DOES INDUSTRIAL CLASSIFICATION OF FIRMS REFLECT CORPORATE PERFORMANCE? EVIDENCE FROM THE EUROPEAN UNION

Julia Koralun-Bereznicka PhD[106]
University of Gdansk, Poland

Abstract

The aim of the study is to compare the industrial classification of firms in Europe (NACE) with the classification results based on corporate performance ratios. The methodology employed for this purpose includes mainly the *k*-means clustering technique as well as a similarity measure for evaluating the resemblance of grouping results. Using the total of almost 90 thousand aggregated observations from a sample of firms from 13 industries, 9 countries and 3 size groups, covering the period 2000-2010, findings provide evidence about the generally poor resemblance between the industrial classification and the grouping results based on financial ratios. There are also some country-wise differences found in the similarity level between the compared classifications.

Keywords: Corporate performance, industrial classification, European Union

Introduction

Industrial classification of firms in an important aspect of categorising corporate activity, as well as a key instrument for cross-sectional comparisons of corporate performance in the area of finance. The primary reason for organising firms into industries is to classify entities into groups of similar objects in terms of products or services. This way of classifying firms, however, might not always be effective and precise due to the fact that e.g. some firms may produce products from more than one category. As a result, the assumedly homogeneous industrial groups may be in fact quite varied internally, whereas firms from different industries may bear more resemblance. Despite the inevitable imperfections of the industrial classification systems, this kind of division seems necessary and remains one of the most ubiquitous systematics of firms. The main question addressed in this study is to find whether and to what extent the industrial classification of firms reflects their financial performance.

Corporate financial health is a complex issue and may be affected by an almost countless number of factors of both internal and external character (Koralun-Bereźnicka, 2013). Industrial classification is one of the most commonly mentioned external determinants of corporate performance. Consequently, one may assume that financial parameters describing this performance should vary along with the industrial sections of firms. In other words, corporate financial performance should, at least to some extent, correspond to the industrial classification of firms. Therefore, the aim of this study is to compare the industrial classification of firms in Europe (NACE) with the classification based on financial performance ratios.

[106] Notice: The project was funded by the National Centre of Science in Poland on the basis of the decision number DEC-2013/09/B/HS4/01936.

The research is based on a sample of all-sized firms from 13 industries, 9 EU countries and covers the period 2000-2010. Financial performance of firms is described with the use of 28 ratios grouped into several economic categories: profitability, working capital, financial income and charges, asset structure and capital structure. The similarity of groupings is evaluated for the whole population, as well as for individual countries covered by the analysis and for three size groups of firms separately.

This study contributes to the corporate finance literature in several ways. Firstly, although the topic of industry effect has already been explored on multiple occasions, this analysis provides deeper, cross-country and cross-size insights into the relationship between industrial classification and performance. Secondly, while most of research in the field tends to focus on large public companies and the easily accessible market returns data, this study extends the empirical work on the industry effect by taking into consideration private companies of various sizes, including SMEs which usually constitute the core of most economies. Thirdly, the range of financial ratios applied for describing corporate performance is much broader than in previous studies from the field. Finally, the variables characterise fundamental condition of companies and not their market value.

Literature review

The literature review provides a number of studies where industrial classifications were used to determine the extent to which industrial specificity is responsible for cross-sectional diversity of corporate financial characteristics. The study by King (1966) is considered as the first major attempt to identify industry effects in corporate performance reflected in market returns. Using principal component analysis and clustering techniques on a sample of companies from different industries, he reported that industry code explains about 10 percent of the variance in rate of returns. Industry effects in stock returns were also found by many other researchers, including Meyers (1973), Lessard (1974), Roll (1992) or Heston and Rouwenhorst (1994, 1995). The influence of industry characteristics was also searched for in other areas of corporate finance, especially in capital structure. The studies by Gupta and Huefner (1972), Remmers, Stonehill, Wright and Beekhuessen (1974), Scott and Martin (1975), Bowen, Daley and Huber (1982), Meric and Meric (1983), Martin and Henderson (1984), Maksimovic, Stomper and Zechner (1999), Hall, Hutchinson and Michaelas (2000) and more recent by Omran and Pointon (2004), Phillips and MacKay (2005), Abor (2007), Das and Roy (2007) or Talberg, Winge, Frydenberg and Westgaard (2008) are just some examples of the empirical evidence documenting the impact of industry features on financial leverage.

