Common Analyst –Based Method for Defining Peer Firms

Markku Kaustia
Aalto University

Ville Rantala
Aalto University

March 11, 2013

Abstract

Most analysts specialize in covering firms that belong to a specific industry or are otherwise related to each other in their firm characteristics. This specialization suggests that analysts' coverage choices contain valuable information on firm similarity: Through their role as information intermediaries in financial markets, analysts provide a natural and market-based source of information on firm relatedness. We develop a method for defining groups of peer firms on the basis of joint analyst coverage. Our main performance tests concentrate on measuring peer group homogeneity. In these tests we pit the analyst-based method against common industry classifications such as SIC codes, Fama-French industries, The North American Industry Classification System (NAICS), and Global Industry Classification Standard (GICS). The analyst-based method consistently outperforms these common classifications: The peer firms identified by the method are more homogenous in terms of different firm characteristic measures and financial ratios, and have a higher correlation in stock returns. Besides industry boundaries, analysts' coverage choices can also reflect other important aspects of relatedness between individual firms. We also find that common analysts can be used very effectively to identify single comparable firms, and provide a competitive alternative to commonly used size and market-to-book matching methods. More generally, our results show that common analysts between firms offer a simple, yet powerful method for finding comparable firms.

Keywords: Peer firms, industry classification

JEL Classification: G19, G39

1. Introduction

Forming groups of similar firms is a rudimentary component of methodology in countless academic studies as well as industry and policy analyses. The most well-known classification system is the Standard Industrial Classification (SIC) developed in the 1930's to organize firms by the type of their primary business activity. Academic studies mostly use 2- or 3-digit SIC codes to group firms. Other alternatives have also emerged. Fama and French (1997) reorganize SIC information to form 48 industry groups. These 'Fama-French industries' are probably the most widely used classification in finance research today. In 1997 The US Office of Management and Budget introduced The North American Industry Classification System (NAICS) to replace the SIC system.

Firm classification is not a trivial task. One can judge a classification system along multiple dimensions. First, homogeneity of the groups is a desirable attribute. Second, the relatedness of firms within a group should be based on genuine links between them, as opposed to being statistical artifacts. Third, a good system is dynamic, promptly reacting to changes in the underlying firms. Fourth, ease of use is naturally one consideration.

In this paper we develop a method for defining peer firms on the basis of firms followed by the same sell side security analysts. Analysts tend to specialize in covering firms in the same industry. This industry specialization is apparent, for example, in analyst rankings, such as the *Institutional Investor* All-America Research Team, and the *Wall Street Journal* Best on the Street survey, which are based on identifying top analysts in different industry sectors. In addition to industries, analysts' coverage choices can also reflect other aspects of firm similarity. Common analysts provide a natural measure of company relatedness, reflecting professionals' views on industry boundaries and related firms.

Our method is based on comparing analysts' actual coverage choices to simulated random coverage choices. In essence, we identify firms that have more common analysts than what would be expected by chance. Peer groups produced by this method are firm- and year-specific, and independent of any conventional industry classification schemes. On an average sample year, we are able to compile an analyst-based peer group for 69% of U.S. based NYSE firms.

How does the analyst-based method perform? Our main tests concentrate on measuring peer group homogeneity. In these tests we pit the analyst-based method against common industry classifications (SIC, Fama-French, NAICS, and GICS). More specifically, we run regressions explaining various firm-specific financial variables with the peer group average of those variables. Note that for each firm-level observation *i*, firm *i* is not included in the calculation of the group average. This method was used by Bhojraj, Lee, and Oler (2003) for comparing the performance of standard industry classification systems.

The adjusted R^2 values from these regressions are high when the distances of individual firms' figures from their group average are small. As an example, consider the market value-based leverage ratio. Running a regression that explains individual firms' leverage ratios by their peer group average results in an adjusted R^2 of 13.5% when using the Fama-French industry classification. SIC code classification produces figures between 12.9% and 16.5%, depending on how many SIC code digits are used. The common analyst-based method achieves an adjusted $R^2$of 22.2%. Other test variables include variables related to the scope of the firm's operations (e.g., sales), valuation (e.g., market-to-book), basic financials (e.g., profitability), corporate finance policies (e.g., dividend payment), and others. The analyst-based method comes out ahead of competition in these tests with remarkable consistency.

We also study the effectiveness of common analysts in identifying a single comparable firm. Individual comparable firms are needed e.g., when constructing matching firm-adjusted long-term stock returns, as recommended by Barber and Lyon (1997) and Lyon, Barber, and Tsai (1999). In these tests, we compare the analyst-based method to using firms' size and market-to-book ratio to find the closest comparables. A firm's common analyst matched comparable firm is the firm with the highest number of shared analysts. Regressions similar to the performance tests with the peer groups show that defining comparables with this simple method produces adjusted R^2 values that outperform alternative matching methods in several variables, including monthly stock return. The method's performance with different test variables could naturally be further improved by combining it with other variables in the matching algorithm, but our findings suggest that common analysts are a very effective matching variable even on their own.

There are two technical factors that contribute to, but do not ultimately drive the superior performance of the analyst-based method. First, part of its success can be attributed to the fact that the method excludes firms with low analyst-coverage. These firms are typically quite small, and there can be more idiosyncrasies among small firms. However, the method also outperforms conventional industry classifications in the vast majority of cases even after we ensure a level playing field by only using firms that have analyst-based peers for all classification systems. Second, the analyst-based peer groups are smaller than the peer groups produced by most conventional industry classifications.[1] The smaller group size may be partially responsible for the method's ability to bundle together closely related firms. However, the analyst-based method also outperforms 4-digit SIC codes, 5-digit NAICS codes, and 8-digit GICS codes, the most fine-

[1] Some classifications, such as Fama-French industries, consist of firm groups that are likely to be too large to effectively identify the most closely related comparables. They are, for example, significantly larger than the typical benchmark peer groups used in executive compensation (Faulkender and Yang, 2010; Bizjak, Lemmon, and Nguyen 2011).

grained standard classifications, and it excels in finding a single comparable firm, as described earlier. Furthermore, statistics on industry classification codes within analyst-based peer groups show that although peers generally have similar industry codes, the peer groups are not subgroups of any existing industry classification.

We suspect that the superior performance of the analyst-based method is due to analysts' better information on firm similarities and boundaries, and their incentives to monitor such issues. Besides industry relatedness, coverage choices can reflect other aspects of firm relatedness, such as domestic versus international scope, ownership structure, or firm maturity. Firms also change over time. There is slow change due to industry migration, but also sudden changes due to mergers and divestitures, or changes in corporate strategy. Analysts likely react to such changes in their coverage choices, perhaps more promptly than standard classification systems do. In some sense one could call our method a market-based solution to classification, while other methods may be more bureaucratic. Conventional industry classifications generally classify firms into pre-defined groups based on their product-market operations. This is often not straightforward, as illustrated by different SIC codes allocated to the same firms by different data vendors (Guenther and Rosman, 1994). Hoberg and Phillips (2010) compile product-market industries based on 10-K product descriptions reported to the Securities and Exchange Commission, and their study suggests that it is possible to use such product market information more efficiently than standard classifications systems do.

Our analyst-based peer groups are particularly suitable for studying hypotheses related to comparable firms' effect on decision making, as analysts' coverage choices are likely to be correlated with managements' views on comparable firms: For example, Porter (1980) mentions security analysts and their reports as a source for competitor information in industry analysis.

Analysts also commonly engage in direct communication with firms' management (see e.g. Francis, Hanna, and Philbrick (1997) and Frankel, Johnson, and Skinner (1999)). Kaustia and Rantala (2012) use the analyst-based method developed in this paper to study peer firms' effect on corporate decision making in stock splits. Other methods for selecting comparable firms are still likely to be more suitable with hypotheses where accurate measurement of product-market relationships is important, or where the ideal number of comparables is larger (such as in industry portfolios for asset pricing tests). Standard classifications are also more convenient for filtering data, for example, excluding financial firms from the sample is much simpler based on SIC codes.

All in all, the analyst-based method outperforms common industry classification systems in producing homogeneous groups of firms. It provides a dynamic measure of firm relatedness based on genuine links between individual firms. It is practical to use, as the number of common analysts between firms can be easily calculated. If necessary, the procedure can be adjusted to generate groups of different sizes to suit different purposes.

