A Latent Class Cluster Analysis Study of Financial Ratios and Industry Classifications between Japan and Malaysia

by

Yong Sook Lu

楊淑茹

PhD Econ, Lancaster University, United Kingdom

M. Econ (Distinction), University of Malaya, Malaysia

B. Econ (First Honour), University of Malaya, Malaysia

Report submitted to Japan Center for Economic Research (JCER)

November 2012

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to Nikkei (Nihon Keizei Shimbun), Inc. and Japan Center for Economic Research (JCER) which funded my scholarship (the Nikkei Asia Scholarship). I am grateful to the University of Malaya (UM) for granting my research leave and providing me with a platform and an opportunity in my career to teach and to undertake research in the Department of Economics.

My sincere thanks to all JCER and Nikkei staff, especially Mr. Shigesaburo Kabe, Mr. Go Yamada and Mr. Masaki Yamazaki for their support, guidance and advice during the entire research period. Ms. Tamami Kikuchi also deserves a special mention for her kind assistance whenever any help was needed.

Last but not least, for those who contributed to this report, a great appreciation goes to them. I have tried my very best to complete this report and I hope it will beneficial to all.

Table of Content

List of Tables

List of Figures

Chapter 1 - Introduction

One of the important aspects of industrial organisation is firm behaviour or characteristics according to which firms could be classified. Problems regarding the classification of industries using economic data have been discussed for the past 50 years. Financial analysts always argued that firms adjust their financial ratios according to industry-wide averages. According to Weiner (2005), the central point of each classification is to determine a balance between aggregation of similar firms and differentiation between industries. Gupta and Huefner (1972) examined the differences in financial ratio average between industries by using the cluster analysis. The conclusion of this study was that differences do exist in ratio means amongst industry groups.

Grogaard *et al.* (2005) examined the role of industry factors in the internationalisation patterns of firms. The study shows that high concentration ratios in the home market do not generally provide an impetus to internationalisation. The authors found that internationalisation patterns are also influenced by industry characteristics that drive or hinder internationalisation.

Interestingly, the focus of these studies is on evaluating the industrial classifications systems. For instance, Guenther and Rosman (1994), and, Kahle and Walking (1996) compared Compustat and CRSP [1] SIC [2] codes; Bhojraj *et al.* (2003) compared

[1] The University of Chicago Center for Research in Securities Prices
[2] Standard Industrial Classification

classifications based on codes developed by Fama and French (1997), GICS[3], NAICS[4] and SIC.

Research on classification of firms listed in a stock exchange is relatively scarce. Security markets make an important contribution to the integration process and the momentum of economic growth. Securities markets are organised exchange plus over-the-counter (OTC) markets in which securities are traded. In the era of K-economy and globalisation, the trading process at the securities markets is changing very rapidly. In the light of all these changes, it is to be expected that the behavioural of investors and the structure of firms will also be changing. But many investors follow strategies based on identifying stocks that are matched along economically relevant dimension but which trade at different valuations, perhaps indicating relative mispricing.

Therefore, financial researchers, analysts, decision-makers, investment researcher and portfolio management frequently conducted along industry classifications and grapple with the issue of identifying homogeneous groups of stocks. Academic researchers and investment practitioners follow a variety of approached to construct homogenous stock groupings. Perhaps the most popular method of delineating sets of economically similar stocks is to follow their industry affiliation. However, stocks can be categorised into homogeneous groups using criteria other than industry affiliation. Alternatively, classifications of stocks can be based on similarity with respect to firm attributes such as market capitalisation, operating performance or valuation ratios, for instance.

[3] Global Industry Classification Standard
[4] North American Industrial Classification System

This study proposes to examine and compare the classification of the companies listed in Bursa Malaysia and Tokyo Stock Exchange. Bursa Malaysia (Bursa), the only stock exchange in Malaysia, has 823 companies listed on the Main Market and 117 companies listed on the Ace[5] Market with a combined market capitalisation of US$429 billion as of March 2011. Meanwhile, the Tokyo Stock Exchange (TSE) is the third largest stock exchange in the world by aggregate market capitalisation of its listed companies. It had 2292 listed companies with a combined market capitalisation of US$3.3 trillion as of December 2011.

According to Clarke (1989), an examination of an industrial organisation entails the study of similarities and differences between industries based on the structure of firm sizes within industries, the manner in which firms conduct their business, and their economic performance. He argued that it is essential that firms be divided into homogeneous economic industries if such comparisons between industries, or within industries over time, are to be valid. He proposed a dummy variable model to test how well the hierarchical SIC separated groups of firms into economically distinct industries. He found that the SIC codes were more effective at dividing firms into broad industrial groups than at dividing firms into three- and four-digit segments to represent economic markets more closely. He also commented ‘until a more useful classification system is proposed and implemented, empirical IO research should be carefully evaluated to determine its sensitivity to the use of a possibly spurious classification system’ (p. 29).

[5] ACE's mean; A is "access", C is "Certainty" and E is "Efficiency".

Elliott et al. (2000) used cluster analysis and analysis of variance on the UK SIC and investigated whether, when data are aggregated from the fourth to the second digit, they are sorted into categories that can be equated meaningfully with industries. They justified for continuing with board SIC groups as a proxy for similarity in factor intensities.

Chan et al. (2007) used hierarchical cluster analysis to assign stocks to groups so as to minimise the average within-group distance between group members, where distance if measured as one minus the correlation coefficient between the two stocks' returns. They found that GICS codes compare favourably to purely statistical classification with respect to producing homogeneous sets of stocks.

Therefore, the present study contributes to the limited literature in the area. This study has taken a similar approach as Elliott et al. (2000) but instead of using the traditional cluster analysis, we employed the latent class (LC) cluster analysis approach to probe the appropriateness of the industry classifications in the SIC/Nikkei, a method which, to our knowledge, has not been used before for this purpose. We use the technique to explore whether disaggregated 3-digit headings can be justifiably aggregated into 2-digit sections. The paper sheds new light on to this problem by addressing alternative means by which to adjudge whether or not an industry can be considered to be homogeneous. The structure of the companies will be examined to answer the question whether the current sectoral classification is sufficient to classified companies into homogeneous groups, pertaining to their economic activities, and more importantly, the firm characteristics that are reflected in the financial ratios for both countries. The results are new and lead us to believe that this is an approach which could be used more widely.

1.1 Objectives and Significance of Study

The key objective of this study is to examine the classification of firms listed in Bursa and TSE according to sectors/industries determined by their key economic activities. Financial ratios of firms are evaluated to investigate if they are good indicators of industry characteristics as suggested by the sectoral classification. The homogeneity of firm in terms of their industrial organisation within each sectoral classification is studied, with the aim of establishing if the current classification system is sufficient to separate firms into homogeneous groups.

The industrial organisation of companies is often an important criterion for the investors in a stock market in their investment decision making process. For portfolio diversification purposes, investors may choose to spread their investments in stocks that are of different firm characteristics. Often, the sectoral classification is a useful guide as a grouping of companies that are homogeneous in economic activities, and to some extent, a reflection of their financial similarities. The results of the study are useful for evaluating the relevance of the current sectoral classification system to meet the purpose as guidance to investors.

Allocating companies to industrial groups is, however, by no means easy in practice. For single product firms this exercise is normally relatively simple. However, for multiproduct firms, the act of classifying them into industries can be very difficult. Of course, sub-groups within each classification systems can also be very varied. Thus, for example, whilst one industry group may genuinely comprise of similar firms all competing for market share, in other groups segments may exist meaning that actual competition between at least some of the companies in a group is minimal. This can also have implications for modelling as it determines to what extent the measured values of market share really do represent market power and the associated ability to extract economic rent.

The work therefore provides a rich menu of hypotheses and propositions for further analysis, some of which are listed below:

(i) How are firm/company characteristics related to firm-level financial ratios?

(ii) How are financial ratios related to the sectoral classification in Bursa/TSE?

(iii) Are the current classification systems successful grouping companies that are homogeneous in terms of firm characteristics?

(iv) Does the current classification systems create competition or harmonisation among companies listed within the same sectoral group?

(v) Is there a need for an alternative classification system? If so, what is the alternative?

The remainder of the report is organised as follows: the next Chapter introduces the related prior studies, the definition of industry (homogeneous versus heterogeneity) and different ways of classification, the third explains our cluster analysis approach, the fourth Chapter gives the empirical results and the final Chapter concludes.

Chapter 2 - Literature Review

This Chapter reviews some of the definitions of industry. The Chapter also provides certain definitions of homogeneity and heterogeneity, and its impact on firm performance. Financial ratios and other methods which are used to classify firms are discussed, as is the issue of whether to use single or multiple indicators to determine company industry groupings.

2.1 The Definition of Industry

The definition of an industry, and a market, is a controversial issue amongst industrial economists. It is very rare for ‘all firms’ within an industry to sell one commodity, normally they sell different commodities or products. This means that it is very difficult to partition firms, according to the nature of the products they produce, into non-overlapping groups. This, in turn, makes industry groupings somewhat problematic.

According to Robinson (1956), an industry may be regarded as a group of firms having certain technical characteristics in common. They may be bound together by using similar methods of production, dealing with the same suppliers or supplying the same dealers, drawing upon the same labour force, or merely by a historically determined sentiment that they are members of an industry (p. 361). However, according to Davies (1955) ‘a precise and meaningful definition of an industry is a vain objective’ (p. 710). According to this author, there is no advantage (and much error) in making definitions of words more precise than the subject matter that they refer to. But rough working demarcations of industries are required, for instance, by a ‘group of business-men considering who is eligible to join a trade association or by the compilers of the Census of Production’ (Robinson, 1956, p. 361).