Contrary to the profusion of empirical studies in the area of corporate finance which use the industrial classifications to find industry effects, the literature related to the very idea of industry classification is rather sparse (Kale, Walking 1996). As to the literature referring to the impact of industry classification on financial research, an important expansion upon King's (1966) and Meyers' (1973) research into the link between industrial codes and variance was performed by Fertuck (1975) who discussed SIC (Standard Industrial Classification) codes and the industry effects of such codes. He examined whether companies can be properly classified according to their expected returns and found that some SIC groupings are more homogeneous in nature than others. The author also addressed the question about the efficiency of using industry indices to forecast returns.

The ability of SIC codes to form homogeneous groups was also examined by Clarke (1989), who measured how well the SIC succeeds at combining firms into homogeneous economic markets. The author assumed that firms in more similar economic markets should display more similar sales changes, profit rates, or stock price changes than firms in less similar economic markets, but found that the industry code is not successful at identifying firms with such similar characteristic variables.

In another study (Guenther, Rosman, 1994) the differences between SIC codes assigned to companies by COMPUSTAT and CRSP were examined. The authors reported significant differences across the two classifications in the variance of some financial ratios. Finally, the aforementioned study by Kale and Walking (1996) should be referred to as an important attempt to compare the inference from studies based on different classifications. One of the objectives of the study was to examine the extent of agreement between the two commonly used classification systems, namely the SIC codes on Compustat and the CRSP (Center for Research in Security Prices). Similarly to Guenther and Rosman (1994), the authors reported significant differences between the two sources of data, which may affect the inference from empirical research depending on which database is used.

Database and methodology

The source of data is the BACH-ESD database (Bank for the Accounts of Companies Harmonised - European Sectoral references Database). The study includes companies of three size groups: small (with the net turnover of less than EUR 10 million), medium (with a turnover of 10 million euros to 50 million euros) and large (with a turnover over EUR 50 million) in thirteen industries according to the NACE classification (Nomenclature Statistique des Activités économiques dans la Communauté Européenne) and in nine European Union countries available in the BACH-ESD database: Austria, Belgium, Germany, Spain, France, Italy, the Netherlands, Poland and Portugal. Table 1. shows the industries covered by the study and the three-letter symbols assigned to each sector used in the remainder of the paper.

Table 1. Industrial sections covered by the analysis

NACE	Section	Symbol
A	Agriculture, forestry and fishing	AGR
B	Mining and quarrying	MIN
C	Manufacturing	MNF
D	Electricity, gas, stream and air conditioning supply	ELE
E	Water supply; sewerage, waste management and remediation activities	WAT
F	Construction	CST
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	TRD
H	Transport and storage	TRS
I	Accommodation and food service activities	HOT
J	Information and communication	INF
L	Real estate activities	RLE
M	Professional, scientific and technical activities	PRF
N	Administrative and support service activities	ADM

Source: Statistical Classification of Economic Activities in the European Community, Rev. 2 (2008)

The harmonised and aggregated data from the annual reports of non-financial firms were used for calculating financial performance ratios for groups of companies in each country, industry, size group and each year of the eleven-year study period covering the years 2000-2010. The diagnostic variables can be grouped into several categories illustrating different economic areas, i.e. profitability, working capital, financial income and charges, asset structure and liabilities structure. Taking into account the data availability, the analysis involves 28 financial ratios, the details of which are shown in Table 2.

Summarising, the subject of the study is formed by the groups of companies of different sizes, from different industries in different countries and years. The corporate performance, measured with the use of financial ratios is the object of the analysis. Thus the study includes 28 financial ratios for the three size groups of enterprises in thirteen sectors and in nine countries for eleven years, which taking into account the missing data gives 88,536 observations (data items).

