]
ACP [EW]	0.102***	0.130***	0.144***	0.028***	0.042***
(N= 1,306)	[0.005]	[0.006]	[0.006]	[0.002]	[0.003]
ACP [VW]	0.116***	0.130***	0.144***	0.014***	0.028***
(N= 1,306)	[0.005]	[0.006]	[0.006]	[0.002]	[0.003]
ACP10 [EW]	0.121***	0.130***	0.144***	0.009***	0.023***
(N= 1,306)	[0.005]	[0.006]	[0.006]	[0.002]	[0.002]
ACP10 [VW]	0.126***	0.130***	0.144***	0.004**	0.018***
(N= 1,306)	[0.005]	[0.006]	[0.006]	[0.002]	[0.002]
Number of Months	120	120	120	120	120

Table 6.
(Continued)

Panel B: SP500 Base Firms

	Alternative (1)	SBP EW (2)	SBP TW (3)	(2)-(1) (4)	(3)-(1) (5)
Panel A1: Google Finance (GOOGLE), Capital IQ (CIQ) Peers					
GOOGLE	0.139***	0.186***	0.216***	0.047***	0.077***
(N= 412)	[0.012]	[0.018]	[0.020]	[0.010]	[0.010]
CIQ	0.034***	0.171***	0.204***	0.136***	0.170***
(N= 402)	[0.005]	[0.016]	[0.017]	[0.015]	[0.016]
Number of Months	24	24	24	24	24
Panel A2: Text Network Industry Classification (TNIC) Peers					
TNIC	0.115***	0.216***	0.240***	0.101***	0.125***
(N= 464)	[0.006]	[0.007]	[0.008]	[0.005]	[0.006]
Number of Months	114	114	114	114	114
Panel A3: Yahoo Finance (YHOO), Analyst Co-Coverage (ACP) Peers					
YHOO	0.153***	0.186***	0.216***	0.034***	0.064***
(N= 413)	[0.020]	[0.018]	[0.020]	[0.011]	[0.012]
ACP [EW]	0.148***	0.212***	0.237***	0.063***	0.089***
(N= 419)	[0.007]	[0.008]	[0.008]	[0.004]	[0.006]
ACP [VW]	0.174***	0.212***	0.237***	0.038***	0.064***
(N= 419)	[0.007]	[0.008]	[0.008]	[0.004]	[0.005]
ACP10 [EW]	0.195***	0.212***	0.237***	0.016***	0.042***
(N= 419)	[0.008]	[0.008]	[0.008]	[0.003]	[0.004]
ACP10 [VW]	0.207***	0.212***	0.237***	0.004	0.030***
(N= 419)	[0.008]	[0.008]	[0.008]	[0.003]	[0.003]
Number of Months	120	120	120	120	120

Table 6.
(Continued)

Panel C: SP1000 Base Firms

	Alternative (1)	SBP EW (2)	SBP TW (3)	(2)-(1) (4)	(3)-(1) (5)
Panel A1: Google Finance (GOOGLE), Capital IQ (CIQ) Peers					
GOOGLE	0.066***	0.088***	0.097***	0.022***	0.031***
(N= 671)	[0.007]	[0.009]	[0.008]	[0.004]	[0.004]
CIQ	0.029***	0.077***	0.086***	0.048***	0.057***
(N= 719)	[0.004]	[0.008]	[0.007]	[0.007]	[0.006]
Number of Months	24	24	24	24	24
Panel A2: Text Network Industry Classification (TNIC) Peers					
TNIC	0.067***	0.106***	0.117***	0.039***	0.050***
(N= 947)	[0.005]	[0.006]	[0.006]	[0.002]	[0.003]
Number of Months	114	114	114	114	114
Panel A3: Yahoo Finance (YHOO), Analyst Co-Coverage (ACP) Peers					
YHOO	0.037***	0.103***	0.113***	0.065***	0.076***
(N= 488)	[0.009]	[0.010]	[0.010]	[0.009]	[0.010]
ACP [EW]	0.089***	0.106***	0.117***	0.017***	0.028***
(N= 886)	[0.005]	[0.006]	[0.006]	[0.002]	[0.003]
ACP [VW]	0.099***	0.106***	0.117***	0.007***	0.018***
(N= 886)	[0.005]	[0.006]	[0.006]	[0.002]	[0.002]
ACP10 [EW]	0.099***	0.106***	0.117***	0.007***	0.018***
(N=886)	[0.005]	[0.006]	[0.006]	[0.002]	[0.002]
ACP10 [VW]	0.102***	0.106***	0.117***	0.004**	0.015***
(N= 886)	[0.005]	[0.006]	[0.006]	[0.002]	[0.002]
Number of Months	120	120	120	120	120