The rest of the paper is organized as follows. Section 2 describes the data and the method for forming the analyst-based peer groups. Section 3 reports group composition statistics for he analyst-based peer groups and compares them to conventional industry classification groups. Section 4 studies peer group homogeneity, and Section 5 examines the effectiveness of common analysts in identifying individual comparable firms. Section 6 concludes.

2. Description of data and the analyst-based peer group method

Stock price and firm data are from CRSP and Compustat, and analyst data are from the IBES Detail database. The sample consists of all U.S. based firms (CRSP Share Code 10 or 11) listed

in the NYSE (Exchange Code 1) and covers years 1983-2009. The NYSE offers a good setting for studying analysts as a measure of company relatedness, as NYSE listed firms are more likely to be covered by analysts (see e.g. Chang, Dasgupta, and Hilary (2006)). The limited availability of analyst data in IBES prior to 1983 significantly reduces the number of firms for which we can form an analyst-based peer group. The industry classification codes which we compare to the analyst-based peer groups are from CRSP and Compustat: SIC codes are from Compustat (if available) and otherwise from CRSP, Fama-French (49) industry groups are formed based on the SIC codes, and NAICS and GICS codes are from Compustat. NAICS codes are available in Compustat starting in 1985 and GICS codes starting in 1999.

The analyst data consists of analysts following the sample firms during the sample years. Individual analysts are separated based on Analyst Code item in IBES Detail. The code is normally assigned to individual analysts, but can also refer to analyst teams. As some codes may refer to teams consisting of several individuals, we exclude codes that are associated with more than 50 different firms in a single year (the excluded year-code combinations account to less than 0.6% of all the analyst observations). An analyst is considered to follow a firm in year *t* if she has provided an estimate figure for the firm in year *t*.

A. Common analysts as a measure of firm relatedness

The analyst-based peer groups are based on the observation that analysts generally focus on following related firms that belong to the same industry. Statistics on the firms followed by individual analysts show that firms followed by the same analyst typically have similar industry classification codes (see e.g. Mikhail, Walther, and Willis (2004) and Boni and Womack (2006)). In addition to the industry classification-based evidence, security analyst literature also offers potential institutional and incentive-based explanations for analysts' industry specialization.

An institutional factor behind the specialization is that brokerage houses are typically organized so that individual analysts focus on a specific industry. Michaely and Womack (1999) report, based on discussions with research and M&A directors of large investment banks, that "most analysts [in the large investment banks] focus on a specific industry, although some are generalists, covering multiple industries or stocks that do not easily fit into industry groupings". Quite similarly, according to Hong and Kubik (2003), large elite brokerage houses employ a large number of individual analysts who cover different industries, whereas smaller brokerage houses specialize in covering specific industries or types of stock. Hong and Kubik explain that smaller brokerage houses base their business on catering to the needs of specific institutional investor clienteles that need research on a particular industry. They also report that the number of smaller "specialist" brokerage houses has increased significantly since the 1980s. Altogether, these findings suggest that individual analysts tend to focus on specific industries regardless of the size of their brokerage house.

Besides institutional reasons, analysts' personal incentives can also contribute to coverage choices that concentrate on a specific industry. One such factor can be that public analyst rankings, such as the *Institutional Investor* All-America Research Team and the *Wall Street Journal* Best on the Street survey, are based on identifying top analysts in different industry sectors. Being selected in the All-America Research Team has a significant effect on analyst compensation (Stickel, 1992; Michaely and Womack, 1999; Hong, Kubik, and Solomon, 2000), which suggests that there is an incentive to be regarded as a specialist in a specific industry. Another motive for industry specialization can be the value of industry-specific knowledge in analyst reports and estimates. Findings of Boni and Womack (2006) suggest that analysts create value in their recommendations mainly through their ability to rank stocks within industries,

which indicates that the value of industry-specific knowledge can also contribute to the observed industry specialization.

Altogether, industry specialization appears to be the general norm among security analysts, and there are incentives to cover firms in the same industry. Evidence also indicates that analysts actively adapt their coverage according to changes in the firm composition of an industry: Das, Guo, and Zhang (2006) report that the number of analysts providing initial coverage for an IPO firm is significantly correlated with the number of analysts covering seasoned firms in its industry. Gilson, Healy, Noe, and Palepu (2001) have similar results for conglomerate breakups: Former conglomerate subsidiaries experience a significant increase in coverage by analysts who cover firms in their industry. These results show that analysts' coverage choices actively adapt to changes in industry composition, which is important for the accuracy of the analyst-based peer groups.

B. Method for forming analyst-based peer groups

When peer firms are identified based on common analysts, a crucial question is what number of common analysts between two firms is a sufficient indication of their relatedness. Even when analysts generally focus on following firms in a specific industry, it is very possible that unrelated firms have common analysts merely by chance, because of individual analysts that cover several industries or groups of unrelated firms. Establishing how many common analysts are sufficient to indicate a nonrandom connection between two firms is therefore important.

Analyst coverage varies greatly between firms. Two widely-followed firms could thus easily have one or more common analysts even if the firms are unrelated. This would not be as likely in the case of firms with less analyst coverage. For any firm, the probability of sharing same analysts with another random firm is increasing in the number of analysts following it, as well as

the number of other firms followed by these analysts. Both variables increase the number of analyst-firm links a firm has with other firms, and the larger the number of links, the larger is the probability that the firm has a specific number of common analysts with another unrelated firm by chance.

Our method for forming analyst-based peer groups is based on deriving the minimum number of analysts for each firm *i* that it must share with another firm *j*, to include *j* in *i*'s peer group. This criterion is based on the number of analysts following *i* in year *t*, as well as the number of other firms followed by each such analyst. Specifically, we run a simulation, in which the analysts following *i* counterfactually choose the other firms they follow at random among NYSE firms that have analysts in year *t*. The criterion *C* for each firm is selected so that the probability of having more than *C* common analysts by chance is less than 1% (i.e., we set *C* sufficiently large so as not to assign unrelated firms as peers). We use 1,000 simulation rounds, so the 1% confidence interval means that *C* is selected so that the firm has more than *C* analysts in common with any other firm in less than 10 simulation rounds. After *C* has been derived, a firm's peer group is defined as all other firms that have at least *C* common analysts with it. The simulation is run separately for all firms in each sample year, and the peer groups are updated annually. Our choice of 1% confidence interval is ad hoc. Tinkering with this parameter affects the size of the peer groups, and increasing the figure would lead to larger peer groups.

C. Features of the analyst-based peer group method

The number of analysts following a firm, and the number of other firms followed by each such analyst both contain information that is relevant when making inferences about nonrandom relationships between firms. Any statistical cluster analysis method that only utilizes the number

of analyst-firm connections a firm has with other firms would not account for all that information.

The simulation-based peer criterion does not only depend on the number of analyst-firm connections a firm has with other firms (i.e., the sum of other firms followed by each analyst following the firm). In the simulation, analysts individually choose at random which other firms to follow, and so the peer criterion also reflects how analyst-firm connections are distributed among different analysts. Suppose, for example, that a firm is followed by five analysts that each follow ten other firms, and another firm is followed by ten analysts that each follow five other firms. Both firms have fifty analyst-firm connections with other firms, but the ex-ante probabilities of sharing a specific number of same analysts with another firm by chance are different. The number of analysts following a firm in the data also imposes a natural upper limit for the number of common analysts a firm can possibly have with another firm.

As the simulated peer criterion is calculated based on the analyst data available in IBES Detail, possible differences between real-life analyst coverage choices and those recorded in IBES detail should have no systematic effect on the composition of the peer groups. Ljungqvist, Malloy, and Marston (2009) report that IBES Recommendation database suffers from ex-post alterations of recommendations, additions and deletions of records, and removal of analyst names. As we only use information on the firms covered by individual analysts, possible alterations in the data should not have any systematic impact on the peer groups. If some real-life analysts or estimates are missing from the IBES data, the smaller number of available analysts will, on average, result in a lower peer criterion, and we can still compile peer groups based on the data available in IBES.

A numerical implication of the method is that the smallest possible peer criterion is two analysts. A firm with a single analyst can never have a peer group. If a single analyst chooses the firms she follows randomly, the method cannot distinguish which firms appear to be nonrandom choices by the analyst. With at least two analysts, the simulation produces a probability distribution for having a specific number of common analysts with another firm by chance. The method does not provide peer groups for all sample firms, as some firms do not have sufficient analyst coverage in the IBES Detail database. Having at least two analysts is a necessary, but not a sufficient condition for having a peer group: A firm does not have a peer group if the number of common analysts with other firms is always below the peer criterion.

