STRATEGIC GROUP ANALYSIS: STRATEGIC PERSPECTIVE, DIFFERENTIATION AND PERFORMANCE IN CONSTRUCTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

CENK BUDAYAN

IN PARTIAL FULLFILMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTORATE OF PHILOSOPHY
IN
CIVIL ENGINEERING

JULY 2008

Approval of the thesis:

STRATEGIC GROUP ANALYSIS: STRATEGIC PERSPECTIVE, DIFFERENTIATION AND PERFORMANCE IN CONSTRUCTION

Submitted by **CENK BUDAYAN** in partial fulfillment of the requirements for the degree of **Doctorate of Philosophy in Civil Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Güney Özcebe _____
Head of Department, **Civil Engineering**

Prof. Dr. M. Talat Birgönül _____
Supervisor, **Civil Engineering Dept., METU**

Assoc. Prof. Dr. İrem Dikmen Toker _____
Co-Supervisor, **Civil Engineering Dept., METU**

Examining Committee Members:

Asst. Prof. Dr. Metin Arıkan _____
Civil Engineering Dept., METU

Prof. Dr. M. Talat Birgönül _____
Civil Engineering Dept., METU

Assoc. Prof. Dr. Canan Sepil _____
Industrial Engineering Dept., METU

Asst. Prof. Dr. Rıfat Sönmez _____
Civil Engineering Dept., METU

Asst. Prof. Dr. Gökhan Arslan _____
Civil Engineering Dept., Anadolu University

Date: 14.07.2008

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Cenk BUDAYAN

Signature :

ABSTRACT

STRATEGIC GROUP ANALYSIS: STRATEGIC PERSPECTIVE, DIFFERENTIATION AND PERFORMANCE IN CONSTRUCTION

BUDAYAN, Cenk

Ph. D., Department of Civil Engineering

Supervisor: Prof. Dr. M. Talat BİRGÖNÜL

Co-Supervisor: Assoc. Prof. Dr. İrem DİKMEN TOKER

July 2008, 271 pages

The aim of strategic group analysis is to find out if clusters of firms that have a similar strategic position exist within an industry or not. In this thesis, by using a conceptual framework that reflects the strategic context, contents and process of construction companies and utilising alternative clustering methods such as traditional cluster analysis, self-organizing maps, and fuzzy C-means technique, a strategic group analysis was conducted for the Turkish construction industry. Results demonstrate that there are three strategic groups among which significant performance differences exist. Self-organising maps provide a visual representation of group composition and help identification of hybrid structures. Fuzzy C-means technique reveals the membership degrees of a firm to each strategic group. It is recommended that real strategic group structure can only be identified by using alternative cluster analysis methods.

The positive effect of differentiation strategy on achieving competitive advantage is widely acknowledged in the literature and proved to be valid for the Turkish construction industry as a result of strategic group analysis. In this study, a framework is proposed to model the differentiation process in construction. The relationships between the modes and drivers of differentiation are analyzed by structural equation modeling. The results demonstrate that construction companies can either differentiate on quality or productivity. Project

management related factors extensively influence productivity differentiation whereas they influence quality differentiation indirectly. Corporate management related factors only affect quality differentiation. Moreover, resources influence productivity differentiation directly whereas they have an indirect effect on quality differentiation.

Keywords: Strategic groups, competitive advantage, performance, cluster analysis, self-organizing map, fuzzy C-means, differentiation.

ÖZ

STRATEJİK GRUP ANALİZİ: İNŞAAT SEKTÖRÜNDE STRATEJİK PERSPEKTİF, FARKLILAŞMA VE PERFORMANS

BUDAYAN, Cenk

Doktora, İnşaat Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. M. Talat BİRGÖNÜL

Y. Tez Yöneticisi: Doç. Dr. İrem DİKMEN TOKER

Temmuz 2008, 271 sayfa

Stratejik grup analizinin amacı, aynı sektörde rekabet eden firmaların stratejik pozisyonlarına göre farklı kümeler oluşturup oluşturmadığının saptanmasıdır. Gerçekleştirilecek analiz sonuçları ışığında, kümelerin performansları arasında farklılıklar bulunup bulunmadığı araştırılarak, bir sektörde başarılı olabilmek için gerekli stratejik seçim ve kümeleri birbirlerinden ayıran temel özellikler irdelenebilmektedir. Bu çalışma kapsamında; inşaat firmalarının stratejik içerik, süreç ve kapsamlarını yansıtmak için geliştirilen kavramsal bir model ile istatistiksel küme analizi, öz düzenleyici haritalar ve bulanık c-ortalamalar yöntemleri kullanılarak, Türk inşaat sektörü için gerçekleştirilen stratejik grup analizinin sonuçları irdelenmiştir. Yapılan analizler sonucunda, elde edilen çıktılar karşılaştırılıp, sentezlenerek, aralarında istatistiksel olarak anlamlı performans farklılıkları olan üç adet küme belirlenmiştir. Bu analizler sonucunda, öz düzenleyici haritalar metodunun görsel gösterim kapasitesi, melez stratejik grupların bulunmasına katkı da bulunmaktadır. Bulanık c-ortalamalar metodu ise her verinin hangi stratejik gruba hangi derecede bağlı olduğunu ortaya koymaktadır. Bunların sonucu olarak, stratejik grupların belirlenmesinde alternatif kümeleme metotlarının kullanımı önerilmektedir.

Farklılaşma stratejilerinin rekabet avantajı yaratmaktaki pozitif etkisi literatürde kabul gören bir gerçektir. Ayrıca bu çalışmada gerçekleştirilen stratejik grup analizinin sonuçları da, Türk inşaat sektörü için de bu çıkarımın geçerli olduğuna kanıt olarak gösterilebilir. Bu çalışma kapsamında, inşaat firmalarının farklılaşmalarını modellemek üzere bir çerçeve önerilerek, önerilen farklılaşma yöntemleri ve belirleyicilerinin istatistiksel olarak geçerlilikleri

araştırılmış ve bunlar arasındaki ilişkiler yapısal denklem modelleme yöntemi ile analiz edilmiştir. Elde edinilen sonuçlar, inşaat firmalarının kaliteye veya verimliliğe dayalı olarak farklılaşabileceklerini ortaya koymaktadır. Araştırmanın bulguları, proje yönetimine yönelik faktörlerin verimlilikle farklılaşmayı yoğun şekilde etkilediklerini ve kalitede farklılaşmayı ise dolaylı olarak etkilediklerini göstermektedir. Şirket yönetimine yönelik faktörlerin ise sadece kaliteyle farklılaşmayı etkiledikleri anlaşılmaktadır. Ayrıca, kaynaklar ise verimlilikle farklılaşmayı yoğun olarak etkilemekle beraber, kaliteyle farklılaşmayı dolaylı olarak etkilemektedirler.

Anahtar kelimeler: Stratejik gruplar, rekabet avantajı, performans, küme analizi, öz düzenleyici haritalar, bulanık c-ortalamalar, farklılaşma.

ACKNOWLEDGEMENT

I would like to thank to my co-supervisor Assoc. Prof. Dr. İrem Dikmen Toker and my supervisor Prof. Dr. Talat Birgönül without whom I would not be able to complete my thesis, for their unconditional guidance, patience, and tolerance at each step of this study.

I owe special thanks to the respondents of the survey.

I should thank all my friends, who motivate me and make me feel strong in the way to achieve this study.

And finally, my mother, Necmiye Budayan and my father, Mehmet Cengiz Budayan deserve the greatest thanks for their unconditional love, support and understanding throughout my studies. They deserve special thanks because without their patience and support this thesis could have never been materialized. I dedicate this study to my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADF	Asymptotic Distribution Free
ADI	Alternative Dunn Index
AIC	Akaike Information Criterion
ANOVA	Analysis of Variances
BIC	Bayesian Information Criterion
BMU	Best-Matching Unit
CF	Cluster Features
CFI	Comparative Fit Indices
CPU	Central Processing Unit
DI	Dunn's Index
dof	Degree of Freedom
EM	Expectation-Maximization
ESOM	Emergent Self Organizing Map
FCM	Fuzzy C-Means
FMLE	Fuzzy Maximum Likelihood Estimation
FPI	Fuzziness Performance Index
GFI	Goodness Of Fit Indices
GG	Gat-Geva
GK	Gustafson-Kessel
GLS	Generalized Least Squares
GPL	General Public License
IFI	Bollen Fit Indices
LS	Least Squares
MFI	Mcdonald Fit Indices
ML	Maximum Likehood
MPE	Modified Partition Entropy
NaN	Not a Number
NFI	Normed Fit Indices
NNFI	Non-Normed Fit Indices

PAM	Partitioning around medoids
RMSEA	Root Mean-Square Error Of Approximation
S	Separation Index
SC	Partition Index
SEM	Structural Equation Modeling
SOM	Self Organizing Map
TCA	Turkish Contractors Association
USD	United States Dollar
XB	Xie and Beni's Index

CHAPTER 1

INTRODUCTION

All companies should have a clear strategic perspective to achieve competitive advantage. The selected strategies should be in accordance with objectives, competencies and competitive rules prevailing in the market. The need for a strategic perspective for construction companies has long been stressed by many researchers (Betts and Ofori 1992; Chinowsky and Meredith 2000; Dikmen and Birgonul 2003; Warszawski 1996). It is clear that a number of companies that compete in the same industry may have similar resources/competencies and develop similar strategic perspectives. The question is whether these companies show similar performance or not. Finding an answer to this question constitutes the major motivation of this study. In order to reveal the competitive structure and strategic groups within the Turkish construction industry, strategic group analysis is decided to be conducted.

This study comprises of basically two parts. The first part is about "strategic group analysis" whereas the second part is about "differentiation strategies" as the first part of the study indicated the significance of differentiation strategy for success in the Turkish construction industry.

For the first part of the thesis, the core objectives are threefold;

1. To carry out strategic group analysis for the Turkish construction industry and identify possible clusters having different strategic positions,

2. To test whether performance differences exist between the firms in different strategic groups,

3. To carry out strategic group analysis by using different cluster analysis methods (namely, traditional clustering method, self organizing maps and fuzzy clustering method) and determine the convergence between the solutions obtained from different methods.

For the second part of the thesis, the core objectives are:

1. To identify the modes and drivers of differentiation that create competitive advantage in the Turkish construction industry,
2. To find out the interrelations between different drivers and modes of differentiation using Structural Equation Modeling (SEM).

Below sections summarize the background and scope of this research, contribution of the study to the literature and finally present the organization of the thesis.

1.1. Background and scope of the research

The strategic group concept was introduced by Hunt (1972) for explaining the performance difference between firms which follow different strategies. Since then, strategic group concept was adopted in different industries for both theoretical and empirical purposes. Like Porter (1979) in multi-industry, Hawes and Crittenden (1984) in retail grocery, Lewis and Thomas (1990) in UK retail grocery industry, Fiegenbaum and Thomas (1990; 1995) in U.S. insurance industry, Houthoofd and Heene (1997) in the Belgian brewing industry and many others. However, the application of strategic group analysis in construction management literature is limited with only two studies, namely Kale and Arditi (2002) and Claver et al. (2003).

According to Porter's (1980; 1985) generic competitive strategies' typology, the companies compute in two modes, namely cost leadership and differentiation. The positive effect of differentiation strategy on the performance of the companies was mentioned by many authors, Porter (1985), Henderson (1981), and Grant (1995). Besides Kale and Arditi (2003) stated the positive relationship between differentiation strategies and performance of the construction companies. This positive relationship can be interpreted as the companies can keep themselves from intensive competition in the industry, create entry as well mobility barriers, and in turn, achieve high returns by using a differentiation strategy. However, there are no studies in the construction management literature that deal with the contents and context of differentiation together with its potential to create competitive advantage.

Consequently, this research aims to fill these two important gaps by carrying out a strategic group analysis for the Turkish construction industry and identifying the dynamics of differentiation process within the Turkish construction companies.

1.2. Problem definition

Although the strategic group is a popular subject in the management literature due to its power for explanation to the performance difference between the companies placed in the same industry, the existence of the strategic groups is questioned by many researchers, Barney and Hoskisson (1990), Hatten and Hatten (1987), Dranove et al. (1998). Also, whether there exists a direct relationship between the strategic group membership and performance was investigated by various empirical studies for different sectors. However, there is no consistency in the findings of these studies. In other words, in some of these studies, strong relationship has been established, whereas some studies demonstrate no clear relationship. The inconsistency of results may be attributed to different factors which will be discussed in the forthcoming parts of this thesis. Strategic group analysis requires choosing or developing a conceptual framework in which the strategic dimensions are defined properly. Yet, there is no consensus on these dimensions. In this respect, a complete and valid conceptual framework should be constructed. This is one of the challenges of this study.

The other problematic issue about strategic group analysis is the clustering methods used in determination of strategic groups. In the literature, traditional cluster analysis methods are preferred and there exist only a limited number of studies that are performed by using other clustering methods. However, determination of the strategic groups by using only one of the clustering methods is one of the criticism points of the work on strategic group analysis (Ketchen and Shook 1996). Therefore in this study, different clustering methods are used for determining the strategic groups, and results are compared and combined to reveal the real strategic group structure.

Despite acceptance of positive effect of differentiation on performance of the companies and benefits of differentiation, there are only a limited number of studies in the construction management literature that investigate the ways of differentiation. Due to this fact, two main difficulties were faced while developing the conceptual model. The first one is to decide the modes of differentiation and the second difficulty is to identify a complete and valid list of determinants of differentiation. Proper constructs about differentiation should be found

as well as the relations between strategies and sources of competitive advantage as a result of differentiation.

1.3. Scope of the study

Within the context of this study, the strategic groups of medium-big Turkish construction companies are determined by using different cluster analysis methods. The convergence of these methods provides evidence for existence of strategic groups in the Turkish construction industry. Also, it is proved that a direct relationship between the strategic group membership and performance exists. The competitive structure is discussed by referring to the performance implications of strategic groups as well as existence of overlaps between these groups leading to a hybrid structure. Results are evaluated so that guidelines can be provided to the companies about how they can position themselves in the competitive environment to increase their performance. The second part of this research contains a structural model that comprises the modes and drivers of differentiation as well as their interrelations, which is applicable to the Turkish construction companies. Recommendations are given to the companies which aim to follow differentiation strategies by providing guidelines about how they can differentiate, which competencies and resources they should develop and how they should design their business processes in order to differentiate successfully.

1.4. Research methodology

A questionnaire survey was administered to construction companies for the strategic group analysis. Three cluster analysis methods, namely traditional cluster analysis methods (such as hierarchical or partitioning methods), self organizing maps and fuzzy clustering methods (such as fuzzy C-means) are used to evaluate data and find out relevant strategic group structures.

Based on the results of the first part, another questionnaire was prepared for investigation of differentiation process within the medium-big size Turkish construction companies. Structural equation modeling is used as a research method to test the validity of the proposed measures of differentiation modes; to test the significance of hypothesis about the relations among determinants of differentiation and to analyze the influences of these determinants on modes of differentiation.

1.5. Organization of the thesis

The thesis comprises of nine chapters. After the introductory information given in Chapter 1, Chapter 2 consists of the literature survey and definitions about strategic groups and strategic group analysis. In the following three chapters, detailed information is depicted about the clustering methods, namely traditional cluster analysis, self organizing maps, and fuzzy clustering method. In Chapter 6, the conceptual model used for strategic group analysis is presented, the procedure followed in determination of strategic groups by using different cluster analysis methods is highlighted, the findings are discussed and some recommendations are provided for construction companies. In Chapter 7, definitions and related literature about differentiation are presented as well as the conceptual model developed to analyze the differentiation process in construction companies. Chapter 8 presents the structural equation modeling results and modes and ways of differentiation are discussed in the light of these findings. In Chapter 9, major findings of the research are summarized and some strategies are proposed for the Turkish contractors. In addition to the main text, this study also includes four appendices, in which a sample of the questionnaire prepared for determination of strategic groups and a sample of another questionnaire related with differentiation, descriptive statistics related with questionnaire study on differentiation as well as the curriculum vitae of the author exists.