Table 7.
Fundamentals Co-movement Test: Comparison with Alternatives

This table compares the average R^2 from quarterly cross-sectional regressions of the form

$$Var_{i,t} = \alpha_t + \beta_t Var_{p_{i,t}} + \epsilon_{i,t}$$

using most recently observable quarterly financial statement data from Compustat and market capitalization data from CRSP on March, June, September, and December of each calendar year.
Odd columns report average R^2s from quarterly cross-sectional regressions, regressing base firm i's Var in a given month t on the concurrent Var of its portfolio p_i. Each row considers a different Var, as defined in Table A2. Column 1 considers equal-weighted portfolios of peers from Google Finance (GOOGLE); column 3 considers equal-weighted portfolios of peers from Capital IQ (CIQ); column 5 considers equal-weighted portfolios of peers from the text industry classification network (TNIC); column 7 considers equal-weighted portfolios of peers from Yahoo Finance (YHOO); column 9 considers equal-weighted portfolios of analyst co-coverage peers (ACP); column 11 considers value-weighted portfolios of ACPs, weighting by each peer firm's *co-coverage fraction*; column 13 considers equal-weighted portfolios of base firms' top 10 ACPs by *co-coverage fraction*; column 15 considers value-weighted portfolios of base firms' top 10 ACPs, weighting by each peer firm's *co-coverage fraction*. Even columns report the differences (Δ) in R^2 generated by cross-sectional regressions of base firm Var on SBP portfolio Var from R^2s generated using the preceding column's peer identification scheme. Columns 12 and 16 compare the difference between traffic-weighted SBP portfolios with the respective value-weighted portfolios of columns 11 and 15. All other differences are with respect to equal-weighted SBP portfolios.

Regressions are performed for the sample of S&P1500 base firms in Panel A, S&P500 base firms in Panel B, and S&P1000 base firms in Panel C. To facilitate comparisons, all the regressions are conducted using the same underlying set of base firms. In addition, for regressions involving *pe* we also drop observations with negative net income before extraordinary items and for regressions involving *rnoa* we drop observations when values are missing for current assets, current liabilities, or property, plant, and equipment. The variable N in parentheses represents the average cross-sectional sample size for each quarterly regression and standard errors are reported in square brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

Table 7.
(Continued)

Panel A: SP1500 Base Firms

	GOOGLE (1)	Δ (2)	CIQ (3)	Δ (4)	TNIC (5)	Δ (6)	YHOO (7)	Δ (8)
Valutaion Multiples								
pb	0.036***	0.101***	0.010***	0.099***	0.039***	0.078***	0.114***	0.022**
	[0.004]	[0.005]	[0.002]	[0.004]	[0.002]	[0.004]	[0.008]	[0.009]
evs	0.320***	0.110***	0.194***	0.245***	0.326***	0.108***	0.145***	0.282***
	[0.006]	[0.006]	[0.007]	[0.007]	[0.005]	[0.005]	[0.013]	[0.014]
pe	0.023***	0.009*	0.017***	0.011	0.028***	0.016***	0.024***	0.011**
	[0.003]	[0.005]	[0.004]	[0.007]	[0.003]	[0.003]	[0.004]	[0.005]
Financial Statement Ratios								
rnoa	0.222***	0.022**	0.113***	0.118***	0.181***	0.045***	0.136***	0.126***
	[0.012]	[0.009]	[0.005]	[0.010]	[0.008]	[0.006]	[0.017]	[0.010]
roe	0.034***	0.042***	0.014***	0.046***	0.025***	0.045***	0.049***	0.028***
	[0.005]	[0.004]	[0.002]	[0.005]	[0.002]	[0.003]	[0.008]	[0.007]
at	0.543***	0.066***	0.284***	0.252***	0.486***	0.073***	0.388***	0.224***
	[0.009]	[0.006]	[0.010]	[0.010]	[0.005]	[0.004]	[0.014]	[0.016]
pm	0.297***	0.033***	0.128***	0.186***	0.246***	0.103***	0.069***	0.255***
	[0.011]	[0.005]	[0.006]	[0.009]	[0.006]	[0.006]	[0.009]	[0.009]
lev	0.049***	0.043***	0.010***	0.082***	0.067***	0.056***	0.071***	0.022***
	[0.005]	[0.005]	[0.001]	[0.006]	[0.006]	[0.004]	[0.008]	[0.004]
Other Financial Information								
ltgrowth	0.133***	0.085***	0.026***	0.125***	0.211***	0.078***	0.190***	0.054***
	[0.008]	[0.007]	[0.003]	[0.007]	[0.016]	[0.005]	[0.010]	[0.008]
salesgrowth	0.169***	0.005	0.050***	0.096***	0.113***	0.070***	0.035***	0.143***
	[0.012]	[0.006]	[0.006]	[0.011]	[0.007]	[0.006]	[0.006]	[0.010]
rdpersales	0.695***	-0.012**	0.549***	0.116***	0.636***	0.060***	0.389***	0.323***
	[0.007]	[0.005]	[0.013]	[0.011]	[0.006]	[0.005]	[0.028]	[0.025]
No. Quarters	20	20	20	20	38	38	20	20

