FAJ
Financial Analysts Journal
Volume 63 • Number 6

Industry Classifications and Return Comovement

Louis K.C. Chan, Josef Lakonishok, and Bhaskaran Swaminathan

A company's industry affiliation is commonly used to construct homogeneous stock groupings for portfolio risk management, relative valuation, and peer-group comparisons. A variety of industry classification systems have been adopted, however, creating disagreements as to companies' industry assignments. This analysis of the Global Industry Classification System (GICS) and Fama–French system indicates that common movement in returns and operating performance resulting from industry effects is stronger for stocks of large companies than for those of small companies. Also, increasingly fine levels of disaggregation improve discrimination up to six-digit GICS codes, after which the benefits tail off. Stock groupings based on industry exhibit stronger out-of-sample homogeneity than groups formed from statistical cluster analysis.

Financial researchers, analysts, and decision makers frequently grapple with the issue of identifying homogeneous groups of stocks. Researchers and analysts, for example, often wish to analyze the consequences of such events as corporate reorganizations or changes in financial and investment policies. A common procedure is to pair stocks in the sample with others that have not experienced the event but are similar in all other respects. The behavior of the sample in question is then assessed against the reference group. In presenting financial statements to the public, corporate managers frequently assess how their companies are faring in terms of peer-group comparisons. And many investors follow strategies based on identifying stocks that are matched along economically relevant dimensions but that trade at different valuations, which perhaps indicates relative mispricing.

Academic researchers and investment practitioners follow a variety of approaches to construct homogeneous stock groupings. Purely statistical procedures can be applied to the problem. Early heuristic approaches for partitioning stocks into similar groups were proposed by Farrell (1974) and Elton and Gruber (1970). Brown and Goetzmann (1997) clustered mutual funds into distinct investment style categories. Alternatively, stocks can be assigned to groups on the basis of *a priori* economic attributes, such as market capitalization or operating performance.

Perhaps the most popular method of establishing sets of economically similar stocks is to follow their industry affiliations. Numerous academic studies provide evidence that industry influences capture a large portion of the extra-market correlations in stock returns.[1] Connor (1995), for instance, found that industry factors dominate fundamental stock attributes in terms of ability to fit the in-sample behavior of returns. In practice, research coverage by financial analysts is typically structured along industry classifications, as is the analysis carried out by such popular publications as the *Value Line Investment Survey*. Many quantitative risk models widely used by asset managers and consultants are based on industry factors. Most notably, when investment managers structure portfolios, they typically take into consideration the industry affiliations of the component stocks. A portfolio that is substantially overweight in an industry relative to the benchmark may experience increased tracking-error volatility.

Nonetheless, deciding which companies belong to an industry is far from straightforward. Standard Industrial Classification (SIC) codes, which aggregate companies selling related end-products or using similar production processes, have traditionally been used for this purpose. Changes in the variety of products, the growing importance of services, and shifts in technology and the makeup of businesses, however, have

Louis K.C. Chan is Hoeft Professor of Finance and Josef Lakonishok is professor of finance at the University of Illinois, Urbana-Champaign. Bhaskaran Swaminathan is partner and director of research at LSV Asset Management, Chicago, and professor of finance at Cornell University, Ithaca, New York.

called into question the usefulness of the SIC system (see Clarke 1989). The Fama and French (1997) system starts from companies' four-digit SIC codes and reorganizes them into 48 industry groupings. For instance, the Fama–French (FF) "automobiles and trucks" industry aggregates some companies from SIC major (two-digit) groups—namely, 22, 23, 30, and 35—with some from the standard "transportation equipment" SIC group, 37. The FF classification has been highly influential and is widely used in academic studies of asset pricing (e.g., Brennan, Wang, and Xia 2004; Daniel and Titman 2006; Ferson and Harvey 1999; Hong, Torous, and Valkanov 2007; Moskowitz and Grinblatt 1999; Pastor and Stambaugh 1999; Purnanandam and Swaminathan 2005), corporate finance (e.g., Flannery and Rangan 2006; Graham and Kumar 2006), accounting (e.g., Chan, Frankel, and Kothari 2004; Francis, LaFond, Olsson, and Schipper 2005; Richardson 2006), and economics (e.g., Bebchuk and Grinstein 2005; Wulf 2002).

Fama and French (1997) did not provide any evidence about how well their classification system produces groups of economically similar companies. It is still an open issue, therefore, as to whether the FF specification of industries deserves its status as the default choice in academic studies.

Another approach to industry classification enjoys widespread use among investment practitioners. Portfolio managers and analysts have gravitated to the Global Industry Classification System (GICS). Its categorization is based not only on a company's operational characteristics but also on information about investors' perceptions of what constitutes the company's main line of business.

Because GICS codes consider investors' attitudes, the SIC and GICS schemes can disagree about a company's industry assignment. For example, GATX Corporation, which leases and operates railroad equipment and ships, is classified in the financial sector by GICS but is placed in the transportation equipment sector by SIC codes. Other instances of conflict arise in the case of several companies that trade as real estate investment trusts (hence, GICS puts them in the financial sector) but also manage timberlands or paper mills (so their SIC codes are those for the wood or paper product industries).

The GICS approach has received relatively less coverage in the research literature, and its validity has not been extensively documented (a notable exception is Bhojraj, Lee, and Oler 2003). Our aim is to resolve the disparity between academic and practitioner approaches to industry classification by comparing the FF industry grouping system, which is based initially on SIC code, with the GICS groupings. We evaluate each method's ability to form groups of stocks whose price movements are related.

The existing literature also provides scant guidance on choosing between the different classification levels available under the GICS standard. On the one hand, a coarse partitioning level (such as two-digit GICS codes) blurs commonalities that may emerge more strongly in subgroups. On the other hand, finer partitions are not necessarily preferable: Past a point, the increasingly specific factors that account for homogeneity may become less important than the general factors affecting the majority of companies in a broadly defined industry. Also, moving to a finer detail of classification produces groupings with fewer companies, making comparisons noisy and unreliable. The relevant question is how much is lost at each level of aggregation. A second objective of this article, therefore, is to provide evidence to aid in such a choice: Specifically, we document how each GICS aggregation level performs with respect to highlighting within-group similarity in stock price movements.

Our analysis measures homogeneity in terms of coincidence in stock price movements. Two companies that are economically similar may not experience strong return covariation, however, over short horizons. Their comovement may be drowned out by the idiosyncratic portion of returns, including noneconomic forces, such as investor sentiment. As a check that our conclusions about industry classification are not limited by our focus on return correlations, we verify that the FF and GICS specifications also capture commonalities in underlying operating performance as proxied by sales growth. Finally, we use statistical cluster analysis to form stock categories with high within-group return correlations. Because this procedure overlooks industry affiliation entirely, its performance provides a yardstick for the usefulness of the FF and GICS grouping methods.