CHAPTER 2

STRATEGIC GROUPS

In this part of the thesis, definitions related with strategic groups and strategic group analysis will be presented to set the theoretical background of the research. The current state of knowledge about strategic group analysis in the management literature as well as the existing studies in the construction management literature will be revisited.

2.1. Strategic groups: Definition

The strategic group concept was first introduced by Hunt (1972) for explaining the performance difference between firms which follow different strategies. The popularity of the term increased after its usage by Porter (1979) and Caves and Porter (1977). Since then, strategic group concept is adopted in different industries for both theoretical and empirical purposes. Like Porter (1979) in multi-industry, Hawes and Crittenden (1984) in retail grocery, Lewis and Thomas (1990) in UK retail grocery industry, Fiegenbaum and Thomas (1990; 1995) in U.S. insurance industry, Houthoofd and Heene (1997) in the Belgian brewing industry and Arditi and Kale (2002) in U.S. construction industry and many others.

The major characteristics of strategic groups are defined by Thomas and Venkatraman (1988) as;

1. Similar strategies are carried out by each group composed of firms,
2. The similarity between the firms within a group with each other is more than the similarity of these firms with other firms outside the group,
3. Firms within a group are more likely to respond similarly to a market opportunity.

According to Porter (1980), strategic group is a group of firms in an industry following the same or similar strategy along a set of strategic dimensions. Dimensions include strategic decision variables that best distinguish the business strategies and competitive positions of firms within an industry such as scope and mode of competition. According to Porter,

strategic groups are persistent structural features of industries that are bounded by mobility barriers, in other words the strategic groups will remain stable due to the presence of the mobility barriers. Mascaranhas (1989) also highlights the importance of mobility barriers and argues that strategic groups may exist only if significant mobility barriers exist between groups. Leask (2004) defines strategic groups as stable intra-industry structures separated by mobility barriers in pursuit of different strategies that may be expected to yield significant performance differences.

Drawing from economic and cognitive theories, researchers have argued that there may be differences between the performances of firms that belong to different strategic groups. The traditional view of strategic groups draws from industrial organization economics and proposes that firms within strategic groups collude to isolate themselves from firms outside their group (Caves and Porter 1977). Due to this collusive action, a favorable competitive environment exists in a strategic group leading to a similar performance among them. For establishing a direct of group membership on performance, strategic interactions among members should be existed. McNamara et al. (2003) argue that rivalry rather than collusion may also exist within strategic groups.

Strategic group analysis is a subset of industry analysis that looks specifically at the different groups of rival firms clustered around a similar competitive approach or strategic position (Bensoussan and Fleisher 2000). Porter (1980) defines strategic group analysis as the first step in structural analysis of industries to characterize the strategies of all significant competitors. It is used to determine the different strategic positions that rival firms occupy intensity of competitive rivalry within and between industry groups, the profit potential of the various strategic groups in an industry and implications for the competitive position of the firm under analysis. However, members of a strategic group, while pursuing similar strategies are not necessarily in competition with one another. Due to differences in locations, sub-markets etc. companies in the same strategic group may not be direct competitors.

2.2. Reasons for carrying out strategic group analysis

Strategic group analysis has attracted the attention of many researchers. The reason of this can be explained by the benefits of identifying the strategic groups. The most important benefits of strategic group concepts is to facilitate strategy researchers and managers when analyzing competitors, making strategic investment decisions, and developing successful

strategies by simplifying the market structure (Mascarenhas and Aaker 1989). Flavian and Polo (1997) mentioned about the six potential applications of strategic group analysis:

- Simplification of analysis of the strategic heterogeneity of the market is one of the oldest applications of strategic group analysis. Analyzing the strategies of the market on a firm basis is a time consuming and expensive process. Also, the analysis of the market on a generic strategic perspective can lead to errors and loss of some significant information. Strategic group analysis can eliminate these problems by providing a global idea about the market. It can also help preservation of the data, which can be lost in a global analysis of industry using averaged and aggregated data (Hatten and Hatten 1987). In conclusion, by discriminating the firms from each other and collecting the similar ones in the same groups, the strategic group analysis makes the analysis of strategic heterogeneity of the market more manageable.

- Due to the relation of the rivalry with the profitability of the industry and the firm (Cool and Dierickx 1993), the rivalry analysis (or competition analysis) is extremely useful for the strategist. The analysis of business rivalry can be markedly simplified through strategic group analysis. For instance, Fiegenbaum et al. (1988) explained rivalry between the companies by referring to the results of strategic group analysis and concluded that the rivalry is related with the number and size of strategic groups. McNamara (2003) argue that, due to the competition for the same resources and consumer segments in a market suffering from shortage in resources, the rivalry within the strategic groups can be intensive than the rivalry between the strategic groups. Consequently, strategic group analysis can be utilized as a starting point for concentrating attention in a favorable manner on the strategic movements of the competitors, and establishing an order of priorities (Flavian and Polo 1997).

- The link between the performance and the strategic groups was concluded empirically by some researchers, like Porter (1979), Dess and Davis (1984) and Reger and Huff (1993). Furthermore, different theoretical approaches about the strategic groups were developed for explaining the existence of differences in performance between the firms of the same industry and their performance in the long term. Consequently, findings of the strategic group analysis can help the companies to conceptualize their strategic positioning and the resources required to pursue alternative strategies.

- The strategic group analysis can be used for determining the strengths and weaknesses of a firm. Strategic group analysis represents a systematic and comprehensive approach to elaborate a strengths and weaknesses analysis in terms of the framework of relative competitive advantage (McGee and Thomas 1986). Through strategic group analysis, the appropriateness between the available resources and the chosen strategy can be identified, by that way, the strengths and weaknesses of companies in terms of resources can be investigated (Flavian and Polo 1997). Additionally, strategic group analysis can help identification of the threats and opportunities offered by the environment to the firms. The strategic group analysis can highlight the white spaces not currently occupied by any group- that represents opportunities of the market in the future (McGee and Thomas 1992). Using the strategic distance of the group to the rest of the market as an indicator, the height of the mobility barriers, as well as the risk of new comers to the group, can be quantified (Flavian and Polo 1997).

- Longitudinal analysis of strategic groups can be a valuable tool for the strategist and managers, because the evolution of the market can be analyzed in a structured format. In this way, the information about changes in the number of the strategies implemented in the market and convergence or divergence of these strategies can be monitored. The changes in homogeneity in competitive behavior of the firms can be observed (Flavian and Polo 1997). In addition, the longitudinal analysis can provide a framework to allow the classification of strategic changes, an impartial analysis of the position of a firm within an industry and a way of evaluating industry evolution (McGee and Thomas 1986). This helps understanding of stability and sustainability of competitive advantage and performance in an industry (Mascarenhas 1989).

- One benefit of strategic group analysis is that, it can be used to alert the companies to the identities of companies which can be easily imitated. Also, it attracts attention to those companies which are different from them but have a capacity to threaten their existence (Hatten and Hatten 1987).

2.3. Existence of the strategic groups: Do they really exist?

Although, strategic group analysis has the above mentioned benefits for analyzing the competitive environment in a market, there are still questions about the existence of groups. It is questioned by some of the researchers, like Barney and Hoskisson (1990) who argue that the theoretical underpinning of the strategic groups does not exist. In addition, the

strategic group analysis is criticized due to the cluster analysis methods used in these analyses which can provide clusters even if no meaningful groups are embedded in a sample. Consequently, cluster analysis can offer arbitrary clusters which do not reflect the real structure of the market; even it can impose artificial groupings (Ketchen and Shook 1996). Hatten and Hatten (1987) claim that strategic groups are more than analytical convenience only if true group-level effects exist. By the same line of thinking, Dranove et al. (1998) state that true effects (strategic group-level) and spurious effects (industry-level) should be distinguished. They argue that "a strategic group exists if the performance of a firm in the group is a function of group characteristics".

As mentioned before, Porter (1980) explain the formation of the strategic groups by the mobility barriers formed due to differences in risk posture and random differences in skills and quality of the assets of the companies. Mobility barriers are structural factors that prevent ease of movement between market positions and protect firms' strategies against imitation and competition (Leask and Parnell 2005). Mobility barriers are the total costs of movement from one group to another by taking account of the operations or financial resources, in other words the operating or variable cost penalty which the firms must overcome for changing their market position (Hatten and Hatten 1987).

The expected costs of changing group membership, which affect the probability of gaining profit with changing a group, increases with the height of the mobility barriers; therefore it is expected to deter entry to groups and deter efforts to change groups (Hatten and Hatten 1987). The types of competitive investment made in the past by the companies determine the heights of each strategic group's mobility barriers. Also, the height of mobility barriers can be shaped by firms' differences in brand identification, product quality, technological leadership, asset specificity, and extent of service offerings, degrees of financial leverage, differences in the use of backward or forward integration or other factors (Harrigan 1985). However, the contribution of these variables to the height of mobility barriers is not equal; for example, sales-force size or advertising largely a function of expenditure can be imitated easily by the competitors which can afford the expenditure of these variables. Whereas, some of the variables reflect accumulated experience, knowledge and the application of leading edge research that may take a very long time to understand and they are very difficult to imitate in a short time (Leask 2005). The height of the mobility barrier determines the competitive advantage of a strategic group; on the other hand, low mobility barriers may be competitive disadvantages. Without mobility barriers, the successful strategies of a strategic group can be imitated by other firms from different groups or industries, therefore firm

groups' profit rates would decrease to the equality at the margin (Zúñiga-Vicente et al. 2004). McGee and Thomas (1986) identified three broad categories of mobility barriers; namely market related strategies, industry supply characteristics, and features specific to the individual firm. They illustrated the sources of the mobility barriers according to these categories shown in Table 2-1. At the end, they concluded that mobility barriers are a counterpart of group structures and arise from strategic decisions. In other words, the mobility barriers are formed by the strategic decisions under management control.

Table 2-1: Sources of mobility barriers (McGee and Thomas 1986)

Market-related strategies	Industry supply characteristics	Characteristics of firm
Product line	Economies of scale:	Ownership
User technologies	Production	Organization structure
Market segmentation	Marketing	Control systems
Distribution channels	Administration	Management skills
Brand names	Manufacturing processes	Boundaries of firms
Geographic coverage	R&D capability	Diversification
Selling systems	Marketing and distribution systems	Vertical integration
		Firm size
		Relationships with influence groups

Apart from the concept of mobility barriers, existence of strategic groups was explained by Rumelt (1984) in his discussion of 'uncertain imitability' as a source of 'isolating mechanisms'. Rumelt (1984) used the isolating mechanism to refer to 'phenomena that limit the ex post equilibration of rents among individual firms'. Isolating mechanisms are features of resources, such as skills, knowledge, and capabilities that are tacit, unique, invisible, complex, or path dependent that prevent other firms from obtaining and replicating them (Oliver 1997). Isolating mechanisms are used for describing the uncertain imitability of resources at firm level in parallel to the concept of mobility barriers at the group level (Mehra and Floyd 1998). In other words, isolating mechanisms are the firm-level equivalent of industry entry barriers and mobility barriers (McNamara et al. 2003). Therefore there are similarities between the concepts of mobility barriers and isolating mechanisms (McGee and Thomas 1986). Table 2-2 illustrates Rumelt's (1984) isolating mechanism which represents a simple theory of strategy (McGee and Thomas 1986). The isolating mechanisms prevent the other companies to imitate even with full knowledge of their strategic choices with respect to

scope and resource deployment (Reed and DeFillippi 1990). Since the companies at the same strategic group pursuit similar strategies, the probability of making similar resource investment is high. Due to the investment in similar resources, isolating mechanisms prevent the firms involved in the strategic group from switching from one strategy to another (Leask and Parnell 2005). In conclusion, isolating mechanisms degrade the mobility barriers from group level to firm level for explaining the existence of the strategic groups.

Table 2-2: Rumelt's isolating mechanisms (McGee and Thomas 1986)

Elements of strategic position	
Sources of potential rents (unexpected events)	Isolating mechanisms
Changes in technology	Causal ambiguity
Changes in relative prices	Sunk costs and limited markets
Changes in consumer tastes	Switching and search costs
Changes in law, tax and regulation	Consumer and producer learning
Discoveries and inventions	Idiosyncratic investment
	Teams embodied skills
	Unique resources
	Special information
	Patents and trademarks
	Reputation and image

Mehra and Floyd (1998) proposed a model of strategic group formation based on the resource-based view of the firm with the industrial organization view of intra-industry heterogeneity. In other words, the strategic groups are formed by the presence of heterogeneous market positions with the inimitable resources. If these conditions do not exist in the market, the groupings of the companies in the market do not have any strategic meaning. Therefore, two other conditions are added to the definition of strategic groups; namely 1) the available strategic positions should be heterogeneous enough that the companies can differentiate along the dimensions of group strategies, and 2) the other companies cannot imitate the resources underlying the shared strategies. In conclusion, according to Mehra and Floyd (1998), the strategic groups in a market will exist, if one of

these conditions is satisfied, however, for existence of the performance difference between these strategic groups, both of these conditions should exist in the market.

Another view that explains the formation of strategic groups is based on the spatial competition and cognitive taxonomy of Tang and Thomas (1992). According to the view of "spatial competition", the reason of the strategic groups' formation is the relocation cost. According to this perspective, with zero or nearly zero relocation cost in the market, a single group will exist in the market, whereas high relocation costs lead to maximum differentiation. In other words, the formation of the strategic groups depends on the existence of the moderate relocation costs; therefore, if the strategic variables are easily changed, no group structure should exist. On the other hand, if the strategic variables are difficult to change, a group structure should be expected.

According to the cognitive perspective, the cognitive classification schemes are developed in the mind of the decision makers to simplify the market for being able to evaluate the differences between his or her firm and competitors. This perspective fills an importance gap in case of insufficient amount of data available to perform a strategic group analysis. On the basis of the direct and indirect imitative tendencies over time, the model of the market is developed in people's mind (Fiegenbaum and Thomas 1995). It is argued that due to the similar threats and opportunities encountered, shared information, interpretations, and expectations, cognitive models of different people become close to each other (Reger and Huff 1993). Therefore, the companies employing similar mental models behave similar in the market and apply the similar strategies in their business which creates group-level beliefs about the marketplace (Tang and Thomas 1992). Due to these group beliefs, the companies are perceived as a reference point in the process of strategic decision making; even they adjust their strategic behavior toward a group reference point (Fiegenbaum and Thomas 1995). Due to the insufficiency of literature on cognitive strategic groups, Peteraf and Shanley (1997) contribute to the cognitive strategic groups theory by developing the concept of a strategic group identity. They defined strategic group identity as "a set of mutual understandings, among members of a cognitive intra industry group, regarding the central, enduring and distinctive characteristics of the group". In this definition, the term of "mutual understandings" implies that the members can predict the behaviors of other members and the underlying logic of their decision making.

Nath and Gruca (1997) conducted a strategic group analysis among the care hospitals in a major metropolitan area by applying three alternative methods. In the first method, archival

data is clustered by factor analysis. The second is multidimensional scaling of managerial perceptions using industry specific attributes. In last method, the groups are identified by the managers in the industry directly. In conclusion, they concluded that the concept of a strategic group is a theoretical construct and it is not created artificially, as there is a convergence between multiple measures of the group structure in a mature, geographically delimited competitive environment, and the relationship between the theory and measurement of strategic groups was confirmed.

Above discussions pinpoint an important point about strategic group analysis: In order to be able to talk about strategic groups in a given industry, some conditions about mobility barriers, isolating mechanisms, reallocation costs and cognitive views should be met. Without a valid strategic differentiation reasons and a conceptual framework for strategic group analysis, there is a risk of identifying artificial strategic groups that actually do not exist. Thus, the prerequisite for strategic group analysis is definition of meaningful strategic variables to categorize firms in a market and using a sound clustering method that has the ability to reveal those categories.