Davies (1955) wrote: ‘any attempt to reconcile the views of those economists who regard the “group and the industry as useless concepts” and those of businessmen, who still seem to think and act as if there were such things is to be commended’ (p. 710). Robinson (1956) also pointed out that ‘questions relating to competition, monopoly and oligopoly must be considered in terms of markets, while questions concerning labour, profits, technical progress, localisation and so forth have to be considered in terms of industries’ (p. 361).

An additional contribution to the debate comes from Bain (1951) who wrote: ‘the industry may be viewed in derivative fashion as a group of firms or divisions thereof, so far as the firms or divisions thereof all produce entirely (or, for rough purposes, almost entirely) within the close-substitute output group’ (p. 298). He also commented that it is quite possible for a firm, or even a plant, to produce simultaneously in several theoretical industries.

Andrews and Brunner (1975) remarked that the very existence of industrial economics demonstrates the usefulness and intuitive appeal of the industry as a concept, but even the practitioners in this field fail to deal effectively with the definitional problem. The authors suggested that an ‘industry’ is any grouping of firms which operate similar processes and could produce technically identical products within a given planning horizon. These groupings by no means exhaust the constraints on the behaviour of any particular firm. They encompass constraints on the input side as each firm would be using very similar inputs of labour, raw material and machinery. A market, by contrast, is the institution within which a firm attempts to sell its output or buy an input. A firm's behaviour is constrained by other firms selling in that market and by the behaviour of buyers in the market [cited from Nightingale (1978, pp. 35-36)].

Researchers have generally used the Standard Industrial Classification (SIC) system for assigning firms to industries. SIC codes, established in 1937, which aggregate firms selling related end-products or using similar production processes into an industry, have traditionally been used for this purpose.

SIC codes are commonly used in empirical research to measure industry membership in cross-sectional regression (e.g. Francis and Reiter (1987), Kim and Schroeder (1990)), to match firms by industry (e.g. Hand et al. (1990), Ghicas (1990)), to control for industry-specific cross-sectional correlation of abnormal returns (e.g. Biddle and Seow (1991), Xiang (1993)), to identify homogeneous industries for study of intra-industry information transfer (e.g. Foster (1981), Baginski (1987), Han et al. (1989), Han and Wild (1990), Freeman and Tse (1992), Szewczyk (1992)), and to include or exclude certain industries from samples (e.g. Mohrman (1993), Thomas and Tung (1992)).

However, changes in the variety of products, the growing importance of services, together with shifts in technology and the makeup of businesses, have called into question the usefulness of the SIC system. In 1999, the major statistical agencies of Canada, Mexico and the United States began implementing the North American Industry Classification System (NAICS). The new scheme changed industry classification by introducing production as the basis for grouping firms, creating 358 new industries, extensively rearranging SIC categories, and establishing uniformity across all North American Free Trade Agreement (NAFTA) nations.

Since 1997, Fama and French (FF) classification has been influential, being widely used in many academic studies on asset pricing. It start from firms' 4-digit SIC codes and reorganise them into 48 industry groupings. Their aim was to address some of the more obvious problems with the SIC codes by forming industry groups with firms that were more likely to share common risk characteristics. FF do not provide any evidence on how well their classification system produces groups of economically similar firms.

In the investment community portfolio managers and analysts have gravitated to the Global Industry Classification System (GICS). The GICS is an industry taxonomy developed by MSCI (Morgan & Stanley Capital International) and Standard & Poor's (S&P). The purpose was to enhance the investment research and asset management process for financial professionals worldwide. Despite using the production-oriented and supply-based approach by SIC in delineating industry categories, GICS classified companies based on their principal business activity.

Table 2.1 provides brief information on the historical development and basic doctrine behind each of the competing classification algorithms.

Table 2.1: Background Information on Different Industry Classification Schemes

SIC (Standard Industry Classification)	Oldest of five. Established in 1937 by an inter-departmental Committee on Industrial Classification operating under the jurisdiction of the Central Statistical Board. The objective was to develop a plan of classification of various types of statistical data by industries and to promote the general adoption of such classification as the standard classification of the Federal Government.
NAICS (North American Industry Classification System)	Jointly developed by governmental statistical agencies in Canada, Mexico and United States in 1999. Aimed to improve the SIC by using a production-based framework throughout to eliminate definitional difference; identifying new industries and reorganizing industry groups to better reflect the dynamics of our economy; and allowing the first-ever industry comparability across North America.
ISIC (International Standard Industrial Classification)	The ISIC of all economic activities is a United Nations (UN) system for classifying economic data according to kind of economic activity in the fields of production, employment, gross domestic product and other statistical areas. ISIC is a basic tool for studying economic phenomena, fostering international comparability of data, providing guidance for the development of national classifications and for promoting the development of sound national statistical systems.
FF (Fama & French) Algorithm)	Developed by financial academics, namely Fama and French in 1997. Their aim was to address some of the more obvious problems with the SIC codes by forming industry groups with firms that were more likely to share common risk characteristics.
GICS (Global Industry Classifications Standard)	Collaboration between Morgan Stanley Capital International (MSCI) and Standard and Poor's (S&P). The purpose was to enhance the investment research and asset management process for financial professionals worldwide. Despite using the production-oriented and supply-based approach by SIC and NAICS in delineating industry categories, GICS classified companies based on their principal business activity.

Source: Bhojraj et al. (2003, pp. 7-9) and Palanyandy and Talha (2002).

2.2 Impact of Industrial Classifications on Firm Performance

Industrial organisation of companies is often an important criterion for industrial economists studying an industry by looking at their profit making process. Although the groupings employed by almost all authors have been the same, i.e. the SIC industrial classification, and the assumption has been made that classification by end product is a suitable technique for grouping in almost all studies (Wippern, 1966; Arditti, 1967; Clarke, 1989), a host of evidence exists that grouping by industries is not particularly suitable for most of the purposes for which it is employed (Elton and Gruber, 1971, p. 434). There are two issues here: the levels of the classification such as 1, 2, 3, 4 and 5, where the lower level industry groups contain firms that are more heterogeneous, and even within a three-digit scheme where some industries will be more homogeneous than others. Therefore, the crucial point here is that both market share and, therefore, market concentration ratios (e.g. CR4 and HI) will be subject to measurement error if the total size of the market is incorrect.

Measurement errors can be divided into two components: random error and systematic error (Taylor, 1982). Experimental uncertainties that can be revealed by repeating the measurements are called random errors; those that cannot be revealed in the way are called systematic. For example, we have to measure market share within an industry. However, there is the possibility that the industry has been incorrectly defined; and this source of uncertainty would probably be systematic. For example, if an industry is very heterogeneous and there is little actual competition between companies assigned to it, then the calculated market share measures for the firms in this industry are going to systematically underestimate companies' market power.

At the economy, industry, and firm levels, one would expect to find different income-influencing events, differences in their potential impact on different groups and differences in individual reactions to the events. Therefore, at the very beginning, industry classifications have been of interest to finance researchers. They investigate the grouping of firms and always call for industry classifications because they believe that if accounting earnings numbers reflect these events, and if industry classifications group together firms which are similar in significant ways, then it would be reasonable to postulate an association between the earnings of any particular firm and both (1) the average earnings of the firms which constitute its industry group, and (2) the average earnings of all firms in all groups, that is, the economy (Brown and Ball, 1967, p. 56). The major results of Brown and Ball's study are that 35-40% of the variance in a firm's annual earning is associated with the market (earnings averaged over all firms) and 10-15% can be explained by the variance in the industry (earnings averaged over firms by a two-digit industry classification). They suggested that for future research, one could redefine an industry in terms of the covariability of earnings of firms and investigate whether there is an association between errors in predicting the earnings of a firm and subsequent price behaviour on the stock market' (p. 68).

The first major study of industry effects on market returns was performed by King (1966). He used principal components analysis and clustering techniques on a sample of 63 companies chosen from six two-digit industries (e.g. tobacco products, petroleum products, metals (ferrous and non-ferrous), railroads, utilities and retail stores) based on Security and Exchange Commission (SEC) codes, which are similar to the SIC codes defined by the US Bureau of the Budget Office of Statistical

Standards. King found that about 10 percent of the variance in rate of return could be explained by the industry codes.

Meyers (1973) repeated the King study and showed that for some other industries the industry effect was not as significant as for the ones which King studied. He critised King's finding due to the empirical methods and the sample used. He modified both methods and concluded that, although the finding generally supports King's observations, there was also evidence of industry relationships that were considerable less persuasive than that based on a similar analysis reported by King. However, his sample did not allow an analysis of the effect in one- and three-digit industry classifications.

Later, Fertuck (1975) extended the King and Meyers studies to determine whether the appropriate level of aggregation for studying industry effects was the one-, two-, or three-digit SIC codes. He found that 'SIC codes provide a useful basis for creating an industry index in some industries, and in some industries the industry effect is trivial and can be safely ignored' (pp. 847-848). He also reported that clustering by similarity in past returns did not provide an improved classification.