In addition to analyzing the GICS approach, Bhojraj et al. (2003) considered various methods for assigning companies to industries, including the FF, SIC, and North American Industry Classification System codes. They focused on the ability of industry indices to capture the cross-sectional dispersion in stock returns, valuation multiples, financial ratios, and growth rates. Unlike us, they did not examine how correlations in returns and in operating performance vary depending on whether companies belong to the same or different industries. As a result, their findings are not directly applicable to issues in portfolio analysis and risk management. Moreover, they did not confront industry-based grouping procedures with alternative classifications that take no account of companies' industry membership, nor did they examine different levels of detail in the GICS code.

Methodology

We first describe our procedures for evaluating the degree to which each classification scheme isolates groups of companies that belong to an industry from companies that do not belong to that industry.

Industry Classification Schemes. Many studies follow the SIC scheme to partition companies into industry groups. The scheme was established to categorize all industries in the U.S. economy and is currently administered by the U.S. Office of Management and the Budget. Starting from 11 categories, successively finer partitions are defined in terms of major groups (corresponding to the first two digits of the code), industry groups (three-digit codes), and industries (four-digit codes). The SIC scheme aggregates into an industry companies that use similar production processes or whose products tend to be used or distributed together (see Economic Classification Policy Committee 1994). Accordingly, the groupings are intended to aid economic and marketing analysis and do not directly address the concerns of investors.[2]

In contrast, the GICS scheme is expressly designed to cater to financial analysts and investment managers. This classification, maintained by Morgan Stanley Capital International together with Standard & Poor's, identifies a company's principal business activity by considering the sources of its revenues, its earnings, and the market perception of its business. The scheme comprises 10 sectors (corresponding to the leading two digits of the code), which are broken down into 24 industry groups (denoted by the first four digits of the code), 64 industries (the first six digits of the code), and then 139 subindustries (the full eight-digit code).

Fama and French (1997) drew up their own set of homogeneous stock groups. Starting from four-digit SIC codes, they categorize companies into 48 industry sectors.[3]

Comparing the Performance of Industry Classification Systems. If equity market participants consider a set of companies closely related, then stocks in the group should experience coincident movements in their stock returns. The comovement in their returns with stocks outside the group should be relatively weaker. Accordingly, we judge each classification system's ability to produce homogeneous groupings by comparing the magnitude of return correlations between stocks in the same industry with the magnitude of correlations between within-industry stocks and outside-industry stocks.[4]

The correlations were calculated as follows. Let K be the number of stocks in the sample. We applied a particular industry classification system (such as four-digit GICS codes) to each stock $i = 1, \ldots, K$ to identify its industry, I. Suppose stock i's industry contains N stocks (inclusive). We averaged the pairwise correlations between stock i's return and the return on each of the other members of its industry:[5]

$$\rho_{iI} = \frac{\sum_{j \in I, j \neq i} \rho_{ij}}{N-1}, \tag{1}$$

where ρ_{ij} is the time-series correlation between the return on stocks i and j. Similarly, the average pairwise correlation between stock i's return and the returns of all other stocks not in its industry is

$$\phi_{iI} = \frac{\sum_{j \notin I} \rho_{ij}}{K-N}. \tag{2}$$

We then defined the average within-industry correlation over all stocks in the sample as

$$\bar{\rho}_I = \frac{\sum_{i=1}^{K} \rho_{iI}}{K} \tag{3}$$

and the average correlation between a stock and other stocks not in its industry as

$$\bar{\phi}_I = \frac{\sum_{i=1}^{K} \phi_{iI}}{K}. \tag{4}$$

By comparing the values of $\bar{\rho}_I$ and $\bar{\phi}_I$, we can assess the degree to which an industry classification distinguishes between similar and dissimilar stocks. Classification schemes that highlight groups with strong commonalities will tend to produce large positive differences between the within-industry and outside-industry correlations.

Data. To calculate correlations, we used return data for all U.S.-listed domestic common equity issues available in the University of Chicago CRSP database for the 1975–2004 sample period. Because the correlation structure of individual stock returns may not be stationary over time, we divided the period into nonoverlapping five-year blocks and report results averaged over the six epochs. For each subperiod, we examined stocks that existed for the full five years. To mitigate measurement error problems in correlations arising from infrequent trading and microstructure-related biases resulting from bid–ask bounce, we required that a stock trade at above \$2 per share at the beginning of the subperiod.

For each subperiod, based on the five years of monthly return data, we calculated the within-industry correlation, $\bar{\rho}_I$, and the outside-industry correlation, $\bar{\phi}_I$. Companies were assigned to an industry under the following classification schemes, based on increasingly finer industry groupings: two-digit GICS code, four-digit GICS code, six-digit GICS code, and eight-digit GICS code. In addition, we report results using the FF industry groups. To improve the reliability of the results, we required that any industry classification consist of at least three companies over the subperiod; any industries not meeting this criterion were dropped from the analysis.

Results

We present in this section the results of the procedures for evaluating classification schemes for large-capitalization and small-capitalization stocks, and to take the perspective of a delegated investment manager who is concerned with tracking error, we also provide evidence at the level of returns net of a market index.

Correlations for Large Companies. Some lines of business tend to be more narrowly defined and uniform, so the contrasts between within-industry and outside-industry correlations should be more pronounced for companies engaged in such activity. Accordingly, in presenting the results, we break out average correlations for subsets of companies. To streamline the formatting across all the tables, we organize the findings for subsets in terms of two-digit GICS sectors.[6] Commonality in return movements is likely to show up more strongly for large companies because of their relatively stable behavior (see Chan, Karceski, and Lakonishok 1999). Also, their return correlations are less likely to be contaminated by infrequent or nonsynchronous trading issues. Finally, because the bulk of equity assets is concentrated in large-cap stocks, the issue of industry diversification in this market segment is of particular concern to investors. Accordingly, we begin our analysis by focusing in **Table 1** on large companies, those that fall in the top six deciles when stocks are ranked by market capitalization based on NYSE breakpoints.

When industries are defined on the basis of two-digit GICS codes, companies in the same industry share correlations averaging between 0.31 (for consumer staples, Sector 30) and 0.48 (for energy, Sector 10). Correlations tend to be high also inside the utilities group (Sector 55) and financials group (Sector 40), which suggests that these lines of business involve activities that are fairly uniform. The mean within-industry correlation is 0.39 when weighted by the number of companies in each industry or 0.38 when each industry is equally weighted.

Correlations between companies in the same industry are uniformly higher than correlations between those companies and companies outside the industry. Evidently, all the classification schemes succeed in identifying clusters of homogeneous large-cap stocks. For example, in the eight-digit GICS classification, the simple mean within-industry correlation is 0.44 whereas the corresponding mean outside-industry correlation is 0.27, yielding an average difference of about 0.18 across the sectors. A test that there is no difference between correlations inside and outside an industry under the eight-digit classification based on the 10 sectors in Table 1 produced a t-statistic of 8.32.