2.4. Strategic groups and performance

Investigation of the link between strategic groups and performance has been a very popular research theme. However, whether there exists a direct relationship between the strategic group membership and performance is still questionable. There are various empirical studies carried out in different sectors to investigate this issue. For example, Frazier and Howell (1983) found no difference in performances among different strategic groups in the medical supply and equipment industry. Cool and Schendel (1987) found significant relationships according to market share but this relationship is not preserved in terms of profitability and risk-adjusted performance in the U.S. pharmaceuticals industry. In some of the studies like Amel and Rhoades (1988), Lewis and Thomas (1990) and Wiggins and Ruefli (1995), no clear relationship has been established. In contrast, Porter (1979) tried to determine the strategic groups according to the “size” of the companies and determined two groups named as “leaders” and “followers”. He proposed that the leaders gain more profit than the followers. Dess and Davis (1984) found significant performance differences in the paint and allied product sectors by using Porter’s (1980) generic strategies (differentiation, overall low cost and focus) as the strategic dimensions. Further, the study of Reger and Huff (1993) about banking industry demonstrated that significant differences exist between strategic

groups when the performance is measured in terms of return on assets. Some examples of studies which show some performance differences between the strategic groups are those by Mascarenhas and Aaker (1989), Fiegenbaum and Thomas (1990), and McGee and Thomas (1986). Kale and Arditi (2003) empirically proved that differentiation strategy is positively related with performance in the US construction industry.

The inconsistency of results may be attributed to different factors. Strategic group analysis requires choosing or developing a conceptual framework in which the strategic dimensions are defined. Yet, there is no consensus on these dimensions. Caves and Porter (1977) suggested that due to differences in risk posture or random differences in skills and quality of assets, firms tend to utilize different strategies, and this creates mobility barriers and different strategic groups. In other words "strategic choices" are the reasons why different strategic groups exist. Consequently, Porter (1980) defined two strategic dimensions which may be used for strategic grouping: mode and scope of competition. Mode of competition refers to a firm's decisions on how to achieve competitive advantage whereas scope of competition refers to a firm's decisions on the breadth of its operations. Thus, according to Porter (1980), the performance differences between firms are explained by different competitive positions resulting from different strategic choices. Since late 1980s, there has been a surge of interest in the role of resources and capabilities of a firm as the basis for strategy and primary determinants of profitability, which is considered under "the resource-based view". Rather than the "generic" sources of competitive advantage, resource-based view concentrates upon the resources and capabilities that underlie these advantages. Accordingly, resources and capabilities are proposed to be considered during strategic grouping. McGee and Thomas (1986) also argue that mobility barriers (which may be due to the resources and capabilities as well) provide a much firmer basis for identifying groups rather than strategies which tend to be loosely defined. Cool and Schendel (1987) argue that mobility barriers alone are insufficient to explain profitability; firm-level characteristics and market factors should also be considered. In their work, strategic resources are considered as well as scope for strategic grouping. Industry specific variables are also proposed to be used as strategic dimensions. For example, Hodgkinson (1997) defines the historical evolution of the industry among the other variables which are; differentiating resources and capabilities, unique goals, different chronological points of entry, segmentation and risk profiles. Relevant strategic dimensions may also change with respect to the characteristics of each industry. As the strategic dimensions are industry specific, the direct comparison of results from different industries may not be possible. Similarly, as different performance indicators

(financial indicators such as return on equity, growth, market share or subjective reporting of overall performance) are used, research results about impact of strategic grouping on performance may not be directly comparable.

Also, there is no universally accepted technique for carrying out strategic group analysis. Most common method for identifying strategic group structures is through cluster analysis. Major challenges of cluster analysis are; choosing the clustering variables, algorithms, number of clusters and validating clusters. Cluster analysis is criticized as it relies heavily on researcher judgment and it does not offer a test statistic that supports results (Ketchen and Shook 1996). Using perceptions of industry professionals to identify the structure of the industry and competitive positions of firms is an alternative way. The study of strategic groups from a cognitive perspective has gained prominence during the past years (Hodgkinson 1997). According to this cognitive view, the cognitive classification schemes developed in the mind of the decision makers to simplify the market for being able to evaluate the differences between his or her business and competitors may be used for strategic grouping (Tang and Thomas 1992). The concept of "cognitive groups" first appeared in Porac et al.'s article (1989) about strategic groups in the Scottish knitwear industry. Construction of a set of cognitive strategic groups based on expert judgment eliminates some of the criticisms about reliability of cluster analysis. Osborne et al. (2001) argue that cognitive and performance-based groups may finally emerge.

As a final note, strategic group analysis has got its benefits and shortcomings. Key question is whether strategic group analysis provides any information that cannot be gained from the study of industries and individual firms. In order to answer this question, more research is needed to investigate behavior of strategic groups and possible group-level impacts on performance. Strategic group analysis is known to be more a descriptive than a predictive tool. It is unlikely to offer much insight into why some firms in an industry perform better than others (Grant 1995). However, it can increase understanding about the structure of the industry, strategic issues and dynamics of the competitive environment. In this research, an attempt is made to understand the dynamics of competition in the construction industry through strategic group analysis and explain the performance difference between the groups (if they exist) so that some recommendations can be made to the companies in different groups.

2.5. Strategic group analysis in construction management

There are a limited number of studies within the construction management literature about strategic grouping and its performance implications. Kale and Arditi (2002) argue that research on competitive positioning in the construction industry appears to be unbalanced in favor of anecdotal or descriptive approaches. Only a few construction management researchers (such as (Jennings and Betts 1996)) have conducted empirical research studies in this area and explored performance implications (Akintoye and Skitmore 1991; El-Mashaleh et al. 2006; Hampson and Tatum 1997). In their study, Kale and Arditi (2002) used one of the most influential generic typologies, Porter's (1980) generic competitive positioning typology to classify US firms. Modes of competition have been identified as quality, product and service innovations, time and cost. A construction company may select one mode or place varying degrees of emphasis on all of the modes (which is called as a hybrid mode). Scope of competition is defined as either narrow or broad market approach. A questionnaire has been prepared in which all items are measured on a 1-5 Likert scale. Performance is measured by using a subjective reporting approach. K-means cluster analysis is used to classify firms according to the identified dimensions (mode and scope). As a result, based on the responses from 107 firms, four clusters have been identified. Statistically significant performance differences have been found between the clusters. Authors claim that differences in construction companies' performances can be partly explained by their choices of mode and scope of competition. It is found that construction companies that outperform their rivals adopt a hybrid mode, rather than a single mode of competition. However no statistically significant finding was found about impact of scope on performance.

Claver et al. (2003) studied the linkage of strategic groups and performance by examining business strategies of Spanish contractors. Using the data regarding 88 housing contractors and defining variables that are based on Porter's generic strategies, authors identified four strategic groups. However, the empirical findings demonstrate that significant differences in performance do not exist among these groups.

The aim of this study is to conduct a strategic group analysis to understand if different strategic groups exist within the Turkish construction industry and to identify the reasons related with strategic positioning of firms to explain the performance differences between different groups. Consequently, in the next chapters, the conceptual framework that is

defined as a basis of strategic grouping is discussed together with the alternative clustering techniques.

CHAPTER 3

CLUSTER ANALYSIS

Clustering is defined as "a mathematical technique designed for revealing classification structures in the data collected in the real world phenomena" (Mirkin 1996). In other words, the main purpose of clustering is to reveal the classification structure of the data. The term "classification" was defined by Platts (1980) as "in the strictest sense, it means ordering or arranging objects into groups or sets on the basis of their similarities or relationships". The purpose of classification is formulated as to shape and keep knowledge; to analyze the structure of phenomenon and to relate different aspects of a phenomenon in question to each other (Mirkin 1996). According to these definitions, the classification can be defined as simplifying a complex structure by ordering or arranging objects into groups according to their similarities or relationships without losing significant information about the data for providing comprehension of the main structure of the data and relations in the data in a short time and minimum effort. There are many different methods available for classifying the data. The frequently used methods are discriminant analysis and cluster analysis. Although these two methods can be used for classifying the observations, there are many differences in application of these methods. Discriminant analysis requires the user to know group membership values for the cases used to derive the classification rule. In other words, at the inception of the analysis, the number of groups should be known and the groups should contain data points which illustrate the general structure of the group. In cluster analysis, there is no prerequisite for starting the analysis. Since the number of the clusters and the members of these clusters are rarely known in the ordinary life, the cluster analysis provides many benefits to the researchers.

In this chapter, one of the widely used clustering analysis method for strategic grouping, namely traditional cluster analysis, will be explained. Following sections summarize the similarity measures, clustering algorithms used in a cluster analysis and factors which can affect the performance of the analysis. Finally, validation techniques used in evaluating the quality of the clusters will be discussed.

3.1. Cluster analysis in the construction management literature

The clustering of the items of any kind was practiced and developed from the beginning of ancient years to present date. The examples of clustering can be seen in grouping of similar sounds into meaningful word, to the simulation of learning, recognition, and prediction processes on computers (Zupa 1982). In construction management literature, classification is also a popular subject. For instance, Chinyio et al. (1998) classified the construction clients according to the client's needs on economy, function, safety, quality, time, running/maintenance costs, and flexibility of the projects by using the Ward's method for identifying contractors who can satisfy the needs of the clients. Holt (1997) classified the construction contractors according to twenty-one prequalification criteria and eight tenderer criteria for reducing a large number of potential bidders to identify only those suitable to tender for a particular project by utilizing the hierarchical and K-means cluster analysis. Acar et al. (2005) used K-means cluster analysis method for classifying the contractors according to their size and the number of the employees. Their purpose in that study is determination of the significance difference between the different sizes of the companies with respect to their information and communication technologies-related attitudes. Also, Kale and Arditi (2003) classified the construction companies in United States according to the strategic variables that are chosen by referring to Porter's competitive generic strategies. Strategic groups in United States construction industry was tried to be identified by using K-means clustering analysis. In the same manner, Claver et al. (2003) tried to determine the strategic groups of Spanish house-building firms by utilizing Ward's method for finding out the number of the clusters as an input of the K-means clustering analysis. The organizations are also classified according to the relative value of their respective safety management index, and "safety management clusters" are formed for examining the relationship between the intensity of safety management commitment and to the overall safety performance, proactiveness and record by using complete linkage method (Mohamed 1999). In addition, Kumaraswamy et al. (2000) classified the Hong Kong construction companies for distinguishing between the strongly positive and negative outcomes from ISO 9000 certification system implementations by using 'between- groups linkage' methods. The other important application area of classification in construction management is determination of market segmentation. For instance, Bourassa et al. (1997) used Ward's and K-means methods to analyze housing submarkets in Sydney and Melbourne, Australia. Kauko (1997) also showed an exploratory application with the self organizing map (SOM) in the context of analyzing housing markets in Finland and Helsinki (Kauko 2003).

3.2. Cluster analysis: Definitions and criticisms

The term cluster analysis was first used by Tryon (1939); since then, the cluster analysis was used in many fields including biology, zoology, marketing, archeology, agriculture, economics, education, geology, political science, market research, genetics, medicine, psychology, data mining and pattern recognition (Everitt et al. 2001). The major aim of the cluster analysis is finding K clusters, so that the objects of one cluster are similar to each other whereas these are dissimilar to the objects of the other clusters (Bacher 2002). Gordon (1999) defines the cluster analysis as "the subject of classification is concerned with the investigation of the relationships within a set of ’objects’ in order to establish whether or not the data can validly be summarized by a small number of classes (or clusters) of similar objects". From that definition, validity of the clusters and small number of clusters are important priorities for successful cluster analysis.

Cormack (1971) estimated the publication rate of articles on clustering and classification at over 1000 publications per year (Seber 2004). The reason of the popularity of cluster analysis is that the goals of cluster analysis are varied and include widely different activities. One of them is reducing the data objectively from an entire population or sample to information about specific, smaller subgroups. Secondly, cluster analysis can be used for developing hypotheses or examining already stated hypotheses (Hair et al. 1995). In addition to these, the cluster analysis can be used for finding true typology, model fitting, prediction based on groups, data exploration (Everitt 1974).

The cluster analysis is an effective instrument for analyzing the data, but the method has some caveats. Aldenderfer and Blashfield (1984) stated these caveats as;

1) Due to the relatively simple procedures, statistical reasoning does not support the most of the cluster analysis methods.
2) Due to the evolvement of cluster analysis methods from many disciplines, methods are affected from the biases of these disciplines.
3) Due to the evolvement of clustering methods from disparate sources that have stressed different rules of group formation, different solutions can be generated from different clustering methods for the same data set.
4) The strategy of cluster analysis is structure seeking although its operation is structure- imposing.

In addition to these, Hair (1998) stated that the variables used as basis for similarity measure determines the cluster analysis solution, so, the changes in the variables cause huge changes in the result of the cluster analysis. Another criticism point of the cluster analysis is that there are no available tests that determine the accuracy of the results of the analysis, and all of the results are verified by the judgment of the researchers. Despite the controversy surrounding cluster analysis, cluster analysis is the best way for analyzing the multiple variables as configuration analysis (Ketchen and Shook 1996).

In this research, different clustering techniques are used to minimize the sensitivity of results to the chosen clustering technique. Fundamental concepts, techniques and some critical success factors for clustering applications that are followed in this research are explained in the next sections.

3.3. Exploratory and confirmatory cluster analysis

Clustering techniques can be used for two purposes, namely exploratory and confirmatory. Since the only requirement of the cluster analysis is the specification of the variables and cases used in the analysis, it can be used for exploratory purposes. In the application of exploratory cluster analysis, the number of clusters is identified according to the outputs of the analysis. Despite the advantages of cluster analysis in confirmatory applications, the clustering techniques are used rarely as a confirmatory method. The reason of this can be explained by the prerequisite regarding the number of clusters and certain characteristics of the clusters. Bacher (2002) tabulated the differences between the exploratory and confirmatory usage of cluster analysis as shown in Table 3.1.

Table 3-1: Differences between exploratory and confirmatory cluster analysis

Exploratory cluster analysis	Confirmatory cluster analysis
• The number of clusters is unknown. => The number of clusters has to be estimated.	• The number of clusters is known. => The number of clusters shall not be estimated.
• The characteristics of clusters (e.g. cluster centers in K-means) are unknown => Clusters have to be interpreted. Finding a substantive interpretation can be difficult.	• Characteristics of clusters are - at least partially - known. => Clusters already have a substantive interpretation.
• The fit to data is maximized.	• The fit to data may be poor.

3.4. Data

Seber (2004) states that three main types of data are used in clustering. The first is nxp data matrix. It is denoted as shown Figure 3-1. In this matrix, the row stands for objects and columns stands for variables.

$$\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

Figure 3-1: nxp data matrix

The second one is nxn proximity matrix $[(c_{rs})]$ or $[(d_{rs})]$, where c_{rs} (d_{rs}) is a measure of the similarity (dissimilarity) between the rth and s^{th} objects. This kind of data occurs frequently in social sciences and in marketing.

The last type of data that is already in a cluster format is what might be called sorting data.

3.5. Dissimilarities

Determining the dissimilarities (or similarities) between the objects is the first and most important step in cluster analysis, since this step affects outputs of the following steps of

cluster analysis. Even, Aldenderfer and Blashfield (1984) stated "to be successful, science must be based upon objective, replicable procedures; therefore the development of statistical procedures to measure objectively the similarity of things is a natural consequence of necessity for replicable and reliable classification". There are different methods for determining the dissimilarities, but how the dissimilarity between two objects is computed depend on the type of the original variable. Variables are classified into eight types according to scale and type of the data.