Table 2. Financial ratios used in the analysis

Ratio category	Ratio structure	Ratio number in BACH-ESD
Profitability	Added value / Net turnover	R01
	Staff costs / Net turnover	R02
	Gross operating profit / Net turnover (ROS)	R03
	Gross Operating profit / Total net debt	R04
	Net operating profit / Net turnover	R05
	Net turnover / Total Assets	R16
	Net operating profit / Total Assets (ROI)	R10
	Profit or loss of the year before taxes / Capital and reserves (ROEBT)	R11
	Profit or loss of the year / Capital and reserves (ROE)	R12
Working capital	Inventories / Net turnover	R17
	Trade accounts receivable / Net turnover	R18
	Trade accounts payable / Net turnover	R19
	Operating working capital / Net turnover	R20
Financial situation	Interest and similar charges / Net turnover	R07
	Interest and similar charges / Gross operating profit	R06
	Financial income net of charges / Net turnover	R09
	Financial income net of charges / Gross operating profit	R08
Assets structure	Financial fixed assets / Total assets	R13
	Tangible fixed assets / Total assets	R14
	Current assets / Total assets	R15
	Current investment and cash in hand or at bank / Total assets	R21
Liabilities structure	Capital and reserves / Total assets	R22
	Provisions / Total assets	R23
	Bank loans / Total assets	R24
	Long and medium-term bank loans / Total assets	R25
	Short-term bank loans / Total assets	R26
	Long and medium-term debt / Total assets	R27
	Short-term debt / Total assets	R28

Source: BACH-ESD database.

The choice of the research methodology to a large extent is conditioned by the nature of the data, which is a relatively large collection of objects (industries, size groups, countries and years), characterised by a few diagnostic variables. The data is four-dimensional, as there is a time series for each object in the three cross-sections (countries, industries, size groups), Therefore the multivariate analysis is a natural tool for simplifying the data structure and identifying the most important regularities within the population. The review of the existing literature (e.g. Cinca, Molinero & Larraz, 2005; Gupta & Huefner, 1972; Sell, 2005; Helg, Manasse, Monacelli & Rovelli, 1995) suggests that multivariate classification often provides an effective solution to this kind of research problems.

The initial phase of the empirical research is the analysis of the descriptive statistics of the financial ratios across industries, which is aimed at the preliminary recognition of the corporate performance diversity in this cross-sections as well as detecting the basic regularities within the population.

In the event of finding differences in ratio means across industries, it should be established whether these differences are statistically significant. Then the analysis of variance (ANOVA) is applicable is as a method of studying observations dependent on one or more factors acting simultaneously. These factors are also known as grouping or manipulative variables. The analysis of variance (Fisher, 1954) allows to assess the significance of differences between many means and explains the probability with which the considered factors may be the reason for the discrepancies between the observed group means. If the means differ significantly from each other, it can be intuitively concluded that the analysed factor affects the dependent variable.

The heterogeneity of the objects from the examined population, as well as some similarities found between them imply the need for organising these objects by classifying them according to certain criteria. The idea of classification can be defined as a process of linking objects into categories, called clusters, based on their properties. Therefore, the grouping procedure is the next step of the analysis. One of the many clustering methods, which allows to extract internally homogeneous groups of objects is the *k*-means grouping, which aims at partitioning observations by creating *k* different, possibly distinct clusters, formed by the relocation of objects between these clusters in a way which minimises the within-group variance while maximising the between-group variance (Wishart, 2001).

The following sets of binominal objects were subject to the *k*-means grouping procedure:

- industries in countries – in individual size groups separately and in all size groups as a total,
- size groups in industries – in individual countries separately and in all countries as a total.

The missing data items were replaced with means. The advantage of the *k*-means algorithm is the ease of application even with large data sets. In addition, the target number of clusters must be determined *a priori*, which can also be helpful when that number can be based on certain criteria.

In order to compare the clustering results of industries in countries and size groups in industries with the NACE classification of industries, i.e. to evaluate the differentiation degree of the grouping results, the adjusted Rand's similarity measure was applied. The calculation method of the measure can be found e.g. in Rand (1971). The higher the value of the measure, the more similar the grouping results. Negative values indicate dissimilarity.

Results

The ratios used in the analysis are continuous variables, which is why they may analysed with the use of descriptive statistics, including mean value, minimum, maximum and standard deviation. The descriptive statistics for the total sample are presented in Table 3.