Table 7.
(Continued)

Panel A: SP1500 Base Firms Continued

	ACP [EW] (9)	Δ [EW] (10)	ACP [VW] (11)	Δ [VW] (12)	ACP10 [EW] (13)	Δ [EW] (14)	ACP10 [VW] (15)	Δ [VW] (16)
Valutaion Multiples								
pb	0.077***	0.039***	0.084***	0.026***	0.079***	0.036***	0.083***	0.027***
	[0.004]	[0.005]	[0.003]	[0.005]	[0.003]	[0.003]	[0.003]	[0.003]
evs	0.279***	0.143***	0.318***	0.147***	0.346***	0.076***	0.365***	0.099***
	[0.005]	[0.003]	[0.005]	[0.004]	[0.005]	[0.002]	[0.004]	[0.003]
pe	0.033***	0.010***	0.039***	0.007*	0.040***	0.003	0.042***	0.005
	[0.003]	[0.004]	[0.003]	[0.004]	[0.003]	[0.003]	[0.004]	[0.003]
Financial Statement Ratios								
rnoa	0.188***	0.038***	0.212***	0.054***	0.219***	0.006*	0.232***	0.035***
	[0.007]	[0.005]	[0.007]	[0.006]	[0.007]	[0.004]	[0.008]	[0.005]
roe	0.039***	0.028***	0.045***	0.026***	0.041***	0.025***	0.046***	0.026***
	[0.003]	[0.004]	[0.004]	[0.003]	[0.004]	[0.003]	[0.004]	[0.003]
at	0.437***	0.120***	0.477***	0.119***	0.528***	0.030***	0.539***	0.056***
	[0.008]	[0.006]	[0.008]	[0.007]	[0.006]	[0.003]	[0.007]	[0.004]
pm	0.230***	0.106***	0.263***	0.117***	0.285***	0.051***	0.301***	0.079***
	[0.005]	[0.005]	[0.006]	[0.004]	[0.006]	[0.004]	[0.006]	[0.004]
lev	0.098***	0.026***	0.113***	0.001	0.096***	0.029***	0.099***	0.015***
	[0.006]	[0.006]	[0.006]	[0.004]	[0.006]	[0.004]	[0.006]	[0.004]
Other Financial Information								
ltgrowth	0.242***	0.043***	0.276***	0.042***	0.293***	-0.009	0.311***	0.007
	[0.015]	[0.006]	[0.017]	[0.007]	[0.018]	[0.005]	[0.019]	[0.006]
salesgrowth	0.110***	0.068***	0.133***	0.067***	0.144***	0.034***	0.150***	0.050***
	[0.007]	[0.005]	[0.008]	[0.005]	[0.008]	[0.004]	[0.008]	[0.004]
rdpersales	0.629***	0.062***	0.675***	0.053***	0.698***	-0.007*	0.709***	0.019***
	[0.008]	[0.007]	[0.006]	[0.005]	[0.005]	[0.003]	[0.005]	[0.003]
No. Quarters	40	40	40	40	40	40	40	40