The peer relationships identified by the method are not always mutual; If firms A and B have different peer criteria, it is possible that firm A is firm B's peer, but B is not A's peer. Suppose, for example, that A's analyst criterion is five, B's criterion is three, and the firms have four common analysts. Because B has a lower analyst criterion, A makes it to B's peer group, but B does not make it to A's peer group. The peer criteria of A and B can be different, if they have different analyst coverage. Non-mutual peer relationships can be intuitively justified in situations where we are trying to find the closest firm-specific comparables among a group of related firms. For example, suppose that firms A, B, and C are the only firms operating in a specific industry, A and B are large firms, and C is small. A and B are likely each others' best comparables because of their similar size. The best comparable for firm C can also be either firm A or firm B from the same industry, or another smaller firm from a related industry. In case it is A or B, C would have a non-mutual comparable firm relationship with A or B.

3. The composition of peer groups

A. Statistics on sample firms and characteristics of the analyst-based peer groups

Table 1 shows annual descriptive statistics on sample firms and analysts. On an average year there are 1,501 firms and 2,077 analysts, and over two thirds of the firms have an analyst-based peer group. The number of analysts is significantly smaller in the 1980s, but over 60% of the firms still have an analyst-based peer group in each of the sample years. The average number of firms followed by an analyst is higher in the earlier years of the sample, and ranges from 6.5 in 2002 to 11.7 in 1986. The size of the peer groups remains relatively stable over the years, although the groups are on average slightly larger during the first half of the sample period. The average annual group size ranges from 9.5 in 2002 to 14.0 in 1987. The groups are generally larger in the years when the average number of firms followed by an analyst is also higher. The composition of firm-specific peer groups does change over the years, but overall remains relatively stable. On average, 78% of firm *i*'s peers in year *t* were its peers also in year *t* - 1 (the median is 86%).

Statistics on the simulation-based analyst criteria illustrate the general probability that firms can have some common analysts merely by chance: the average analyst criterion in the data is 4.12, and the median is four. 61% of the criteria are between three and five, and the highest criterion in the data is eight. The simulated criteria are not highly dependent on the specific 99% confidence level that is used in the analysis; over 60% of the firm observations would have the same criterion at the 95% confidence level, and over 83% would have the same criterion at the 99.5% confidence level. Absolute change in the value of the criterion would not be larger than one unit for any firm with either of the two alternative confidence levels. Altogether, the simulation statistics suggest that firms can have several common analysts even by chance, and

the number of common analysts that is deemed non-random can vary significantly across firms. The simulation results support the view that it is necessary to account for firm-level variation in analyst coverage when defining groups of comparables based on common analysts.

Table 2 compares firms with analyst-based peers to all NYSE firms. As expected, the statistics show that firms with analyst-based peers are larger in terms of market capitalization and book equity. The average market capitalization in the sample is $4.4 billion for all NYSE firms and $6.2 billion for firms with analyst-based peers, and the average common equity (based on Compustat Item 60) of firms with analyst-based peers is also 36% higher than the NYSE average.

As analyst-based peer groups can only be formed for firms with analysts, small firms without analyst coverage in the IBES Detail database are excluded from the sample. In the highest market capitalization quartile, 95% of the NYSE firms have an analyst-based peer group, whereas the same number for the lowest quartile is 29%. A potential additional implication of the firm size statistics is that analyst-based peers are more similar in terms of firm size compared to the comparable firms suggested by conventional industry classifications.

B. Group composition of analyst-based peer groups vs. industry classification groups

The industry classifications used in the comparisons are SIC codes, Fama-French (49) industries, and the more recently introduced NAICS and GICS classifications. The comparable firm groups based on SIC codes are formed at two-, three-, and four-digit levels, groups based on NAICS codes are formed at three- and five-digit levels, and groups based on GICS codes are formed at six- and eight-digit levels. Five-digit NAICS codes are the most detailed general industry classification level in the six-digit NAICS system, as the sixth digit refers to the country of the firm. Six-digit GICS codes are also referred to as GICS industries, and eight-digit GICS

codes are referred to as sub-industries. The eight-digit codes are the most detailed level of the GICS classification.

Panel A of Table 3 reports group size statistics for the analyst-based peer groups and the industry classification groups. The figures for the industry classification groups are calculated based on firm-year observations on the number of other NYSE firms sharing a firm's industry classification code in the data during the sample years, and the analyst-based peer group statistics are based on annual peer groups. With the industry groups, we measure the typical number of comparable firms for each firm in the data (e.g., average number of firms with the same classification code), rather than the typical size of an industry (e.g., average number of firms in an industry group), because the focus of our analysis is on comparing different methods for identifying the comparable firms of an individual firm.

The average size of an analyst-based peer group is 11.7 firms, and the median is 10. Among the industry classifications, only 4-digit SIC codes and 8-digit GICS codes provide a smaller number of comparables (11.6 and 11.3) on average. The median number of comparable firms is smaller with 4-digit SIC codes (6 firms), 5-digit NAICS codes (8 firms), and 8-digit GICS codes (8 firms). All other classification schemes lead to a larger number of comparables than the analyst-based peers. In terms of average group size, the industry-classification closest to the analyst-based peer groups is 4-digit SIC codes, and in terms of median group size, the closest classification is 3-digit SIC codes. Fama-French industries have the largest group size among the industry classifications with average group size of 54.9 firms and median of 44.

The analyst-based method produces the least amount of variation in the number of comparable firms. The distribution of the number of firms is also the least skewed: the ratio of average to median group size is 1.17. Figures on relative variation (standard deviation / average)

show a lowest figure for Fama-French industries (0.67), followed by analyst-based groups (0.761) and 3-digit NAICS codes (0.764). Compared to most classifications reported in Table 3, the analyst-based peers suggest less comparables, and the number of comparables is more homogeneous across firms.

Panel B of Table 3 describes the extent to which conventional classification schemes overlap with analyst-based peers. In general, analyst-based peers have similar industry codes: the average number of codes in a peer group ranges from 1.74 when using 6-digit GICS codes to 4.4 when using 4-digit SIC codes. The median figures for the number of industry codes are close to the averages. However, statistics on the percentage of analyst-based peers sharing a firm's own industry classification code show that the analyst-based peer groups are not merely subgroups of existing industry classifications: On average, less than half of a firm's analyst-based peers share its own 4- or 3-digit SIC code, or 5-digit NAICS code. With less detailed classifications the industry classification codes of peers are more similar, for example the average number of peers in the same Fama-French group is 64.8% and the median is 80.0%.

The analyst-based peer groups can also be used to compare conventional industry classifications with each other. Although different industry classifications may serve different purposes and measure different aspects of company relatedness, it is nevertheless interesting to identify which classification is the most correlated with analysts' coverage choices. Based on the statistics in Table 3, GICS has the highest overlap with the analyst-based peer groups, to the extent that 6-digit GICS groups have more overlap than the much larger Fama-French industries.

4. Homogeneity of peer groups

In this section we study how similar the firms within a peer group are given different classification methods. The SIC and Fama-French industry classification groups that we compare to the analyst-based peer groups are based on annual observations from the entire 1983-2009 period. The NAICS was introduced in 1997 and is replacing SIC codes with US Federal statistical agencies. The system was backfilled to cover historical data as well, Compustat covers the years 1985-2009 for NAICS. The GICS classification was developed jointly by MSCI and Standard & Poors and it was introduced in 1999. GICS codes are available in Compustat from the year 1999, and our comparisons with GICS codes are based on the years 1999-2009.

Note that one could always come up with new classification schemes ex post, and backfill to cover historical data. In fact, when designing a classification system based on any measurable firm characteristics, it is rational to consider historical data to separate meaningful clusters in the firm characteristics space. While there is nothing wrong in doing so, it can introduce a backward looking bias in classification performance evaluation: future firms may not organize as neatly according to the criteria as past or current firms do. The common analyst method only uses information on who is covering whom, it does not involve any judgment of industry boundaries, or use of historical data. Analysts themselves can of course employ any subjective criteria, and also be biased in their coverage choice, but our method nevertheless takes coverage choices as given. Therefore, unlike with backfilled classification systems, the analyst-based peer groups at any historical date *t* are based only on information available at date *t*.