In the FF classification system, the simple mean within-industry correlation is 0.40, compared with an outside-industry correlation of 0.26. The difference of 0.14 is of the same magnitude as that based on a four-digit GICS classification. The FF procedure achieves this level of discrimination by using 48 categories, however, or twice as many as the number of groups at the four-digit GICS level. Assigning companies to industries on the basis of six-digit GICS codes is more effective, in that it generates a higher average within-industry correlation (0.43) and a starker contrast (0.17) between the inside- and outside-industry correlations (versus 0.13 for the FF system). Put differently, based on the FF definitions, about 16 percent (the square of the within-industry correlation) of a stock's return volatility, on average, is accounted for by the return on an industry peer. In six-digit GICS groups, the percentage accounted for is 18.5 percent, representing a relative improvement in explanatory power of 16 percent.

Up to a point, finer industry partitions generally sharpen the differences between within-industry and outside-industry correlations. Moving from two-digit GICS codes to four-digit GICS codes has a meager effect on the difference (the simple mean differences are 0.13 and 0.14, respectively), but the mean difference rises to 0.17 under a six-digit GICS classification and 0.18 under an eight-digit GICS classification. Notably, for the more homogeneous sectors (energy, financials, and utilities), the partitioning based on six-digit GICS codes is generally associated with a jump in the differences with respect to within- and outside-industry correlations. For example, in the case of companies within the utilities sector, the spread between the two correlations jumps from 0.26 to 0.28.

Table 1. Average Pairwise Correlations between Individual Stocks' Returns and Returns of Within-Industry or Outside-Industry Stocks: Large-Cap Stocks (size deciles 5–10), 1975–2004

Description	Average Correlation															
	Within Industry					Outside Industry					Difference					Average No.
(GICS sector)	GICS8	GICS6	GICS4	GICS2	FF48	GICS8	GICS6	GICS4	GICS2	FF48	GICS8	GICS6	GICS4	GICS2	FF48	of Companies
Energy (10)	0.55	0.50	0.48	0.48	0.45	0.23	0.22	0.22	0.21	0.22	0.32	0.29	0.26	0.26	0.23	35
Materials (15)	0.45	0.43	0.39	0.39	0.41	0.29	0.29	0.29	0.28	0.29	0.16	0.14	0.10	0.10	0.12	60
Industrials (20)	0.43	0.42	0.38	0.37	0.38	0.30	0.30	0.30	0.30	0.30	0.13	0.12	0.08	0.07	0.08	96
Consumer discretionary (25)	0.44	0.41	0.39	0.35	0.39	0.29	0.29	0.29	0.29	0.29	0.14	0.12	0.10	0.06	0.09	92
Consumer staples (30)	0.36	0.36	0.34	0.31	0.34	0.25	0.25	0.25	0.25	0.25	0.11	0.10	0.09	0.06	0.08	47
Health care (35)	0.38	0.37	0.34	0.32	0.36	0.24	0.24	0.24	0.24	0.24	0.14	0.13	0.10	0.08	0.12	42
Financials (40)	0.48	0.46	0.46	0.42	0.45	0.30	0.30	0.30	0.29	0.30	0.17	0.16	0.16	0.13	0.16	101
Information technology (45)	0.46	0.46	0.44	0.41	0.41	0.28	0.28	0.28	0.27	0.28	0.19	0.18	0.16	0.14	0.14	48
Telecommunication services (50)	0.40	0.40	0.37	0.37	0.33	0.25	0.25	0.26	0.25	0.25	0.14	0.14	0.11	0.11	0.08	13
Utilities (55)	0.50	0.50	0.45	0.45	0.44	0.22	0.22	0.19	0.19	0.19	0.28	0.28	0.26	0.26	0.25	86
Weighted mean	0.45	0.43	0.41	0.39	0.41	0.27	0.27	0.27	0.26	0.27	0.18	0.16	0.14	0.13	0.14	
Simple mean	0.44	0.43	0.40	0.38	0.40	0.27	0.26	0.26	0.26	0.26	0.18	0.17	0.14	0.13	0.13	
t-Statistic											8.32	8.10	6.62	5.35	7.00	

Notes: In each nonoverlapping five-year block of the period, each company in a sector was assigned to an industry on the basis of one of five classification systems. The average correlation between the company's monthly stock return and the return of every other member of its industry was calculated for the five-year period and averaged across all companies in the sector. The average correlation between the stock's monthly return and the return of every stock outside its industry was similarly calculated and averaged across members of the sector. The difference between the two average correlations was calculated. The results were averaged over the six five-year periods. Sectors were defined by two-digit GICS codes. Companies were assigned to an industry based on their eight-digit GICS code, six-digit GICS code, four-digit GICS code, two-digit GICS code, or FF 48 industry groups. "Weighted mean" refers to the average (weighted by average number of companies) over sectors; the "Simple mean" is the unweighted average over sectors. The *t*-statistic tested that the correlation within an industry and the correlation outside an industry are equal, and it is based on the sample of mean differences from the 10 sectors.

Table 2 reports the mean correlations and the differences for the subperiods. The individual subperiod results confirm the findings from Table 1. In every subperiod, the FF and four-digit GICS schemes are roughly equivalent. Singling out the turbulent period of 2000–2004, for instance, we find the simple mean within-industry correlations to be 0.34 for FF and 0.36 for the four-digit GICS codes, representing differences of 0.17 and 0.19, respectively, from the outside-industry correlations. An improvement occurs by going to the six-digit GICS codes, but little is gained from further detail. For the 2000–04 period, the within-industry correlation for six-digit GICS codes is 0.41 (a gain of 0.23 over the outside-industry correlation) and for eight-digit GICS codes, 0.43 (a difference of 0.25 over the outside-industry correlation).

Consistent with the findings of Campbell, Lettau, Malkiel, and Xu (2001), Table 2 provides some evidence that the average stock's correlation with other stocks has declined over time. For the 1975–79 epoch, for example, the average outside-industry correlation under a two-digit GICS classification is 0.36. For the 2000–04 epoch, the mean outside-industry correlation drops to 0.17. Fama and French (2004) documented that the jump after 1980 in the number of newly listed stocks, which tend to exhibit lower profitability and higher growth, on average, than more established stocks, is a contributing factor to the increase in idiosyncratic volatility.