Table 7.
(Continued)

Panel B: SP500 Base Firms

	GOOGLE (1)	Δ (2)	CIQ (3)	Δ (4)	TNIC (5)	Δ (6)	YHOO (7)	Δ (8)
Valutaion Multiples								
pb	0.064***	0.092***	0.016***	0.086***	0.037***	0.079***	0.191***	-0.037***
	[0.005]	[0.006]	[0.003]	[0.006]	[0.004]	[0.005]	[0.014]	[0.012]
evs	0.340***	0.089***	0.176***	0.249***	0.354***	0.099***	0.267***	0.169***
	[0.007]	[0.012]	[0.009]	[0.011]	[0.012]	[0.008]	[0.008]	[0.008]
pe	0.041***	-0.006	0.018***	0.020**	0.041***	0.010**	0.038***	-0.003
	[0.007]	[0.007]	[0.003]	[0.009]	[0.005]	[0.005]	[0.009]	[0.008]
Financial Statement Ratios								
rnoa	0.279***	-0.001	0.136***	0.118***	0.212***	0.025***	0.255***	0.027***
	[0.016]	[0.014]	[0.008]	[0.008]	[0.011]	[0.008]	[0.012]	[0.006]
roe	0.045***	0.046***	0.017***	0.050***	0.035***	0.044***	0.067***	0.023***
	[0.006]	[0.005]	[0.003]	[0.006]	[0.004]	[0.005]	[0.008]	[0.008]
at	0.567***	0.075***	0.291***	0.295***	0.507***	0.047***	0.447***	0.184***
	[0.014]	[0.007]	[0.013]	[0.015]	[0.007]	[0.006]	[0.013]	[0.005]
pm	0.331***	-0.017**	0.117***	0.174***	0.239***	0.076***	0.149***	0.164***
	[0.016]	[0.007]	[0.006]	[0.014]	[0.008]	[0.007]	[0.012]	[0.009]
lev	0.072***	0.054***	0.019***	0.101***	0.060***	0.096***	0.067***	0.060***
	[0.009]	[0.007]	[0.005]	[0.010]	[0.005]	[0.006]	[0.008]	[0.006]
Other Financial Information								
ltgrowth	0.139***	0.112***	0.012***	0.160***	0.237***	0.090***	0.270***	-0.024**
	[0.014]	[0.009]	[0.002]	[0.009]	[0.020]	[0.007]	[0.011]	[0.010]
salesgrowth	0.214***	-0.002	0.047***	0.128***	0.145***	0.073***	0.089***	0.123***
	[0.018]	[0.011]	[0.007]	[0.012]	[0.010]	[0.008]	[0.011]	[0.011]
rdpersales	0.754***	-0.036***	0.535***	0.153***	0.678***	0.047***	0.486***	0.232***
	[0.010]	[0.007]	[0.011]	[0.008]	[0.007]	[0.005]	[0.007]	[0.006]
No. Quarters	20	20	20	20	38	38	20	20

Table 7.
(Continued)