A. Method

We adopt the comparison method of Bhojraj, Lee, and Oler (2003) who compare conventional industry classification schemes against each other. Besides being a sensible method, following Bhojraj et al. has the advantage of eliminating any potential bias: since we are advocating our analyst-based method, it might seem suspicious if we also developed the criteria by which we judge its performance against other systems. Based on similar reasoning we adopt all the test variables used by Bhojraj et al. (2003), and then augment that list to cover more variables of interest to finance researchers.

Specifically, we compare the adjusted R^2 values from the following regression:

$$Variable_{i,t} = \alpha_1 + PeerAverage_{i,t} + \varepsilon_{i,t} \tag{1}$$

where $Variable_{i,t}$ refers to the variable value for the firm i, and $PeerAverage_{i,t}$ refers to the average value among the peers of firm i. Importantly, the peer average does not include firm i itself. The adjusted R^2 value shows the explanatory power of the peer average on the variable value of the firm. It is a distance metric: adjusted R^2 is higher when the squared distance of each firms' observation from its own peer group average is lower.

We use two different sets of test variables for the comparison. The first group of variables consists of those that are commonly of interest in finance research, including variables related to stock return correlation, firm size, valuation, dividend policy, and capital structure. The second set of variables consists of valuation multiples, financial statement ratios, and other financial information variables that Bhojraj, Lee, and Oler (2003) use in a comparison of conventional industry classification schemes.

The finance variables are Monthly Stock Return, Monthly Stock Return with Month Fixed Effects (FE), Beta (estimated from a single-index model with S&P 500 based on the most recent 36 monthly returns from the end of the observation year), Market Value of Equity, Total Assets (Compustat Item 6), Net Sales (Item 12), Market-to-Book (book assets (Item 6) minus book equity plus market capitalization all divided by book assets), Dividend Payment (binary variable which takes value one if a firm has non-zero dividends per share by the ex date (Item 26), and zero otherwise), Book Leverage (book debt to total assets) and Market Leverage (book debt divided by the result of total assets minus book equity plus market equity).

Share prices for the market capitalization are from CRSP and shares outstanding are from Compustat (Item 25 if available) and otherwise from CRSP. Book equity is stockholders' equity (216) (or first available of common equity (60) plus preferred stock par value (130) or book assets (6) minus liabilities (181)) minus preferred stock liquidating value (10) (or first available of redemption value (56) or par value (130)) plus balance sheet deferred taxes and investment tax credit (35) if available and minus post-retirement assets (330) if available. Book debt is defined as total assets minus book equity. Regressions using the dividend payment variable and the two leverage variables exclude utilities (SIC codes 4900 to 4949) and financial firms (SIC codes 6000 to 6999).

The second set of variables is calculated similarly as in Bhojraj et al. (2003) and consists of Price-to-Book[2], Enterprise Value-to-Sales (the sum of market cap, long-term debt (Item 9), and debt in current liabilities (Item 34) all divided by net sales (Item 12)), Price-to-Earnings (market cap divided by net income before extraordinary items (Item 18), Return on Net Operating Assets (net operating income after depreciation (Item 178) divided by the sum of property, plant, and

[2] The variable definition is different from the market-to-book defined earlier in the set of finance variables. For completeness, we also include Price-to-Book calculated according to the variable definition of Bhojraj et al. in the second set of variables. It is calculated as market cap divided by total common equity (Computat Item 60)

equipment (Item 8) and current assets (Item 4), less current liabilities (Item 5)), Return on Equity (net income before extraordinary items divided by total common equity), Asset Turnover (total assets (Item 6) divided by net sales), Profit Margin (net operating income after depreciation divided by net sales), Leverage (total liabilities (Item 9) divided by total stockholders' equity (Item 216), Sales Growth (net sales one year in the future divided by current value), and Scaled R&D Expense (research and development expense (Item 46) divided by net sales).[3] In analyses with the second set of test variables, we require that firms must have positive common equity and total stockholders' equity and net sales above $10 million. When calculating Price-to-Earnings we additionally require that net income before extraordinary items be positive.

All test variables except monthly stock returns are based on annual values at the end of the year and financial statement information at the end of year *t* is based on the fiscal year ending in year *t*. To eliminate the effect of possible outliers, we exclude observations with the lowest and highest 1% for all variables except Monthly Stock Return and Dividend Payment. Also, we do not exclude the lowest 1% for Scaled R&D Expense where many firms report zero. Regressions with SIC codes and Fama-French industries are based on annual observations between 1983 and 2009 and regressions with NAICS and GICS codes are based on annual observations from the years for which their industry codes are available in Compustat (1985 to 2009 for NAICS and 1999 to 2009 for GICS). To obtain comparable performance results with NAICS and GICS, we run the regressions with analyst-based peer averages separately for the period 1985-2009 and 1999-2009.

[3] Bhojraj et al. (2003) also include analysts' long-term growth forecast as a test variable but we do not. The analyst-based peer groups are based on firms with common analysts in IBES and there might be an analyst fixed effect in growth forecasts. A test based on this variable could thus be biased against the standard classification schemes.

In our analyst-based method, the number of shared analysts is a measure of the strength of the similarity relation between two peer firms. Standard classification schemes either do or do not assign two firms as peers, while we can measure the degree of relatedness in a less lumpy way. In this paper we utilize this property for generating analyst-weighted peer averages, calculated by weighting the test variable values for each peer firm with the number of common analysts with the peer. Other applications can also benefit from using information on the degree of relatedness. We run all regressions using these analyst-weighted peer averages as well as equally-weighted averages that ignore this information on the strength of the peer relation.

We calculate the industry code averages separately among all NYSE firms and among NYSE firms with analyst-based peers. The relative performance of the industry classification average among firms with analyst-based peers indicates to what extent the performance of the analyst-based peer groups can be attributed to the exclusion of small firms with low analyst coverage.

B. Results

Tests using the common finance variables are reported in Table 4 for SIC codes and Fama-French industries and in Table 5 for NAICS and GICS codes. When industry classification averages are calculated based on all NYSE firms, the analyst-weighted peer average receives the highest adjusted R^2 values with all test variables and the equally-weighted analyst-based peer variable always receives the second highest values. If the classification averages are calculated based on firms having analyst-based peers, the equally-weighted analyst-based peers still outperform the industry classifications on all test variables except Net Sales (5-digit NAICS groups have higher adjusted R^2), Book Leverage (8-digit GICS groups have higher adjusted R^2), and Market Leverage (4-digit SIC and 8-digit GICS groups have higher adjusted R^2). The

analyst-weighted analyst-based peer average is only inferior to 8-digit GICS codes with the two leverage variables, and outperforms all other classifications with every variable.

As stock return co-movements are the focus of attention in numerous studies, it is particularly interesting to note that analyst-based peers outperform the industry classification groups in explaining monthly stock returns: Both equally-weighted and analyst-weighted peer averages outperform all industry classification averages in the stock return regressions. Ramnath (2002) studies analysts' earnings forecast revisions for later announcers after one of the firms they cover announces its earnings. It is conceivable that firms with shared analysts may have more synchronous stock price movements in part due to analysts facilitating the incorporation of common information into stock prices (see also Piotroski and Roulstone, 2004; Chan and Hameed, 2006).

Comparing the performance of 2-, 3-, and 4-digit SIC codes in Table 4 gives insight regarding the possible advantage of more fine-grained industry classification. Surprisingly, the results do not uniformly improve as one moves from 2-digit groups to more granular groups. When using all NYSE firms, 3-digit SIC codes are the best performing SIC code scheme in 4 out of 10 tests, and both 2- and 4-digit codes are the best performing scheme in three cases. When limiting to firms having analyst-based peers, the 4-digit SIC code steals an additional victory from the 3-digit scheme in the Market-to-book ratio category, but only with a very narrow margin: it has an adjusted R^2 of 28.96% vs. 28.74%. With NAICS and GICS codes, the more detailed classification levels perform relatively better on most tests, but there are exceptions with individual variables.