Table 3. Descriptive statistics for all years, countries, industries and size groups

Ratio	N	Mean value	Median	Minimum value	Maximum value	Standard deviation
R01	3317	0,357	0,368	0,000	0,848	0,129
R02	3317	0,220	0,213	0,000	0,590	0,109
R03	3317	0,137	0,111	-0,304	0,668	0,093
R04	2993	0,228	0,184	-11,65	7,967	0,340
R05	3317	0,065	0,050	-0,442	1,282	0,068
R16	3314	0,890	0,797	0,000	3,891	0,571
R10	3314	0,046	0,042	-0,249	0,475	0,036
R11	3307	0,129	0,117	-7,495	2,603	0,214
R12	3307	0,097	0,088	-6,800	2,369	0,183
R17	3314	0,131	0,067	0,000	4,823	0,278
R18	3005	0,234	0,198	0,000	1,890	0,169
R19	2381	0,186	0,169	0,000	1,499	0,103
R20	2381	0,204	0,143	-1,189	4,851	0,333
R07	3238	0,051	0,022	0,000	5,122	0,176
R06	3231	0,385	0,190	-72,25	97,55	2,837
R09	3238	0,011	-0,005	-0,857	2,169	0,143
R08	3231	0,249	-0,046	-16,79	100,9	3,671
R13	3317	0,151	0,103	0,000	0,962	0,141
R14	3317	0,345	0,317	0,000	0,881	0,201
R15	3317	0,452	0,447	0,000	0,914	0,189
R21	3006	0,082	0,074	0,000	0,487	0,047
R22	3317	0,349	0,331	-0,079	0,921	0,135
R23	3317	0,062	0,034	0,000	0,697	0,073
R24	2891	0,183	0,168	0,000	0,762	0,106
R25	2928	0,115	0,092	0,000	0,590	0,089
R26	2970	0,069	0,057	0,000	0,740	0,055
R27	3317	0,212	0,187	0,000	0,699	0,119
R28	3317	0,352	0,350	0,000	0,865	0,138

Source: author's calculations based on BACH-ESD database.

It is also relevant and informative to look at the means of the ratios by industries, as shown in Table 4.

Table 4. Mean values of ratios by industry

Ratio	Industry												
	AGR	MIN	MNF	ELE	WAT	CST	TRD	TRS	HOT	INF	RLE	PRF	ADM
R01	0,250	0,391	0,280	0,338	0,396	0,290	0,138	0,383	0,447	0,429	0,451	0,400	0,476
R02	0,152	0,206	0,191	0,104	0,227	0,212	0,091	0,266	0,330	0,280	0,127	0,304	0,346
R03	0,098	0,186	0,089	0,233	0,170	0,077	0,047	0,117	0,117	0,149	0,324	0,095	0,130
R04	0,161	0,447	0,230	0,248	0,265	0,133	0,170	0,198	0,196	0,313	0,132	0,231	0,235
R05	0,039	0,098	0,044	0,107	0,059	0,047	0,030	0,040	0,051	0,060	0,181	0,073	0,047
R16	0,884	0,698	1,147	0,446	0,551	0,978	1,984	0,850	0,910	0,879	0,247	0,710	1,025
R10	0,035	0,074	0,052	0,043	0,034	0,044	0,059	0,032	0,042	0,052	0,033	0,042	0,046
R11	0,066	0,198	0,134	0,114	0,096	0,165	0,173	0,065	0,117	0,127	0,066	0,165	0,174
R12	0,062	0,148	0,098	0,083	0,077	0,125	0,127	0,049	0,087	0,081	0,058	0,125	0,126
R17	0,225	0,169	0,146	0,041	0,046	0,348	0,116	0,023	0,034	0,048	0,517	0,074	0,024
R18	0,212	0,249	0,196	0,225	0,343	0,285	0,147	0,192	0,102	0,251	0,212	0,383	0,253
R19	0,196	0,188	0,163	0,204	0,206	0,264	0,150	0,155	0,126	0,186	0,214	0,245	0,130
R20	0,250	0,305	0,192	0,065	0,230	0,374	0,129	0,074	0,019	0,132	0,591	0,233	0,171
R07	0,022	0,039	0,018	0,077	0,030	0,023	0,012	0,030	0,031	0,026	0,147	0,206	0,025
R06	0,233	0,622	0,201	0,627	0,171	0,300	0,241	0,253	0,274	-0,153	0,410	1,679	0,192
R09	-0,009	0,012	0,002	0,048	-0,001	-0,004	-0,002	-0,013	-0,010	-0,001	-0,060	0,171	-0,004
R08	-0,150	0,833	0,023	0,561	-0,003	-0,031	-0,039	-0,105	-0,076	0,156	-0,159	2,184	-0,020
R13	0,098	0,174	0,154	0,164	0,111	0,098	0,114	0,099	0,163	0,192	0,151	0,308	0,141
R14	0,401	0,370	0,258	0,524	0,502	0,174	0,174	0,508	0,470	0,202	0,536	0,111	0,339
R15	0,478	0,413	0,553	0,259	0,330	0,696	0,673	0,345	0,302	0,490	0,269	0,538	0,458
R21	0,067	0,084	0,077	0,059	0,070	0,093	0,088	0,078	0,070	0,105	0,059	0,121	0,085
R22	0,412	0,418	0,375	0,393	0,354	0,259	0,316	0,333	0,327	0,347	0,389	0,401	0,249
R23	0,025	0,113	0,064	0,078	0,091	0,054	0,045	0,071	0,049	0,071	0,021	0,069	0,041
R24	0,200	0,136	0,165	0,170	0,199	0,189	0,165	0,212	0,233	0,106	0,301	0,103	0,216
R25	0,111	0,076	0,080	0,132	0,139	0,100	0,060	0,159	0,180	0,058	0,239	0,059	0,120
R26	0,091	0,062	0,088	0,039	0,061	0,092	0,108	0,056	0,057	0,051	0,060	0,043	0,096
R27	0,197	0,156	0,161	0,276	0,232	0,165	0,127	0,260	0,308	0,167	0,343	0,159	0,235
R28	0,354	0,301	0,388	0,228	0,271	0,496	0,501	0,306	0,297	0,381	0,217	0,347	0,434