Panel B: SP500 Base Firms Continued

	ACP [EW] (9)	Δ [EW] (10)	ACP [VW] (11)	Δ [VW] (12)	ACP10 [EW] (13)	Δ [EW] (14)	ACP10 [VW] (15)	Δ [VW] (16)
Valutaion Multiples								
pb	0.097***	0.022***	0.106***	0.002	0.096***	0.023***	0.098***	0.010**
	[0.005]	[0.006]	[0.006]	[0.005]	[0.007]	[0.004]	[0.007]	[0.005]
evs	0.285***	0.160***	0.351***	0.151***	0.387***	0.058***	0.410***	0.092***
	[0.007]	[0.006]	[0.007]	[0.005]	[0.008]	[0.004]	[0.009]	[0.004]
pe	0.043***	0.012**	0.052***	0.011**	0.051***	0.004	0.053***	0.010**
	[0.005]	[0.005]	[0.005]	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]
Financial Statement Ratios								
rnoa	0.236***	0.006	0.263***	0.025***	0.259***	-0.017***	0.268***	0.020***
	[0.009]	[0.007]	[0.010]	[0.008]	[0.008]	[0.005]	[0.010]	[0.006]
roe	0.056***	0.022***	0.068***	0.008*	0.057***	0.022***	0.063***	0.013***
	[0.005]	[0.005]	[0.006]	[0.004]	[0.006]	[0.004]	[0.006]	[0.004]
at	0.412***	0.138***	0.459***	0.139***	0.533***	0.017***	0.549***	0.049***
	[0.011]	[0.008]	[0.011]	[0.009]	[0.007]	[0.005]	[0.008]	[0.006]
pm	0.171***	0.128***	0.217***	0.137***	0.257***	0.043***	0.272***	0.082***
	[0.005]	[0.007]	[0.007]	[0.005]	[0.008]	[0.005]	[0.007]	[0.005]
lev	0.108***	0.044***	0.124***	0.007	0.097***	0.054***	0.098***	0.033***
	[0.005]	[0.007]	[0.006]	[0.007]	[0.006]	[0.007]	[0.006]	[0.006]
Other Financial Information								
ltgrowth	0.230***	0.102***	0.275***	0.105***	0.316***	0.016**	0.342***	0.039***
	[0.014]	[0.011]	[0.015]	[0.011]	[0.018]	[0.007]	[0.020]	[0.007]
salesgrowth	0.133***	0.084***	0.166***	0.082***	0.181***	0.037***	0.188***	0.060***
	[0.010]	[0.007]	[0.011]	[0.007]	[0.011]	[0.006]	[0.011]	[0.006]
rdpersales	0.668***	0.054***	0.723***	0.038***	0.742***	-0.020***	0.751***	0.010**
	[0.009]	[0.007]	[0.007]	[0.005]	[0.005]	[0.004]	[0.005]	[0.004]
No. Quarters	40	40	40	40	40	40	40	40

Table 7.
(Continued)

Panel C: SP1000 Base Firms

	GOOGLE (1)	Δ (2)	CIQ (3)	Δ (4)	TNIC (5)	Δ (6)	YHOO (7)	Δ (8)
Valutaion Multiples								
pb	0.018***	0.085***	0.005***	0.088***	0.039***	0.052***	0.068***	0.033***
	[0.004]	[0.006]	[0.001]	[0.004]	[0.003]	[0.005]	[0.008]	[0.009]
evs	0.258***	0.117***	0.178***	0.227***	0.286***	0.089***	0.090***	0.281***
	[0.008]	[0.009]	[0.008]	[0.008]	[0.005]	[0.006]	[0.010]	[0.012]
pe	0.019***	0.014**	0.021***	0.008	0.026***	0.016***	0.022***	0.017**
	[0.004]	[0.006]	[0.006]	[0.009]	[0.004]	[0.004]	[0.004]	[0.006]
Financial Statement Ratios								
rnoa	0.172***	0.021**	0.102***	0.086***	0.160***	0.035***	0.074***	0.131***
	[0.011]	[0.008]	[0.007]	[0.017]	[0.008]	[0.006]	[0.011]	[0.009]
roe	0.025***	0.025***	0.016***	0.023***	0.025***	0.027***	0.038***	0.017**
	[0.004]	[0.005]	[0.002]	[0.005]	[0.003]	[0.004]	[0.008]	[0.007]
at	0.501***	0.053***	0.246***	0.201***	0.447***	0.091***	0.351***	0.226***
	[0.008]	[0.007]	[0.010]	[0.009]	[0.006]	[0.005]	[0.020]	[0.020]
pm	0.214***	0.036***	0.114***	0.126***	0.225***	0.088***	0.041***	0.207***
	[0.010]	[0.006]	[0.009]	[0.008]	[0.008]	[0.008]	[0.005]	[0.011]
lev	0.040***	0.040***	0.008**	0.073***	0.083***	0.036***	0.079***	-0.002
	[0.004]	[0.006]	[0.004]	[0.009]	[0.008]	[0.006]	[0.010]	[0.005]
Other Financial Information								
ltgrowth	0.120***	0.053***	0.049***	0.074***	0.195***	0.058***	0.131***	0.096***
	[0.007]	[0.009]	[0.006]	[0.008]	[0.015]	[0.005]	[0.012]	[0.012]
salesgrowth	0.143***	0.009	0.054***	0.070***	0.097***	0.067***	0.019***	0.133***
	[0.011]	[0.007]	[0.008]	[0.013]	[0.008]	[0.007]	[0.004]	[0.012]
rdpersales	0.620***	0.017**	0.562***	0.075***	0.595***	0.062***	0.325***	0.377***
	[0.006]	[0.007]	[0.016]	[0.014]	[0.006]	[0.006]	[0.031]	[0.025]
No. Quarters	20	20	20	20	38	38	20	20