Analyst-based peer groups also perform strongly with the variables used for standard industry classification comparison by Bhojraj et al. (2003). The results of these tests are reported in Table

6 for SIC codes and Fama-French industries and in Table 7 for NAICS and GICS codes. On each variable, equally-weighted and analyst-weighted analyst-based peer averages both receive higher adjusted R^2 values than any industry classification average calculated based on all NYSE firms. Again, the analyst-weighted average always performs better than the equally-weighted average. When the industry classification averages are calculated based on firms having analyst-based peers, the analyst-weighted analyst-based peer average outperforms all the classifications on seven out of the ten variables. It is outperformed by the 3- and 4-digit SIC codes and 5-digit NAICS codes in the Enterprise Value-to-Sales–ratio, where the adjusted R^2 figures are 48.88%, 48.74%, 48.68% and 47.77% for the 3- and 4-digit SIC codes, 5-digit NAICS codes, and analyst-based peer groups, respectively. The 4-digit SIC codes narrowly outperform it in asset turnover (80.38% vs. 80.27%), and the 3-digit SIC codes in profit margin (44.61% vs. 44.36%). Both the equally-weighted and analyst-weighted analyst-based peer average outperform the GICS codes in every variable of the second variable set.

Altogether, the results show that the analyst-based method outperforms conventional industry classifications in producing homogenous groups of comparable firms. Both the analyst-based method, as well as conventional classifications, perform better when they are calculated over firms having analyst-based peers. Comparing classical industry classification schemes with each other, we find that none of them perform systematically better than others on all the different variables, and more detailed classifications are not systematically better than broader ones. These results also imply that the strong performance of the analyst-based peer groups cannot be explained by the smaller number of comparable firms. 4-digit SIC codes, which produce on average smaller groups of comparables than the analyst-based peer method, do not systematically outperform other SIC classifications.

Naturally, strong explanatory power on any set of variables does not suggest that one comparable firm selection method is categorically better than another. The ideal characteristics of a system for selecting peer firms depend on the research problem at hand. The analyst-based peer method appears to be very good at identifying peer firms that are close to each other in terms of different financial firm characteristics, and suits research questions where such characteristics are important. Other company relatedness measures can be more accurate in, e.g., identifying product-market relationships between firms or identifying product-market competitors.

5. Selecting a single matching firm

In this section we study the effectiveness of common analysts in identifying individual comparable firms. Individual comparable firms selected based on different matching algorithms, such as size and market-to-book matching, are sometimes used in the finance and accounting literatures to control for the effect of firm-specific characteristics. One important application is the calculation of long-run abnormal returns (Barber and Lyon, 1997; Lyon et al., 1999). They use size and market-to-book matched comparable firms to control for firm characteristics that explain stock returns. Our results so far show that common analysts are effective in identifying groups of homogenous firms, and the method could also work well in identifying an individual comparable firm.

In the analysis, we compare individual comparable firms matched based on common analysts to firms matched based on market capitalization and market-to-book. The common analyst-method works as follows: firm *i*'s common analyst matched comparable in year *t* is the firm with the greatest number of common analysts with firm *i* in year *t*. If there are several potential

comparables with the same number of common analysts, one is selected randomly. The matching is conducted annually for all NYSE firms and the sample period covers years 1983-2009. 89% of the firm-year observations in the sample have common analysts with other firms, so firm coverage in single-firm matching is larger than in the forming of peer groups. For comparison, we report statistics also based on four subsets of firms with additional analyst coverage requirements (a minimum of 2 – 5 common analysts with their closest match),

The matching methods that we compare to common analyst matching are size (market capitalization) matching, market-to-book matching, and joint size/market-to-book matching. These variables are defined as in Section 4. Size and market-to-book matching are conducted so that firm *i*'s comparable in year *t* is the firm with the closest size or market-to-book value in year *t*. We run the matching separately among all NYSE firms, as well as among firms within the same 4-digit SIC code. Joint size/market-to-book matching is conducted as in Barber and Lyon (1997): firm *i*'s comparable is the firm with the closest market-to-book ratio among the firms with market capitalization between 70% and 130% of firm *i*'s market capitalization.

We use the same methodology as in the earlier comparisons to compare the relative performance of different matching methods. I.e., we compare adjusted R^2 values from firm characteristic regressions where the dependent variable is the firm's own variable value, and the independent variable is the variable value of the matched comparable. The set of firm characteristics variables used in the regressions consists of the finance variables defined earlier in Section 4. In tests where comparables are matched based on market capitalization or market-to-book, we do not report adjusted R^2 values for the matching variables, as they are being used in selecting the comparables.

Results from the firm characteristic regressions are reported in Table 8. Common analyst matching among firms with at least five common analysts with their matched comparable produces the highest adjusted R^2 values in seven of the ten test variables, and joint size/market-to-book matching produces the highest values in the remaining three variables. As can be expected, matching methods that include market capitalization as matching variable produce the highest adjusted R^2 values with Total Assets and Net Sales, which are the two alternative firm size measures among the test variables. Matching methods with Market-to-Book have the strongest performance with market leverage, which includes market equity and book equity in its variable definition.

The adjusted R^2 values among common analyst matched comparables increase almost monotonically with the minimum analyst criterion (the number of common analysts a firm must have with its comparable to be included in the sample). This is likely due to two factors: First, a higher minimum analyst criterion can result in more informative analyst-based matches, as it excludes firms that have low analyst coverage or high dispersion among the other firms followed by the analysts covering them. Such firms would more likely have common analysts with unrelated firms merely by chance. Second, part of the R^2 increase can be caused by a more homogenous pool of potential comparables, as the excluded low analyst coverage firms are likely to be smaller firms.

The performance of the common analyst-based matches relative to the other matching methods also increases with the minimum analyst requirement. When no requirement is imposed, common analyst matching outperforms all non-analyst-based matching methods in four out of the ten test variables. With a minimum requirement of three common analysts the method outperforms the other methods in six of the variables, and with a requirement of four or five

analysts the number increases to seven variables. This suggests that some analyst-based links between firms can be more informative than others.

The finding that R^2 values increase with the minimum analyst criterion suggests that it may be useful to impose minimum requirements on analyst coverage when using common analyst-matching. The downside of the analyst requirements is, of course, a smaller sample of firms that can be matched: For example, 82% of firm-year observations that have common analysts with other firms have at least three common analysts with another firm and 62% have at least five common analysts with another firm.

Among size- and market-to-book-based matching methods, no single method systematically outperforms the others. Size matching within a 4-digit SIC code produces the highest adjusted R^2 value in five variables, joint size/market-to-book matching in three variables, and market-to-book matching within the 4-digit SIC code in two cases. With size and market-to-book matching, matches within the 4-digit SIC code generally have much higher adjusted R^2 values than matches among all sample firms, and the full sample match has higher value than the 4-digit SIC match in only one case.

Altogether, the results show that common analyst -based matching is a powerful method for identifying individual comparable firms. For example, looking at monthly stock returns, the analyst-based match is better than any of the non-analyst-based matches even when no requirements are imposed on analyst coverage. The performance of common analyst matching with specific variables could of course be improved further by combining it with other matching variables, such as size or market-to-book. The results of this analysis are illustrative of the effectiveness of the method, but the ideal matching algorithm for a specific research problem could naturally include other features as well.

6. Conclusion

We have developed a method for identifying comparable firms on the basis of common analysts. The method is independent of conventional industry classification systems and outperforms them in producing homogeneous groups of peer firms. Part of the method's success likely derives from coverage choices reflecting, and reacting to, the emergence of new types of firm clusters as well as changes within individual firms.

The analyst-based method can be particularly useful in empirical studies where identifying closely related comparable firms is important to the research question. For many empirical research questions, the number of comparable firms suggested by conventional industry classifications may be too large to identify the closest peers effectively, or the detailed product-market based classification may not produce comparables that are sufficiently homogeneous in terms of other firm characteristics.

In general, our results suggest that identifying common analysts is a powerful tool for selecting comparable firms, and could also be used effectively as part of other comparable firm selection algorithms.

References

Barber, B.M., & Lyon J.D., (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics* 43, 341-372.

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

Bizjak, J., Lemmon, M., & Nguyen, T. (2011). Are all CEOs above average? An empirical analysis of compensation peer groups and pay design. *Journal of Financial Economics* 100 , 538-555.

Boni, L., & Womack, K. L. (2006). Analysts, industries, and price momentum. *Journal of Financial and Quantitative Analysis* 41, 85-109.

Chan, K. & A. Hameed. (2006). Stock price synchronicity and analyst coverage in emerging markets. *Journal of Financial Economics* 80, 115-147.

Chang, X., Dasgupta, S., & Hilary G. (2006). Analyst coverage and financing decisions. *Journal of Finance* 61, 3009-3048.