The table indicates that the diagnostic variables are far from homogeneous across industries. There are several ratios with clearly better discriminating properties, i.e. the most varied in this cross-section, namely the interests to gross operating profit (R6), financial income to gross operating profit (R8) and the net turnover to total assets (R16).

The one-way ANOVA procedure was carried out in two sections, for which the qualitative predictors were: industry and year. The discrimination power of the ratios can be analysed on the basis of the *F* statistic and probability *p* calculated for the entire data set and presented in Table 5.

Table 5. Univariate significance tests

Ratio	Industry		Year	
	F	p	F	p
R01	F(2,404)= 297,56*	0,000	F(0,007)=0,425	0,935
R02	F(1,798)= 327,26*	0,000	F(0,012)=1,041	0,406
R03	F(1,263)= 304,06*	0,000	F(0,005)=0,619	0,799
R04	F(1,549)=14,097*	0,000	F(0,393)=3,422*	0,000
R05	F(0,371)=111,26*	0,000	F(0,015)=3,243*	0,000
R16	F(45,06)=276,54*	0,000	F(0,317)=0,972	0,465
R10	F(0,035)=29,089*	0,000	F(0,016)=12,24*	0,000
R11	F(0,496)=11,202*	0,000	F(0,139)=3,046*	0,001
R12	F(0,247)=7,521*	0,000	F(0,171)=5,168*	0,000
R17	F(5,021)=84,662*	0,000	F(0,056)=0,727	0,700
R18	F(1,312)=55,715*	0,000	F(0,016)=0,549	0,856
R19	F(0,320)=35,232*	0,000	F(0,004)=0,380	0,956
R20	F(3,861)=41,956*	0,000	F(0,108)=0,973	0,465
R07	F(0,821)=29,398*	0,000	F(0,019)=0,612	0,805
R06	F(48,289)=6,114*	0,000	F(0,008)=0,411	0,942
R09	F(0,703)=39,517*	0,000	F(7,327)=0,910	0,523
R08	F(105,74)=8,052*	0,000	F(10,73)=0,796	0,633
R13	F(0,824)=48,178*	0,000	F(0,016)=0,818	0,611
R14	F(6,035)=325,01*	0,000	F(0,021)=0,517	0,879
R15	F(5,436)=335,02*	0,000	F(0,067)=1,871*	0,045
R21	F(0,072)=37,933*	0,000	F(0,008)=3,707*	0,000
R22	F(0,743)=47,418*	0,000	F(0,179)=10,06*	0,000
R23	F(0,154)=31,802*	0,000	F(0,008)=1,486	0,138
R24	F(0,613)=69,800*	0,000	F(0,031)=2,796*	0,002
R25	F(0,631)=116,45*	0,000	F(0,013)=1,620	0,095
R26	F(0,117)=44,998*	0,000	F(0,016)=5,202*	0,000
R27	F(1,100)=106,83*	0,000	F(0,018)=1,270	0,242
R28	F(2,139)=190,23*	0,000	F(0,145)=7,812*	0,000

Note: The table presents the results of the one-way ANOVA procedure performed for all the ratios in the two cross-sections, i.e. across industries and across years. Values significant at p=0,05 are marked with *.
Source: author's calculations based on the BACH-ESD database.