Table 7.
(Continued)

Panel C: SP1000 Base Firms Continued

	ACP [EW] (9)	Δ [EW] (10)	ACP [VW] (11)	Δ [VW] (12)	ACP10 [EW] (13)	Δ [EW] (14)	ACP10 [VW] (15)	Δ [VW] (16)
Valutaion Multiples								
pb	0.052***	0.037***	0.057***	0.037***	0.053***	0.036***	0.057***	0.037***
	[0.003]	[0.006]	[0.003]	[0.005]	[0.003]	[0.004]	[0.003]	[0.004]
evs	0.261***	0.101***	0.279***	0.115***	0.282***	0.080***	0.302***	0.092***
	[0.006]	[0.005]	[0.006]	[0.006]	[0.005]	[0.004]	[0.005]	[0.004]
pe	0.032***	0.008*	0.038***	0.005	0.038***	0.002	0.040***	0.003
	[0.004]	[0.005]	[0.004]	[0.005]	[0.005]	[0.004]	[0.005]	[0.004]
Financial Statement Ratios								
rnoa	0.165***	0.027***	0.185***	0.043***	0.187***	0.004	0.200***	0.028***
	[0.007]	[0.006]	[0.008]	[0.006]	[0.008]	[0.005]	[0.008]	[0.005]
roe	0.030***	0.018***	0.036***	0.023***	0.032***	0.017***	0.034***	0.025***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
at	0.430***	0.106***	0.463***	0.105***	0.495***	0.041***	0.506***	0.063***
	[0.007]	[0.005]	[0.008]	[0.006]	[0.007]	[0.004]	[0.007]	[0.005]
pm	0.237***	0.063***	0.260***	0.078***	0.265***	0.035***	0.281***	0.057***
	[0.008]	[0.006]	[0.008]	[0.005]	[0.009]	[0.004]	[0.009]	[0.005]
lev	0.101***	0.019***	0.116***	-0.002	0.105***	0.016***	0.109***	0.005
	[0.009]	[0.006]	[0.010]	[0.007]	[0.009]	[0.006]	[0.009]	[0.006]
Other Financial Information								
ltgrowth	0.237***	0.005	0.264***	0.003	0.267***	-0.025***	0.279***	-0.012*
	[0.017]	[0.006]	[0.019]	[0.006]	[0.020]	[0.007]	[0.020]	[0.007]
salesgrowth	0.099***	0.059***	0.118***	0.059***	0.126***	0.032***	0.132***	0.045***
	[0.006]	[0.006]	[0.007]	[0.006]	[0.008]	[0.005]	[0.008]	[0.005]
rdpersales	0.598***	0.054***	0.630***	0.056***	0.646***	0.006	0.659***	0.027***
	[0.007]	[0.007]	[0.007]	[0.006]	[0.006]	[0.005]	[0.007]	[0.005]
No. Quarters	40	40	40	40	40	40	40	40

Table 8.
Base Firm Characteristics and Agreement between SBP and ACP

This table reports results from three different regression specifications. Observations are at the firm-year level. Columns 1 and 2 report ordinary least squares ("OLS") regressions of the % agreement between a base firm's top 10 SBPs and ACPs on base firm characteristics. Columns 3 and 4 reports OLS regressions similar to 1 and 2, but uses as the dependent variable: log of 1 + % agreement ("Log"). Columns 5 and 6 report results of ordered logit ("OLogit") regressions of the number of firms disagreed upon among the top 10 SBPs and ACPs on base firm characteristics. *"log size"* is the log of the base firm's market capitalization. *"complexity"* is the number of reported segments with distinct SIC2 classifications. Year fixed effects are included throughout but coefficients are suppressed for ease of reporting. Even columns include additional controls that require availability of data from I/B/E/S. Standard errors are two-way clustered at the base-firm and year level. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