Because the correlations reported in Tables 1 and 2 are based on raw returns, they are directly relevant to investors whose concern is how a portfolio performs in terms of gross returns. Professional investment managers are held to a benchmark, however, such as the S&P 500 Index. Consequently, they are focused on structuring a portfolio's return and risk versus the baseline index. From this standpoint, if stocks in an industry covary strongly in terms of their deviations from the benchmark, then any overweighting of that industry will tend to raise the manager's exposure to the risk of straying from the reference portfolio (tracking error). Our approach is easily modified to conform to a typical investment manager's perspective. Specifically, we measured returns in excess of the return on a general market index and calculated correlations based on monthly excess returns. We chose the equally weighted index of stocks in the top six deciles by market capitalization based on NYSE breakpoints so as not to attach more importance to some companies than others in the benchmark. **Table 3** reports how well each classification system captures within-industry commonality in return movements net of the index.

Deducting the market index from each stock's return partially removes a prime source of pervasive return movement. As a result, the average correlations of excess returns in Table 3 tend to be lower than the raw return correlations in Table 1. For instance, the simple mean correlation of raw returns is 0.38 across sectors in the two-digit GICS classification; the simple mean correlation is 0.17 for excess returns.

After the effect of the market is removed, the covariation resulting from industry is more apparent and the consequences of choosing different classification methods show up more clearly. In these circumstances, the GICS classification displays a bigger advantage over the FF procedure. Correlations of excess returns under four-digit GICS codes are slightly higher than those in the FF grouping method. Larger improvements in excess return correlations occur when six-digit and eight-digit GICS classifications are used. The simple mean correlations move from 0.18 based on FF industries to 0.20 based on four-digit codes, and they rise to 0.24 for six-digit codes. In terms of the proportion of an average stock's extra-market return variance explained by an industry peer's excess return, FF industries yield 3.24 percent whereas six-digit GICS codes yield 5.26 percent, a relative improvement of about 78 percent.

To summarize, industry groups based on four-digit GICS codes do about as well as the FF industry aggregations in terms of capturing the commonality in the returns of large-cap stocks. At the same time, the four-digit GICS scheme is more parsimonious with respect to the number of classifications (24 for four-digit GICS codes compared with 48 for FF). If a finer breakdown is the objective, six- or eight-digit GICS categories are roughly equivalent. They deliver similar levels of improvement over FF groupings on the basis of higher correlations between industry peers and larger differentiation between the inside-industry and outside-industry correlations.

Correlations for Small Companies. **Table 4** reports correlations between an average small-cap stock and other small-cap stocks inside and outside its industry and also provides the corresponding differences for the five industrial assignment criteria. (Small-cap stocks are defined as those falling below the median market capitalization of NYSE issues.) In the case of small-cap stocks, Table 4 shows that the comovement in returns associated with common industry affiliation is weak. Even at the finest partitioning level of eight-digit GICS codes, the within-industry correlations for small-cap stocks average only 0.23, as opposed to 0.44 for

Table 2. Average Pairwise Correlations between Individual Stocks' Returns and Returns of Within-Industry or Outside-Industry Stocks: Large-Cap Stocks (size deciles 5–10) by Subperiod, 1975–2004

	Average Correlation														
	Within Industry					Outside Industry					Difference				
Statistic	GICS8	GICS6	GICS4	GICS2	FF48	GICS8	GICS6	GICS4	GICS2	FF48	GICS8	GICS6	GICS4	GICS2	FF48
1975–1979 (average 60 companies in sector)															
Weighted mean	0.49	0.48	0.46	0.45	0.46	0.36	0.36	0.36	0.36	0.36	0.13	0.12	0.10	0.09	0.10
Simple mean	0.49	0.48	0.46	0.45	0.46	0.36	0.36	0.36	0.36	0.36	0.13	0.12	0.11	0.10	0.10
t-Statistic											8.60	7.41	6.15	4.77	6.18
1980–1984 (average 56 companies in sector)															
Weighted mean	0.44	0.43	0.39	0.38	0.40	0.26	0.26	0.25	0.25	0.26	0.18	0.18	0.14	0.13	0.15
Simple mean	0.44	0.43	0.40	0.38	0.39	0.26	0.26	0.25	0.25	0.26	0.18	0.17	0.15	0.13	0.14
t-Statistic											6.87	6.32	5.79	4.98	7.21
1985–1989 (average 56 companies in sector)															
Weighted mean	0.52	0.51	0.49	0.48	0.50	0.38	0.38	0.37	0.37	0.37	0.15	0.14	0.13	0.11	0.13
Simple mean	0.51	0.50	0.48	0.47	0.49	0.37	0.37	0.37	0.36	0.37	0.13	0.13	0.12	0.11	0.12
t-Statistic											6.45	6.27	5.71	4.98	6.13
1990–1994 (average 66 companies in sector)															
Weighted mean	0.41	0.40	0.37	0.35	0.38	0.24	0.24	0.24	0.23	0.24	0.17	0.16	0.13	0.12	0.14
Simple mean	0.41	0.39	0.37	0.35	0.37	0.24	0.24	0.23	0.23	0.23	0.17	0.16	0.13	0.12	0.14
t-Statistic											6.71	7.02	5.16	4.60	6.13
1995–1999 (average 67 companies in sector)															
Weighted mean	0.39	0.38	0.35	0.32	0.34	0.20	0.19	0.19	0.19	0.19	0.20	0.18	0.16	0.14	0.15
Simple mean	0.39	0.38	0.34	0.32	0.33	0.19	0.19	0.18	0.18	0.18	0.20	0.19	0.16	0.14	0.15
t-Statistic											6.24	5.91	4.86	3.81	4.72
2000–2004 (average 68 companies in sector)															
Weighted mean	0.42	0.40	0.36	0.32	0.34	0.18	0.18	0.18	0.17	0.18	0.24	0.22	0.19	0.15	0.16
Simple mean	0.43	0.41	0.36	0.34	0.34	0.18	0.18	0.17	0.17	0.18	0.25	0.23	0.19	0.17	0.17
t-Statistic											9.55	9.80	7.86	6.01	8.43

Note: See notes to Table 1.

