Economics Letters 118 (2013) 77–80

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

A comparison of industry classification schemes: A large sample study

Karel Hrazdil *, Kim Trottier, Ray Zhang

Simon Fraser University, Canada

ARTICLE INFO

Article history:
Received 21 August 2012
Received in revised form
15 September 2012
Accepted 19 September 2012
Available online 28 September 2012

JEL classification:
C15
G1

Keywords:
Industry classification
Homogeneity
Performance
GICS

ABSTRACT

In a large sample setting, we compare four broadly available industry classification schemes in their effectiveness to group stocks with similar operating characteristics. We demonstrate the advantage of the Global Industry Classification Scheme to be consistent across different application schemes common to capital market research and across different groups of firms.

1. Introduction

Accurate industry classification is essential for drawing valid statistical inferences from empirical samples. Controlling for industry effects is used in a variety of research and regulatory settings, such as determining the cost of capital, comparing firm to peer group performance, or designing effective compensation contracts based on comparable industry benchmarks. Capital market research predominantly utilizes the Standard Industrial Classification (SIC) system to identify homogenous groups of stocks in order to be consistent with prior studies and because, prior to 1999, it was the only industry classification system available through major data vendors. However, the reliability of SIC is a long-standing problem in financial research, as the codes differ across databases and can fail to identify firms with similar operating characteristics (Kahle and Walkling, 1996).

The aim of this note is to compare the properties of the SIC system with three alternatives: the Global Industry Classification Standard (GICS), developed jointly by Standard & Poor's (S&P) and Morgan Stanley Capital International to construct their widely followed indices; the North American Industry Classification System (NAICS), developed by US, Canadian, and Mexican governments to establish uniformity across North American countries in terms of the reporting of government statistics; and the Fama–French (FF) industry classification, derived from the SIC system by Fama and French (1997) to provide a manageable number of distinct industries. In particular, we evaluate these four classifications based on the intra-industry homogeneity of twelve applications common to capital market research using a comprehensive sample of US firms. We contribute to the literature by demonstrating the superiority of the GICS for grouping stocks with similar operating characteristics across most application schemes and across different groups of firms.

2. Extant research

The four industry classification systems are compared by a handful of studies, all of which conclude that the GICS outperforms the other three systems in specific subsamples and research settings, such as among S&P 1500 firms (Bhojraj et al., 2003, hereafter BLO), high-tech firms (Kile and Phillips, 2009), manufacturing firms in estimating concentration ratios (Hrazdil and Zhang, 2012), partitioning industries based on analyst coverage choices (Boni and Womack, 2006), explaining return comovements (Chan et al., 2007), and in its ability to estimate abnormal accruals (Hrazdil and Scott, forthcoming).[1]

Of particular importance to our analysis is the work of BLO, who examine the degree of intra-industry homogeneity of twelve applications common to capital market research and demonstrate that the GICS significantly outperforms the other systems in explaining variations in valuation multiples, growth rates, and

* Correspondence to: Faculty of Business Administration, Beedie School of Business, Simon Fraser University, 8888 University Drive, Burnaby, B.C., V5A 1S6, Canada. Tel.: +1 778 782 6790; fax: +1 778 782 4920.
E-mail address: karel_hrazdil@sfu.ca (K. Hrazdil).

[1] For further details on the evolution, availability, limitations, and uses of the different industry classifications, please see Hrazdil and Scott (forthcoming).

0165-1765/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2012.09.022

Table 1
Descriptive statistics.

	SIC (2-digit)	GICS (6-digit)	FF (2-digit)	NAICS (3-digit)
Average number of functional industry categories	67	61	49	85
Minimum number of firms per industry/year	5	6	7	5
Maximum number of firms per industry/year	1564	929	1603	1194
Mean number of firms per industry/year	145	159	199	115
Median number of firms per industry/year	64	113	129	52
Standard deviation (industry-years)	206	149	221	178
Skewness (industry-years)	2.8	1.7	2.6	3.3
Kurtosis (industry-years)	9.3	4.1	9.0	12.5

various financial ratios. The implication of BLO's result is that the GICS is the preferred method to group firms by industry in most research settings. Before their results can be generalized to other research settings and utilized beyond S&P 1500 firms, further analysis needs to be conducted on a more comprehensive sample. Exploring and validating the overall effectiveness of the GICS as a superior industry categorization has thus important implications for capital market research.