The calculations show that all of the considered ratios demonstrate good discriminating abilities across industries. However, the opposite is the case for the majority of ratios when the other factor – year – is taken into account. Even in the few cases where the ratio means do differ significantly in time, their discriminatory power is much poorer in this cross-section, as indicated by the values of the F statistic. The results of the analysis of variance across time are important from the methodological point of view of the further analyses, since significant variation in time would mean that it is purposeful to perform clustering procedures separately for each year. However, the lack of significant differences indicates that for most ratios the time means of variables can be considered as typical ratio levels in the analytical period.

The grouping procedure was carried out in several versions. In all cases the number of clusters was established at 13 so that it corresponds to the number of industries analysed and therefore enables more reliable comparison of groupings. First, the *k*-means clustering was applied with the use of all financial ratios for grouping binominal objects: industries in countries (all size groups). Than the same was repeated for individual countries, where the binominal objects in the form of size groups in industries were clustered, and for individual size groups, where the grouping was again performed on industries in countries. The above steps were then repeated for each category of financial ratios in order to find out which group of diagnostic variables best reflects the industrial classification. Table 6 shows the details of each version of the clustering procedure.

Table 6. Items subject to k-means clustering analysis.

Population	Ratio category					
	All ratios	Profitability	Working capital	Financial situation	Asset structure	Capital structure
Total	Industries in countries (IND_CT)					
AT	Size groups in industries (IND_S)					
BE						
DE						
ES						
FR						
IT						
NL						
PL						
PT						
S	Industries in countries (IND_CT)					
M						
L						

Note: IND – industry, CT – country, S – size group

It is clear from the number of rows and columns in Table 6 that the grouping procedure was performed 78 times. Therefore, due to the abundance of clustering results, the details of the grouping results were presented only for the total population and for all financial ratios. They are shown in Table 7.

Table 7. Clustering results of industries in countries
for the total population and all ratios.
Source: author’s calculations based on the BACH-ESD database.

Cluster number												
1	2	3	4	5	6	7	8	9	10	11	12	13
At_agr	At_inf	Be_min	Be_agr	Fr_min	Es_prf	At_ele	At_rle	De_min	Be_wat	At_mnf	Be_prf	At_trd
At_min	At_adm	Be_mnf	Be_cst	Pl_agr	Fr_prf	Es_ele	Be_rle	Fr_wat	Es_cst	At_cst	Es_min	Be_trd
At_wat	De_ele	Be_ele	De_cst	Pl_min	Pt_prf	Es_wat	De_wat	It_min	It_wat	At_prf	Es_rle	De_trd
At_trs	De_trs	Be_inf	Es_mnf	Pl_ele		Fr_ele	De_rle	Nl_mnf	It_cst	De_mnf	Pt_rle	De_hot
At_hot	De_prf	Be_adm	Fr_agr	Pl_wat		It_ele	Es_trs	Nl_prf	It_rle	Es_trd		Fr_trd
Be_trs	Es_adm	De_inf	It_agr	Pl_inf		It_trs	Fr_rle	Pt_inf	Pt_agr	Fr_mnf		Pl_trd
Be_hot	Fr_trs	Es_inf	It_mnf	Pl_prf		Nl_ele	Nl_wat		Pt_cst	Fr_cst		
De_adm	Fr_hot	It_inf	It_prf			Pl_rle	Pt_wat			It_trd		
Es_agr	Fr_inf	Nl_inf	It_adm			Pt_min	Pt_trs			Nl_min		
Es_hot	Fr_adm		Pl_cst			Pt_ele	Pt_hot			Nl_cst		
It_hot	Nl_adm		Pt_mnf							Nl_trd		
Nl_agr										Pl_mnf		
Nl_trs										Pt_trd		
Nl_hot												
Pl_trs												
Pl_hot												
Pl_adm												
Pt_adm												

If the grouping results corresponded ideally to the industrial classification, than each cluster would be made up of nine items representing the same industry but different countries. Obviously, such clustering results are highly unlikely. In fact, there are only two clusters (number 6 and 13) where the items belong to just one industrial section. Other clusters, however, are mixtures of both different countries and industries and in a number of cases it is not obvious which feature prevails. Therefore, in order to evaluate the similarity between the clustering results and the NACE classification, it is convenient to apply a more formal measure. The adjusted Rand's measure was calculated for all the 78 clustering results. The values of the similarity measure for the total population, as well as for each country and size group are shown in Figure 1.