1. **Interval-scaled variables**: are continuous measurements on a linear scale. Typical examples are temperature, height, weight and energy.
2. **Continuous ordinal variables**: are continuous measurements on an unknown scale, or where only the ordering is known but not actual magnitude.
3. **Ratio scaled variables**: are positive continuous measurements on a nonlinear scale, such as an exponential scale. For example, the growth of a bacterial population. With this model, equal time intervals multiply by the population by the same ratio.
4. **Discrete ordinal variables**: Have m possible values which are ordered.
5. **Nominal variables**: have m possible values, which are not ordered.
6. **Symmetric binary variables**: have two possible values, coded 0 and 1, which are equally important. For example, male and female, animal or vegetation.
7. **Asymmetric binary variables**: have two possible values, one of which carries more importance than the other. The most meaningful outcome is coded as 1, and the less meaningful outcome as 0. Typically, 1 stands for the presence of a certain attribute, 0 for its absence.
8. **Variables of mixed type**: many data sets contain variables of different types.

3.5.1. Similarity measures for numerical variables

For interval scaled variables, the following distance measurements are used directly,

- **Euclidean distance**: for continuous variables, the most common used dissimilarity measure is Euclidean distance. The formula of Euclidean distance is $d_{ij}=\sqrt{\sum_{k=1}^{p}(x_{ik}-x_{jk})^2}$,where d_{ij} is the distance between i and j cases, and x_{ik} is the value of kth variable for i^{th} case.

- **Manhattan distance**: $d_{ij}= \sum_{k=1}^{p}|x_{ik} - x_{jk}|$ This distance is simply defined as the average difference across dimensions. This distance tends to have low sensitivity to outliers according to Euclidean distance, since it does not use the square of the difference. On the other hand, no correlation between the variables is assumed in this distance; therefore this leads to some questions about the validity of the findings when the correlation exists between the variables.

- **Square Euclidean distance**: $d_{ij}= \sum_{k=1}^{p}(x_{ik} - x_{jk})^2$The formula of this similarity measure and Euclidean distance formula is approximately same, except the square root. This similarity measure provides benefits in analysis of complex data by lessening the computation durations through increasing computation speed due to the extrusion of the square root, on the other hand, this leads to high sensitivity to the outliners.

- **Chebychev distance**: $d_{ij}=|x_{ik} - x_{jk}|$ This distance measure may be appropriate in cases when one wants to define two objects as "different" if they are different on any one of the dimensions.

- **Minkowski distance**: $d_{ij}= (\sum_{k=1}^{p}(|x_{ik} - x_{jk}|)^p)^{1/p}$ In this method, varying p in changes the weight assigned to larger and smaller distances.

- **Power distance**: $dij= (\sum_{k=1}^{p}(|x_{ik} - x_{jk}|)^p)^{1/r}$ Sometimes one may want to increase or decrease the progressive weight that is placed on dimensions on which the respective objects are very different. Parameter *p* controls the progressive weight that is placed on differences on individual dimensions; parameter *r* controls the progressive weight that is placed on larger differences between objects.

Other available method used for calculating similarity measures of quantitative variables is correlation coefficients. Correlation coefficient is the standard measure of the linear relationship between two variables. The most popular correlation coefficient is product moment correlation coefficient, and it is formulated as;

$$r_{jk}=\frac{\sum(x_{ij} - \bar{x}_j).(x_{ik} - \bar{x}_k)}{\sqrt{\sum(x_{ij} - \bar{x}_j)^2.(x_{ik} - \bar{x}_k)^2}} \qquad 3.1$$

where x_{ij} is the value of the variable i for case j, and x_j is the mean value of the variables for case j. This correlation coefficient can be used with binary variables by transforming the data into familiar phi coefficient. The value of r_{jk} is between -1 and +1 and does not depend on the choice of measurement units (Kaufman and Rousseeuw 1990), therefore the values obtained

from this formula are independent from scale difference between the variables. This measure should be preferred when the similarity of profile shapes is considered more important than the similarity of the average profile levels, because the correlation is calculated as unity whenever two profiles are parallel, irrespective of how far apart they are. Bacher (2002) advised that the correlation coefficients can be better for clustering variables.

This distance measure is sensitive to shape at the expense of the magnitude of differences between variables. This means that two cases can have a correlation of +1, but this does not mean that they are identical. As a result, this measure can lead to misleading results. In conclusion, Seber (2004) concluded that dissimilarities based on metrics are better proximity measures than correlations, and Cormack (1971) states that when variables are uncoded, comparable measurements or counts, then the correlation coefficients should be used; as it is not invariant under scaling of variables, or even under alterations in the direction of coding some variables.

3.5.2. Similarity measure for binary variables

Binary variables have only two possible outcomes, and in data matrix, they are coded as zero or one. Often 1 indicates a certain attribute is present, whereas 0 means the absence of the attribute. Different similarity measures are proposed for binary variables and these entire measures base on matches and mismatches of the entries in cross-classification in the p variable for two individuals (Everitt et al. 2001). The general version of this cross-classification is shown in Table 3-2.

Table 3-2: Counts of binary outcomes for two individuals

	Individual i			
	Outcome	1	0	Total
Individual j	1	a	b	a+b
	0	c	d	c+d
	Total	a+c	b+d	P=a+b+c+d

Bacher (2002) tabulated the coefficients and their formulas and properties for binary variables as follow;

Table 3-3: Coefficients for Binary Variables (Bacher 2002)

Similarity coefficient	Formula	Properties
Jaccard	$\frac{a}{a+b+c}$	Conjoint absence (0,0) is ignored.
Dice	$\frac{2a}{2a+b+c}$	Conjoint absence (0,0) is ignored, conjoint presence (1,1) is double weighted.
Sokal and Sneath	$\frac{a}{a+b+c}$	Conjoint absence (0,0) is ignored, mismatches are double weighted.
Russel and Rao	$\frac{a}{a+b+c+d}$	Conjoint absence (0,0) is not evaluated as similarity, but used in the denominator.
Simple matching	$\frac{a+d}{a+b+c+d}$	Absence and presence as well as matches and mismatches have equal weights.
Sokal – Sneath II	$\frac{2(a+d)}{2a+b+c+2d}$	Matches (conjoint absence and presence) are weighted double.
Rogers and Tanimoto	$\frac{a+d}{a+2(b+c)+d}$	Mismatches are weighted double.

There are two types of binary similarity coefficients, for the binaries like sex whose outputs are male and female; there is no preference which outcome should be coded as 0 or 1. For these variables, double zeros become meaningful and in these cases zero-zero matches are completely equivalent to one-one matches. Such a variable is called symmetric. But for asymmetrical variables, double zeros become meaningless, for that reason it can be excluded from the analysis (Kaufman and Rousseeuw 1990). In Table 3-3, first three coefficients are used for asymmetrical coefficients, and the remaining coefficients are for symmetrical coefficients.

3.5.3. Similarity measure for nominal variables

The variables which can take on more than two states are called nominal variables, for example marital status, nationality of the people. These states are usually denoted by the numbers, but these are not ordered in any way, in other words the state denoted by two is not greater than the state denoted by one. The code numbers are only used to facilitate data

handling, but the states can be coded by letters or other symbols (Kaufman and Rousseeuw 1990).

The simple matching is the most common way of measuring the similarity between the objects for nominal variables. In this method, firstly a score s_{ijk} is allocated to zero or one to each variable k, depending on whether the two individuals i and j are the same on that variable. These scores are then simply averaged over all p variables to give the required similarity coefficient as (Everitt et al. 2001);

$$s_{ij} = \frac{1}{p} \sum_{1}^{p} s_{ijk} \qquad 3.2$$

In addition to this measure, other matching coefficients for nominal data are possible. These matching coefficients are determined by extending from binary measures to nominal data. these measures are tabulated in Table 3-4.

Table 3-4: Matching coefficients for nominal variables (Gan et al. (2007))

Measure	S(x,y)	Weighting of matches, mismatches
Russell and Rao	$\frac{N_{a+d} - N_d}{N_{a+d} + N_{b+c}}$	Equal weights
Simple matching	$\frac{N_{a+d}}{N_{a+d} + N_{b+c}}$	Equal weights
Jaccard	$\frac{N_{a+d} - N_d}{N_{a+d} - N_d + N_{b+c}}$	Equal weights
Dice	$\frac{2N_{a+d} - 2N_d}{2N_{a+d} - 2N_d + N_{b+c}}$	Double weight for matched pairs
Rogers-Tanimoto	$\frac{N_{a+d}}{N_{a+d} + 2N_{b+c}}$	Double weight for unmatched pairs
Kulczynski	$\frac{N_{a+d} - N_d}{N_{b+c}}$	Matched pairs excluded from denominator

In this table, N_{a+d} represents the number of the states on which the two records match, N_d is the number of states on which two records match in a “not applicable” category, and N_{b+c} is the number of the states on which the two records do not match.

3.5.4. Similarity measure for mixed variables

The variables in a data set are usually described by more than one type of attribute. Gower (1971) proposed a measure of similarity which has been widely implemented and used to measure the similarity for two mixed-type data points. This coefficient is defined as;

$$s_{ij} = \frac{\sum_{k=1}^{p} S_{ijk} * W_{ijk}}{\sum_{k=1}^{p} W_{ijk}} \qquad 3.3$$

where W_{ijk} is a weighting variable valued at 0 or 1 depending on whether the comparison is considered valid for variable k, and it is set zero when it is not valid. S_{ijk} is a similarity 'score' based upon the outcome of the comparison of the variable k across cases i and j. For quantitative attributes, S_{ijk} is defined as;

$$S_{ijk} = 1 - \frac{|i_k - j_k|}{R_k} \qquad 3.4$$

where R_k is the range of kth attribute; W_{ijk}=0 if data points has missing values at the kth attribute; otherwise W_{ijk}=1. For binary attributes, S_{ijk}=1 if both data points have k^{th} attribute "present", otherwise S_{ijk}=0, and W_{ijk} is zero when variable k is not known for one or both individuals under comparison (Everitt 1980). For nominal attributes, S_{ijk}=1 if i_k=j_k; otherwise S_{ijk} becomes zero. W_{ijk} is determined as the other types of attributes.

From this formula, S_{ijk} gets values between zero (when the two data points are identical) and one (when the two data points are extremely different) (Gan et al. 2007).

3.6. Critical succes factors for cluster analysis

The clusters obtained at the end of the cluster analysis can be affected from some factors, namely missing values, scale differences in the variables, and multi-linearity between the variables. Therefore, at the inception of the analysis, the data set should be examined aganist these factors in order to obtain optimum solutions.

3.6.1. Missing values

The missing values in a data set are a common problem in multivariate analysis. The missing values can occur for a variety of reasons like problems in measurement, availability of information, impossibility of measurement (Kaufman and Rousseeuw 1990), therefore

methods are developed for handling the missing values. Bacher (2002) mentioned about three methods for handling the missing values in cluster analysis.

- Listwise deletion: in this method, if one or more variables of a case are missed, then this case is excluded from the analysis. If many variables are used to cluster cases, the number of cases may be reduced dramatically. This method is not preferred due to losing of the information regarding this case; especially in cases when the number of the data points is limited.
- Pairwise deletion: this method tries to use all available information, and assigns cases to clusters based on distances computed from all variables with non-missing values. In this method, Gower's general similarity coefficient can be used to construct a proximity matrix for individuals who have at least one attribute value (Everitt et al. 2001).
- Estimating missing values with imputation techniques: this method is not recommended for cluster analysis, and only it should be used if it is suspected that the individuals belong to different groups (Everitt et al. 2001).

Kaufman (1985) studied the effect of different treatments of missing values for Ward's method; he concluded that listwise deletion results in fewer misallocations of cases than pairwise deletion. However, the differences between the two methods were small.

3.6.2. Standardization

The scale differences between the variables can lead to incoherent clustering outputs. For instance, the scale difference between age and income variables is high. Therefore, the income dominates the results throughout the clustering process, and if these two variables have same importance in determination of clusters, then this will affect the reliability of the results. Consequently, the data should become dimensionless by using standardization methods. Standardization is advised by the some authors, like Milligan and Cooper (1988), Hair et al. (1998) and Harrigan (1985) in cluster analysis applications. However, due to the elimination of meaningful differences among members (Ketchen and Shook 1996), it is not preferred in some studies; moreover, some experts like Edelbrock (1979), Milligan (1980) stated that standardization has no significant effect. According to Aldenderfer and Blashfield (1984), standardization should be decided on a problem to problem basis, and the users should be aware of the changes in the results based on the standardization. Ketchen and

Shook (1996) advised that standardization decision should be made according to the outputs obtained by using and not using standardization. According to the consistency of the results, the validity of each should be assessed; the output which shows the highest validity is chosen as the solution for this data.

Two types of standardization is used essentially, namely global standardization and within-cluster standardization. In global standardization, the variables are standardized across all elements in the data set. But within-cluster standardization standardizes the variables within cluster on each variable, so this standardization requires the knowledge of the clusters which is unknown at the beginning of the analysis. Overall and Klett (1972) proposed an iterative approach to overcome this problem (Gan et al. 2007).

The most common way of standardization is the conversion of each variable to standard scores by subtracting the mean and dividing by standard deviation for each variable. This process called z-score standardization converts each raw data score into a standardized value with a zero mean and a unit standard deviation. This transformation eliminates the bias introduced by the differences in the scales of the several attitudes or variables in used in the analysis (Hair 1998).This standardization can be applied in global standardization, however not in within-cluster standardization (Milligan and Copper 1988).

3.6.3. Multi-linearity

High correlations between variables can cause flawed results by overweighting one or more underlying constructs. The remedy for this issue can be using the Mahalanobis distance which both standardizes variables and adjust for high correlations in calculations of similarity measures (Ketchen and Shook 1996). The Mahalanobis distance is formulated as follows (Seber 2004);

$$\Delta(x_r, x_s) = \{(x_r, x_s)'.S^{-1}(x_r, x_s)\}^{1/2} \qquad 3.5$$

Also, since this distance measure puts away the standardization requirement, the problems related to the standardization is also eliminated. Unfortunately, this distance measure is not available in popular statistical package programs, such as S-plus and statistical package for the social sciences (SPSS).

The another solution of the multi-linearity is subjecting the variables to factor analysis in which principal components analysis with orthogonal rotation is advised by Ketchen and

Shook (1996) and using uncorrelated factor scores for each observation as the basis for cluster analysis. However, this application can cause exclusion of the factors which may provide unique information (Dillon et al. 1989).

These two methods have a cost; therefore the researchers should check the existence of multi-linearity by performing correlation analysis before the application of these methods. If the multi-linearity between the variables is obvious, then for every cluster analyses, one of the methods addressing the multi-linearity should be applied, and then the stability of the outputs should be examined (Ketchen and Shook 1996).

3.7. Traditional cluster analysis methods

There are numerous traditional cluster analysis methods, but three of them are mainly used, namely hierarchical, iterative partitioning and two-step cluster analysis.

3.7.1. Hierarchical cluster analysis

Hierarchical clustering algorithms operate by grouping data objects into a tree of clusters. This method can be divided into two main approaches;

- Divisive hierarchical techniques
- Agglomerative hierarchical techniques

Hierarchical agglomerative algorithms start by initially assigning each object to its own cluster and at every step merging the pairs of cluster for forming a new cluster according to the similarities between the clusters until reaching a cluster which contains all objects or a certain stopping criterion is met. Hierarchical divisive algorithms start inversely, namely by initially forming a cluster made of all objects, and at each step most dissimilar objects are split off and made into smaller clusters until a certain stopping criterion is met.