3. Data and descriptive statistics

Due to limited data availability, BLO examine only S&P 1500 firms between 1994 and 2001. Our sample selection begins with all NYSE and NASDAQ firms during the 20-year period from 1990 to 2009. We obtain the GICS codes directly from S&P, which provides the most accurate record of the company's industry classification as of a given historical date.[2] The remaining variables are obtained from Compustat, Center for Research and Security Prices (CRSP), and Institutional Brokers' Estimate System (I/B/E/S).

Similar to BLO, we examine the homogeneity of twelve variables: calendar month-end returns (*RET*), year-end price-to-book (*P/B*, market capitalization divided by total common equity), enterprise-value-to-sales (*EVS*, the sum of market capitalization and debt divided by net sales), price-to-earnings (*P/E*, market capitalization divided by income before extraordinary items), return-on-assets (*ROA*, income before extraordinary items divided by total assets), return-on-equity (*ROE*, income before extraordinary items divided by total common equity), profit margin (*PM*, operating income after depreciation divided by net sales), leverage (*LEV*, total liabilities divided by total common equity), asset turnover (*AT*, total assets divided by net sales), current ratio (*CR*, total current assets divided by total current liabilities), one-year-ahead realized sales growth (*SGR*), and *R&D* (research and development expense divided by net sales). In order to retain a comprehensive sample, some of these variables differ slightly from definitions in BLO.[3]

We follow BLO in identifying primary equivalent levels for each industry category and compare the first two-digit (major group) SIC codes with the first six-digit (industry) GICS codes, the first three-digit (subsector) NAICS codes, and the 48 FF industry classifications. We winsorize all variables at extreme percentiles, retain observations with non-missing variables, positive sales, total book equity, and price greater than 1, and analyze industry groups with at least five firms (hereafter, functional categories). Collectively, these filters yield a sample of 16,329 firms across twenty years (123,020 firm-year observations). Table 1 provides univariate statistics across functional categories of each industry classification.

Similar to BLO, we find that the GICS codes distribute firms more evenly across industry categories. The average GICS group has the lowest standard deviation of 149, the least skewed distribution (skewness of 1.7), lowest excess kurtosis (4.1), and the smallest difference between mean and median values among the four industry classification schemes.

4. Methodology, results, and concluding remarks

For each functional category, we estimate the adjusted R^2 values from the following ordinary least squares (OLS) regressions: $vble_{i,t} = \alpha_1 + \beta vble_{ind,t} + \varepsilon_{i,t}$, where $vble_{i,t}$ is the tested variable for firm i within a particular industry ind at year t, and the independent variable is the yearly average for that variable for all firms in a particular functional industry classification. The R^2 value for the returns variable (*RET*) is obtained similarly using monthly data.

Since, for a fixed number of firms, a classification system having more industry categories will mechanically produce greater R^2 values, we modify the results with simulated values of R^2 to obtain revised R^2 values. We follow BLO and conduct Monte Carlo simulations by first randomly assigning our sample into the same number of industry categories with the same number of firms per industry. We then estimate the regressions on the simulated data, repeating the procedure 500 times to obtain a simulated R^2 value (a performance benchmark) for each classification standard.

The last column of Table 2 compares the revised adjusted R^2 values across the twelve performance variables. For each variable, we consider the GICS to be superior if it generates a revised R^2 value that is greater than the R^2 value of each remaining schemes and when its homogeneity is significantly higher than that of at least two other classifications. We find that the GICS system offers an advantage in seven out of twelve variables (stock return comovements in Panel A, EVS multiple in Panel B, four financial statement ratios in Panel C, and R&D in Panel D).[4] The revised R^2 values of the other three systems differ little from one another across most ratios, which accentuates the GICS advantage, and the similarity of results in last two columns further indicates that structural differences among classification schemes have little effect on our main conclusion.[5]

The GICS is developed by analysts with the purpose of enhancing investment research and asset management processes, and therefore may not offer improvement for firms that do not have analyst following. Furthermore, BLO point out that their results may be induced by the fact that the GICS classifications and analyst perceptions are endogenous (since analyst perception play a role in GICS categorization). To address this concern, we extend the analysis of BLO and partition the comprehensive sample into three groups based on analyst following.[6]

[2] The GICS History is available only on a subscription basis, and it provides the most comprehensive historical coverage for more than 25,000 active and inactive North American firms going back to June 1985. The results in this note are based on the GICS History as of December 31, 2011.