There have been additional studies that have focussed on evaluating the industrial classifications systems and how differences in SIC code assignment may influence the results of empirical research. For instance, Guenther and Rosman (1994), and Kahle and Walkling (1996) compared COMPUSTAT[6] and CRSP[7] SIC codes. The first of these papers, studied determinants of monthly stock returns and financial statement

[6] The Standard & Poor's Compustat database.
[7] The University of Chicago's Center for Research in Securities Prices (CRSP) database.

ratios and revealed that ‘differences between the database in SIC code assigned to companies by COMPUSTAT and CRSP influenced findings in empirical research’ (p. 127). Kahle and Walkling (1996) also analysed the impact of industrial classification on financial research. They reported that four-digit comparisons generally performed better than two-digit comparisons, and that Compustat SIC codes tended to give more accurate classifications than CRSP SIC codes.

Bhojraj et al. (2003) compared classifications based on codes developed by Fama and French (1997), Global Industry Classifications Standard (GICS), the North American Industry Classification System (NAICS) and the Standard Industrial Classification (SIC). These authors found that the GICS classifications were significantly better than the others at explaining stock return co-movements, as well as cross-sectional variations in valuation-multiples, forecast and realised growth rates, R&D expenditures, and various key financial ratios.

Therefore, the grouping of firms in terms of their industrial organisation within each sectoral/industry classification is important. If the current classification system is not sufficient to separate firms into homogeneous groups, the consequence it that the estimation of firm/industry performance will be biased.

2.3 Industry Homogeneity in Practice

In economics, the term homogeneity usually refers to goods for which consumers do not have material, spatial or personal preference. That is, that they are ready to exchange products on a one to one basis. Homogeneous goods do not have to be identical. In industrial economics, a homogeneous group usually refers to firms that have similar characteristics in terms of their market structure, profits, firm size etc. In contrast, heterogeneous is an adjective used to describe an object, or system, consisting of multiple items having a large number of structural variations. It is the opposite of homogeneous. In reality, economic units are typically quite heterogeneous with respect to their economic performance (Fritsch and Stephan, 2006, p. 17).

The problem of how firms should be grouped together into separate industries is not a new issue, as discussed earlier in this work. One important reason for defining markets and grouping firms into industries is in order to investigate whether any relationship exists between the characteristics of markets and the terms under which transactions are conducted in those markets (Needham, 1978, p. 110). Whether the relationship between price and marginal cost influences the level of aggregate satisfaction in the community is important to predict, or attempt to control, the terms of transactions in any sector of the economy. The need to define markets and to group firms into industries also arises in connection with the application of anti-trust policy.

Although the industry concept is part of everyday life, when one attempts to define an industry operationally, matters are not so simple. People usually group together all those firms that produce the same product or service, but this requires a definition of what constitutes the same product or service. All firms produce different products because the products of two different firms are produced at different geographical locations, but a definition yielding single-firm industries is too narrow for most purposes. In some sense, products and services are the same in that they compete for buyers' purchasing power, but, again, a definition which yields a single economy-wide industry is too wide for most purposes (Needham, 1978, p. 110).

Furthermore, some big and established companies have diversified their investments into different businesses. Tew and Handerson (1959) pointed out that 'the classification of large enterprises engaged in multiple activities according to their main activity must tend to blur the industrial outlines' (p.14). For example, YTL Corporation Berhad is one of the largest companies listed on the Bursa Malaysia, its core-businesses include utilities (power supply and communication), operation and management activities, cement manufacturing, property development, hotels and resorts, technology incubation, construction contracting, real estate investment trust (REIT) as well as carbon consulting.

In all of the standard industrial classifications mentioned earlier, primary emphasis in defining an industry is on the supply side of the economic picture. Most of the industries are defined in terms of establishments primarily engaging in producing a product or group of products that are related by technical process or raw materials used in their manufacture (Needham, 1978, p. 116).

In general, different groupings of firms into industries are likely to result, depending on whether the 'same' product or service means physically identical, using the same process in its manufacture, using the same inputs, performing the same function, having the same price range, is sold in the same geographic location, or some combination of criteria for grouping (Needham, 1978, p. 110).

No single industrial classification could possibly suit all purposes, and criticism of existing standard industrial classifications on the grounds that they do not suit a particular purpose amounts to little more than arid argument concerning which aspect of industry is the most important. Behaviour itself has many different aspects, and different groupings may be appropriate for a study of different aspects of firm behaviour. The important requirement which a standard industrial classification must fulfill is that it should be as complete and detailed as possible, in order that the information contained can be re-grouped to suit the particular purpose of anyone wishing to use the data (Needham, 1978, p. 117).

According to the Australian Bureau of Statistics (ABS), to use statistical information about business units effectively, it is first necessary to organise that information into categories suitable for economic analysis. This can be done by using different classifications depending on the particular interests of users. An industry classification is one way of organising data from a business unit perspective. It provides a standard framework under which units carrying out similar productive activities can be grouped together, with each resultant group being referred to as an industry. The term industry is used in its widest context, covering the full range of economic activities undertaken to produce both goods and services.

Each individual class is defined in terms of a specified range of activities. It is common for a business unit to engage in a range of activities wider than those designated as belonging to a particular class, and when this occurs, the classification is based on its predominant activity. Any activities undertaken which belong to classes other than that to which the unit is primarily classified, are described as its secondary activities. The secondary activities of a unit play no part in assigning the class to which the unit is classified, but are useful for coverage and specialisation ratio analysis. The specialisation ratio shows what percentage of production (or inputs) of units classified in a specific branch are produced (or used) in “typical” products. The coverage ratio shows how many percent of a specific product (or input for a product) come from units classified with the classification code of the producing activity.

All classifications, particularly those that conform to international or national statistical standards, should satisfy a number of fundamental principles. These include that the classification: 1) is comprehensive in its coverage; 2) has categories which are mutually exclusive; 3) has categories which can be readily understood by users and data providers; 4) is hierarchical to support its use for different statistical purposes; 5) should remain stable over a period of time, or be designed so that it can easily be updated; and 6) be based on a strong and consistently applied conceptual framework.

In an industrial classification, each unit has to be classified uniquely to one class, so that only those units with the same predominant activities are brought together to form a class. As indicated above, industry classes should be comprised of businesses that undertake similar economic activities (i.e. they should be as homogeneous as possible).

In most of the official European structural business statistics, the statistical units “enterprise” and “local units” are used to obtain statistical results. But it is obvious that a given statistical unit is often not homogeneous in a sense that it performs more than one activity; i.e. it is diversified. In this case, all variables of the unit are counted under the activity code of the main activity, even if there are other, so-called secondary activities. In order to get consistent results for all levels of an activity classification, the main activity has to be carried out by the “top-down method”. In cases where individual data not only for the main activity, but also for secondary activities can be obtained by surveys, it is possible to investigate if a given classification is suitable for the description of a national economy. In Germany, for example, figures were constructed for a "unit of homogeneous production[8]" by aggregating production-related variables surveyed for main and secondary activities from local units to enterprises[9].

Therefore, we need to disaggregate economic data into meaningful groups in order to better understand and forecast the future course of economic phenomena, and to illustrate with a specific example that such disaggregation can lead to improved results (Elton and Gruber, 1999, p. 3).

2.4 Means of Classification

Companies differ in their financial success, partly because of differences in their environment and partly because of differences in the way they have adapted themselves to their environment. The ostensible purpose of classifying companies to industrial groups is to bring together companies operating in similar environments: an industrial grouping which fully achieved this objective would ensure that all variation between companies in the same industrial group could be attributed to the different managements having responded in different ways, and with varying degrees of success, to identical environment circumstances.

The homogeneity of firms in terms of their industrial organisation within each industry group is an important issue, not only amongst industrial economists but also

[8] Defined by production-related variables, in which input-output tables are aggregated to form homogeneous branches.
[9] Although input-out tables must be supplied to European Statistical Office only every five years, they are compiled and published by the Federal Statistical Office for every reference year (source: Statistisches Bundesamt Deutschland).

finance researchers. ‘Capital market research often calls for firms to be divided into more homogeneous groups, and the most common method for achieving this end is through industry classifications’ (Bhojraj et al., 2003, p. 1). The industry classification might not be the best classification but the main purposes in grouping similar companies in capital markets are the evaluation of the risk-return characteristics of their securities and portfolio diversification decisions. In business, numerous classifications of business firms (or operating units within firms) into industries, regions, risk classes, etc. exist (e.g. Jensen (1971)). Investment services provide a wide variety of lists of recommended securities classified into groups, where companies within a given grouping are perceived by the analyst as ‘similar’ with respect to anticipated price appreciation, yield, and risk (Jensen, 1971, p. 37). An industry classification scheme is used to separate firms into finer partitions, with the expectation that these partitions will then offer a better context for financial and economic analysis (Bhojraj et al., 2003).

Methods of grouping or classifying ‘similar’ entities based upon measured characteristics possessed by each entity appear in the literature under various synonyms, e.g. cluster analysis, grouping methods, classification theory, numerical taxonomy, and clump theory (Jensen, 1971).

Classification techniques group similar objects. An important requirement of a classification scheme is that it considers the salient properties of objects. Such considerations may vary. Therefore, the characteristics deemed relevant and important depend on the task at hand; only the relevant characteristics can provide a meaningful profile of an object (Narayanaswamy, 1996, p. 3).

Firm behaviour, or characteristics, are important aspects of IO. Problems regarding the classification of industries using economic data have been widely discussed. Table 2.2 summarises past studies that have attempted to group companies by homogeneous industrial characteristics, using different methods and financial ratios.

Therefore, industrial economists are always interested in analysing a group of companies smaller, and more homogeneous, than any broad industrial group. The question is how are we going to classify the companies into more homogeneous group? Besides using the financial ratios, what are the other important indicators to determine a company grouping? Should companies be classified into industries using a single indicator or multiple indicators?