Hierarchical cluster methods are conceptually simple to understand; in fact Groth (1998) mentioned about the greatest benefit of hierarchical methods as their understandability. The algorithm of the hierarchical cluster method is the simplest among the algorithms of other clustering methods. These clustering methods produce non-overlapping clusters; therefore the clusters obtained at the end of application of these methods are nested. Due to these properties, computation time of these methods is smaller than other cluster algorithms. Despite fast computation speed, the data storage requirements cause problems when dealing

with large sample sizes. For example, a sample of 400 cases requires storage of approximately 80.000 similarities, and this number increases to 125.000 for a sample of 500 (Hair 1998).

The results of agglomerative and divisive cluster methods are represented by using different graphical formats. These are n-tree, dendrogram, banner, pointer representation, packed representation and icicle plot. Dendrogram (tree diagram) shown in Figure 3-2 is the most used representation method among these methods.

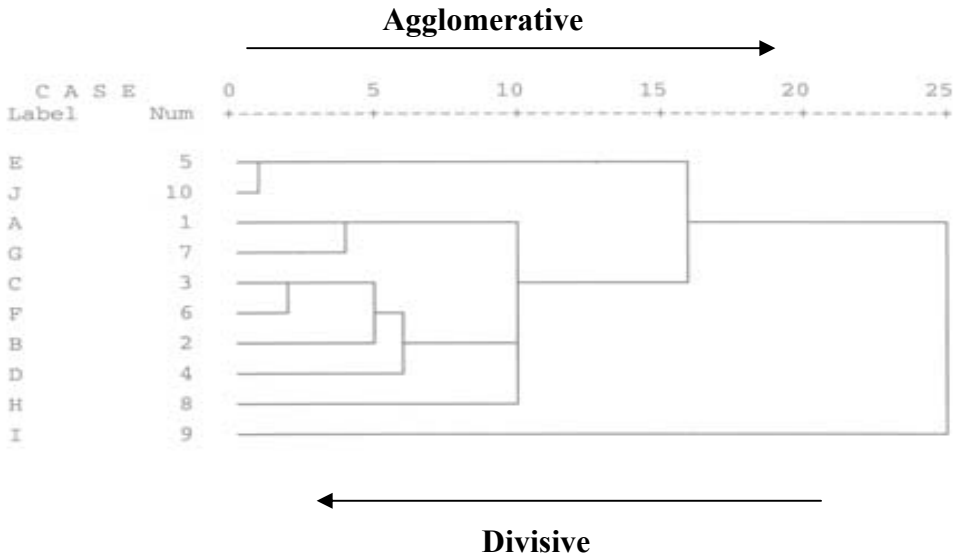


Figure 3-2: Dendrogram example

Dendrogram is defined as the graphical presentation of the results of a hierarchical procedure how each object is merged each other at the steps of the hierarchical procedure until all are contained in a single cluster (Hair 1998). The nodes of the dendrogram term are used for clusters and the lengths of the stems (height) represent the distances at which clusters are joined (Everitt et al. 2001). Dendrogram shows how the clusters are similar to each other by considering heights of the connection of the nodes, for example in the figure height of the connection point of cases 5 and 10 is the lowest one, that means that these two cases are the closest case among all the cases. After connection of those cases, cases 3 and 6 are combined and form a new cluster. Finally, case 9 joins the other cases to form the final cluster, therefore this case can be considered as an outline. The dendrogram can be a useful tool to comprehend the structure of the clusters. Another important point is that these methods require n-1 steps to cluster a similarity matrix. Different appearances of

dendrograms can be provided for the same data and clustering procedure, depending on the order in which the nodes are displayed (Everitt et al. 2001).

Another method for showing the results of the hierarchical cluster analysis is agglomeration schedule, the agglomeration schedule is a numerical summary of the cluster solution and does not inform which objects belong to a cluster and how many clusters are formed, but it gives information about the process and hints about the number of the clusters. An example of the agglomeration schedule can be seen in Table 3-5.

Table 3-5 : Agglomeration schedule, nearest neighbor (single linkage) and squared Euclidean distance (standardized).

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	8	11	1,260	0	0	7
2	6	7	1,579	0	0	4
3	2	9	1,625	0	0	7
4	1	6	2,318	0	2	6
5	3	5	2,619	0	0	8
6	1	10	3,670	4	0	10
7	2	8	4,420	3	1	8
8	2	3	4,505	7	5	9
9	2	4	4,774	8	0	10
10	1	2	5,718	6	9	0

In the coefficients column, the values of the distance (or similarity) statistic are used to form the cluster. If dissimilarity measures are used, small coefficients show that fairly homogeneous clusters are being attached to each other. Large coefficients show that dissimilar clusters are combined. For similarity measures, the opposite is valid.

The actual value shown depends on the clustering method and the distance measure used. These coefficients can be used to decide how many clusters should represent the data. Cluster formation can be stopped according to scale of the increase (for distance measures) or decrease (for similarity measures) in the coefficients column between two adjacent steps.

In this example, at the first stage cases 8 and 11 are combined. The next stage column shows that the created cluster will appear in stage 7. When there are many cases, this table becomes rather long, but it may be easier to scan the coefficients column for large gaps rather than scan the dendrogram. For this example, the largest steps in the coefficients column occur between stages 5 and 6, indicating a 6-cluster solution and stages 9-10, indicating a 2- cluster solution.

The agglomeration levels should continuously increase or decrease depending on using similarities or dissimilarities, unfortunately the agglomeration schedule does not behave like that in all techniques, namely median linkage, centroid linkage and within average linkage (Bacher 2002), in addition the agglomeration schedule does not indicate the ties, however ties can influence the results (Gordon 1999).

The agglomeration schedule can be visualized by establishing inverse scree tree. In this visualization, the x-axis shows the number of the clusters and y-axis shows the agglomeration levels. According to an elbow knick in the graph which shows the sharp increase in the agglomeration schedule, the number of the clusters is determined. However, the methods usually do not result in a unique solution. It may be difficult to determine the number of hills or to decide if a cluster is a small hill. Very often two or more sharp increase exists. In this situation it is difficult to select one cut. Bacher (2002) advised to choose the first elbow, but this decision may result in too many clusters.

3.7.1.1. Agglomerative methods

The most widely used method among hierarchical methods is agglomerative procedures (Everitt et al. 2001). There are different algorithms used for developing the cluster analysis, and these algorithms differ in how modified the distance between clusters. Aldenderfer and Blashfield (1984) mentioned that there are at least twelve different possible linkage forms, and six of these methods are widely used in the literature. The agglomerative hierarchical methods are divided into two groups, namely graph and geometric methods. Single-linkage, complete-link, average methods are included into the first group, and second group contains Ward's, centroid and median methods. A sub-graph or interconnected points represents a cluster in graph methods, whereas in geometric methods a cluster can be represented by a center point (Gan et al. 2007).

Hierarchical cluster analysis methods have some caveats, firstly the procedure makes only one pass through the data, and a poor early partition of data may persist throughout the analysis and lead to artificial results. Also, there is no possibility of correcting a poor initial partition. Another important problem is that the results of the cluster analysis can be changed by reordering the data in the similarity matrix and by dropping the cases from the analysis, in other words the stability is an important problem for this method. The problem of stability is especially an important problem when dealing with small samples of cases (Aldenderfer and Bashfield 1984). In addition, these methods are failed to recover non-spherical clusters (Everitt et al. 2001).

3.7.1.1.1. Single linkage (nearest neighbor)

The single-link method was first introduced by Florek et al. (1951) and independently by McQuitty (1957) and Sneath (1957) (Gan et al. 2007). In this method, the distance between two clusters is determined by comparing the similarity distances between all the objects in one cluster and all objects placed in the other cluster and the largest distance is used for the distance between these two clusters. The following steps illustrate the process of single linkage method, and Table 3-6 shows the similarity matrix of the objects. This example is taken from Seber (2004).

Table 3-6: Proximity similarity matrix

	A	B	C	D
A	1			
B	0.500	1		
C	0.250	0.400	1	
D	0.600	0.350	0.300	1

Steps of the single linkage method are as follows:

Step 1: According to Table 3-6, A and D are determined as the most similar cases, so these two objects are combined for forming a new cluster.

Step 2: New proximity matrix is established by determining the larger similarity distance between the members of the cluster and the other cases. The distance of B-A is 0.500 and B-

D is 0.350, so the distance between the new cluster which contains A-D and B is determined as 0.500. This procedure repeats between new cluster and C object, and the distance is determined as 0.300. The modified proximity similarity is shown in Table 3-7.

Table 3-7: Modified proximity similarity matrix

	A-D	B	C
A-D	1		
B	0.500	1	
C	0.300	0.400	1

Step 3: The similarities between the objects are compared, and the pair more similar than the other pairs are combined for forming a new cluster or joining an established cluster. According to the Table 3-7, the most similar pair is determined as A-D cluster and B, so B joins the A-D cluster. This procedure contains until placing all objects into one cluster.

The advantage of single linkage method is its invariant to monotonic transformations of the similarity matrix, and also it is not affected from ties in the data. This means that single linkage will not be affected by any data transformation that retains the same relative ordering of values in the similarity matrix (Aldenderfer and Blashfield 1984). The disadvantage of this method is the possibility of chaining which means formation of unbalanced and straggly clusters, so too few, large and heterogeneous clusters may be formed at the end of the process (Bacher 2002). Also, cluster structure is not taken account during the process (Everitt et al. 2001).

3.7.1.1.2. Complete linkage (furthest neighbor)

This method is the opposite of the single linkage method, because the similarity between two clusters is defined as the smallest similarity between an object of the one cluster and an object of the other. For illustrating the process of this method, the example used in the single linkage can be used, according to the second step of process, the similarity between B-A and B-D are determined as 0.500, 0.350 respectively. The similarity between the established cluster and B is determined as 0.350 in complete linkage method.

This method is also invariant against monotonic transformation of the dissimilarities or similarities, but it leads to dilatation and many clusters with small within-cluster dissimilarities are formed (Bacher 2002). Consequently, complete linkage method has tendency to find relatively compact, hyper-spherical clusters composed of highly similar cases (Aldenderfer and Blashfield 1984). Like single linkage, the cluster structure is also not taken into account in this method (Everitt et al. 2001).

3.7.1.1.3. Average method

This algorithm proposed by Sokal and Michener (1958) evaluates the similarity of two clusters by considering all members of the clusters and is an antidote to the property of single and complete linkages based on extreme values. There are different methods for taking the average of the members of the clusters, but the most commonly used method of average linkage is the arithmetic similarities among the cases. Timm (2002) formulated this method as;

$$d_{rs} = \sum\sum(d_{rs})/(n_r * n_s) \qquad 3.6$$

where R and S show two clusters, and r ∈ R, s ∈ S, also n_r and n_s represent the number of objects in each cluster. Hence, the dissimilarities are replaced by an average of n_r and n_s dissimilarities between all pairs of elements r ∈ R, s ∈ S.

This algorithm tends to join clusters with small variances, but it is relatively robust (Everitt et al. 2001).

3.7.1.1.4. Ward's method (minimum variance)

Ward's linkage method forms the new clusters by trying the minimum increase of sum of squared Euclidean distance. According to the algorithm, firstly the means for all variables are computed for each cluster, and then, for each object, the squared Euclidean distance to cluster means is calculated for summing these distances for all of the cases. At each step, by checking the increase in overall sum of the squared within-cluster distances for all possible clusters merging, two clusters are determined according to the smallest increase (Norusis 2004). The following formula shows the process of Ward's method.

$$SSe = x_i^2 - \frac{1}{n(Sx_i)^2} \qquad 3.7$$

where x_i is the score of the ith object. SSe equals to zero at the beginning of the clustering process, because each object is in cluster of its own. And SSe increases with merging of the cluster with each other. At the final stage it becomes equal to the total sum of squares of squares. In this method only the squared Euclidean distance is used, so the outliers affect the outputs of this method. In addition, this means that interval scaled variables or variables that can be treated as interval scaled are necessary and the distances are weighted implicitly (Bacher 2002). In addition, this method assumes that points can be represented in Euclidean space for geometrical interpretation (Everitt et al. 2001), and it tends to find clusters of relatively equal sizes and shapes as hyperspheres (Aldenderfer and Blashfield 1984). In addition, a set of k clusters produced by Ward's method may or may not give the minimum possible SSe over all possible set of k clusters formed from n objects. However, the results of Ward's method are usually very good approximations of the optimal one (Gan et al. 2007).

3.7.1.1.5. Centroid method

This method was originally developed by Sokal and Michener (1958), and by King (1967). In the centroid method, the distance between two clusters is defined as the distance between cluster centroids. Suppose cluster R contains n_R elements and cluster S contains n_S elements. Then, the centroids for the two item clusters can be calculated as;

$$\bar{y_r} = \sum \frac{y_r}{n_r} \qquad \text{and} \qquad \bar{y_s} = \sum \frac{y_s}{n_s} \qquad 3.8$$

Namely, the cluster centroids are the mean values of the objects on the variables in the cluster variety. In this method, the centroid of the group is computed for every individual added to the group. In other words, the cluster centroid is changed according to the individuals or groups of individuals added to the cluster (Kiran et al. 2007). Due to the reversals, confusing results can be provided at the end of the analysis, in other words the distance where clusters are combined can actually decrease from one step to the next leads that clusters merged at later stages become more dissimilar than those merged at early stages. Another problem of this method is that if the sizes of the two groups which were combined are very different, the centroid of the new group will be very close to that of the larger groups and may remain within that group; the characteristic properties of the smaller groups are then virtually lost. In this method, it is assumed that the objects can be represented in Euclidean distance (Everitt et al. 2001), so only squared Euclidean distance can be used for

calculation of dissimilarities (Bacher 2002). The advantage of this method is that it is less affected by outliers than the other methods.

3.7.1.1.6. Median method

This method is developed by Gower (1967) in order to eliminate a shortcoming of centroid method which is losing of the identity of smaller group when merged with larger group, in other words the size of the groups that form the new group have the equal effect in determination of the centroid of a new group (Gan et al. 2007). But, in this case, the objects of smaller group will carry much larger weight than those of larger group in the dissimilarity (Kaufman and Rousseeuw 1990). In addition, it is assumed that the objects can be represented in Euclidean distance (Everitt et al. 2001), so the other measures such as correlation coefficients are not suitable for this method (Gan et al. 2007).

In the median method, the distances between newly formed groups and other groups are computed as;

$$D\left(C_k, C_i \cup C_j\right) = \frac{1}{2} D\left(C_k, C_i\right) + \frac{1}{2} D\left(C_k, C_j\right) - \frac{1}{4} D\left(C_i, C_j\right) \qquad 3.9$$

3.7.1.1.7. Comparing agglomerative clustering methods

Six agglomerative methods are described in this study, all of which have weaknesses and strengths over each other. Also, for the same data set, the results from these algorithms can be varied and there is no single “best” clustering procedure (Kaufman and Rousseeuw 1990). According to the structure of the clusters, the suitable method can be varied, for instance, the single linkage method is the most suitable method for long-shaped clusters, and compact clusters can be determined more accurately by using complete linkage. Accuracy of determined clusters of the data set depends on a good knowledge of the properties of these methods, because of the risk of imposition of their own structure on the data (Kaufman and Rousseeuw 1990).

There are a number of ways to compare the various hierarchical algorithm methods. One is to analyze the transformation of the relationships between the points in the multivariate space in these methods. According to this, the methods are divided into three types, namely space contracting, space dilating and space conserving methods. Firstly, space contracting methods; the relationships are affected by reducing the space between any groups in data and

the new encountered data points are consolidated rather than to use these data to start new points. Space dilating linkage forms are the opposite, so smaller and more distinct clusters are formed in these method. Complete linkage and Ward's method can be said to be space dilating. Finally, space conserving methods, such as average linkage, try to conserve the nature of the original space (Aldenderfer and Blashfield 1984).