[3] For example, since many small firms have negative total shareholders' equity, we use total common equity when defining leverage.

[4] BLO find that the GICS system is superior to other schemes in eight out of twelve variables. Compared to our results, some of the variables found to be superior by BLO are not superior in our study (i.e. *P/B* and *SGR*). This is likely due to differences in our sample size (over 16,000 firms compared to 1500 by BLO) and time period (1990–2009 compared to 1994–2000 by BLO) examined.

[5] The untabulated results further confirm that the GICS superiority is consistent from year to year.

[6] BLO found that the GICS's superiority diminishes as the firm size gets smaller. Our untabulated results and conclusions based on three size subsamples are the same as those based on analyst following.

Table 2
Tests of homogeneity: actual, simulated, and revised R^2 values.

	I (Actual) (%)	II (Simulated) (%)	I–II (Revised) (%)	GICS versus competitor actual (%)	GICS versus competitor revised (%)
Panel A: Returns					
RET					
SIC	12.46	0.10	12.36	1.13***	1.13***
NAICS	12.92	0.10	12.82	0.67***	0.67***
FF	12.25	0.09	12.16	1.34***	1.33***
GICS	13.59	0.10	13.49		Superior
Panel B: Valuation multiples					
P/B					
SIC	4.45	0.63	3.82	0.31	0.26
NAICS	4.32	0.84	3.48	0.43	0.59
FF	4.56	0.49	4.08	0.19	−0.01
GICS	4.75	0.68	4.07		
EVS					
SIC	3.84	1.07	2.77	2.72***	2.65**
NAICS	3.94	1.40	2.54	2.62***	2.89**
FF	5.14	0.80	4.34	1.42***	1.09**
GICS	6.56	1.13	5.43		Superior
P/E					
SIC	0.91	1.07	−0.15	0.11	0.05
NAICS	1.23	1.40	−0.16	−0.21	0.06
FF	1.65	0.80	0.85	−0.62	−0.96
GICS	1.02	1.13	−0.11		
Panel C: Financial statements ratios					
ROA					
SIC	4.86	1.05	3.80	3.03***	2.94***
NAICS	7.74	1.40	6.34	0.15	0.41
FF	5.64	0.80	4.84	2.25***	1.91***
GICS	7.88	1.14	6.75		Superior
ROE					
SIC	2.71	1.06	1.65	0.63***	0.56***
NAICS	3.20	1.40	1.80	0.14	0.41
FF	2.56	0.80	1.76	0.78***	0.45**
GICS	3.34	1.13	2.21		Superior
PM					
SIC	4.11	1.07	3.05	3.98***	3.92***
NAICS	4.24	1.40	2.84	3.85***	4.12***
FF	5.82	0.81	5.01	2.28***	1.96***
GICS	8.09	0.13	6.96		Superior
LEV					
SIC	4.11	1.05	3.05	−0.20	−0.28
NAICS	3.93	1.41	2.52	−0.02	0.25
FF	3.13	0.80	2.33	0.78*	0.44
GICS	3.91	1.14	2.77		
AT					
SIC	8.39	1.07	7.32	1.02***	0.95***
NAICS	8.35	1.38	6.97	1.06***	1.31***
FF	8.26	0.80	7.47	1.14***	0.81
GICS	9.41	1.13	8.27		Superior
CR					
SIC	7.39	1.29	6.09	0.85*	0.75
NAICS	8.60	1.72	6.89	−0.37	−0.05
FF	7.26	0.98	6.28	0.98***	0.56*
GICS	8.23	1.39	6.84		
Panel D: Other financial information					
SGR					
SIC	0.96	1.09	−0.13	0.05	−0.04
NAICS	1.35	1.43	−0.08	−0.34**	−0.09
FF	0.90	0.81	0.08	0.11	−0.25
GICS	1.01	1.18	−0.17		

(*continued on next page*)

Table 2 (*continued*)

	I (Actual) (%)	II (Simulated) (%)	I–II (Revised) (%)	GICS versus competitor actual (%)	GICS versus competitor revised (%)
R&D					
SIC	5.20	2.34	2.86	4.87***	4.75***
NAICS	5.18	3.03	2.15	4.89***	5.46***
FF	7.38	1.76	5.61	2.69***	2.00***
GICS	10.07	2.46	7.61		Superior

*** Denote 1% significance level, (based on two-tailed *t*-test on the differences between the GICS and other classifications based on the time series of differences from 1990 to 2009).
** Denote 5% significance level, (based on two-tailed *t*-test on the differences between the GICS and other classifications based on the time series of differences from 1990 to 2009).
* Denote 10% significance level, (based on two-tailed *t*-test on the differences between the GICS and other classifications based on the time series of differences from 1990 to 2009).