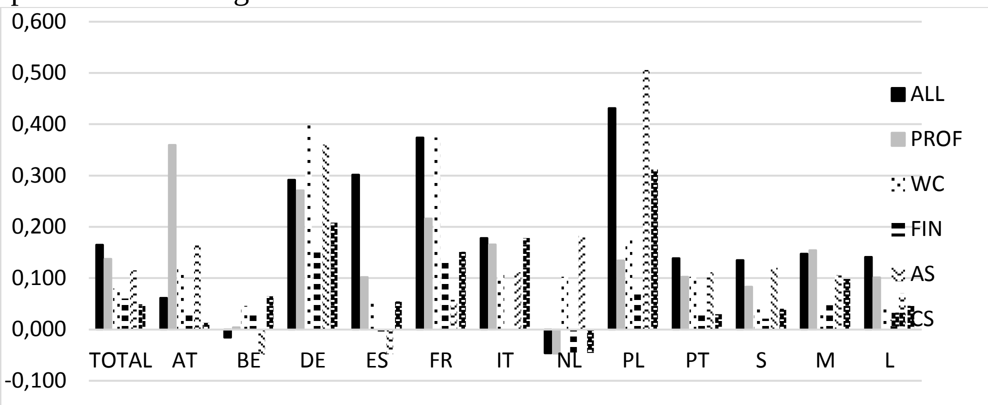

Figure 1. Similarity evaluation between individual grouping results and the NACE - adjusted Rand's measure values.

The higher the values of the adjusted Rand's measure, the more similar the compared groupings, i.e. the higher the correspondence between the NACE classification and the classification of industries in countries or industries in size groups based on corporate performance ratios. The values below 0,5 indicate low similarity, which is the case for the vast majority of the above comparisons. The negative values indicate that the compared clustering results are dissimilar, as in the case of e.g. assets structure ratios for Belgium and most of the ratio categories for the Netherlands. When the Rand's measure is around 0,5, as in the only case of Poland for the assets structure ratios, the similarity can be considered as moderate.

It should not be very surprising that the groupings based on all ratios correspond better to the NACE industrial classifications than the groupings based on narrower categories. The broader the range of financial features, the more detailed the object characteristics. However, this is not always the rule, as there are exceptions in terms of both countries (Austria, Belgium, Germany, the Netherlands and Poland) and size groups (medium enterprises), where the clustering results based on some individual categories of financial ratios bear more resemblance to the industrial code. On average, the assets structure and profitability are those categories of financial ratios which best reflect the industrial diversity, contrary to the financial situation ratios.

Conclusion

One of the most general conclusions resulting from the analytical comparison of the grouping results between the NACE industrial classification and the cluster analysis based on financial ratios is the poor resemblance between these two categorisation systems. It is also noticeable that, in general, a broader range of ratios involved in the classification procedure results in higher similarity level between the industrial code and clustering results. There are also clear differences in terms of the similarity of grouping results across countries. Germany and Poland are the two countries, where the clustering results based on corporate performance ratios are the most similar (though still weakly) to the NACE classification. The opposite is true for Belgium and the Netherlands, where the groupings are most dissimilar when compared with the industrial code. No significant differences in terms of similarity level with NACE were observed across size groups of firms.

References:

Abor, J. (2007), Industry classification and the capital structure of Ghanaian SMEs. *Studies in Economics and Finance*, 24(3), 207-219.

BACH-ESD database, http://www.bach.banque-france.fr

Bowen, R. M., Daley, L. A. & Huber, C. C. (1982), Evidence on the Existence and Determinants of Inter-Industry Differences in Leverage. *Financial Management*, 11(4), 10-20.