	(1) OLS	(2) OLS	(3) Log	(4) Log	(5) OLogit	(6) OLogit
log size	0.0542***	-0.0059	0.0409***	-0.0030	0.4432***	-0.0419
	[0.0046]	[0.0047]	[0.0034]	[0.0033]	[0.0398]	[0.0392]
pb	-0.0001	0.0000	-0.0001	0.0000	-0.0006	0.0001
	[0.0001]	[0.0000]	[0.0001]	[0.0000]	[0.0007]	[0.0003]
evs	-0.0000*	-0.0000*	-0.0000*	-0.0000**	-0.0000	-0.0001*
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
pe	-0.0000*	-0.0000*	-0.0000*	-0.0000*	-0.0000**	-0.0000*
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
rnoa	0.0014*	0.0016***	0.0011**	0.0012***	0.0117**	0.0141***
	[0.0008]	[0.0005]	[0.0005]	[0.0004]	[0.0057]	[0.0036]
at	0.0002	-0.0020	0.0004	-0.0010	0.0036	-0.0075
	[0.0042]	[0.0053]	[0.0030]	[0.0037]	[0.0318]	[0.0435]
lev	0.0002	0.0000	0.0001	0.0000	0.0013	0.0002
	[0.0001]	[0.0001]	[0.0001]	[0.0000]	[0.0012]	[0.0004]
salesgrowth	0.0080	0.0228	0.0052	0.0157	0.0772	0.1904
	[0.0101]	[0.0214]	[0.0071]	[0.0153]	[0.0718]	[0.1877]
rdpersales	0.0129	-0.1409*	0.0116	-0.0881*	0.1189	-0.9245
	[0.0144]	[0.0751]	[0.0108]	[0.0524]	[0.1128]	[0.6360]
complexity	-0.0245***	-0.0128***	-0.0180***	-0.0093***	-0.1929***	-0.1012***
	[0.0037]	[0.0043]	[0.0027]	[0.0032]	[0.0307]	[0.0380]
coverage		0.0124***		0.0088***		0.1068***
		[0.0013]		[0.0009]		[0.0118]
ltgrowth		-0.0016***		-0.0011***		-0.0138***
		[0.0006]		[0.0004]		[0.0052]
eps spread		0.0057		-0.0018		0.0844
		[0.0530]		[0.0360]		[0.5138]
ltgrowth spread		0.0019**		0.0014**		0.0151*
		[0.0009]		[0.0006]		[0.0090]
Observations	11,789	7,999	11,789	7,999	11,789	7,999

Table 9.
Explaining Top Peers: SBP vs. ACP

In columns 1∼2 (3∼4), the sample consists of the top 10 SBPs(ACPs) to each base firm in our sample as well as all other firms from the same GICS2 sector with the dependent variable an indicator for being a top 10 SBP(ACP) to a base firm in our sample. Explanatory variables are decile ranked absolute percentage difference in firm fundamentals between the base firm and the peer firm ($|\frac{peer}{base} - 1|$). "*Supply Chain*" is an indicator variable equaling 1 when the base and peer firm are supply chain partners. "*Different GICS4*" is an indicator variable equaling 1 when the base firm and the peer firm belong to different 4-digit GICS industry groupings. Year fixed effects are included throughout. Marginal effects ("*MFX*") from probit estimation is reported, and are evaluated at 1 for all explanatory variables with the exception of "*Supply Chain*," "*Different GICS4*" and the year fixed effects, which are evaluated at 0. Standard errors are clustered at the base-firm level. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