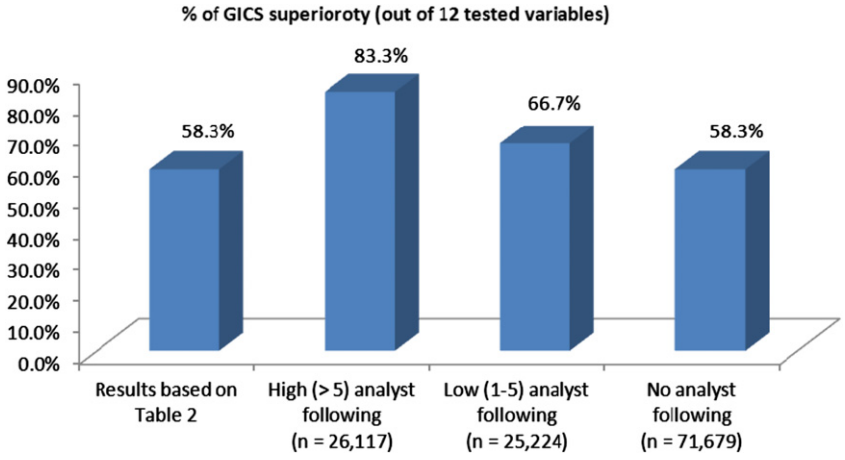

Fig. 1. Percentage of variables where the GICS is superior in capturing homogeneity.

Although the GICS's superiority varies across the portfolios of sample companies formed on the basis of analyst coverage and the GICS's superiority diminishes as the analyst coverage gets smaller, the results in Fig. 1 suggest that the GICS remains superior in capturing homogeneity in more than half of the tested variables for the group of firms with no analyst coverage. The robustness of our finding has thus important implications to researchers interested in non-S&P 1500 firms, which are more likely to be neglected by market intermediaries and whose information environment is less rich than that of large and widely followed firms.

In conclusion, we find that the GICS continues to outperform the other three systems in a large sample setting as well as among smaller and less followed firms. It provides more homogeneous and hence better industry classifications, and therefore is expected to yield a better and more precise empirical analysis. Accordingly, the GICS has important implications for academics, regulators, and practitioners, in that it should be used as the primary industry classification system or as an important validity consideration.

Acknowledgments

The authors are grateful to B. Krauth, S. Mongrain, and an anonymous referee for very helpful comments on earlier versions of the paper. The Global Industry Classification Standard system (GIGS History) was licensed from the Standard and Poor's. All other data were obtained from publicly available sources cited in the study. The authors acknowledge financial support of the Canadian Academic Accounting Association and CICA. Any errors are ours.

References

Bhojraj, S., Lee, C., Oler, D., 2003. What's my line? a comparison of industry classification schemes for capital market research. Journal of Accounting Research 41, 745–769.
Boni, L., Womack, K., 2006. Analysts, industries, and price momentum. Journal of Financial and Quantitative Analysis 41, 85–109.
Chan, L., Lakonishok, J., Swaminathan, B., 2007. Industry classifications and return comovement. Financial Analyst Journal 63, 56–70.
Fama, E., French, K., 1997. Industry cost of equity. Journal of Financial Economics 43, 153–193.
Hrazdil, K., Scott, R., 2013, The role of industry classification in estimating discretionary accruals, Review of Quantitative Finance and Accounting 40 (1) (forthcoming).
Hrazdil, K., Zhang, R., 2012. The importance of industry classification in estimating concentration ratios. Economics Letters 114, 224–227.
Kahle, K., Walkling, R., 1996. The impact of industry classifications on financial research. Journal of Financial and Quantitative Analysis 31, 309–332.
Kile, C., Phillips, M., 2009. Using industry classification codes to sample high-technology firms: analysis and recommendations. Journal of Accounting, Auditing and Finance 24, 35–58.