Table 2.2 : Summary of the Methods and Financial Ratios used in Classification

Authors	Methods	Financial Ratios	Results
Farrell Jr. (1974)	Stepwise clustering procedure[10], correlation matrix of stock return residuals[11], index procedure[12], and proportion of variance due to group factors[13]	Covariance of returns	It was considered appropriate to assign a factor to the explanation of the variance of returns of a common stock additional to market, industry, and company, based upon a system of classification corresponding to (1) growth, (2) stable, (3) cyclical, and (4) oil stocks.
Gupta and Huefner (1972)	Cluster analysis[14]	Financial ratio average	Differences found in ratio means amongst industry groups
Jensen (1971)	Cluster analysis	Input characteristics and risk characteristics	(1) Single-entity clusters in later stages of the clustering program contained mostly companies which performed 'better-than-average' in the post 1954-65 period, and (2) companies within given multiple-entity clusters performed 'more alike' ex-post in the 1954-65 period than companies clustered on the basis of a random technique.
Pinches et al. (1973)	Factor analysis[15]	Financial ratios such as total income/sales, cash flow/total assets, EBIT/total assets, current liabilities/net plant, quick assets/total assets etc.	Grouped financial ratios into seven patterns: return on investment, capital intensiveness, inventory intensiveness, financial leverage, receivables intensiveness, short-term liquidity and cash position. The results indicated that meaningful empirically-based classifications of financial ratios can be determined and that the composition of these groups are reasonably stable over time, even when the magnitude of the financial ratios are undergoing change.

[10] A stepwise clustering procedure is able to divide a large number of variables into sub-groups from highly intercorrelated to less intercorrelated.
[11] A correlation matrix of stock return residuals is a 'direct inspection of the correlation matrix of the residuals of the stock returns'.
[12] The index procedure is the 'forward selection procedure of determining the most appropriate variables to include in a multiple regression equation.
[13] The proportion of variance due to group factors mean to examine the effect of the magnitude from each classification system groups in explaining the variance in rate of return of common stock.
[14] The cluster analysis approach has the ability to search for natural groupings among companies and partition them into clusters with similar industrial characteristics.
[15] Factor analysis has been used to combine two or more variables into a single factor, and, identification of groups of inter-related variables, to see how they are related to each other.

Sudarsanam and Taffler (1985)	Multiple discriminant analysis (MDA)[16]	Operating scale, fixed capital intensity, labour-capital intensity, profitability, asset turnover, short term asset intensity, net trade credit and financial leverage	There were significant differences between the 14 Stock Exchange Industrial Classification (SEIC) industries considered, several appeared to be largely non homogeneous with respect to certain of their fundamental economic and structural characteristics.
Gupta (1969)	Inter-industry variations analysis	Activity ratios, leverage ratios, liquidity ratios and profitability ratios	1) Activity ratios and leverage ratios decreased with an increase in the size of the corporation, but they increased with the growth of the corporation; 2) liquidity ratios rose with an increase in the size of the corporation but they fell with growth rates, and 3) the growth rate and the profitability ratios showed no regular pattern, but the larger-sized corporations tended to have higher profit margins on sales than the smaller-sized corporations (pp. 528-529).
Lev (1969)	Partial adjustment model	Short-term liquidity ratios, long-term solvency ratios, short-term capital turnover ratios, long-term capital turnover ratios and return on investment ratios	Financial ratios were periodically adjusted to their industry means.
Rayment (1976)	Analysis of variance	Factor intensity such as value adder per head, wages and salaries per head, and non-wage value added per head	He concluded that if total manufacturing had been disaggregated in the United Kingdom census into more than 144 separate branches, then the within branch variance of factor intensity at the 18 and 28 branch levels of aggregation may have turned out to be even greater.
Bradley et al. (1984)	Analysis of variance	Average firm leverage ratios	Leverage ratios were strongly influence by industrial classification.

[16] The objective is to derive functions of these variables which jointly maximize group separation. The maximum number of functions is the smaller of the number of groups minus one or the number of discriminating variables.

Chapter 3 - Data and Methodology

This Chapter discusses the data sources, the variables and the methodology that are used to test the industry homogeneity, with the empirical results following in Chapter four.

3.1 Description of Companies, Data and Variable construction

Bursa Malaysia (Bursa), the only stock exchange in Malaysia, does not use the Security and Exchange Commission Codes[17]. Many people including economists, investment analysts, financial analysts, market analysts, shareholders and investors have difficulty figuring out the exact nature of a company's business based on current classification. On the other hand, the Tokyo Stock Exchange (TSE) is the Japan's largest stock exchange, and third largest stock exchange in the world by market capitalisation.

To construct the sample for the proposed study, two different databanks are employed: (a) Bursa Malaysia database and (b) Financial QUEST 2.0 database. The financial data for Japanese listed companies is obtained from Financial QUEST 2.0 database. The sample used here comprises the population of the 416 and 1511 firms listed in Bursa and TSE, respectively, for the year 2010. The information stems from balance sheet items, and from profit and loss and cash flow statements. Financial information

[17] US Securities and Exchange Commission has the Division of Corporation Finance whose mission is to see that investors are provided with material information in order to make informed investment decisions – both when a company initially offers its stock to the public, and on a regular basis as it continues to give information to the marketplace. The Division uses the Standard Industrial Classification (SIC) code list to indicate the company's type of business and as a basis for assigning review responsibility for the company's filings. http://www.sec.gov/info/edgar/siccodes.htm (Accessed date: 25 October 2009)

retrieved from these databases included total sales, total assets, current assets, current liabilities, revenues, shareholder equity, gross profit, and earnings before interest, taxes, depreciation and amortisation (EBITDA).

In this study, profitability (profit), asset turnover (AT), the current ratio (CR) and stockholder's equity (SE) are used to group together companies who share similar characteristics. These four financial ratios are selected because they represent the firm from different perspectives, such as the quality of the firm in terms of its ability to generate gains or losses; its market liquidity towards debt payment; its efficiency towards investment and its net worth. Brief definitions of these financial ratios are presented in Table 3.1.

The samples are next classified into nine (for Malaysia) and 16 (for Japan) industries based on SIC and Nikkei industrial classification, respectively.

Table 3.1: Financial Ratios Definition

Financial ratio	Definition	Notes
Profitability	Measures the ability of a firm, or an investment to make a profit after costs, overheads, etc.	In this thesis, it calculated as earnings before interest, taxes, depreciation and amortization / total assets.
Current ratio	Measures whether or not a firm has enough resources to pay its debts over the next 12 months or an indication of a firm's market liquidity and ability to meet creditor's demand. It compares a firm's current assets to its current liabilities.	For most industrial companies, 1.5 is an acceptable CR. If CR<1, the company may have problems meeting its short-term obligations. If CR is too high, the company may not be efficiently using its current assets and may be doing a poor job of investing it.
Asset turnover	Measures the efficiency of a company's use of its assets in generating sales revenue or sales income to the company. It is calculated by dividing revenue or sales by assets.	The higher the number, the better, although investors must compare a business to its industry. The higher a company's asset turnover, the lower its profit margin tends to be (and vice versa).
Stockholder equity	Measures the ownership equity spread among shareholders. It is calculated as share capital plus retained earnings minus treasury shares.	The individual investor is interested not only in the total changes in equity, but also in the increase/decrease in value of his own personal share of the equity.

3.2 Latent Class Model

This study provides a micro firm level analysis of companies listed in Bursa and TSE. Financial ratios of companies (e.g. profitability ratio, current ratio, asset turnover ratio and stockholder's equity) will be used as measures of company characteristics.

An attempt will be made to group the companies by homogeneous industrial characteristics using cluster analysis. Generally, the cluster analysis approach has the ability to search for natural groupings among companies and partition them into clusters with similar industrial characteristics. Going beyond the traditional cluster analysis, the latent class (LC) cluster analysis is employed to separate firms into homogeneous and heterogeneous groups.

The recent increase in interest in LC models is due to the development of extended computer algorithms, which allow today's computers to perform latent class analysis on data containing more than just a few variables. In addition, researchers are realising that the use of LC models can yield powerful improvements over traditional approaches to cluster, factor, regression/segmentation, as well as to multivariable bi-plots and related graphical displays. Traditional models used in regression, discriminant and log-linear analysis contain parameters that describe only relationships between the observed variables. LC models (also known as finite mixture models) differ from these by including one or more discrete unobserved variables. LC models do not rely on the traditional modelling assumptions which are often violated in practice (linear relationships, normal distributions, homogeneity). Hence, they are less subject to biases associated with data not conforming to model assumptions (Magidson and Vermunt, 2003).

LC models can be useful in several other areas, such as its probabilistic cluster-analysis tool for continuous observed variables, an approach that offers many advantages over traditional cluster techniques such as K-means clustering (Wolfe, 1970; McLachlan and Peel, 2000; Vermunt and Magidson, 2002). Another application is when dealing with unobserved heterogeneity, as happens in mixture regression analysis of multilevel or repeated measurement data (Wedel and DeSarbo, 1994; Vermunt and Van Dijk, 2001) [cited in Vermunt (2010)]. LC models have recently been extended to include variables of mixed scale (nominal, ordinal, continuous and/or count variables) in the same analysis. Also, for improved cluster or segment description, the relationship between the latent classes and external variables (covariates) can be assessed simultaneously with the identification of the clusters. This eliminates the need for the usual second stage of analysis where a discriminant analysis is performed to relate the cluster results to demographic and other variables (Magidson and Vermunt, 2003).