Some authors like Williams (1971) thought that space contracting methods are not preferable, especially in applied science, whereas some authors like Jardine and Sibson (1968) thought that this method should be applied because of their desirable mathematical properties. Kaufman and Rousseeuw (1990) mentioned about three conditions for comparing the methods. These are the monotonity of similarity between the merging clusters, ambiguity of the similarities between clusters and statistical consistency of the dissimilarities between clusters. According to these conditions, the group average method is advised. Group average method is also suggested by some of the authors, like Milligan and Isaac (1980) and El-hamdouchi and Willett (1989), whereas, Bacher (2002) advised the Ward's method among these methods. Everitt et al. (2001) mentioned that the success of these methods depends on the a priori conception of the data.

3.7.1.2. Divisive methods

The divisive method proceeds the opposite way of agglomerative hierarchical method. While agglomerative clustering starts with many groups and combines them to form one group, divisive analysis starts with one group and repeatedly divides groups to form many groups, so the agglomerative hierarchical methods are started with details, and the unfortunate decision in the first steps of the algorithm affects the results. On the other hand, oppositely the divisive methods deal with the main hunks in the first steps, so this algorithm is less affected from the unfortunate decisions (Kaufman and Rousseeuw 1990). Everitt et al. (2001) mentioned that outset of this algorithm provides advantages for the users who interested in the main structure of their data set, consisting of few large clusters. However, the use of the divisive algorithms in social sciences is limited (Ketchen and Shook 1996).

The reason of this can be the computational problems faced during the algorithm. In the first step of an agglomerative method, there are $\binom{n}{2} = \frac{(n(n-2))}{2}$ possible ways to merge two clusters. But in the first step of a divisive method, we are faced with 2^{n-1}-1 possibilities to split up the data set into two clusters. At each stage, this formula is valid, so the number of

possibilities increases in each stage, and in practice it is not feasible to try all possible splits.

Another problem with divisive methods is monotonicity. In a divisive method, one cluster is divided at a time. The selection of the next cluster to be divided is a problem, and this depends on the definition of a level. Such a level must be meaningful and monotone, which means that level of the subcluster is smaller than the level of its parent cluster (Gan et al. 2007).

There are two types of divisive algorithms: monothetic and polythetic. The term monothetic refers to the use of a single variable on which to base the split at a given stage; polythetic methods use all the variables at each stage. Monothetic divisive methods are used with the data set which contains binary variables. Monothetic analysis produces a hierarchy of clusters in which at each step a group is split in two based on the value of one of the binary variables. The choice of the variable in monothetic methods on which a split is made depends on optimizing a criterion reflecting either cluster homogeneity or association with other variables. Due to the usage of all variables simultaneously and proximity matrix in the algorithm, polythetic methods are more akin to the agglomerative methods. Polythetic divisive methods, in essence, are the logical opposite or 'mirror image' of agglomerative methods. Although, the methods are different at start point and end point, the number of groups identified should be the same regardless of which one is used (Ketchen and Shook 1996).

3.7.1.2.1. Monothetic divisive method

Monothetic analysis produces a hierarchy of clusters in which at each step a group is split in two based on the value of one of the binary variables. In monothetic methods, the variable on which a split is made is chosen according to a criterion reflecting either cluster homogeneity or association with other variables, so the number of splits that have to be made is minimized. An example of the homogeneity criterion which is developed by Lance and Williams (1968) is the information content, C, defined by p variables and n objects (Everitt et al. 2001).

$$C = pn * \log(n) - \sum_{k=1}^{p}[f_k \log f_k - (n - f_k)\log(\mathrm{n} - f_k)] \qquad 3.10$$

where f_k is the number of individuals having the kth attribute. Clusters are split at each stage according to possession of the attribute which leads to the greatest reduction in C. The most important part of the algorithm is determination of the variable which will separate a subset.

Kaufman and Rousseeuw (1990) mentioned that the variable should be selected according to the largeness of sum of the similarities to all other variables.

In the algorithm of the monothetic method, it is assumed that no missing values existed in the data, so all missing values in the binary data matrix are required to be replaced by estimated values. Suppose that some values of x_{if} are missing. Then any other variable g for which all values are available is determined, in other words for estimating the missing values of a variable, at least one variable should have no missing values, after that the contingency table is constructed.

Table 3-8: Contingency table

f/g	1	0
1	a_{fg}	b_{fg}
0	c_{fg}	d_{fg}

The association between f and g is then defined as;

$$A_{fg} = |a_{fg} * d_{fg} - b_{fg} * c_{fg}| \qquad 3.11$$

The variable t for which $A_{ft} = \max_g A_{fg}$ is the most correlated with variable f. The missing values of f are then estimated by means of variable t in the following way:

put $x_{if} = x_{it}$ when $a_{ft} * d_{ft} - b_{ft} * c_{ft} > 0$

put $x_{if} = 1 - x_{it}$ when $a_{ft} * d_{ft} - b_{ft} * c_{ft} < 0$

When all missing values have been replaced, the actual splitting can begin. In each step, each available cluster is divided according to one variable, namely one cluster with all objects having value 1 for that variable, and another cluster with all objects having value 0 for that variable. The clusters are divided in this way until each cluster consists of objects having identical values for each variable.

The monothetic divisive methods lead to easy classification of new members. Also, in these methods, the data which contains missing values is not eliminated. Another advantage of the monothetic divisive method is the easiness of the determination of variables which produce the split at any stage of the process. However, a general problem with these methods is that the possession of a particular attribute, which is either rare or rarely found in combination with others, may take an individual down the wrong path (Everitt et al. 2001).

3.7.1.2.2. Polythetic divisive methods

Since all variables are used simultaneously, and a proximity matrix is established for the process, polythetic methods are more similar to the agglomerative methods. The procedure developed by Macnaughton-Smith et al. (1964) avoids considering all possible splits for increasing the speed of the computation. It proceeds by finding the object that is the farthest away from the others within a group, and using that as the seed for a splinter group, in other words the data is divided into two groups, namely splinter group and remainder. For each object in the remainder, its average dissimilarity with the other objects in the remainder is calculated and its average dissimilarity with the objects in the splinter group is subtracted. According to the results, the process can be stopped or the object with the largest positive value is shifted to the splinter group and the process is repeated (Seber 2004).

3.7.2. Non-hierarchical cluster procedures

A distance or similarity matrix between all pairs of cases should be established in hierarchical clustering. According to the number of the objects and variables in the data, this matrix becomes enormous (Norusis 2004). However, in non-hierarchical cluster analysis, all possible distances are not required to be calculated. Briefly, most partitioning methods work in the following fashion:

1. Begin with an initial partition of the data set into some specified number of clusters; compute the centroids of these clusters.
2. Allocate each data point to the cluster that has the nearest centroid.
3. Compute the new centroids of the clusters; clusters are not updated until there has been a complete pass through the data.
4. Alternate steps 2 and 3 until no data points change clusters (Aldenderfer and Blashfield 1984)

Determining the initial partitioning is an important issue in non-hierarchical cluster analysis. Different methods are proposed for determining the initial partitioning. Two alternatives are available for determining the initial partitioning: the cases can be randomly assigned to k clusters, or the starting values for the k cluster centers can be specified by the user. In addition to these, another possibility is using the cluster centers obtained from the hierarchical cluster analysis for initial partitioning. Milligan (1980) demonstrated that the k-means pass, using an initial starting partition derived from Ward's method, provided superior

recovery of known data structure when compared to the performance of other iterative and hierarchical clustering methods. Hair (1998) mentioned that the major problem of this process is selecting the initial cluster seeds. Pena et al. (1999) criticized this process, because this process does not contain only initialization methods, but also contains a clustering algorithm, and when used with non-hierarchical cluster methods result in hybrid clustering algorithm. Bacher (2002) determined that the worst results are provided from randomly generated starting values among all studied methods. Pena et al. (1999) compared four initialization methods; namely random, Forgy, MacQueen, and Kaufman, and determined that the random and the Kaufman initialization method outperform among these methods according to the effectiveness and robustness for k-means algorithm.

The non-hierarchical cluster methods can be divided into two methods; K-means and Partitioning around medoids.

3.7.2.1. K-means cluster analysis

One of the most well-known partitioning methods is K-means. K-means method was developed in the 70's by Forgy (Bacher 2002). K-means cluster analysis is started with initial partition of the data set into some specified number of clusters and the centroid of this cluster is calculated. After initial partition, for each object, its similarity to these k clusters is computed, and it is assigned to the cluster corresponding to its most similar cluster. The centroid of clusters is recalculated only after an entire pass through the data has been completed. This forms the initial k-way clustering. After the initial k-way clustering, the objects are reassigned to the clusters according to the distance between centroids of the redefined clusters and their centroids. After all objects are assigned to the clusters, the centroids of the clusters are recalculated. And this process is repeated until reaching the convergence of the cluster centers or minimal increase in squared error. One other important distinction is that k-means passes can also be either exclusive or inclusive. Exclusive methods remove the case under consideration from the parent cluster when a centroid is computed, whereas inclusive methods include them.

K-means procedure attempts to optimize the trW (square-error) criterion. This criterion can be formulated as;

$$TrW = \sum_{i=1}^{K} \sum_{j=1}^{K_i} \left\| w_{ij} - \overline{w}_{ij} \right\| \qquad 3.12$$

where K is the number of the clusters, K_i the number of objects of the cluster I, w_{ij} is the jth object of the ith cluster and $\overline{w}_i$ is the centroid of the ith cluster.

The most important advantage of K-means is handling distinctly large data sets, because of working upon directly raw data. Finally, because of more than one pass through the data, this algorithm can compensate for poor initial partition of the data; in addition it can work very well for compact and hyper-spherical clusters. Whereas K-means algorithm has also some drawbacks; firstly it suffers from initial partitioning due to the iterative procedure of the K-means algorithm, so the convergence centroids vary with different initial points. In addition, the number of clusters should be known before starting the analysis, in real life, this is a unusual occasion, in addition no universal and efficient method is developed for identifying the number of the clusters. Another drawback of this method is the convergence of the process; the algorithm will converge to a local minimum point after a finite number of iterations (Selim and Ismail 1984). Also, K-means is sensitive to outliners and noise. Even if an object is quite far away from the cluster centroid, it is still forced into a cluster and, thus, distorts the cluster shape. Finally, the numerical variables can be used in the analysis due to the definition of "means" (Xu and Wunsch 2005).

3.7.2.2. Partitioning around mediods

The function partitioning around medoids (PAM) developed by Kaufman and Rousseeuw (1990) is based on the searching medoids defined as the representative objects among the objects of the data set (Kaufman and Rousseeuw 1987). These medoids are computed according to the total dissimilarity of all objects to their nearest medoid. The objective function can be formulated as;

$$OF = \sum_{i=1}^{n} d(i, m_{v_i}) \qquad 3.13$$

The algorithm of PAM proceeds in two steps:

1. **Build-step:** An initial clustering is obtained by the successive selection of representative objects until k objects have been found. The first object is selected according to the sum of the dissimilarities to other objects, in other words this object is the most centrally located in the set of the objects. Subsequently, at each steps another object which decreases the objective function as much as possible is selected. For finding this object, the following steps are performed (Kaufman and Rousseeuw 1990):

a. Consider an object i which has not yet been selected.

b. Consider a non selected object j and calculate its dissimilarity Dj with the first object chosen and calculate its dissimilarity d(j, i) with object i. Calculate the difference between Dj and d(j, i). If the difference is positive, then object j will contribute in the selection of object i. Then calculate,

$$C_{ji} = \max\,(D_j - d(j,i), 0) \qquad 3.14$$

c. Calculate the total gain obtained if object i is selected.

d. Choose the object i that maximizes.

2. **Swap-step:** If the objective function can be reduced by interchanging (swapping) a selected object with an unselected object, then the swap is carried out. This is valid for all pairs of objects, pairs are shown by i for selected objects and h for non-selected objects. The following calculations are carried out for calculating the effect of a swap between i and h on the value of clustering (Kaufman and Rousseeuw 1990).

   a. First, consider an object j that has not been selected. Then calculate its contribution C_{jih} to the swap:

      i. If j is closer from one of the other representative objects than from both i and h then the contribution of object j to the swap is $C_{jih} = 0$.

      ii. Consider these two situations if j is not further from i than from any other selected representative object.

         1. If j is closer to h than to any other representative object, d(j, h) < Ej where Ej is the dissimilarity between j and the second most similar representative object, then the contribution of object j to the swap is;

$$C_{ijh} = d(j,h) - d(j,i) \qquad 3.15$$

         2. If j is at least as distant from h than from the second closest representative object, d(j,h) ≥ Ej, then the contribution of object to the swap is;

$$C_{ijh} = E_j - d(j,i) \qquad 3.16$$

      iii. If j is more distant from object i than from at least one of the other representative objects, but closer to h than to any representative object, then the contribution to the swap is;

$$C_{ijh} = d(j,h) - d(j,i) \quad 3.17$$

b. Second, add the contributions C_{jih} to calculate the total result of the swap:

$$T_{ih} = \sum_j C_{ijh} \quad 3.18$$

c. Select the pair (i, h) according to $\min_{i,h} T_{ih}$.

According to the value of minimum T_{ih}, it is decided whether to carry out a swap, if the value is determined negative, then the swap is iterated, otherwise it is decided that the objective cannot be decreased by carrying out a swap, so the algorithm stops (Kaufman and Rousseeuw 1990).

The dissimilarity matrix should be used in calculation of the PAM procedure, but PAM function can also convert data matrix into dissimilarity matrix before starting the procedure. Any specified distance metric can be used for calculating the dissimilarities; so this makes PAM a flexible algorithm. This flexibility for the definition of similarity is not valid for many clustering algorithms. K means could be performed with respect to any metric, but allows only Euclidean and Manhattan distance in current implementations. In addition, PAM has the advantage of identifying clusters by the medoids. Medoids are robust representations of the cluster centers that are less sensitive to outliers than other cluster profiles, such as the cluster means of K-means. When many elements do not belong well to any cluster, this robustness becomes important (Van der Laan et al. 2003). Also, the order in which the objects are examined does not affect the output of the algorithm. Furthermore, they are invariant with respect to translations and orthogonal transformations of data points (Ng and Han 1994). Moreover PAM does not need initial guesses for the cluster centers, contrary to K-means (Insightful 2007).

PAM is performed for small data sets very well; however the performance of PAM for medium and high sized data set is not satisfactory (Ng and Han 1994). In addition, PAM does not succeed in finding the relatively small clusters in the presence of one or more larger clusters (Van der Laan et al. 2003).

Finally PAM provides a graphical output, the silhouette plot, and a corresponding quality index allowing selecting the number of clusters.

It can be said that the non-hierarchical methods have three advantages over hierarchical agglomerative, firstly unlike hierarchical agglomerative methods, iterative methods work directly upon the raw data. They therefore offer opportunity of handling distinctly larger data sets than hierarchical methods. Secondly, non-hierarchical methods are less impacted from

the outliers because it allows data points to switch cluster membership. Moreover, iterative methods make more than one pass through the data and can compensate for poor initial partition of data. These methods also produce single-rank clusters that are not nested and therefore are not part of hierarchy. Most iterative methods don't permit overlapping clusters.

Aldenderfer and Blashfield (1984) pointed out a limitation for that method, the most straightforward way to discover the optimal partition of a data set by means of an iterative method is to form all possible partitions of that data set, but this leads to many unique partitions, for example this approach requires the examination of 217.945.728.000 unique partitions for 15 cases and 3 clusters.

This method's main difficulty is that it requires the number of the clusters to be known at the beginning of the analysis, but in many fields, cluster analysis is used for exploratory purposes.

3.7.3. Two-step cluster analysis

SPSS Two-step cluster method was developed by Chiu et al. (2001) for analysis of large data sets. This method can handle both continuous and categorical variables by extending the model-based distance measure used by Banfield and Raftery (1993) to situations with both continuous and categorical variables, also it utilizes a two-step clustering approach similar to BIRCH (Zhang et al. 1996), also the number of the clusters are determined automatically and the two-step algorithm can analyze large amount of data very rapidly by constructing a cluster features (CF) tree that summarizes the records.