Table 3. Average Pairwise Correlations between Individual Stocks' Excess Returns and Excess Returns of Within-Industry Stocks: Large-Cap Stocks (size deciles 5–10), 1975–2004

Description (GICS sector)	Average Correlation within Industry					Average No. of Companies
	GICS8	GICS6	GICS4	GICS2	FF48	
Energy (10)	0.44	0.40	0.35	0.35	0.32	35
Materials (15)	0.19	0.16	0.11	0.11	0.14	60
Industrials (20)	0.14	0.13	0.07	0.05	0.11	96
Consumer discretionary (25)	0.17	0.13	0.11	0.05	0.11	92
Consumer staples (30)	0.19	0.18	0.17	0.12	0.15	47
Health care (35)	0.21	0.20	0.16	0.12	0.18	42
Financials (40)	0.22	0.20	0.19	0.13	0.19	101
Information technology (45)	0.23	0.22	0.20	0.16	0.15	48
Telecommunication services (50)	0.28	0.28	0.24	0.24	0.13	13
Utilities (55)	0.47	0.47	0.41	0.41	0.39	86
Weighted mean	0.24	0.22	0.19	0.16	0.18	
Simple mean	0.25	0.24	0.20	0.17	0.18	

Notes: In each nonoverlapping five-year block of the period, each company in a sector was assigned to an industry on the basis of one of five classification systems. The average correlation between the company's monthly excess stock return and the excess return of every other member of its industry was calculated for the five-year period and averaged across all companies in the sector. The difference between the two average correlations was calculated. The results were averaged over the six five-year periods. Excess returns are measured as the stock's return net of the return on the equally weighted market index. For definitions and industry assignment, see Table 1.

large-cap stocks (Table 1). Even for a fairly homogeneous set of stocks, such as utilities, the within-industry correlation across small-cap stocks is low (0.25) compared with the correlation for large utility companies (0.50). Perhaps more telling is that the average correlation between two small-cap stocks belonging to the same industry is not as strong as the average correlation between two large-cap stocks in different industries: The simple mean correlations in an eight-digit GICS classification are 0.23 for within-industry small-cap stocks but 0.27 for outside-industry large-cap stocks.

The low correlations suggest that the idiosyncratic component of returns is the main source of variation in small-cap returns. Another possibility is that small-cap stocks may not be as efficiently priced as large-cap stocks. Investors may thus miss any underlying economic commonality in small-cap stocks, which dilutes the correlation in the stocks' returns. In any event, altering the fineness of the industry partitions yields negligible changes in the within-industry correlations. The difference between the within- and outside-industry correlations in simple means is 0.04 under a four-digit GICS classification and 0.05 under six-digit or eight-digit GICS classifications.

Average within-industry correlations based on excess returns for the small-cap sample are reported in **Table 5**. Returns are measured in excess of the return on the equally-weighted portfolio of all stocks with market capitalizations less than the median NYSE breakpoint. Here as well as in Table 4, successively finer partitions by GICS codes do not produce large differences in correlations for stocks within an industry. The simple mean correlation varies from 0.08 for two-digit GICS codes to 0.09 for eight-digit codes. FF industry groupings yield roughly the same magnitude of excess return correlations (0.07) for the small-cap sample.

The basic message from the small-cap sample is that, compared with the large-cap sample, the covariation in both raw and excess returns is much attenuated. Given the limited degree of comovement, the various levels of industry disaggregation all do about the same with respect to capturing the within-group correlations.

Industry Groupings and Economic Relatedness

This section offers additional evidence to confirm that the GICS and FF classification systems yield groups of economically related stocks. In particular, we check whether a mechanical grouping procedure that ignores industry affiliation can do better in terms of capturing the commonality in returns. We also show that the commonality in return movements within GICS-based and FF-based groups mirrors commonality in their operating performance.

Table 4. Average Pairwise Correlations between Individual Stocks' Returns and Returns of Within-Industry or Outside-Industry Stocks: Small-Cap Stocks (size deciles 1–5), 1975–2004

	Average Correlation															
	Within Industry					Outside Industry					Difference					
Description (GICS sector)	GICS8	GICS6	GICS4	GICS2	FF48	GICS8	GICS6	GICS4	GICS2	FF48	GICS8	GICS6	GICS4	GICS2	FF48	Average No. of Companies
Energy (10)	0.29	0.28	0.27	0.27	0.25	0.15	0.15	0.15	0.15	0.15	0.14	0.13	0.12	0.12	0.10	43
Materials (15)	0.22	0.21	0.20	0.20	0.21	0.19	0.19	0.19	0.19	0.19	0.03	0.03	0.02	0.02	0.02	78
Industrials (20)	0.21	0.21	0.21	0.20	0.21	0.19	0.19	0.19	0.19	0.19	0.03	0.02	0.02	0.01	0.02	273
Consumer discretionary (25)	0.23	0.21	0.21	0.20	0.21	0.19	0.19	0.19	0.19	0.19	0.04	0.03	0.02	0.02	0.03	271
Consumer staples (30)	0.18	0.18	0.17	0.17	0.18	0.17	0.17	0.17	0.17	0.17	0.01	0.01	0.00	0.00	0.01	76
Health care (35)	0.21	0.21	0.20	0.20	0.20	0.18	0.18	0.18	0.18	0.18	0.03	0.03	0.03	0.02	0.03	92
Financials (40)	0.24	0.24	0.23	0.22	0.23	0.18	0.17	0.17	0.17	0.17	0.06	0.06	0.06	0.05	0.06	248
Information technology (45)	0.25	0.25	0.24	0.24	0.23	0.19	0.19	0.20	0.20	0.19	0.06	0.05	0.05	0.04	0.04	163
Telecommunication services (50)	0.22	0.22	0.22	0.22	0.21	0.18	0.18	0.18	0.18	0.18	0.04	0.04	0.03	0.03	0.03	6
Utilities (55)	0.25	0.25	0.23	0.23	0.22	0.14	0.14	0.14	0.14	0.14	0.11	0.11	0.10	0.10	0.09	55
Weighted mean	0.23	0.22	0.22	0.21	0.21	0.18	0.18	0.18	0.18	0.18	0.04	0.04	0.04	0.03	0.04	
Simple mean	0.23	0.22	0.22	0.21	0.21	0.18	0.18	0.17	0.17	0.17	0.05	0.05	0.04	0.04	0.04	
t-Statistic											4.02	3.90	3.69	3.32	4.16	

Note: See notes to Table 1.

Table 5. Average Pairwise Correlations between Individual Stocks' Excess Returns and Excess Returns of Within-Industry Stocks: Small-Cap Stocks (size deciles 1–5), 1975–2004

Description (GICS sector)	Average Correlation within Industry					Average No. of Companies
	GICS8	GICS6	GICS4	GICS2	FF48	
Energy (10)	0.19	0.18	0.16	0.16	0.14	43
Materials (15)	0.03	0.03	0.02	0.02	0.03	78
Industrials (20)	0.02	0.02	0.01	0.01	0.01	273
Consumer discretionary (25)	0.04	0.03	0.02	0.01	0.03	271
Consumer staples (30)	0.03	0.03	0.03	0.02	0.02	76
Health care (35)	0.04	0.03	0.03	0.03	0.03	92
Financials (40)	0.13	0.12	0.12	0.10	0.12	248
Information technology (45)	0.06	0.06	0.05	0.04	0.03	163
Telecommunication services (50)	0.06	0.06	0.06	0.06	0.04	6
Utilities (55)	0.34	0.34	0.33	0.33	0.29	55
Weighted mean	0.07	0.07	0.06	0.05	0.06	
Simple mean	0.09	0.09	0.08	0.08	0.07	

Note: See notes to Table 3.