The underlying theory of the LC model posits that individual behaviour depends on observable attributes and on latent heterogeneity that varies with factors that are unobserved by the analyst. Greene and Hensher (2003) proposed to analyse this heterogeneity through a model of discrete parameter variation. Thus, it is assumed that individuals are implicitly sorted into a set of Q classes, but which class contains any particular individual, whether known or not to that individual, is unknown to the analyst[18]. One weakness of the LC model is that it does not readily extend to autocorrelation (Greene and Hensher, 2003).

Three common statistical application areas of LC analysis are those involve clustering of cases (LC cluster models), variable reduction and scale construction (LC factor models), and prediction (LC regression models). In the next sub-section, we discuss Latent Class Cluster Analysis in more detail.

3.2.1 Latent Class Cluster Analysis

Latent Class (LC) cluster analysis is a different form of the traditional cluster analysis algorithms. The old cluster analysis algorithms are based on optimisation according to

[18] Refer to Greene and Hensher (2003), pp. 682-684 for a full explanation.

some criterion[19], but LC cluster analysis is based on the probability of classifying the cases[20]. An important difference between standard cluster analysis techniques (i.e. K-means, hierarchical) and LC clustering is that the latter is a model-based clustering approach which uses estimated membership probabilities to classify cases into the appropriate cluster. This means that a statistical model is postulated for the population from which the sample under study is drawn. More precisely, it is assumed that the data is generated by a mixture of underlying probability distributions.

The two main methods to estimate the parameters of the various types of LC cluster models are maximum likelihood (ML) and maximum posterior (MAP). ML is based on the log-likelihood function derived from the probability density function defining the model. On the other hand, Bayesian MAP estimation involves maximising the log-posterior distribution, which is the sum of the log-likelihood function and the logs of the priors for the parameters. The advantage of MAP over ML is that the latter method can prevent the occurrence of boundary or terminal solutions, i.e. probabilities and variances cannot become zero (Vermunt and Magidson, 2002, p. 96).

LC clustering consistently recovers true structural groups where the traditional algorithms fail. LC cluster analysis is able to test the assumptions of local independence and equal within class variance, and to relax them if they are found to be invalid. LC clustering allows for variables to be nominal, ordinal, continuous, count or any mixture of these, any of which may contain missing values. Different

[19] These criteria typically involve minimising the within-cluster variation and/or maximising the between-cluster variation. Distance measures normally used are Euclidean distance, squared Euclidean distance, city-block (Manhattan) distance, Chebychev distance, power distance and percent disagreement.

[20] http://www.statisticssolutions.com/methods-chapter/statistical-tests/latent-class-analysis/ (Accessed date: 20 May 2010)

scale types (i.e. ranks, partial ranks and discrete choice data) are handled by automatically specifying the appropriate distribution. LC cluster analysis also allows the inclusion of covariates for simultaneous parameter estimation (based on indicators) and descriptive profiling based on covariates. Other advantages of LC clustering are that it allows both simple and complicated distributional forms for the observed variables within clusters. This means that more formal criteria to make decisions about the number of clusters and other model features, as well as scaling/normalised of the observed variables, are not needed. Since LC is based on a statistical model, diagnostics are available to help determine the number of clusters. Table 3.2 shows the differences between the traditional clustering and LC clustering.

Table 3.2: Differences between Traditional and LC Clustering

Cluster method	**Adequate assumptions**	**Allow for different scale types**	**Covariate-based profiling**	**Optimal determination of number of clusters**
K-means	No	No	No	No
Hierarchical	No	No	No	No
Latent Class	Yes	Yes	Yes	Yes

The model selection issue is one of the main research topics in LC clustering. The most popular set of model selection tools in LC cluster analysis are information criteria like Akaike information criterion (AIC) and Bayesian information criterion (BIC) as these statistics weight fit and parsimony by adjusting the log-likelihood (LL) to account for the number of parameters in the model. When comparing the models, the lower the value, the better the fit of the model to the data.

Chapter 4 - Empirical Results

This Chapter provides analysis on the implication of Latent Class (LC) cluster analysis on industry classification. As mentioned earlier, the objective of our analysis is to re-cluster the 3-digit industries according to profitability (profit), asset turnover (AT), the current ratio (CR) and stockholder's equity (SE) to investigate whether members of each 2-digit group retain association after re-clustering.

For Malaysia, inspection of Table 4.1 indicates that primary metal industry (SIC33) has the highest AT. SIC 37 (transportation equipment industry) which is comprised of many automobile firms has the highest SE. Lumber and wood products (SIC 24) has the highest CR and, chemicals and allied products industry (SIC 28) the highest profit rate.

Table 4.1: Summary Mean of Two-Digit Industries for 2010 (Malaysia)

	AT	***CR***	***SE***	***Profit***
SIC20	117.32	3.45	584.86	0.11
SIC24	92.07	7.01	349.60	0.09
SIC26	116.85	1.39	292.78	0.09
SIC28	105.38	3.89	361.90	0.12
SIC32	68.50	4.22	309.54	0.07
SIC33	240.05	3.15	178.07	0.05
SIC34	95.92	1.85	351.42	0.10
SIC36	176.08	2.05	352.70	0.08
SIC37	117.53	2.19	1357.22	0.10

For Japan, Table 4.2 shows that food has the highest AT and profit rate. Drugs have the highest CR and, motor vehicles and auto parts the highest SE.

Table 4.2: Summary of Mean for Two-Digit Industries for 2010 (Japan)

Industry	Profit	CR	AT	SE
foods (01)	0.0471	2.8087	1.1987	65292.45
textile products (03)	0.0089	2.6736	0.6244	37633.78
pulp and paper (05)	0.0389	1.4468	0.8100	64330.39
chemicals (07)	0.0333	2.1541	0.7441	61105.04
drugs (09)	-0.0141	5.6870	0.5504	142370.44
petroleum (11)	0.0372	1.6164	1.0303	197508.75
rubber products (13)	0.0298	1.5606	0.7785	86265.00
stone, clay & glass products (15)	0.0143	1.7665	0.7232	45117.66
iron & steel (17)	0.0058	1.9088	0.6465	99981.67
nonferrous metal & metal products (19)	0.0168	2.3026	0.7576	39375.19
machinery (21)	-0.0038	2.5936	0.6183	47308.95
electric & electronic equipment (23)	-0.0030	3.1882	0.7743	103119.63
motor vehicles & auto parts (27)	0.0032	1.6181	0.9405	275253.31
transportation equipment (29)	0.0466	1.7286	0.7947	32251.08
precision equipment (31)	-0.0338	3.0713	0.6082	39829.78
other manufacturing (33)	0.0083	2.6063	0.8445	33604.58

A LC cluster analysis, where the number of clusters is selected based on the lowest BIC value, clearly shows the four distinct groups for Malaysia (Table 4.3). However, the lack of standardisation means the vast majority of cases (46%) have been put into cluster one. Entering the variables singularly helps us obtain more evenly defined clusters.

Table 4.3: BIC Values from LC Cluster Analysis (Malaysia)

	LL	BIC(LL)	Npar	Class. Err.
1-Cluster	-2350.16	4742.46	8	0
2-Cluster	-2103.85	4297.24	17	0.1015
3-Cluster	-2023.56	4184.08	26	0.1573
4-Cluster	-1990.41	4165.20	35	0.2165
5-Cluster	-1971.51	4174.81	44	0.2406
6-Cluster	-1953.42	4186.04	53	0.2668
7-Cluster	-1932.84	4192.29	62	0.2756
8-Cluster	-1917.67	4209.35	71	0.2896
9-Cluster	-1912.69	4246.81	80	0.272
10-Cluster	-1914.05	4266.93	89	0.2676

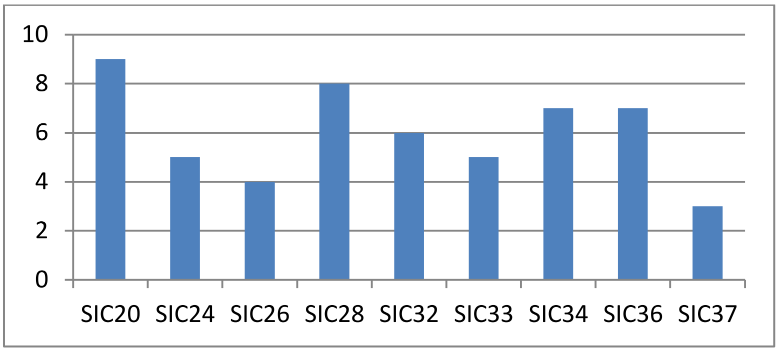
For Japan, Table 4.4 indicates that 16 distinct groups are the most appropriate groupings, and interestingly, it reflects the same number of industry groups as the one grouped under the Nikkei industrial classification.