Cinca, C. S., Molinero, C. M. & Larraz, J. L. (2005), Country and size effects in financial ratios: A European perspective. *Global Finance Journal*, 16(1), 26-47.

Clarke, R. N. (1989), SICs as Delineators of Economic Markets. *Journal of Business*, 62(1), 17-31.

Das, S. & Roy, M. (2007), Inter-Industry Differences in Capital Structure: The Evidence from India. *Finance India: The Quarterly Journal of Indian Institute of Finance*, 21(2), 395-402.

Fertuck, L. (1975), A test of industry indices based on SIC codes. *Journal of Financial & Quantitative Analysis*, 10(5), 837-848.

Fisher, R. A. (1954), *Statistical methods for research workers*, Edinburgh: Oliver and Boyd.

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(1), 115-128.

Gupta, M. C. & Huefner, R. J. (1972), A Cluster Analysis Study of Financial Ratios and Industry Characteristics. *Journal of Accounting Research*, 10(1), 77-95.

Hall, G. C., Hutchinson, P. J. & Michaelas, N. (2000), Industry Effects on the Determinants of Unquoted SMEs' Capital Structure. *International Journal of the Economics of Business*, 7(3), 297-312.

Helg, R., Manasse, P., Monacelli, T. & Rovelli, R. (1995), How much (a)symmetry in Europe? Evidence from industrial sectors. *European Economic Review*, 39(5), 1017-1041.

Heston, S. L. & Rouwenhorst K. G. (1994), Does industrial structure explain the benefits of international diversification. *Journal of Financial Economics*, 36(1), 3-27.

Heston, S. L. & Rouwenhorst, K. G. (1995), Industry and country effects in international stock returns. *Journal of Portfolio Management*, 21(3), 53-58.

Kale, K. M. & Walking, R. A (1996), The Impact of Industry Classification on Financial Research. *Journal of Financial and Quantitative Analysis*, 31(3), 309-355.

King, B. F. (1966), Market and Industry Factors in Stock Price Behavior. *Journal of Business*, 39(1), 139-190.

Koralun-Bereźnicka, J. (2013), *Corporate Performance: A Ratio-Based Approach to Country and Industry Analyses*, Springer.

Lessard, D. (1974), World, national and industry factors in equity returns. *Journal of Finance*, 29(2), 379-391.

Maksimovic, V., Stomper, A. & Zechner, J. (1999), Capital Structure, Information Acquisition and Investment Decisions in an Industry Framework. *European Finance Review*, 2(3), 251-271.

Martin, L. J. & Henderson, G. V. (1984), Industry influence on financial structure: A matter of interpretation. *Review of Business and Economic Research*, 19(2), 57-67.

Meric, G. & Meric, I. (1983), Industry influence on financial structure under conditions of inflation. *Financial Review*, 18(3), 45.

Meyers, S. L. (1973), A re-examination of market and industry factors in stock price behavior. *Journal of Finance*, 28(3), 695-705.

Omran, M. & Pointon, J. (2004), The Determinants of the Cost of Capital by Industry within an Emerging Economy: Evidence From Egypt. *International Journal of Business*, 9(3), 238-258.

Phillips, N. M. & MacKay, P. (2005), How Does Industry Affect Firm Financial Structure? *The Review of Financial Studies*, 18(4), 1433-1466.

Rand W. M. (1971), Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846-850.

Remmers, L., Stonehill, A., Wright, R. & Beekhuessen, T. (1974), Industry and Size as Debt Ratio Determinants in Manufacturing Internationally. *Financial Management*, 3(2), 24-32.

Roll, R. (1992), Industrial structure and the comparative behavior of international stock market indices. *Journal of Finance*, 47(1), 3-42.

Scott, D. F. & Martin, J. D. (1975), Industry influence on financial structure. *Financial Management*, 4(1), 67-73.

Sell, C. W. (2005), The Importance of Country versus Sector Characteristics. *Managerial Finance*, 31(1), 78-95.

Talberg, M., Winge, Ch., Frydenberg, S. & Westgaard, S. (2008), Capital Structure Across Industries. *International Journal of the Economics of Business*, 15(2), 181-200.

Wishart, D. (2001), k-Means Clustering with Outlier Detection, Mixed Variables and Missing Values, Proceedings of the German Classification Society, www.clustan.com (12.9.2010.)