Comparisons with Statistical Clusters. Economic intuition supports the idea that companies in the same industry share higher return correlations than companies in different industries. Criteria other than industry affiliation can be used, however, to categorize stocks into homogeneous groups. For instance, classifications can be based on such company attributes as market capitalization or valuation ratios. Whether a partition based on industry yields strong within-group correlations relative to competing classifications is an open question. To give some assurance that it does, we checked that the within-industry comovement of stock returns is at least as strong as the results from a mechanically based classification scheme based on statistical cluster analysis.

For this comparison, we divided the 1975–2004 sample period into six nonoverlapping five-year periods. For each period, we used hierarchical cluster analysis to assign stocks to groups so as to minimize the average within-group distance between group members. Distance was measured as 1.00 minus the correlation coefficient between the two stocks' returns. The number of clusters was set to match roughly the number of industries under the FF classification. Relative to the pseudo-industry classification produced by cluster analysis, GICS industry groupings can be determined *a priori* without requiring a separate estimation period. To put the comparison on an even footing, therefore, we calculated correlations on an out-of-sample basis by using returns for the following five years. This approach let us assess whether GICS groupings sharpen the contrast between the within-group and outside-group correlations relative to the results from cluster analysis. We repeated the process for each nonoverlapping five-year block. Results averaged over all the periods appear in **Table 6**.

By design, the clusters have high in-sample and within-group correlations of returns relative to the outside-group correlations. For the sample of large-cap stocks (Panel A), over the "Estimation Period," the average within-cluster correlation is 0.51 and the average outside-cluster correlation is 0.29, for a difference of 0.22. In the five-year "Testing Period" for the large-cap sample, within-cluster correlations fall sharply, to 0.39 on average, and exceed the outside-cluster correlations by only 0.13. The magnitude of the within-group correlation, as well as the level of discrimination relative to the outside-cluster correlation, is on a par with that provided by the two-digit GICS groupings in Table 1. The statistical procedure requires 40 categories, however, to obtain this level of differentiation (versus 10 sectors when the two-digit GICS code is used).

For the small-company sample (Panel B in Table 6), a spread of 0.15 is observable in the estimation period between within-cluster and outside-cluster correlations. On an out-of-sample basis, however, statistical clusters perform worse than GICS codes. The average within-cluster correlation is 0.18, which is only marginally larger than the average outside-cluster correlation of 0.16. Because the variability in small-cap returns is predominantly idiosyncratic in nature, the assignment of stocks to statistical clusters is likely to be driven by the correlations of company-specific returns. The correlation structure of idiosyncratic returns is unlikely to be

Table 6. Average Pairwise Correlations between Individual Stocks' Returns and Returns of Stocks within and outside Statistical Clusters, 1975–2004

	Average Correlation						
	Estimation Period			Testing Period			
Sample Period	Within Cluster	Outside Cluster	Difference	Within Cluster	Outside Cluster	Difference	Average No. of Companies
A. Large-cap stocks							
1975–1984	0.57	0.37	0.20	0.39	0.26	0.13	536
1980–1989	0.50	0.26	0.24	0.52	0.39	0.13	468
1985–1994	0.58	0.38	0.20	0.37	0.25	0.12	538
1990–1999	0.47	0.25	0.22	0.32	0.19	0.13	509
1995–2004	0.44	0.19	0.25	0.34	0.21	0.13	511
Average	0.51	0.29	0.22	0.39	0.26	0.13	512
B. Small-cap stocks							
1975–1984	0.45	0.30	0.15	0.24	0.22	0.02	611
1980–1989	0.35	0.20	0.15	0.26	0.25	0.01	682
1985–1994	0.36	0.23	0.13	0.14	0.12	0.02	814
1990–1999	0.26	0.11	0.15	0.13	0.12	0.01	748
1995–2004	0.27	0.12	0.15	0.12	0.10	0.02	771
Average	0.34	0.19	0.15	0.18	0.16	0.02	725

Notes: In each nonoverlapping five-year block over the sample period, hierarchical cluster analysis was used to classify stocks on the basis of their monthly returns into one of 40 groups. Clusters were formed to maximize within-group Pearson correlations of returns and minimize correlations in returns between groups. Each cluster contains at least three companies. Given the assignment of companies to clusters, the correlation between each company's stock return and the return of every other member of its cluster, as well as its correlation with all other stocks not in its cluster, was calculated and averaged across all companies. This procedure was done for returns in the five-year block over which the clustering algorithm was applied (the estimation period) and in the subsequent five years (the testing period). The difference between the two average correlations was calculated. The results were then averaged over all estimation or testing periods.

stable over time. In comparison, within-industry correlations based on four-digit GICS codes average 0.22 and are larger than outside-industry correlations by 0.04.

In summary, GICS codes compare favorably with purely statistical classifications with respect to producing homogeneous sets of stocks.

Commonality in Operating Performance. As further evidence that GICS groupings and FF industries yield sets of economically related stocks, we look behind the commonality of return movements to see whether there is commonality in operating performance across companies in an industry. Our measure of operating performance is the growth rate of sales, measured as the year-to-year percentage change in a trailing sum of sales over the previous four quarters. Sales growth is an important driver of profitability and behaves less erratically than other fundamental indicators, such as growth in earnings or dividends.

Specifically, suppose for company i in month t, the most recently available financial statement data are for quarter q (assuming a publication delay of three months). Sales growth for four quarters, $GS4Q$, is then

$$GS4Q_{it} = \frac{\sum_{j=0}^{3} sales_{iq-j}}{\sum_{j=0}^{3} sales_{iq-4-j}}, \tag{5}$$

where $sales_{iq}$ is net sales for company i in quarter q.[7] As in previous tests, we calculated correlations of sales growth rates between every pair of companies and averaged correlations within and outside an industry for various levels of GICS codes and for the FF classification. **Table 7** reports the results for the large-company sample. (We are not reporting results for small companies because of the limited commonality in returns of small-cap stocks within an industry.)