Table 4.4: BIC Values from LC Cluster Analysis (Japan)

	LL	BIC(LL)	Npar	Class.Err.
1-Cluster	-25374.61	50807.78	8	0
2-Cluster	-20890.19	41904.83	17	0.0133
3-Cluster	-20064.84	40319.99	26	0.0578
4-Cluster	-19655.24	39566.68	35	0.0935
5-Cluster	-19392.80	39107.68	44	0.1047
6-Cluster	-19180.23	38748.43	53	0.1281
7-Cluster	-19096.57	38646.98	62	0.1362
8-Cluster	-18971.55	38462.82	71	0.1688
9-Cluster	-18904.14	38393.87	80	0.1633
10-Cluster	-18837.94	38327.35	89	0.1725
11-Cluster	-18755.15	38227.66	98	0.1773
12-Cluster	-18692.93	38169.09	107	0.1964
13-Cluster	-18682.12	38213.35	116	0.1979
14-Cluster	-18624.35	38163.69	125	0.2024
15-Cluster	-18592.39	38165.66	134	0.2114
16-Cluster	-18537.74	38122.22	143	0.2265
17-Cluster	-18555.01	38222.65	152	0.1942
18-Cluster	-18555.18	38288.87	161	0.202
19-Cluster	-18458.73	38161.84	170	0.2121
20-Cluster	-18460.39	38231.04	179	0.2089
21-Cluster	-18388.10	38152.34	188	0.2295
22-Cluster	-18385.06	38212.15	197	0.2115
23-Cluster	-18412.19	38332.29	206	0.234
24-Cluster	-18336.24	38246.27	215	0.2104
25-Cluster	-18359.84	38359.33	224	0.2197

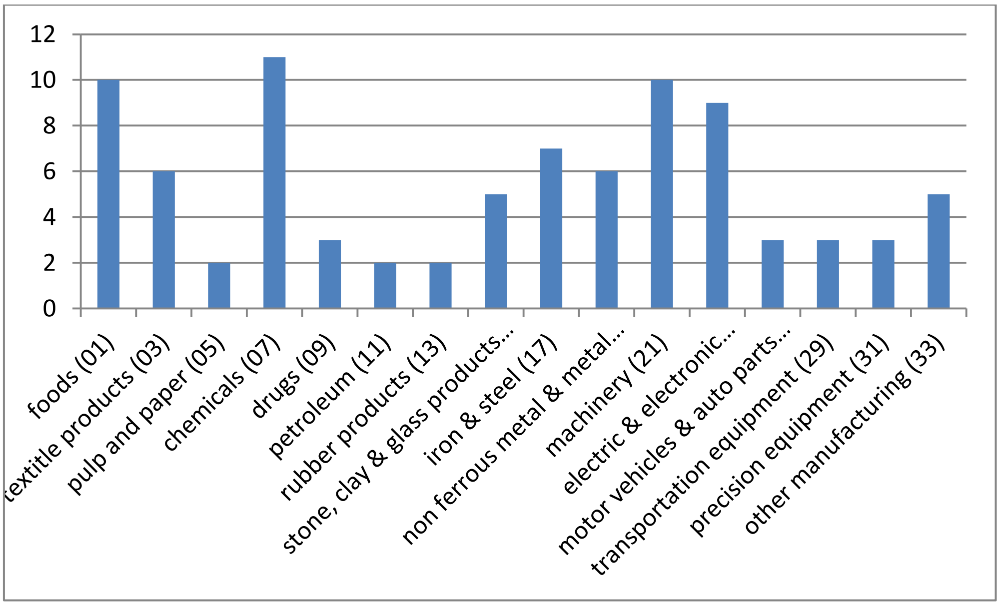
The next step is to examine the existing composition of the 2-digit industries on which to base our forced clustering structure (see Figure 4.1 and Figure 4.2). For Malaysia (Figure 4.1), the two largest groups are SIC 20 and SIC 28 with 9 and 8 sub-industries, respectively, followed by medium-to-high groups that contain 7 sub-industries and medium-to-low groups that contain 5 or 6 sub-industries. The remainder average 3 or 4 sub-industries. The mean group size is 6.

Figure 4.1: Number of 3-Digit Industries in 2-Digit Groups (Malaysia)



For Japan (Figure 4.2), food; chemicals; machinery; and electronic and electrical equipment are the largest groups (9-11 sub-industries), followed by textile products; stone, clay and glass products; iron and steel; non ferrous metal and metal products, and other manufacturing are the medium groups (5-7 sub-industries). The remainder average between 2-3 sub-industries contain pulp and paper; drugs; petroleum, rubber products; motor vehicles and auto parts; transport equipment and precision equipment. The mean group size is 5.5.

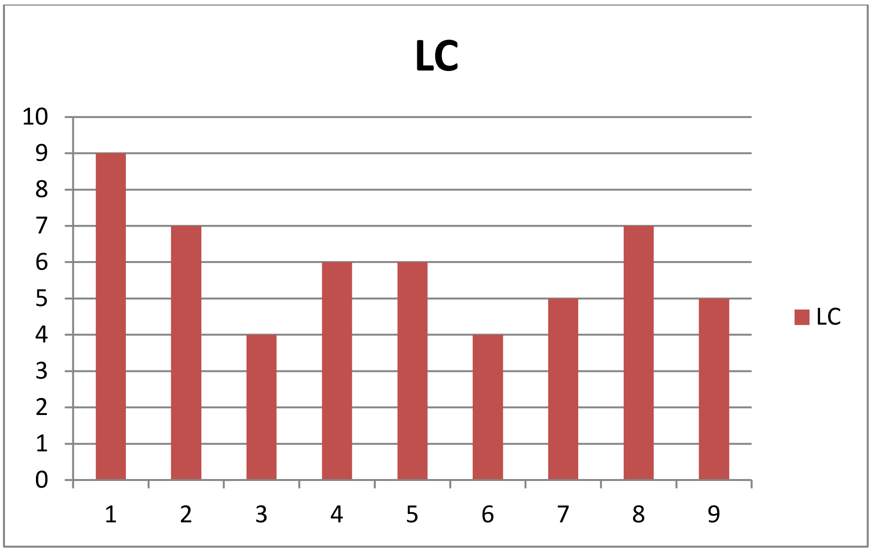
Figure 4.2: Number of 3-Digit Industries in 2-Digit Groups (Japan)



For direct comparison we present in Figure 4.3 and Figure 4.4 the LC cluster analysis results at the 9-cluster and 16-cluster level using the four financial ratios, namely AT, CR, SE and Profit variables for Malaysia and Japan, respectively.

This is just one of the clustering sets that have been generated[21]. Interestingly, the cluster pattern generated using the LC cluster analysis is quite similar to the one in Figure 4.1 and Figure 4.2. The differences that we noted from Tables 4.1 and 4.2 can be clearly seen from the size of certain clusters. At this level, for Malaysia (Figure 4.3), the largest cluster is group one, followed by groups two and eight (medium-to-high), groups four, five, seven and nine (medium-to-low), and groups three and six (low). As the LC clustering continues, these groups become more defined until the four clusters predicted from Table 4.3 begin to exert themselves.

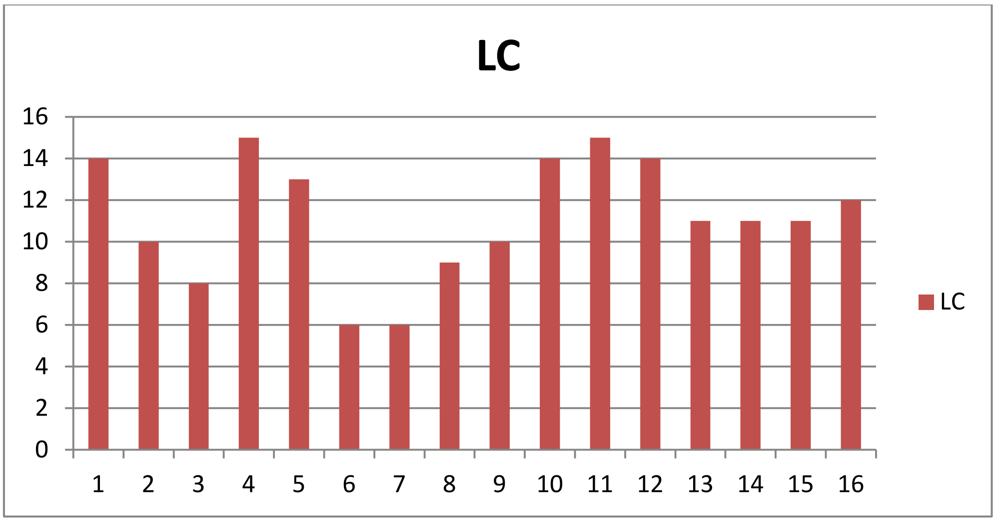
Figure 4.3: Latent Class Cluster Analysis with 9 Clusters (Malaysia)



[21] For robust checking, we entered each variable separately and used another set of financial ratios (i.e. long-term debt-assets, returns on assets, price-earnings, return on capital employed, net earnings per share, operating profit-sales and book value per share) in our study, and they showed the same results.

For Japan (Figure 4.4), the largest cluster is groups one, four, five, ten, eleven and twelve followed by groups two, eight, nine, thirteen, fourteen, fifteen and sixteen (medium), groups three, six and seven (low). As mentioned before, the number of clusters generated from the LC clustering approach is the same as the existing number of groups under the Nikkei industrial classification, which is 16. These are the most appropriate number of groups and it won't exert into finer groups.

Figure 4.4: Latent Class Cluster Analysis with 16 Clusters (Japan)



In order to justify the above findings, we investigated and compared the composition of the firms between each SIC/Nikkei two-digit industries and the LC groupings. Table 4.5 and Table 4.6 show the percentage of the companies overlapping in the two different classifications for Malaysia and Japan, respectively. We found that the overlap between the two classifications is quite high but they do not necessarily belong to the same SIC/Nikkei three-digit branch.