Das, S., Guo, R.-J., & Zhang, H. (2006). Analysts' selective coverage and subsequent performance of newly public firms. *Journal of Finance* 61, 1159-1185.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics* 43, 153-193.

Faulkender, M., & Yang, J. (2010). Inside the black box: The role and composition of compensation peer groups. *Journal of Financial Economics* 96, 257-270.

Francis, J., Hanna D.J., & Philbrick, D.R., (1997). Management communications with securities analysts. *Journal of Accounting and Economics* 24, 363-394.

Frankel, R., Johnson, M., & Skinner, D., (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research* 37, 133-150.

Gilson, S. C., Healy, P. M., Noe, C. F., & Palepu, K. G. (2001). Analyst specialization and conglomerate stock breakups. *Journal of Accounting Research* 39, 565-582.

Guenther, D.A., & Rosman, A.J., (1994). Differences between COMPUSTAT and CRSP SIC Codes and related effects on research. *Journal of Accounting and Economics* 18, 115-28.

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

Hong, H., & Kubik, J. D. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *Journal of Finance* 58, 313-351.

Hong, H., Kubik, J. D., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *RAND Journal of Economics* 31, 121-144.

Kaustia, M., & Rantala, V. (2012). Social norms and corporate peer effects. Working paper, Aalto University.

Ljungqvist, A., Malloy C., & Marston F. (2009). Rewriting history. *Journal of Finance* 64, 1935-1960.

Lyon, J.D., Barber, B.M., & Tsai, C. (1999). Improved methods for tests of long-run abnormal stock returns. *Journal of Finance* 54, 165-201.

Michaely, R., & Womack, K. L. (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *Review of Financial Studies* 12, 653-686.

Mikhail, M. B., Walther, B. R., & Willis, R. H. (2004). Do security analysts exhibit persistent differences in stock picking ability? *Journal of Financial Economics* 74, 67-91.

Piotroski, J. D. & D. T. Roulstone. (2004). The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. *Accounting Review* 79, 1119-1151.

Porter, M. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. The Free Press, New York.

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

Stickel, S. E. (1992). Reputation and performance among security analysts. *Journal of Finance* 47, 1811-1836.

Table 1

Descriptive Statistics of the Sample

This table reports annual statistics on sample firms, analysts, and analyst-based peer groups. The annual number of NYSE firms is based on firms that have CRSP share code 10 or 11, and have item PRC at the end of the year. The sample analysts are analysts that have provided at least one estimate for a sample firm during the year, and individual analysts are separated based on Analyst Code item in the IBES Detail database. The analyst-based peer groups are formed using a simulated peer criterion: A firm's peer group in year *t* consists of all firms that are followed by at least the criterion number of same analysts in year *t*. The criterion is calculated as the number of analysts a firm shares with another firm with a probability that is smaller than 1% in a simulation in which the firm's analysts choose the other firms they follow randomly among NYSE firms that have analysts.

Year	Sample NYSE Firms	NYSE Firms with Peers	Analysts per Year	Average Number of Firms Followed by an Analyst	Average Peer Group Size	Median Peer Group Size
1983	1391	908	1298	10.7	12.4	11
1984	1385	848	1290	11.1	12.1	12
1985	1357	913	1308	10.9	13.4	13
1986	1330	985	1317	11.7	13.5	12
1987	1348	908	1319	11.0	14.0	12
1988	1309	885	1241	10.7	13.6	12
1989	1300	895	1209	10.4	14.0	12
1990	1320	849	1907	9.7	12.5	11
1991	1393	860	1726	10.2	13.0	11
1992	1488	902	1625	11.1	13.3	11
1993	1579	1004	1815	11.3	13.9	12
1994	1653	1093	1965	10.6	13.2	12
1995	1715	1108	2065	10.0	13.0	11
1996	1825	1194	2185	9.6	12.3	11
1997	1886	1269	2430	8.9	11.7	10
1998	1856	1278	2600	8.3	10.9	9
1999	1768	1261	2707	8.0	11.2	9
2000	1614	1146	2594	7.6	10.5	9
2001	1530	1046	2585	6.9	9.9	9
2002	1501	1036	2711	6.5	9.5	9
2003	1486	1025	2693	6.8	9.9	9
2004	1487	1050	2597	6.9	9.9	9
2005	1468	1078	2587	7.0	10.1	9
2006	1446	1100	2596	7.0	10.1	9
2007	1405	1117	2568	7.1	10.0	9
2008	1351	1097	2580	7.0	9.9	9
2009	1330	1029	2551	7.0	9.8	9
Average	1500.8	1032.7	2076.6	9.0	11.8	10.4

Table 2

Firms with Analyst-Based Peers Compared to All NYSE Firms Based on Firm Size

This table reports comparative statistics between firms with analyst-based peers and all NYSE firms. The statistics are based on annual observations of firms between 1983 and 2009. The analyst-based peer groups are formed based on a simulated peer criterion. The criterion is calculated as the number of analysts a firm shares with another firm with a probability that is smaller than 1% in a simulation in which the firm's analysts choose the other firms they follow randomly among NYSE firms that have analysts. A firm's peer group in year *t* consists of all firms that are followed by at least the criterion number of same analysts in year *t*. Market capitalization is calculated as price times shares outstanding at the end of the year. Share prices are from CRSP and shares outstanding are from Compustat (Item 25 if available) and otherwise from CRSP. Common Equity is based on Compustat Item 60. Market capitalization and common equity values are in millions of dollars. The market capitalization quartile statistics are calculated using year-specific market capitalization breakpoints in the data.

	Market Capitalization		Common Equity	
	Average	Median	Average	Median
All NYSE Firms	4,377	848	1,761	423
Firms with Analyst-Based Peers	6,174	1,537	2,402	743

Percentage of Firms with Analyst-Based Peers in Different Market Capitalization Quartiles			
Quartile 1	Quartile 2	Quartile 3	Quartile 4
29.24%	59.75%	82.01%	95.37%

Table 3

Analyst-Based Peer Groups Compared to Conventional Industry Classification Groups

This table reports comparative statistics between the analyst-based peer groups, SIC Codes, Fama-French (49) industry groups, NAICS and GICS Codes. The analyst-based peer groups are formed based on a simulated peer criterion. The criterion is calculated as the number of analysts a firm shares with another firm with a probability that is smaller than 1% in a simulation in which the firm's analysts choose the other firms they follow randomly among NYSE firms that have analysts. A firm's peer group in year *t* consists of all firms that are followed by at least the criterion number of same analysts in year *t*. Panel A reports average and median group sizes and standard deviations of group size for the peer groups and the industry classifications. The industry classification statistics are calculated based on the number of other NYSE firms sharing the same classification code in year *t*. Panel B reports statistics on the industry classification codes of analyst-based peer firms. The statistics include the average and median number of different classification codes in the peer groups and the average and median number of peer firms sharing the firm's own classification code. All statistics are based on annual values between 1983 and 2009, with the exception of NAICS codes which are available from 1985 and GICS codes which are available from 1999. The industry classification statistics are based on classification codes at the end of the calendar year.

Panel A: Group Size Statistics Based on Analyst-Based Peer Groups or Other Firms with the Same Industry Code

	Analyst-Based Peer Groups	4-Digit SIC	3-Digit SIC	2-Digit SIC	Fama-French 49 Industries	5-Digit NAICS	3-Digit NAICS	8-Digit GICS	6-Digit GICS
Average Group Size	11.7	11.6	15.8	52.9	54.9	15.5	40.2	11.3	23.5
Median Group Size	10	6	10	42	44	8	31	8	15
Average/Median Group Size	1.17	1.94	1.58	1.26	1.25	1.94	1.30	1.41	1.57
Standard Deviation of Group Size	8.9	13.6	14.9	40.5	36.9	18.1	30.7	10.4	20.6
St. Dev. / Average Group Size	0.76	1.17	0.94	0.77	0.67	1.17	0.76	0.93	0.87