Table 7 shows that companies in an industry experience lower comovement in their sales growth rates than the covariation in their stock returns (Table 1). In every industry, the average sales growth correlation is smaller than the average return correlation. In the case of utility stocks based

Table 7. Average Pairwise Correlations between Individual Stocks' Sales Growth and Sales Growth of Within-Industry and Outside-Industry Stocks: Large-Cap Stocks (size deciles 5–10), 1975–2004

	Average Correlations															
Description	Within Industry					Outside Industry					Difference					Average No.
(GICS sector)	GICS8	GICS6	GICS4	GICS2	FF48	GICS8	GICS6	GICS4	GICS2	FF48	GICS8	GICS6	GICS4	GICS2	FF48	of Companies
Energy (10)	0.51	0.51	0.44	0.44	0.43	0.16	0.16	0.16	0.16	0.16	0.35	0.35	0.28	0.28	0.27	35
Materials (15)	0.39	0.35	0.30	0.30	0.32	0.15	0.15	0.14	0.14	0.15	0.24	0.20	0.16	0.16	0.17	60
Industrials (20)	0.24	0.21	0.16	0.16	0.18	0.12	0.12	0.12	0.11	0.12	0.12	0.09	0.04	0.05	0.06	96
Consumer discretionary (25)	0.15	0.15	0.11	0.09	0.13	0.07	0.07	0.07	0.06	0.07	0.08	0.08	0.04	0.03	0.06	92
Consumer staples (30)	0.15	0.13	0.14	0.11	0.13	0.09	0.09	0.09	0.09	0.09	0.06	0.04	0.05	0.02	0.04	47
Health care (35)	0.18	0.18	0.15	0.14	0.19	0.08	0.08	0.08	0.08	0.08	0.10	0.10	0.07	0.06	0.10	42
Financials (40)	0.32	0.32	0.29	0.19	0.32	0.12	0.12	0.12	0.12	0.12	0.20	0.20	0.17	0.07	0.20	101
Information technology (45)	0.28	0.26	0.26	0.22	0.23	0.14	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.08	0.09	48
Telecommunication services (50)	0.15	0.15	0.14	0.14	0.14	0.06	0.06	0.07	0.07	0.06	0.09	0.09	0.07	0.07	0.08	13
Utilities (55)	0.25	0.25	0.22	0.22	0.21	0.06	0.06	0.05	0.05	0.05	0.19	0.19	0.17	0.17	0.16	86
Weighted mean	0.26	0.25	0.21	0.19	0.23	0.11	0.10	0.10	0.10	0.10	0.15	0.15	0.11	0.09	0.13	
Simple mean	0.26	0.25	0.22	0.20	0.23	0.11	0.10	0.10	0.10	0.10	0.15	0.15	0.12	0.10	0.13	
t-Statistic											5.47	4.98	4.70	4.63	5.23	

Notes: In each nonoverlapping five-year block, each company in a sector was assigned to an industry based on one of five classification systems. The average correlation between the company's sales growth rate and the sales growth of every other member of its industry was calculated and averaged across all companies in the sector. Sales growth is the monthly percentage change in a trailing sum of net sales over the previous four quarters relative to its value a year ago. The average correlation between the stock's sales growth rate and the sales growth rate of every stock outside its industry was similarly calculated and averaged across members of the sector. The difference between the two average correlations was calculated. The results were then averaged over the six five-year periods. See also notes to Table 1.

on four-digit GICS codes, for example, the correlation of sales growth rates is 0.22 whereas the correlation of returns is 0.45. The simple mean correlation of sales growth for companies in an industry classified by four-digit GICS codes is 0.22, compared with a simple mean return correlation of 0.40.

Nevertheless within-group correlations of sales growth rates exceed outside-group correlations. For four-digit GICS groups, for instance, the simple mean within-group correlation exceeds the outside-industry correlation by 12 percent. These findings boost confidence in the argument that GICS classifications create homogeneous groups of companies that share similar underlying economic features.

The FF industries capture roughly the same degree of commonality in sales growth as four-digit GICS groups do. For the FF categories, the simple mean within-group correlation of sales growth is 0.23, or 0.13 higher than the outside-group correlation. Moving to six-digit GICS codes improves discrimination over FF groups by raising the spread in correlations to 0.15; going to eight-digit GICS codes provides little additional improvement.

In general, therefore, the results for correlations in sales growth echo the results in Table 1 based on return correlations.

Conclusion

We have provided evidence to resolve the divergence in defining industry groups between academic research and investment practice. We compared the GICS and FF grouping procedures in terms of their ability to isolate common return movements of stocks within an industry relative to comovements with stocks outside the industry. In addition, our analysis documented the gains from successively finer industry partitioning. We also verified that the industry groups correspond to collections of economically similar companies in two respects: (1) Industries are better at capturing out-of-sample return covariation than are statistical clusters formed without regard to industry affiliation and (2) industries reflect common movement in companies' underlying operating performance, as measured by sales growth.

Large-cap stocks that belong to the same industry classification, as defined by two-digit GICS codes, share a simple mean correlation of 0.38 in returns, compared with a mean correlation of 0.26 for stocks that belong to different industries. Measured net of an equally weighted market index, return correlations average 0.17 for large-cap stocks inside an industry defined by two-digit codes. The magnitudes of within-industry commonality in movements of raw and excess returns highlight the potential benefit of using industry affiliation as one dimension for managing portfolio risk and tracking-error volatility in the case of large companies. For smaller companies, however, the comovement in returns associated with commonality in industry membership is much less pronounced. In part, their common response to industry effects may be drowned out by the higher volatility of small-cap returns.

The Fama–French categories are comparable to four-digit GICS groups in terms of the magnitude of average within-industry correlation. Both procedures deliver a simple mean within-group correlation of 0.40 for large-cap stocks, corresponding to an improvement of 0.13 over the mean outside-group correlation. The FF method requires 48 industries, however, whereas there are only 24 four-digit GICS groups.

Finer levels of industry disaggregation tend to do better with respect to spreading out within-industry correlations versus outside-industry correlations. The benefits generally tail off, however, beyond six-digit GICS codes. For the sample of large-cap stocks, correlations between industry members differ from correlations between non-member companies by, on average, 0.13 at the two-digit level, 0.14 at the four-digit level, 0.17 at the six-digit level, and 0.18 at the eight-digit level.

Companies in an industry share weaker comovement, on average, in their sales growth rates than in their returns. Nevertheless, affiliation in the same industry generally translates into heightened covariation in sales growth. The mean correlation in sales growth between two large-cap companies with the same four-digit GICS code is 0.22, versus a correlation of 0.10 for two companies in different groups. The FF procedure, despite using twice the number of categories as four-digit GICS groups, yields about the same distinction between within- and outside-industry correlations. Pseudo-industry groups formed from statistical cluster analysis of stock returns do not match the performance of industry classifications on an out-of-sample basis. Both GICS and FF industries have larger within-group correlations and sharper discrimination over outside-group correlations.

This article qualifies for 1 PD credit.