Table 4.5: Percentage of Companies Overlapping between the Two Different Classifications (Malaysia)

SIC	% of companies overlapping in the two groups	LC
SIC 20	100%	Cluster 1
SIC 24	88%	Cluster 2
SIC 26	92%	Cluster 3
SIC 28	95%	Cluster 4
SIC 32	86%	Cluster 5
SIC 33	95%	Cluster 6
SIC 34	93%	Cluster 7
SIC 36	94%	Cluster 8
SIC 37	65%	Cluster 9

Table 4.6: Percentage of Companies Overlapping between the Two Different Classifications (Japan)

Industry	% of companies overlapping in the two groups	cluster
foods (01)	93	Cluster 1
textile products (03)	94	Cluster 2
pulp and paper (05)	50	Cluster 3
chemicals (07)	98	Cluster 4
drugs (09)	33	Cluster 5
petroleum (11)	42	Cluster 6
rubber products (13)	67	Cluster 7
stone, clay & glass products (15)	87	Cluster 8
iron & steel (17)	80	Cluster 9
nonferrous metal & metal products (19)	85	Cluster 10
machinery (21)	94	Cluster 11
electric & electronic equipment (23)	84	Cluster 12
motor vehicles & auto parts (27)	53	Cluster 13
transportation equipment (29)	70	Cluster 14
precision equipment (31)	62	Cluster 15
other manufacturing (33)	86	Cluster 16

Table 4.7 and Table 4.8 show the mean for AT, CR, SE and profit rate for the 9 LC and 16 LC groups for Malaysia and Japan, respectively. For Malaysia (Table 4.7), when compared with Table 4.1, the mean does not change much except for Cluster 9, especially the SE which dropped from 1357 to 229. As expected, this is the group which we would expect to be problematic because it contains only 65% of overlapping companies when compared to SIC37. For this group, a few large automobile companies (e.g. Proton Holdings[22], Tan Chong Motor[23], DBR-HICOM[24]) which are grouped in SIC37 are not classified into Cluster 9.

Table 4.7: Mean for AT, CR, SE and Profit Rate for the LC Classifications (Malaysia)

	AT	CR	SE	profit
Cluster 1	117.32	3.45	584.86	0.11
Cluster 2	105.35	4.29	320.80	0.11
Cluster 3	127.41	1.32	156.97	0.10
Cluster 4	112.27	3.37	221.10	0.13
Cluster 5	71.98	3.87	270.10	0.07
Cluster 6	240.05	2.11	191.62	0.07
Cluster 7	106.56	1.83	289.58	0.10
Cluster 8	180.04	1.85	340.40	0.08
Cluster 9	119.11	1.79	229.89	0.12

[22] It is a Malaysian automobile manufacturer founded in 1983 to manufacture, assemble and sell motor vehicles and related products, including accessories, spare parts and other components.
[23] It was a small motor vehicle distributor and has grown into a conglomerate involved in a variety of activities ranging from the assembly and marketing of motor vehicles and auto parts manufacturing to property development and trading in various heavy machinery, industrial equipment and consumer products, both locally and abroad.
[24] It is one of Malaysia's leading corporations, playing an integral role in the automotive manufacturing, assembly and distribution industry. Originally, it was incorporated in 1980 as the Heavy Industries Corporation of Malaysia (HICOM) and it experienced rapid growth and in 1996 merged with Diversified Resources Berhad (DBR) to form the biggest conglomerate in Malaysia.

For Japan (Table 4.8), due to we have four out of sixteenth groups which the overlapping companies are less than 60 percent, therefore, as expected, the mean might be different. Especially for cluster 13, the SE dropped from 275253 to 14438. For this group, a few large automobile companies (e.g. Toyota motor[25] and Honda motor[26]) which are grouped in the motor vehicles and auto parts industry are not classified into Cluster 13. This is because these are the important automobile companies in Japan/world, and they monopolised the Japan and world automobile share or economy. Their exception high SE might cause some bias in the clustering process.

Table 4.8: Mean for Profit Rate, CR, AT and SE for the LC Classifications (Japan)

	profit	CR	AT	SE
Cluster 1	0.0431	1.7513	1.2348	34610.46
Cluster 2	0.0075	1.8765	0.6194	34443.37
Cluster 3	0.0507	1.6899	0.8538	10614.17
Cluster 4	0.0338	2.1030	0.7476	51710.23
Cluster 5	0.0401	2.7704	0.7049	12906.11
Cluster 6	0.0537	1.1909	0.8663	14954.80
Cluster 7	0.0283	1.4211	0.8285	13861.57
Cluster 8	0.0111	1.7715	0.7489	16038.57
Cluster 9	0.0014	1.6410	0.6983	35548.44
Cluster 10	0.0171	1.9264	0.7855	19857.37
Cluster 11	0.0005	2.4930	0.6321	28854.98
Cluster 12	0.0093	2.6652	0.7584	31291.23
Cluster 13	-0.0063	1.5329	0.8850	14438.54
Cluster 14	0.0336	1.3885	0.7313	21326.56
Cluster 15	0.0227	2.4433	0.6971	12963.58
Cluster 16	0.0142	1.9734	0.8949	16179.43

[25] It is a Japanese multinational automaker headquartered in Toyota, Aichi, Japan. In 2010, Toyota employed 300,734 people worldwide, and was the largest automobile manufacturer in 2010 by production. It is the eleventh largest company in the world by revenue.

[26] It has been the world's largest motorcycle manufacturer since 1959, as well as the world's largest manufacturer of internal combustion engines measured by volume, producing more than 14 million internal combustion engines each year. Honda was the seventh largest automobile manufacturer in the world behind Toyota, General Motors, Volkswagen AG, Hyundai Motor Group, Ford, and Nissan in 2010.

Therefore, the above results give us some insights into the SIC/NIKKEI groupings and demonstrate that there is some justification for continuing to work with broad SIC/NIKKEI groups as proxies for similarity in the four financial ratios used in the study. Hence, three-digit SIC/NIKKEI industries do appear to have an economically meaningful basis for being grouped together. Our results suggest that the system of classification used is economically more meaningful than many previous researchers have feared.

However, although statistically correct, it is apparent from examining the three-digit data in detail that different production structures are grouped together by the SIC/NIKKEI at the two-digit level. To an extent this weakens our conclusions. Nevertheless, the evidence reinforces the need to work at as high a level of disaggregation as possible and to investigate further the possibility of re-classifying the data according to financial ratios using LC cluster analysis.

Chapter 5 - Conclusion

Industrial organisation and the classification of firms into industries is a complicated process. As we saw the outset, the issues of whether statistical classifications of industries map on to groups of financial ratios that can be aggregated together in an economically meaningful way is an important one.

Our focus in this study has been on the Malaysia SIC and Japan Nikkei Industrial classification systems. In particular we investigated whether, when data are aggregated from the third to the second digit level, they are sorted into categories that can be equated meaningfully with industries. We proposed using latent class (LC) analysis as alternative vehicles to classify companies into groups (clusters), such that the companies within a group are sufficiently homogeneous whilst companies in different groups are less homogeneous. LC analysis is considered to be an improved cluster analysis tool insofar as it uses statistical (rather than mathematical) methodology to construct the results.

The importance of LC cluster analysis is that it does not rely on the traditional modelling assumptions which are often violated in practice (e.g. linear relationship, normal distribution and homogeneity). In other words, the LC cluster model develops a segmentation of firms based upon the types of characteristics they have. The industrial classification is often used to divide firms into homogenous markets and firms classified into the same (*n+1*)-digit are thought to be more homogeneous than firms sharing only the same *n*-digit.

The implications from the LC cluster analysis results showed that the SIC/Nikkei system seems to be more effective at dividing firms into coarse industrial groups than at dividing firms into finer three, four or more digit segments thought to more closely represent economic market. This is a complementary method and it gives some reassurance to the extent that differences between two-digit groups are greater that differences within them. However, it also revealed, as one would expect, that there is a degree of coarseness in this classification system and therefore it should always be used with care. Latent class models, therefore, offer researchers an additional tool which to investigate industry homogeneity and they remain, at the time of writing, under utilised in empirical industrial organisation work.

Limitations and Directions for Future Research

The limitations of this research have to be kept in mind when interpreting the results of this study. First, is the availability of data, in particular the level of disaggregation which is arguably insufficient in the data set employed but was all that was publically available to conduct the work. In addition, we only look at large enterprises for the year 2010, and future researchers may want to extend this study by including small and medium sized enterprises and a longer time period to see if the model results are robust when additional years of data and a broader sample of firms are used.

Second, the latent class model has the virtue of being a semi-parametric specification, which frees the analyst from possibly strong, or unwarranted, distributional assumptions about individual heterogeneity. We therefore encourage other researchers in this field to consider employing such models when attempting to predict firm performance.

Recent development (e.g. Ellison et al., 2010 and Delgado et al., 2012) is including geographical clustering at the industry level. Geographic clusters spur information creation, dissemination and learning. It believes that firms those are located in the same industry and geographic area are sharing similar characteristics, furthermore their financial ratios and firm performance. Two firms that are economically similar may not experience strong return covariation over short horizons, however. Future research might need to control this variable when analyse industry classification.

References

ANDREWS, P. W. S. & BRUNNER, E. 1975. *Studies in Pricing,* London, Macmillan.

ARDITTI, F. 1967. Risk and the Required Return on Equity. *Journal of Finance,* 22**,** 19-36.

BAGINSKI, S. P. 1987. Intraindustry Information Transfers Associated with Management Forecasts of Earnings. *Journal of Accounting Research,* 25**,** 196-216.

BAIN, J. S. 1951. Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-1940. *Quarterly Journal of Economics,* 65**,** 293-324.

BHOJRAJ, S., LEE, C. M. C. & OLER, D. K. 2003. What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research. *Journal of Accounting Research,* 41**,** 745-774.