Panel B: Distribution of Industry Codes with Analyst-Based Peer Groups

	4-Digit SIC	3-Digit SIC	2-Digit SIC	Fama-French 49 Industries	5-Digit NAICS	3-Digit NAICS	8-Digit GICS	6-Digit GICS
Average Number of Different Industry Codes in Analyst-Based Peer Group	4.35	3.69	2.49	2.32	3.70	2.48	2.44	1.74
Median Number of Different Industry Codes in Analyst-Based Peer Group	4	3	2	2	3	2	2	1
Average Percentage of Peers Sharing the Firms' Own Industry Code	33.50 %	41.49 %	60.72 %	64.80 %	38.18 %	59.61 %	55.43 %	72.84 %
Median Percentage of Peers Sharing the Firms' Own Industry Code	23.53 %	35.71 %	70.00 %	80.00 %	27.27 %	68.75 %	57.14 %	92.31 %

Table 4

Comparison between Analyst-Based Peer Groups, SIC Codes, and Fama-French Industry Classification Based on a Set of Common Finance Variables

This table compares the explanatory power of analyst-based peer averages to the explanatory power of SIC and Fama-French industry classification-based peer averages in a set of common finance variables. The reported figures are adjusted R^2 values from a regression where the dependent variables is a firm's variable value, and the independent variable is either the average variable value among analyst-based peers or among other firms sharing the same industry classification code. The regression includes a constant. The sample covers years between 1983 and 2009 and includes all NYSE firms with CRSP share codes 10 or 11. Analyst-based peer regressions are also run separately using analyst-weighted peer average values, with the number of common analysts between firms as weights. Industry classification regressions are run separately based on classification groups among all NYSE firms and among NYSE firms with analyst-based peers. The test variables are Monthly Stock Return (based on prices and dividends in CRSP), Monthly Stock Return with Month Fixed Effects (FE), Beta from the single index model based on monthly returns over the previous 36 months, Market Capitalization (price times shares outstanding, prices are from CRSP and shares outstanding are from Compustat (Item 25 if available) and otherwise from CRSP), Total Assets (Compustat Item 6), Net Sales (Item 12), Market-to-Book (book assets (Item 6) minus book equity plus market capitalization all divided by book assets), Dividend Payment (binary variable which takes value one if a firm has non-zero dividends per share by the ex date (Item 26) and value zero otherwise), Book Leverage (book debt to total assets) and Market Leverage (book debt divided by the result of total assets minus book equity plus market equity). Book equity is stockholders' equity (216) (or first available of common equity (60) plus preferred stock par value (130) or book assets (6) minus liabilities (181)) minus preferred stock liquidating value (10) (or first available of redemption value (56) or par value (130)) plus balance sheet deferred taxes and investment tax credit (35) if available and minus post-retirement assets (330) if available. Book debt is defined as total assets minus book equity. All variable values except the monthly returns are annual values at the end of the year. Financial statement items for year *t* are based on the fiscal year ending in year *t*. The highest adjusted R^2 value for each variable is boldface.

	Analyst-Based Peers		Industry Classification Groups Based on All NYSE Firms				Industry Classification Groups Based on Firms with Analyst-Based Peers			
	Peer Average	Analyst-Weighted Average	4-Digit SIC	3-Digit SIC	2-Digit SIC	Fama-French 49 Industries	4-Digit SIC	3-Digit SIC	2-Digit SIC	Fama-French 49 Industries
Monthly Return	26.51	**27.11**	16.15	16.00	18.92	19.39	23.38	24.07	25.27	25.65
Monthly Return with Month FE	28.73	**29.33**	20.32	19.40	19.74	19.89	27.01	26.86	26.15	26.21
Beta (36 Months)	37.28	**38.18**	24.26	24.52	22.05	23.30	31.95	32.06	27.58	29.00
Market Capitalization	13.90	**18.72**	7.59	8.08	7.70	7.85	7.86	8.93	8.09	8.47
Total Assets	24.35	**28.93**	14.97	15.24	16.20	14.89	15.42	15.88	16.64	15.85
Net Sales	13.42	**18.77**	9.54	9.97	8.29	5.57	10.98	11.31	9.27	5.45
Market-to-Book	29.08	**30.44**	18.78	19.21	16.44	19.11	28.96	28.74	21.65	23.94
Dividend Payment	17.09	**17.95**	6.58	8.03	8.81	9.99	11.58	12.93	13.74	14.39
Book Leverage	12.18	**13.06**	7.94	6.81	5.46	5.42	11.53	11.11	7.99	8.32
Market Leverage	22.21	**23.37**	16.47	15.94	12.94	13.49	22.34	21.21	15.34	15.75

Table 5

Comparison between Analyst-Based Peer Groups, NAICS and GICS Codes Based on a Set of Common Finance Variables

This table compares the explanatory power of analyst-based peer averages to the explanatory power of NAICS and GICS codes in a set of common finance variables. Sample firms include all NYSE firms with CRSP share codes 10 or 11. Sample period for comparisons with NAICS codes covers years between 1985 and 2009 and period for comparisons with GICS codes covers years between 1999 and 2009 (statistics for analyst-based peer groups are reported separately for both time periods). The reported figures are adjusted R^2 values from a regression where the dependent variables is a firm's variable value, and the independent variable is either the average variable value among analyst-based peers or among other firms sharing the same industry classification code. The regression includes a constant. The test variables and averages are defined as in Table 4. All variable values except the monthly returns are annual values at the end of the year. Financial statement items for year *t* are based on the fiscal year ending in year *t*. The highest adjusted R^2 value for each variable is boldface.

	Analyst-Based Peers (Period 1985-2009)		NAICS Groups Based on All NYSE Firms		NAICS Groups Based on Firms with Analyst-Based Peers		Analyst-Based Peers (Period 1999-2009)		GICS Groups Based on All NYSE Firms		GICS Groups Based on Firms with Analyst-Based Peers	
	Peer Average	Analyst-Weighted Avg.	5-Digit NAICS	3-Digit NAICS	5-Digit NAICS	3-Digit NAICS	Peer Average	Analyst-Weighted Avg.	8-Digit GICS	6-Digit GICS	8-Digit GICS	6-Digit GICS
Monthly Return	26.64	**27.24**	17.80	19.51	24.48	25.42	26.96	**27.67**	22.67	22.94	26.91	27.47
Monthly Return with Month FE	28.85	**29.45**	21.68	20.99	27.83	26.80	29.12	**29.84**	23.88	23.41	28.40	28.02
Beta (36 Months)	37.45	**38.29**	18.61	16.64	31.78	29.34	40.09	**40.97**	26.29	25.91	37.97	37.94
Market Capitalization	13.23	**18.06**	10.67	7.70	12.42	7.98	12.26	**17.04**	5.24	3.71	4.86	3.36
Total Assets	24.05	**28.65**	14.91	14.57	17.06	16.51	21.72	**26.83**	11.98	10.70	11.47	10.36
Net Sales	13.17	**18.56**	12.48	9.07	14.61	10.03	12.96	**18.56**	9.79	7.02	10.05	7.88
Market-to-Book	28.59	**29.97**	19.88	16.03	28.26	21.16	29.08	**30.44**	17.28	16.72	21.77	20.98
Dividend Payment	16.22	**17.14**	7.64	7.90	12.88	11.81	17.09	**17.95**	9.71	8.84	11.53	10.43
Book Leverage	11.56	**12.42**	7.54	5.75	11.04	8.75	12.18	13.06	10.49	7.61	**13.86**	10.36
Market Leverage	22.28	**23.41**	17.20	14.45	22.24	17.31	22.21	23.37	21.25	16.08	**24.02**	18.40

Table 6

Comparison between Analyst-Based Peer Groups, SIC Codes, and Fama-French Industry Classification Based on Financial Ratios and Accounting-Based Financial Information Variables

This table compares the explanatory power of analyst-based peer averages to the explanatory power of SIC and Fama-French industry classification group-based peer averages in a set of financial ratios and accounting-based financial information variables. The reported figures are adjusted R^2 values from a regression where the dependent variables is a firm's variable value, and the independent variable is either the average variable value among analyst-based peers or among other firms sharing the same industry classification code. The regression includes a constant. The sample covers years between 1983 and 2009 and includes all NYSE firms with CRSP share codes 10 or 11. Analyst-based peer regressions are also run separately using analyst-weighted peer average values withthe number of common analysts between firms as weights. Industry classification regressions are run separately based on classification groups among all NYSE firms and among NYSE firms with analyst-based peers. The test variables are Price-to-Book (market cap divided by total common equity (Computat Item 60), Enterprise Value-to-Sales (the sum of market cap, long-term debt (Item 9), and debt in current liabilities (Item 34) all divided by net sales (Item 12)), Price-to-Earnings (market cap divided by net income before extraordinary items (Item 18), Return on Net Operating Assets (net operating income after depreciation (Item 178) divided by the sum of property, plant, and equipment (Item 8) and current assets (Item 4), less current liabilities (Item 5)), Return on Equity (net income before extraordinary items divided by total common equity), Asset Turnover (total assets (Item 6) divided by net sales), Profit Margin (net operating income after depreciation divided by net sales), Leverage (total liabilities (Item 9) divided by total stockholders' equity (Item 216), Sales Growth (net sales one year in the future divided by current value), and Scaled R&D Expense (research and development expense (Item 46) divided by net sales). All variable values are annual values at the end of the year. Financial statement items for year *t* are based on the fiscal year ending in year *t*. The highest adjusted R^2 value for each variable is boldface.