Notes

1. Studies of the influence of industry include Grinold, Rudd, and Stefek (1989), King (1966), Lessard (1974), and Roll (1992).
2. Government agencies and data vendors use their own criteria to determine which SIC code applies to a given company. As noted in Guenther and Rosman (1994) and in Kahle and Walkling (1996), this practice frequently produces classifications for the same company that conflict among data providers.
3. Recently, the number of industries was expanded to 49 (computer software and computer hardware were split). The composition of the industries is described in detail on Kenneth French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/changes_ind.html).
4. Other criteria for judging classification schemes are possible. For example, Bhojraj et al. (2003) compared the ability of different industry classification schemes to account for the variability in the cross-section of returns, valuation multiples, earnings growth, and financial ratios.
5. To avoid clutter, we do not write $I(i)$ to reflect the dependence of the industry assignment on the stock or N_I to reflect the variation in the number of companies among industries.
6. Note that this breakdown is solely for expository purposes and does not bias the findings for or against any of the schemes. To be specific, for each company in a two-digit GICS sector, we used one of the industry assignment schemes to identify its peers in the same industry. All pairwise correlations between the company and its industry cohorts were averaged (Equation 1). The mean of these statistics over all companies in the two-digit GICS sector was then computed (Equation 3) and is reported as the within-industry correlation for this subset. In the same way, we applied Equation 2 and Equation 4 to calculate the mean of the average pairwise correlations between a company and all others not in its industry; this result is reported as the outside-industry correlation for the subset.
7. If quarterly data were unavailable, we measured the percentage change in the most recently available annual net sales relative to a year ago, still under the assumption of a three-month delay before the release of accounting data.

References

Bebchuk, Lucian, and Yaniv Grinstein. 2005. "The Growth of Executive Pay." *Oxford Review of Economic Policy*, vol. 21, no. 2 (Summer):283–303.

Bhojraj, Sanjeev, Charles M.C. Lee, and Derek K. Oler. 2003. "What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research." *Journal of Accounting Research*, vol. 41, no. 5 (December):745–774.

Brennan, Michael J., Ashley W. Wang, and Yihong Xia. 2004. "Estimation and Test of a Simple Model of Intertemporal Capital Asset Pricing." *Journal of Finance*, vol. 59, no. 4 (August): 1743–1776.

Brown, Stephen J., and William N. Goetzmann. 1997. "Mutual Fund Styles." *Journal of Financial Economics*, vol. 43, no. 3 (March):373–399.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu. 2001. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk." *Journal of Finance*, vol. 56, no. 1 (February):1–43.

Chan, Louis K.C., Jason Karceski, and Josef Lakonishok. 1999. "On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model." *Review of Financial Studies*, vol. 12, no. 5 (Winter):937–974.

Chan, Wesley S., Richard Frankel, and S.P. Kothari. 2004. "Testing Behavioral Finance Theories Using Trends and Consistency in Financial Performance." *Journal of Accounting and Economics*, vol. 38 (December):3–50.

Clarke, Richard N. 1989. "SICs as Delineators of Economic Markets." *Journal of Business*, vol. 62, no. 1 (January):17–31.

Connor, Gregory. 1995. "The Three Types of Factor Models: A Comparison of Their Explanatory Power." *Financial Analysts Journal*, vol. 51, no. 3 (May/June):42–46.

Daniel, Kent, and Sheridan Titman. 2006. "Market Reactions to Tangible and Intangible Information." *Journal of Finance*, vol. 61, no. 4 (August):1605–1643.

Economic Classification Policy Committee. 1994. *Report No. 1: Economic Concepts Incorporated in the Standard Industrial Classification Industries of the United States*. Washington, DC: Bureau of Economic Analysis, U.S. Department of Commerce.

Elton, Edwin J., and Martin J. Gruber. 1970. "Homogeneous Groups and the Testing of Economic Hypotheses." *Journal of Financial and Quantitative Analysis*, vol. 4, no. 5 (January):581–602.

Fama, Eugene F., and Kenneth R. French. 1997. "Industry Costs of Equity." *Journal of Financial Economics*, vol. 43, no. 2 (February):153–193.

———. 2004. "New Lists: Fundamentals and Survival Rates." *Journal of Financial Economics*, vol. 73, no. 2 (August):229–269.

Farrell, James, Jr. 1974. "Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings." *Journal of Business*, vol. 47, no. 2 (April):186–207.

Ferson, Wayne E., and Campbell R. Harvey. 1999. "Conditioning Variables and the Cross Section of Stock Returns." *Journal of Finance*, vol. 54, no. 4 (August):1325–1360.

Flannery, Mark J., and Kasturi P. Rangan. 2006. "Partial Adjustment toward Target Capital Structures." *Journal of Financial Economics*, vol. 79, no. 3 (March):469–506.

Francis, Jennifer, Ryan LaFond, Per Olsson, and Katherine Schipper. 2005. "The Market Pricing of Accruals Quality." *Journal of Accounting and Economics*, vol. 39, no. 2 (June):295–327.

Graham, John R., and Alok Kumar. 2006. "Do Dividend Clienteles Exist? Evidence on Dividend Preferences of Retail Investors." *Journal of Finance*, vol. 61, no. 3 (June):1305–1336.

Grinold, Richard, Andrew Rudd, and Dan Stefek. 1989. "Global Factors: Fact or Fiction?" *Journal of Portfolio Management*, vol. 16, no. 1 (Fall):79–88.

Guenther, David A., and Andrew J. Rosman. 1994. "Differences between COMPUSTAT and CRSP SIC Codes and Related Effects on Research." *Journal of Accounting and Economics*, vol. 18, no. 1 (July):115–128.

Hong, Harrison, Walter Torous, and Rossen Valkanov. 2007. "Do Industries Lead Stock Markets?" *Journal of Financial Economics*, vol. 83, no. 2 (February):367–396.

Kahle, Kathleen M., and Ralph A. Walkling. 1996. "The Impact of Industry Classifications on Financial Research." *Journal of Financial and Quantitative Analysis*, vol. 31, no. 3 (September): 309–335.

King, Benjamin F. 1966. "Market and Industry Factors in Stock Price Behavior." *Journal of Business*, vol. 39, no. 1, pt. 2 (January):139–190.

Lessard, Donald R. 1974. "World, National, and Industry Factors in Equity Returns." *Journal of Finance*, vol. 29, no. 2 (May):379–391.

Moskowitz, Tobias J., and Mark Grinblatt. 1999. "Do Industries Explain Momentum?" *Journal of Finance*, vol. 54, no. 4 (August): 1249–1290.

Pastor, Lubos, and Robert F. Stambaugh. 1999. "Costs of Equity Capital and Model Mispricing." *Journal of Finance*, vol. 54, no. 1 (February):67–121.

Purnanandam, Amiyatosh, and Bhaskaran Swaminathan. 2005. "Do Stock Prices React to SEO Announcements? Evidence from SEO Valuation." Working paper, Cornell University.

Richardson, Scott. 2006. "Over-Investment of Free Cash Flow." *Review of Accounting Studies*, vol. 11, no. 2–3 (June–September): 159–189.

Roll, Richard. 1992. "Industrial Structure and the Comparative Behavior of International Stock Market Indices." *Journal of Finance*, vol. 47, no. 1 (March):3–41.

Wulf, Julie. 2002. "Internal Capital Markets and Firm-Level Compensation Incentives for Division Managers." *Journal of Labor Economics*, vol. 20, no. 2, pt. 2 (April):S219–C262.