BIDDLE, G. C. & SEOW, G. S. 1991. The Estimation and Determinants of Associations Between Returns and Earnings: Evidence from Cross-industry Comparisons. *Journal of Accounting, Auditing & Finance,* 6**,** 183-232.

BRADLEY, M., JARRELL, G. A. & KIM, E. H. 1984. On the Existence of an Optimal Capital Structure: Theory and Evidence. *Journal of Finance,* 39**,** 857-878.

BROWN, P. & BALL, R. 1967. Some Preliminary Findings on the Association between the Earnings of a Firm, Its Industry, and the Economy. *Journal of Accounting Research,* 5**,** 55-80.

CHAN, L. K. C., LAKONISHOK, J. & SWAMINATHAN, B. 2007. Industry classifications and return comovement. *Financial Analysts Journal,* 63**,** 56-70.

CLARKE, R. N. 1989. SICs as Delineators of Economic Markets. *The Journal of Business,* 62**,** 17-31.

DAVIES, J. H. 1955. The Industry and the Representative Firm. *The Economic Journal,* 65**,** 710-712.

DELGADO, M., PORTER, M. E. & STERN, S. 2012. Clusters, Convergence and Economic Performance. *NBER Working Paper Series,* Working Paper 18250.

ELLIOTT, R., GREENAWAY, D. & HINE, R. 2000. Tests for Factor Homogeneity and Industry Classification. *Review of World Economics,* 136**,** 355-371.

ELLISON, G., GLAESER, E. L. & KERR, W. R. 2010. What Causes Indsutry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic review,* 100**,** 1195-1213.

ELTON, E. J. & GRUBER, M. J. 1971. Improved Forecasting Through the Design of Homogeneous Groups. *The Journal of Business,* 44**,** 432-450.

ELTON, E. J. & GRUBER, M. J. 1999. Improved Forecasting Through the Design of Homogeneous Groups. *Investments,* 44**,** 432-450.

FAMA, E. F. & FRENCH, K. R. 1997. Industry Costs of Equity. *Journal of Financial Economics,* 43**,** 153-193.

FARRELL JR, J. L. 1974. Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings. *Journal of Business,* 47**,** 186-207.

FERTUCK, L. 1975. A Test of Industry Indices Based on SIC Codes. *Journal of Financial and Quantitative Analysis,* 10**,** 837-848.

FOSTER, G. 1981. Intra-Industry Information Transfers Associated with Earnings Releases. *Journal of Accounting and Economics,* 3**,** 201-232.

FRANCIS, J. R. & REITER, S. A. 1987. Determinants of Corporate Pension Funding Strategy. *Journal of Accounting and Economics,* 9**,** 35-59.

FREEMAN, R. & TSE, S. 1992. An Earnings Prediction Approach to Examining Intercompany Information Transfers. *Journal of Accounting and Economics,* 15**,** 509-523.

FRITSCH, M. & STEPHAN, A. 2006. Measuring Performance Heterogeneity Within Groups – A Two-Dimensional Approach. *Applied Economics Letters,* 13**,** 17-20.

GHICAS, D. C. 1990. Determinants of Actuarial Cost Method Changes for Pension Accounting and Funding. *Accounting Review,* 65**,** 384-405.

GREENE, W. & HENSHER, D. 2003. A latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. *Transportation Research Part B,* 37**,** 681-698.

GROGAARD, B., GIOIA, C. & ROBERTSTAD GARCIA BENITO, G. 2005. An Empirical Investigation of the Role of Industry Factors in the Internationalization Patterns of Firms. *Available at SSRN 982085*.

GUENTHER, D. A. & ROSMAN, A. J. 1994. Differences between COMPUSTAT and CRSP SIC Codes and Related Effects on Research. *Journal of Accounting and Economics,* 18**,** 115-128.

GUPTA, M. C. 1969. The Effect of Size, Growth, and Industry on the Financial Structure of Manufacturing Companies. *The Journal of Finance,* 24**,** 517-529.

GUPTA, M. C. & HUEFNER, R. J. 1972. A Cluster Analysis Study of Financial Ratios and Industry Characteristics. *Journal of Accounting Research,* 10**,** 77-95.

HAN, J. C. Y. & WILD, J. J. 1990. Unexpected Earnings and Intraindustry Information Transfers: Further Evidence. *Journal of Accounting Research,* 28**,** 211-219.

HAN, J. C. Y., WILD, J. J. & RAMESH, K. 1989. Managers' Earnings Forecasts and Intra-Industry Information Transfers. *Journal of Accounting and Economics,* 11**,** 3-33.

HAND, J. R. M., HUGHES, P. J. & SEFCIK, S. E. 1990. Insubstance Defeasances: Security Price Reactions and Motivations. *Journal of Accounting and Economics,* 13**,** 47-89.

JENSEN, R. E. 1971. A Cluster Analysis Study of Financial Performance of Selected Business Firms. *Accounting Review,* 46**,** 36-56.

KAHLE, K. M. & WALKLING, R. A. 1996. The Impact of Industry Classifications on Financial Research. *Journal of Financial and Quantitative Analysis,* 31**,** 309-335.

KIM, K. & SCHROEDER, D. A. 1990. Analysts' Use of Managerial Bonus Incentives in Forecasting Earnings. *Journal of Accounting and Economics,* 13**,** 3-23.

KING, B. F. 1966. Market and Industry Factors in Stock Price Behavior. *The Journal of Business,* 39**,** 139-190.

LEV, B. 1969. Industry Averages as Targets for Financial Ratios. *Journal of Accounting Research,* 7**,** 290-299.

MAGIDSON, J. & VERMUNT, J. K. 2003. Latent Class Models, available at http://www.statisticalinnovations.com/articles/sage11.pdf (accessed 28 June 2010).

MCLACHLAN, G. & PEEL, D. 2000. *Finite Mixture Models,* New York, Wiley.

MEYERS, S. L. 1973. A Re-Examination of Market and Industry Factors in Stock Price Behavior. *The Journal of Finance,* 28**,** 695-705.

MOHRMAN, M. B. 1993. Debt Contracts and FAS No. 19: A Test of the Debt Covenant Hypothesis. *Accounting Review,* 68**,** 273-288.

NARAYANASWAMY, C. R. 1996. A Re-examination of Homogeneous Stock Grouping in the Context of the APT: An Application of Discriminant Analysis. *Quarterly Journal of Business & Economics,* 35**,** 3-16.

NEEDHAM, D. 1978. *Economics Analysis and Industrial Structure,* London, Holt, Rinehart, and Winston.

NIGHTINGALE, J. 1978. On the Definition of 'Industry' and 'Market'. *Journal of Industrial Economics,* 27**,** 31-40.

PALANYANDY, C. & TALHA, R. 2002. The Malaysian Standard Industrial Classification (MSIC). *United Nations Workshop on International Economics and Social Classifications.* Hanoi, Vietnam, 10-14 June 2002.

PINCHES, G. E., MINGO, K. A. & CARUTHERS, J. K. 1973. The Stability of Financial Patterns in Industrial Organizations. *The Journal of Finance,* 28**,** 389-396.

RAYMENT, P. B. W. 1976. The Homogeneity of Manufacturing Industries with Respect to Factor Intensity: The Case of the United Kingdom. *Oxford Bulletin of Economics & Statistics,* 38**,** 203-209.

ROBINSON, J. 1956. The Industry and the Market. *The Economic Journal,* 66**,** 360-361.

SUDARSANAM, P. S. & TAFFLER, R. J. 1985. Industrial Classification in UK Capital Markets: A Test of Economic Homogeneity. *Applied Economics,* 17**,** 291-308.

SZEWCZYK, S. H. 1992. The intra-industry transfer of information inferred from announcements of corporate security offerings. *Journal of Finance,* 47**,** 1935-1945.

TAYLOR, J. R. 1982. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements,* Mill Valley, California, University Science Books.

TEW, B. & HENDERSON, R. F. 1959. *Studies in Company Finance,* Cambridge, Cambridge University Press.

THOMAS, J. K. & TUNG, S. 1992. Cost Manipulation Incentives Under Cost Reimbursement: Pension Costs for Defense Contracts. *Accounting Review,* 67**,** 691-711.

VERMUNT, J. & MAGIDSON, J. 2002. Latent Class Cluster Analysis. *In:* HAGENAARS, J. A. A. M., ALLAN L. (ed.) *Applied Latent Class Analysis.* Cambridge: Cambridge University Press.

VERMUNT, J. & VAN DIJK, L. 2001. A Nonparametric Random-Coefficients approach: The Latent Class Regression Model. *Multilevel Modelling Newsletter,* 13**,** 6–13.

VERMUNT, J. K. 2010. Latent Class Models. *International Encyclopedia of Education,* 7**,** 238-244 (Oxford, UK: Elsevier).

WEDEL, M. & DESARBO, W. 1994. A Review of Recent Developments in Latent Class Regression Models. *In:* BAGOZZI, R. P. (ed.) *Advanced Methods of Marketing Research.* Cambridge, MA: Blackwell.

WEINER, C. 2005. The Impact of Industry Classification Schemes on Financial Research. *SFB 649 Discussion Paper 2005-062.*

WIPPERN, R. 1966. Financial Structure and the Value of the Firm. *Journal of Finance,* 21**,** 615-633.

WOLFE, J. 1970. Pattern Clustering by Multivariate Mixture Analysis. *Multivariate Behavioral Research,* 5**,** 329-350.

XIANG, B. 1993. The Choice of Return-Generating Models and Cross-Sectional Dependence in Event Studies. *Contemporary Accounting Research,* 9**,** 365-365.