The procedure of the two-step cluster analysis consists of two steps, namely pre-cluster and clustering.

3.7.3.1. Pre-cluster

The goal of pre-clustering is to reduce the size of the distance matrix between all possible pairs of cases. Pre-clusters are just clusters of the original cases that are used in place of the raw data in the hierarchical clustering (Norusis 2004).The algorithm starts by scanning the data records one by one and every new data is examined for merging the existing cluster or forming a new cluster.

The procedure is implemented by constructing a modified CF tree. The CF tree consists of levels of nodes, and each node contains a number of entries. A leaf entry represents a final sub-cluster. New records are placed into the correct leaf nodes according to the non-leaf nodes and their entries. CF characterizes the each entry placed in this CF according to the entry's number of records, mean and variance of each continuous variable and counts of each category of each categorical variable. This process is started with an initial threshold value, and then the appropriate leaf is identified for each object by choosing the closest child node according to a close distance metric through descending the CF tree. Upon reaching a leaf node, the object is absorbed into the leaf entry and CF of that leaf entry is updated according to the threshold distance of the closest leaf entry. If the object is not within the threshold distance, it starts its own leaf entry. The leaf node is split into two for satisfying more space for new objects, if there is no space in leaf node to create a new leaf entry. If the CF-tree grows beyond allowed maximum size, the CF-tree is rebuilt based on the existing CF-tree by increasing the threshold distance criterion. This process continues until a complete data pass is finished. The detailed information about the algorithm can be found in BIRCH by Zhang et al. (1996). After completion of pre-cluster process, the entry's CF collectively represents all records falling in same entry. The size of the distance matrix is no longer dependent on the number of cases but on the number of pre-clusters. When a new record is added to the data set, the new CF can be computed from the old CF without knowing the individual records in the entry. These properties make it possible to maintain only entry's CFs, rather than the sets of individual records (SPSS 2001).

3.7.3.2. Clustering

The inputs of this step are taken as sub-clusters resulting from the first step, and these are grouped into the clusters automatically. Since the number of the sub-clusters are much less than the original data, the traditional cluster methods can be used effectively. The agglomerative hierarchical cluster method is used by the SPSS due to the work capability with the auto-cluster procedure (SPSS 2001). But, in contrast to agglomerative hierarchical techniques, an underlying statistical model is used (Bacher et al. 2004). According to the number of the sub-clusters, the algorithm can be performed effectively. If the number of the sub-clusters is chosen small, this can cause declining in the accuracy of the clusters, whereas the number of the sub-clusters is determined as high, then the second step clustering is

slowed down. The maximum number of sub-cluster should be chosen by considering these issues.

As mentioned before, the number of the clusters is determined automatically by the algorithm according to the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). BIC proposed by Fraley and Raftery (1998) is a criterion statistic for determination of the appropriate number of cluster in their clustering method based on EM (expectation-maximization) algorithm. The clustering criterion is computed for each potential number of clusters. Smaller values of the BIC or AIC indicate better models, and in this situation, the best cluster solution has the smallest BIC and AIC. However, there are clustering problems in which the BIC and AIC will continue to decrease as the number of clusters increases, but the improvement in the cluster solution is not worth the increased complexity of the cluster model. In such situations, the changes in BIC and changes in the distance measure are evaluated to determine the best cluster solution. A good solution will have a reasonably large "Ratio of BIC Changes" and a large "Ratio of Distance Measures" (Chiu et al. 2001). The following table shows a process of a two-step cluster, as it can be seen in the table three clusters solution has the smallest BIC, and when the ratio of distance measure is examined, it can be seen that the highest change occurs at the three clusters solution, so that for this analysis the best solution is three.

Table 3-9: Two-step cluster analysis with BIC

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	Ratio of BIC Changes (b)	Ratio of Distance Measures (c)
1	1792,917		
2	1621,684	1,000	1,639
3	1575,975	,267	2,368
4	1643,692	-,395	1,085
5	1717,917	-,433	1,059

BIC and AIC for j clusters are defined as;

$$BIC(J) = -2\sum_{j=1}^{j} \xi_j + m_j \log(N) \qquad 3.19$$

$$AIC(J) = -2\sum_{j=1}^{j} \xi_j + 2m_j \qquad 3.20$$

where $m_j = J\left\{2K^A + \sum_{k=1}^{K^B}(L_k - 1)\right\}$.

K^A = total number of continuous variables used in the procedure

K^B = total number of categorical variables used in the procedure

N= number of data records in total

L_k = Number of categories for the kth categorical variable

In this analysis, the log-likelihood distance measure and Euclidean distance can be used. The log- likelihood can handle both continuous and categorical variables. It is probability based distance. The distance between two clusters is related to the decrease in log- likelihood as they are combined into one cluster. In calculating log-likelihood, normal distributions for continuous variables and multinomial distributions for categorical variables are assumed. It is also assumed that the variables are independent of each other. The log- likelihood distance between two clusters i and s is defined as;

$$d(i,s) = \xi_i + \xi_s - \xi_{(i,s)} \qquad 3.21$$

where

$$\xi_i = -n_i\left(\sum_{j=1}^{p}\frac{1}{2}\log(\hat{\sigma}_{ij}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^{q}\sum_{l=1}^{m_j}\hat{\pi}_{ijl}\log(\hat{\pi}_{ijl})\right) \qquad 3.22$$

$$\xi_s = -n_s\left(\sum_{j=1}^{p}\frac{1}{2}\log(\hat{\sigma}_{sj}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^{q}\sum_{l=1}^{m_j}\hat{\pi}_{sjl}\log(\hat{\pi}_{sjl})\right) \qquad 3.23$$

$$\xi_{\langle i,s\rangle} = -n_{\langle i,s\rangle}\left(\sum_{j=1}^{p}\frac{1}{2}\log(\hat{\sigma}_{\langle i,s\rangle j}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^{q}\sum_{l=1}^{m_j}\hat{\pi}_{\langle i,s\rangle jl}\log(\hat{\pi}_{\langle i,s\rangle jl})\right) \qquad 3.24$$

Euclidean distance measure can only be applied if all variables are continuous. The distance between the two clusters is here defined by the Euclidean distance between the two cluster centers. A cluster center is defined as the vector of cluster means of each variable.

The most important disadvantage of this algorithm is getting the best result depends on some conditions, like all variables should be independent, and continuous variables should have a normal distribution, and categorical variables should have a multinomial distribution. Although these are seen rarely in real life, nevertheless the algorithm is thought to behave reasonably well when the conditions are not satisfied (Norusis 2004).

The another important point of the two-step cluster analysis is that this process is order sensible for that reason the application of that analysis should be performed many times for different data orders.

The performance of the algorithm is affected from the outliers, for solving that problem an outlier's cluster that contains all cases that do not fit well with the rest should be created. SPSS will do this automatically for you if you select "Outlier Treatment" in the "Options" dialog box. Bacher et al. (2004) evaluated the performance of the two-step cluster algorithms for different data sets, and determined that the performance of this algorithm is satisfactory when the all variables are continuous, but it cannot show the same performance for the mixed type variables.

3.8. Cluster validation techniques

In the literature, a lot of algorithms have been proposed for different applications with different shapes and sizes of data. Even Blashfield and Aldenderfer (1978) pointed out that over 100 different clustering methods have been proposed, with different methods that generate quiet different solutions to the same data set. Also, clustering a data set is an unsupervised process, in other words no predefined classes and no examples that would show what kind of desirable relations should be valid among the data are not available (Halkidi et al. 2001) at the beginning of the analysis. In addition, many clustering algorithms require the number of clusters being given beforehand, and determination of number of the clusters for the other cluster algorithms become difficult when the data set contains many data points. For overcoming these problems, it is necessary to apply validity methods.

The validity indices are discussed for hierarchical and non-hierarchical cluster methods separately.

3.8.1. Evaluation of hierarchical cluster methods

The clusters obtained by using hierarchical cluster methods should be evaluated according to hierarchical structure, rand index, silhouette index, and agglomerative coefficient in order to determine the most suitable solution for the data set.

3.8.1.1. Testing hierarchical structure

The obtained clusters should be stable against minor modifications performed on the data set. The stability of the obtained clusters can be validated by the so-called stability methods,

the analysis of reduced data set and modifications of the data set of variables used in the clustering (Gan et al. 2007).

In the stability method, the original clusters are compared with clusters of modified versions of the data. The data can be modified by adding random errors to the data set or perturbing the elements of the dissimilarity matrix. The differences between the original clusters and clusters of modified versions are compared each other for obtaining the accuracy of the clustering results (Gordon 1999).

In the analysis of reduced data set, the modification of the data set is performed by deleting one or more data points from original data set for determining the influence of the deleted data points. Lanyon (1985) proposed a method as an example of the analysis of a reducing the data set, the difference between the normal procedure and method proposed is making N runs of clustering analysis instead of a single run and in each step performing the analysis by reducing a different data point from the data set. The aim of performing N times cluster analysis is synthesizing these results for obtaining an optimal result.

Lastly, the modification of variables can be used for determining the stability of the obtaining clusters. These methods are motivated by phylogenetic concerns (Gan et al. 2007).

3.8.1.2. Rand index

The stability between the clusters obtained by using different cluster methods or between the original data set and data sets obtained by different modification methods can be examined by using Rand index developed by Rand (1971). This index is based on how pairs of data points are clustered. Three cases are available in evaluation of stability of two data points. Firstly, these two points can be assigned to the same clusters by two cluster methods, or they are assigned to different clusters by these methods, this represents a similarity between the clustering solutions. Lastly, one of the cluster methods can assign these two data points into the same cluster, whereas in other clustering, they are assigned in different clusters, therefore the stability between the outputs of these two clustering methods is deteriorated.

These cases are illustrated as following;

$$\gamma_{ij}=\begin{cases} 1 \text{ if there exist k and k' such that both } X_i \text{ and } X_j \text{ are in both } Y_k \text{ and } Y'_{k'} \\ 1 \text{ if there exist k and k' such that } X_i \text{ is in both } Y_k \text{ and } Y'_{k'} \text{ while } X_j \text{ is in neither } Y_k \text{ or } Y'_{k'} \\ 0 \text{ otherwise} \end{cases}$$

where $Y_k = \{Y_1, Y_2, \dots, Y_{k_1}\}$ and $Y' = \{Y'_1, Y'_1, \dots, Y'_{k_2}\}$

Rand index is formulated as follows;

$$c(Y, Y') = \left(\sum_{i<j}^{N} \gamma_{ij}\right) / \binom{N}{2} \qquad 3.25$$

Rand index is a measure of similarity, ranging from c=0 to c=1. c=0 means that the two clustering solutions have no similarities, and when the c is equal to 1, it means a perfect fit. Values over 0.7 are considered as sufficient for the stability of the cluster analysis (Bacher 2002).

3.8.1.3. Silhouette index

This quality index is determined by silhouette width s(i), a composite index reflecting the compactness and separation of the clusters, and can be applied to different distance metrics, and can be calculated by using following formula;

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \qquad 3.26$$

where a (i) shows the average dissimilarity of i to all other objects of A, C shows any cluster different from A and d (i,C) shows the average of i to all objects of C, and for all clusters not equal to A, d (i,C) is calculated and the smallest one is determined as b(i).

Clearly s (i) always lies between −1 and 1. The value s (i) may be interpreted as follows:

- s(i) =1 object i is well classified (in A);
- s(i) =0 object i lies intermediate between two clusters (A and B);
- s(i) =−1 object i is badly classified (closer to B than to A).

The quality index is the overall average silhouette width of the silhouette plot, defined as the average of the s (i) over all objects i in the data set. The number of cluster is determined by

performing the PAM analysis several times, each time with a different k, and comparing the resulting silhouette plots. The value of k is selected according to the highest average silhouette width, over all k, which is called the silhouette coefficient (Struyf et al. 1997) and they tabulated the subjective interpretation of the silhouette coefficient as given in Table 3-10.

Table 3-10: Interpretation of the silhouette coefficient for partitioning methods.

Silhouette Coefficient	**Proposed Interpretation**
0.71-1.00	A strong structure has been found
0.51-0.70	A reasonable structure has been found
0.26-0.50	The structure is weak and could be artificial,
< 0.25	No substantial structure has been found

3.8.1.4. Agglomerative coefficient

Agglomerative coefficient is a quality index for an agglomerative clustering of the data and it measures the amount of clustering structure found. For each object i denote by d (i) its dissimilarity to the first cluster it is merged with, divided by the dissimilarity of the merger in the last step of the algorithm. The agglomerative coefficient is then defined as the average of all 1 − d (i). Another property of the agglomerative coefficient is that it tends to increase with the number of objects, so it should not be used to compare data sets of very different sizes (Struyf et al. 1996).

3.8.2. Evaluation of non-hierarchical clustering

Different validation indexes are developed for evaluation the clusters obtained from non-hierarchical clustering methods, such as Krzanowski-Lai, Hartigan, Davies-Bouldin, Calinski-Harabasz, and Dunn's index. These indexes can be used for evaluating the suitability of the clusters for the data set and determining the number of the clusters of the data set. All of these indexes evaluate the aspects of the clusters; therefore all of the indexes should be evaluated in order to reach a decision regarding the data set used in the analysis.

3.8.2.1. Krzanowski-Lai index

This index developed by Krzanowski and Lai (1988) by following the general approach of Marriott (1971) is formulated as follows;

$$diff_k = (k-1)^{2/p} trW_{k-1} - n^{2/p} trW_k \qquad 3.27$$

where k shows the number of clusters, W is the pooled within-group covariance matrix for any given partition of the sample, tr (W) is the sum of squares. According to this formula, a stopping criterion shown as following formula is developed.

$$C_k = |iff_k| / |diff_{k+1}| \qquad 3.28$$

The optimum value of k is the value maximizes C_k. Krzanowski and Lai (1988) mentioned that if the particular data set is inappropriate for the sum of squares objective function, then this criterion does not yield optimum results. Secondly, the frequent occurrence of multiple local maxima of C_k should be checked for unusual features, so they advised that the results should never be accepted uncritically but should always be examined for their meaningfulness. Finally, it is reasonable to accept several local maxima of C_k for large data sets.

3.8.2.2. Hartigan index

This index developed by Hartigan (1975) is formulated as follows;

$$H(k) = (n-k-1)\left(\frac{trW_k}{trW_{k+1}} - 1\right) \qquad 3.29$$

where k shows the number of clusters, n shows the number of the data points. W is the pooled within-group covariance matrix for any given partition of the sample, tr (W_k) is the total square of error for *k* cluster partitioning. Since tr (W) is monotonically non-increased with increasing *k*, the ratio is a relative measure of the reduction of square error when number of clusters increases from *k* to *k*+1. The multiplier correction term of *(n-k-1)* is a penalty factor for large number of cluster partitioning. The optimal *k* number is the one that maximizes the *H(k).* As mentioned above, this index also benefits from the sum of squares, so the caution about the particular data set which is inappropriate for the sum of squares objective function is also valid for this index.

3.8.2.3. Davies-Bouldin index

Davies-Bouldin (DB) index developed by Davies and Bouldin (1979) is not affected from the number of the clusters and the clustering algorithms, nonetheless it is based on a measure of dispersion of a cluster and a dissimilarity measure between two clusters (Halkidi et al. 2001). Firstly, the dispersion measure and the cluster similarity measure are required to be defined for determining the DB index.

The dispersion measure S of a cluster C is defined to satisfy the following conditions (Gan et al. 2007).

1. $S \geq 0$;
2. $S=0$ if and only if $x=y \ \forall \ x, y \in C$.

S is the average distance of all patterns in a cluster C to its cluster center, so different measures can be used for determining the S. S was formulated as following by Bezdek and Pal (1998);

$$S_i = \left(\frac{1}{|C_i|} \sum_{x \in C_i} \|x - v_i\|_2^q\right)^{1/q} \qquad 3.30$$

where $|C_i|$ is the number of the data points in Cluster C_i , v_i is the center (or representative data point) of cluster C_i , x are the data points placed in Cluster C_i.