	(1) SBP Top 10 *MFX*	(2) SBP Top 10 *MFX*	(3) ACP Top 10 *MFX*	(4) ACP Top 10 *MFX*
size	0.0038***	0.0069***	0.0006	0.0024***
	(0.000)	(0.001)	(0.000)	(0.000)
pb	0.0004*	0.0007**	0.0010***	0.0010***
	(0.000)	(0.000)	(0.000)	(0.000)
pe	-0.0052***	-0.0051***	-0.0046***	-0.0024***
	(0.000)	(0.001)	(0.000)	(0.000)
rnoa	-0.0037***	-0.0027***	-0.0047***	-0.0025***
	(0.000)	(0.000)	(0.000)	(0.000)
roe	-0.0003	0.0010***	-0.0009***	0.0008***
	(0.000)	(0.000)	(0.000)	(0.000)
at	-0.0054***	-0.0052***	-0.0047***	-0.0029***
	(0.000)	(0.001)	(0.000)	(0.000)
evs	-0.0029***	-0.0028***	-0.0022***	-0.0014***
	(0.000)	(0.000)	(0.000)	(0.000)
lev	-0.0012***	-0.0020***	-0.0019***	-0.0017***
	(0.000)	(0.000)	(0.000)	(0.000)
salesgrowth	-0.0037***	-0.0021***	-0.0041***	-0.0014***
	(0.000)	(0.000)	(0.000)	(0.000)
rdpersales	-0.0022***	-0.0010**	-0.0058***	-0.0037***
	(0.000)	(0.000)	(0.001)	(0.001)
coverage		-0.0034***		-0.0024***
		(0.001)		(0.001)
eps spread		-0.0007**		-0.0011***
		(0.000)		(0.000)
ltgrowth		-0.0035***		-0.0033***
		(0.000)		(0.000)
ltgrowth spread		-0.0023***		-0.0020***
		(0.000)		(0.000)
Supply Chain	0.4001***	0.3052***	0.3584***	0.1900***
	(0.027)	(0.033)	(0.031)	(0.027)
Different GICS4	-0.0336***	-0.0783***	-0.0643***	-0.1567***
	(0.002)	(0.006)	(0.003)	(0.009)
Observations	3,446,653	842,448	3,467,536	845,112
Pseudo R^2	0.0670	0.0878	0.0870	0.1237

Table 10.
Performance of Composite Peers

This table compares the average R^2 values from monthly cross-sectional regressions of the form

$$R_{i,t} = \alpha_t + \beta_t R_{p_{i,t}} + \epsilon_{i,t}$$

using CRSP returns data from January 2004 to December 2013. Columns 1∼2 of Panel A and columns 1∼3 of Panel B report average R^2s from monthly cross-sectional regressions, regressing base firm i's returns in a given month t on the concurrent returns of the relevant peer portfolio p_i. Column 1 considers an equal-weighted portfolio top 10 SBP firms, ranked by the prior calendar year's *Annual Search Fraction* f_{ij}, defined as the fraction of daily-users searching for both firm i and j's' information on the same day conditional on searching for firm i and any other firm $\neq i$, aggregated over the course of a calendar year; Column 2 considers an equal-weighted portfolio of peer firms that belong to both the top 10 SBP and ACP portfolios; Column 3 considers an equal-weighted portfolio of peer firms that belong to either the top 10 SBP or ACP portfolios.

The results are reported for the sample of base firms that belonged to the S&P1500, S&P500, or S&P1000 index at the beginning of each calendar year. To facilitate comparisons, all the regressions are conducted using the same underlying set of firms. Panel A considers the subset of base firms that have both ACPs and SBPs; Panel B considers the subset of Panel A base firms that have overlapping ACPs and SBPs. The variable N in parentheses represents the average cross-sectional sample size for each monthly regression and standard errors are reported in square brackets. Significance levels are indicated by *, **, *** for 10%, 5%, and 1%, respectively.

	SBP (1)	SBP∩ACP (2)	SBP∪ACP (3)	(2)-(1) (4)	(3)-(1) (5)
Panel A: Base Firms with ACP and SBP					
SP1500 Base Firms	0.130***		0.141***		0.012***
(N= 1,311)	[0.006]		[0.006]		[0.001]
SP500 Base Firms	0.211***		0.218***		0.008***
(N= 422)	[0.008]		[0.008]		[0.002]
SP1000 Base Firms	0.106***		0.117***		0.011***
(N= 889)	[0.005]		[0.006]		[0.001]
Panel B: Base Firms with Overlapping Peers					
SP1500 Base Firms	0.143***	0.113***	0.153***	-0.030***	0.010***
(N= 1,199)	[0.006]	[0.005]	[0.006]	[0.002]	[0.001]
SP500 Base Firms	0.214***	0.192***	0.220***	-0.022***	0.006***
(N= 413)	[0.008]	[0.008]	[0.008]	[0.003]	[0.002]
SP1000 Base Firms	0.119***	0.091***	0.130***	-0.028***	0.010***
(N= 785)	[0.006]	[0.005]	[0.006]	[0.002]	[0.001]
Number of Months	120	120	120	120	120