	Analyst-Based Peers		Industry Classification Groups Based on All NYSE Firms				Industry Classification Groups Based on Firms with Analyst-Based Peers			
	Peer Average	Analyst-Weighted Average	4-Digit SIC	3-Digit SIC	2-Digit SIC	Fama-French 49 Industries	4-Digit SIC	3-Digit SIC	2-Digit SIC	Fama-French 49 Industries
Price-to-Book	19.12	**20.36**	12.86	13.14	11.68	13.59	17.54	18.59	15.34	16.78
Enterprise Value-to-Sales	46.31	47.77	44.45	44.56	38.19	38.43	48.74	**48.88**	41.66	41.35
Price-to-Earnings	6.55	**6.84**	4.37	4.28	4.61	5.20	5.80	5.55	6.01	6.08
Return on Net Operating Assets	28.12	**29.51**	15.56	15.67	17.73	17.52	22.17	21.89	23.81	23.21
Return on Equity	9.13	**9.89**	4.79	5.19	4.78	4.60	6.52	6.53	6.00	5.45
Asset Turnover	79.40	80.27	77.61	76.29	69.53	66.27	**80.38**	79.18	72.62	69.51
Profit Margin	43.18	44.36	38.52	38.31	33.48	33.18	44.51	**44.61**	37.79	36.44
Leverage	9.77	**10.72**	8.81	8.36	6.95	6.16	9.58	9.69	8.10	7.38
Sales Growth	17.72	**18.17**	11.28	11.91	12.37	12.23	14.80	14.87	14.82	14.51
Scaled R&D Expense	56.53	**58.44**	47.90	36.48	18.95	37.82	53.48	40.24	21.65	41.25

Table 7

Comparison between Analyst-Based Peer Groups, NAICS Codes, and GICS codes Based on Financial Ratios and Accounting-Based Financial Information Variables

This table compares the explanatory power of analyst-based peer averages to the explanatory power of NAICS and GICS codes is a set of financial ratios and accounting-based financial information variables. Sample firms include all NYSE firms with CRSP share codes 10 or 11. Sample period for comparisons with NAICS codes covers years between 1985 and 2009 and period for comparisons with GICS codes covers years between 1999 and 2009 (statistics for analyst-based peer groups are reported separately for both time periods). The reported figures are adjusted R^2 values from a regression where the dependent variables is a firm's variable value, and the independent variable is either the average variable value among analyst-based peers or among other firms sharing the same industry classification code. The regression includes a constant. The test variables and averages are defined as in Table 6. All variable values are annual values at the end of the year. Financial statement items for year *t* are based on the fiscal year ending in year *t*. The highest adjusted R^2 value for each variable is boldface.

	Analyst-Based Peers (Period 1985-2009)		NAICS Groups Based on All NYSE Firms		NAICS Groups Based on Firms with Analyst-Based Peers		Analyst-Based Peers (Period 1999-2009)		GICS Groups Based on All NYSE Firms		GICS Groups Based on Firms with Analyst-Based Peers	
	Peer Average	Analyst-Weighted Avg.	5-Digit NAICS	3-Digit NAICS	5-Digit NAICS	3-Digit NAICS	Peer Average	Analyst-Weighted Avg.	8-Digit GICS	6-Digit GICS	8-Digit GICS	6-Digit GICS
Price-to-Book	18.03	**19.29**	13.31	10.76	16.92	13.71	14.01	**15.25**	11.80	10.34	13.52	12.11
Enterprise Value-to-Sales	45.94	47.42	44.75	37.49	**48.68**	41.10	48.74	**50.10**	46.12	40.74	48.88	42.77
Price-to-Earnings	6.14	**6.43**	5.01	4.44	4.58	5.82	6.72	**6.95**	5.05	5.02	5.11	5.46
Return on Net Operating Assets	27.97	**29.35**	17.82	18.72	22.96	23.55	25.92	**27.48**	19.73	16.40	27.43	25.04
Return on Equity	8.99	**9.73**	5.28	5.05	6.66	7.30	9.19	**10.02**	7.50	5.88	8.61	7.39
Asset Turnover	79.06	**79.94**	74.80	65.89	78.13	68.96	75.03	**76.22**	71.36	66.49	73.64	68.30
Profit Margin	43.11	**44.29**	40.04	34.66	43.35	38.68	44.97	**46.22**	40.64	34.40	45.21	39.00
Leverage	9.41	**10.35**	8.50	6.59	8.02	7.55	7.60	**8.62**	8.23	5.67	8.18	5.89
Sales Growth	18.12	**18.57**	12.81	13.47	14.02	16.34	27.21	**27.71**	24.85	24.20	27.32	27.19
Scaled R&D Expense	56.66	**58.61**	39.98	30.30	57.62	32.20	57.12	**59.50**	48.97	48.27	51.02	51.87

Table 8

Effectiveness of Common Analyst Matching in Identifying a Single Comparable Firm

This table compares the explanatory power of common analyst matched individual comparable firms to size and market-to-book matched comparable firms based on a set of common finance variables. The sample covers years between 1983 and 2009 and includes all NYSE firms with CRSP share codes 10 or 11. Common analyst matched comparable firm for fim *i* is the firm with the highest number of common analysts with firm *i*. In case there are several firms with the same number of common analysts, one of them is selected randomly. Statistics for the analyst matched firms are reported separately for different minimum analyst criteria ranging from 1 to 5, so that firms with less than the criterion number of common analysts with any other firm are excluded from the analysis. Common analyst matched comparable firms are compared to comparable firms matched based on the closest market cap or market-to-book value. The matches are based on annual market capitalization and market-to-book values defined as in Table 4. The matching is conducted separately both among all firms and among firms with the same 4-digit SIC code. We also run comparisons with comparable firms selected using joint size and market-to-book matching, so that market-to-book matching is conducted among firms whose market capitalization is between 70% and 130% of the market capitalization of firm *i*. The reported figures are adjusted R^2 values from a regression where the dependent variable is a firm's variable value, and the independent variable is the variable value for the matched comparable firm. The regression includes a constant. All variable values except the monthly returns are annual values at the end of the year. The test variables are defined as in Table 4.

	Comparables Matched Based on the Highest Number of Common Analysts					Market Cap Matching		Market-to-Book Matching		Size and M/B Matching
	Minimum Number of Common Analysts									
	1	2	3	4	5	Among All Firms	Among 4-Digit SIC Group	Among All Firms	Among 4-Digit SIC Group	Among All Firms
Monthly Return	12.94	13.89	15.75	17.36	**18.72**	3.27	9.74	3.36	9.25	4.40
Monthly Return with Month FE	21.59	22.45	24.05	25.36	**26.47**	15.88	18.45	15.85	18.26	16.44
Beta (36 Months)	20.87	22.82	24.89	26.74	**28.38**	0.66	18.00	1.40	17.97	1.31
Market Capitalization	14.16	14.31	14.58	14.53	**14.72**	—	—	0.69	4.79	—
Total Assets	18.85	19.26	19.75	20.57	20.82	13.74	45.05	0.37	7.35	**55.27**
Net Sales	16.37	16.62	17.02	16.90	16.78	31.09	40.16	0.14	5.61	**44.27**
Market-to-Book	17.26	18.45	20.13	21.75	**23.19**	1.17	18.01	—	—	—
Dividend Payment	6.63	6.90	7.84	8.75	**9.17**	3.28	8.57	0.44	4.61	4.94
Book Leverage	5.47	6.19	7.20	8.18	**8.62**	0.03	5.07	1.51	6.36	2.62
Market Leverage	10.06	11.04	12.06	13.32	13.94	2.15	14.47	42.75	28.75	**44.29**