The cluster similarity measure between two clusters is defined based on the dispersion measures of these clusters and satisfies the following conditions;

1. $R_{ij} \geq 0$;
2. $R_{ij} = R_{ji}$;
3. $R_{ij} = 0$ if and only if $S_i = S_j$;
4. If $S_j = S_k$ and $D_{ij} < D_{ik}$, then $R_{ij} > R_{ik}$;
5. If $S_j > S_k$ and $D_{ij} = D_{ik}$, then $R_{ij} > R_{ik}$.

D_{ij} shows the distance between the two clusters C_i and C_j, which can be defined as the distance between the centroids of the clusters (Gan et al. 2007). Different similarity measures mentioned before can be used for determining the distance between the clusters. According to the above conditions, a simple choice for R_{ij} is proposed by Davies and Bouldin (1979);

$$R_{ij} = \frac{S_i + S_j}{D_{ij}} \quad 3.31$$

Then DB is defined as;

$$V_{DB} = \frac{1}{k} \sum_{i=1}^{k} R_i \quad 3.32$$

where k is the number of the clusters and R_i is defined as;

$$R_i = \max_{j \neq i} R_{ij} \quad 3.33$$

According to equation 3.31, small DB shows the compact clusters and their centers are far away from each other. Consequently, the cluster solution which shows smallest DB should be searched for determining the most suitable solution. In addition, this index has small computational complexity, and work well for the clusters having spherical shape in feature space (Günter and Bunker 2003).

3.8.2.4. Calinski-Harabasz index

The Calinski-Harabasz index developed by Calinski and Harabasz (1974) is based on the traces of the between-clusters and within-cluster scatter matrices. The Calinski-Harabasz index is formulated as follows;

$$V_{ch} = \frac{(n-k)Tr(B)}{(k-1)Tr(W)} \quad 3.34$$

Where n is the number of the data points in the data set and k is the number of the clusters. Also, Tr(B) and Tr(W) are the traces of matrices B and W, and B and W are the between-clusters and within-cluster scatter matrices (Gan et al. 2007).

3.8.2.5. Dunn's index

Dunn index developed by Dunn (1974) attempts to identify sets of clusters that are compact and well separated as DB index (Bezdek and Pal 1998). The Dunn index is defined as for a specific number of clusters;

$$V_D = \min_{1 \leq i \leq k} \left\{ \min_{i+1 \leq j \leq k} \left(\frac{D(C_i, C_j)}{\max_{1 \leq l \leq k} diam(C_l)} \right) \right\} \quad 3.35$$

Where k is the number of the clusters, $D(C_i , C_j)$ the distance between clusters C_i and C_j is defined as following;

$$D(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y) \quad 3.36$$

diam (C_l) the diameter of cluster C_l may be considered as a measure of dispersion of the clusters, is defined as follows;

$$diam\ (C_l) = \max_{x,y \in C_l} d(x, y) \quad 3.37$$

If the distance between the clusters is large and the diameter of the clusters is small, then the data set contains compact and well-separated clusters. Thus, based on the equation 3.34, it can be concluded that large values of the index indicate the presence of compact and well-separated clusters. Unfortunately, this index is very sensitive to the presence of noise due to the increase in diam (C_l) due to the noise, in addition it requires considerable amount of computational time (Halkidi et al. 2001).

3.9. Final notes about traditional cluster analysis

As mentioned earlier, traditional cluster analysis is the most widely used technique for strategic group analysis. However, there are alternative methods which can be used to check the reliability of results obtained from traditional cluster analysis. Additional benefits can be gained by using alternative methods such as self organizing maps and fuzzy clustering method together with traditional clustering method. In the forthcoming chapters, alternative clustering methods will be introduced and their benefits as well as bottlenecks will be discussed in comparison with traditional clustering method.

CHAPTER 4

SELF ORGANIZING MAPS

One of the model based clustering techniques that can be used for strategic grouping is self organizing maps (SOM). Forthcoming sections summarize its basic properties, available software packages for SOM analysis as well as application areas, algorithms and validation measures used in SOM. Differences of SOM from other clustering methods will be discussed after presenting the fundamentals of fuzzy clustering concepts in chapter 5.

4.1. Introduction

The SOM developed by Teuvo Kohonen between 1979 and 1982 is an unsupervised neural network computational mapping technique that project high dimensional input data items to a low, often one or two dimensional grid nonlinearly. The SOM is generally regarded as a form of neural network, but it can also be used as a cluster tool, since the nonlinear statistical relationships between high dimensional data is transformed into simple geometric relationships of their image points on a low-dimensional display, and by that way, the data points which show similar properties are collected at the same place of the output provided by the SOM algorithm (Kohonen 2001). Also, SOM compresses information embedded in the data set while preserving the most important topological and metric relationships of the primary data elements, so the topology of the data can be obtained at the end of the process (Kohonen 2001). The SOM also can also be used for prediction purposes, but for that; a hidden layer should be inserted into Kohonen layer output as a predictive or categorical output layer. After the Kohonen layer is stabilized, the back-propagation learning rule can be used to train the hidden layer and output layer for prediction. But 0 or 1 are the only output values of this process, so values from the Kohonen layer should be interpolated against coarseness of the input to the predictive network layer (NeuralWare 2000).

The network of SOM usually consists of two layers of neurons: an input layer and output layer. The neurons on input layer are fully connected to the neurons on output layer, whereas

the neurons on each layer do not connect to the neurons in their layer regardless of their relative position. The following figure shows the layers and connections of these neurons in SOM.

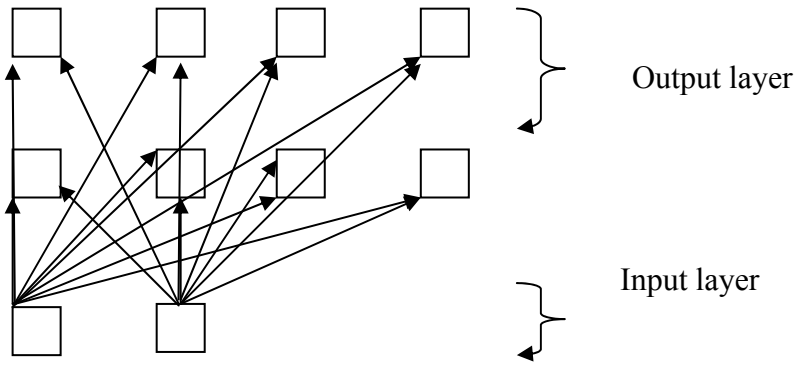


Figure 4-1: Layers and connections of these layers of SOM

Usually, neurons on the output layer are arranged in either a rectangular or hexagonal lattice (Lansiluoto 2004), and following figure shows the illustration of these two lattices. From the figure, a neuron in an output layer arranged by rectangular lattice is surrounded by four neighbors, whereas a neuron in a hexagonal grid is surrounded by six neighbors, except for the ones at the edges of the grid (Lansiluoto 2004). Consequently, the more neighbors are modified for the hexagonal lattice if the same neighborhood radius is used in both of the analysis. Besides, hexagonal lattice is effective for visual display because it does not favor horizontal and vertical directions as much as rectangular array (Kohonen 2001). When the edges of the grid is considered, the neuron placed at the edge of a rectangular grid has only two neighbors for the rectangular grid, on the other hand, this could be two or three neighbors in the hexagonal grid depending on the location of the edge for the edge neurons.

Figure 4-2: Rectangular and Hexagonal lattice

Each node in the grid is a prototype in the sense that it possesses a set of weights which are values for the set of inputs. The position of each node in the grid *vis-à-vis* its neighboring nodes are of major importance, particularly during the training process. The actual spacing of the Kohonen nodes has no meaning: what is important is their grouping together. This is because each node is regarded as a 'prototype', a set of cognate values of the attributes of the input data (Curry et al. 2001).

Data points are presented randomly to the network and at each stage the nearest Kohonen node is identified. This is referred to as the 'winning' node and the learning mechanism itself as 'competitive learning' or 'winner takes all learning'. The weights of the winning node are adjusted so as to move towards the current data point, in which case the training process involves allowing the weights of each node to reflect or describe the data. The topological structure of the data is preserved because not only is the winning node updated, but also its neighboring nodes are. The adjustment of the neighboring node is instrumental in preserving the order of the input space.

4.2. Applications of SOM

The SOM algorithm has attracted a great deal of interest among researchers and practitioners in a wide variety of fields. Many studies can be found in the literature in many fields ranging from engineering sciences and natural sciences to medicine, humanities, economics, physics, mathematics and chemistry. Kaski et. al. (1998) has conducted a bibliography concerning the applications of SOM. They found 3343 scientific papers from the years 1981-1997, after that publication, Oja et. al. (2003) extended that study for the years 1998-2001, they added 2092 publication to this bibliography. These studies are clustered according to their topics into thirteen categories; image and video, pattern recognition, mathematical techniques, artificial intelligence, software, engineering in biology and medicine, information theory and coding, speech, control, signal processing, circuits, information science and documentation and business and administration. Finally, the application of the SOM in the business and administration field is determined as limited.

4.3. The available software applications for SOM

The available software tools can be divided into two groups, non-commercial and commercial software packages. The examples of the non-commercial software applications

can be SOM_PAK, SOM toolbox for Matlab and Databionics ESOM tools. These tools can be used for academic purposes with no fee. Also, these software packages can be used with the special software applications according to needs for pre- and post processing of data to meet the specific needs. For example, Carlson and Tulkki (1998) used the SOM_PAK with GIS and other software applications to analyze real estate markets (Deboeck and Kohonen 1998).

Several commercial software packages are available in the market for implementing SOM. The example of these software packages are Viscovery® SOMine 4.0, NeuroSolutions v5.0, Trajan 6.0 Neural network Simulator, NeuroShell2 /NeuroWindows and Professional II+ from NeuralWorks. The major benefit of these commercial software is their simplicity in usage due to the user-friendly interfaces, in other words, they allow the users without knowledge regarding SOM to implement SOM. In addition, the commercial available software packages are designed for running on commonly used environments and platforms and have better imbedded capabilities for pre and post-processing of the data.

The non-commercial software applications are decided to be used in this research as these software applications are appropriate for the purposes of this study. Firstly the visualization capability is enough to analyze the strategic groups. Finally and most importantly, these software applications are free for the academic studies and the commercial software applications do not outperform them.

There are three alternatives of non-commercial software packages; namely SOM_PAK, SOM toolbox for Matlab 5 and Databionics ESOM tools. These software applications are compared with each other according to their advantages and disadvantages in order to decide which software will be used in this study.

4.3.1. SOM_PAK

SOM_PAK is released for the first time in 1990 by the Laboratory of Computer and Information Science of Helsinki University of Technology. This software can run under the UNIX and MS-DOS platforms, therefore this software provides advantages in application of SOM by using old-configured computers, however this is not valid for this study.

The only algorithm included in this package is the standard incremental-learning SOM. The source code modules of the software have been given as C programming text files, which allow the users to make their own modifications. The SOM_PAK contains only rather simple

graphics software applications of its own. The visualization options of SOM_PAK are the component planes with trajectories and U-matrix, but any more demanding visualization can be implemented with general-purpose graphics or visualization software applications.

The SOM array topology can be selected as either rectangular or hexagonal, and the map size and dimensionality of the vectors are unlimited, being only restricted by the computer's own resources. And, initialization of the models can be made at random or along the row principal axes of the data distribution.

The neighborhood function can have either the 'bubble' or "Gaussian" form and many different combinations can be defined for the learning sequences. For monitoring the quality of the outputs of SOM, the evaluation of the average quantization error during the process and Sammon weights are available as default.

4.3.2 SOM toolbox for Matlab

The first version of the SOM toolbox was released by the researchers of the Laboratory of Computer and Information Science of Helsinki University of Technology in 1996.

In order to utilize this toolbox, at least Matlab 5.0 must be available, in other words, the toolbox is running with the Matlab 5.0.

In addition to standard incremental SOM algorithm, the Batch map algorithm is available in this toolbox. The map can be rectangular or hexagonal, and its size is unlimited like SOM_PAK. Another resemblance to the SOM_PAK is the initialization methods, they both initialize their SOM algorithm by randomly or linear initialization method. Also the missing values are taken into account in similar ways to the SOM_PAK. But, the toolbox can apply more neighborhood functions and learning sequences when it is compared with the SOM_PAK. In addition to "bubble" and "Gaussian" neighborhood functions, "cutgauss" and "epanechikov" functions are available. Also, the toolbox uses "power" and "inv" learning rate functions besides "linear' learning rate function.

For preprocessing the data set, the toolbox provides two options for eliminating the effect of range differences of the variables; these are normalization and histogram equalizations of the numeric variables.

For visualization, the component planes with trajectories, the U-matrix, and histogram of hits of the samples in the map units can be drawn. For preliminary data analysis and monitoring, the Sammon projection and the average quantization error can be plotted.

When these two software applications are compared, it can be seen that the usage of SOM toolbox is easier than the SOM_PAK, because of the graphical user interface of the toolbox, but the SOM_PAK can be used only from the shell command line. But SOM toolbox required Matlab 5.0 for running, so the necessary memory size and processor speed for carrying out the analysis is higher. Also Vesanto et al. (1999) compared the speed of these two software applications in the same computers for the sequential algorithm, and found out that SOM_PAK is approximately three times faster than the SOM toolbox, and this can provide advantage when the data size is huge by decreasing the required computational time.

The Matlab also provides advantages to the user of the SOM toolbox; firstly Matlab's strong graphics capabilities can be utilized for producing attractive figures. In addition to these, the algorithm of SOM toolbox can be easily modified according to the purpose of the study, because the code of Matlab for modification requires less expertise and time, when it is compared with the C code of the SOM_PAK.

4.3.3. Databionics ESOM

The ESOM Tools are being developed by the Databionics Research Group at the University of Marburg, Germany. The ESOM Tools are written in Java for maximum portability and published under the terms of the General Public License (GPL). For running this software, Java runtime must be installed on the computer, but this software can be downloaded from the internet. However, this software uses a different kind of SOM which is emergent SOM (ESOM). In ESOM, the size of the map should be set very large, in other words, the number of the nodes should be selected as at least a few thousands.

Five training algorithms are available in this software, which are online (standard incremental SOM), batch map, slow batch, k-batch training and hybrid-batch training algorithms. More training alternatives are available in this software when compared with the other two software packages. The calculations of ESOM are performed by using the rectangular grids, whereas the hexagonal grid is only utilized in the training part of the Databionics ESOM Tools but not in the visualizations part. In this software, like other software applications, the algorithm initializes with random or linear initialization method. However, in this software three alternatives for random initialization are available.

Five neighborhood functions, bubble, Gaussian, mexican hat, cone and epanechikov, can be used in this software. In addition, four cooling strategies for learning rate; namely linear,

exponential, linear with lead in/out, and no cooling, are available. In addition, eight visualization alternatives are available in this software. As it can be seen, the alternatives of these software applications are more than the other methods. It can provide advantages in analysis of different data structures. The quality measurement is not available in the software; therefore the quality of the solutions cannot be evaluated.

According to the available features of these software applications, the Databionics ESOM tool is determined as the best alternative among these software applications. However, the algorithm used in this software is not suitable for the data sets in which the number of available data points is limited. Unfortunately, the data set used in this study is consisted of 84 data points. Due to the simplicity in usage of SOM toolbox, it is decided to be used in this study.

4.4. Data requirements for SOM

The input data is denoted by an m x n matrix X shown in Figure 4-3, each row contains a sample point and each column contains all values of a variable, in other words m shows the number of the data points and n shows the number of the variables. Each node k in the SOM grid is characterized by a 1 x n vector W (k) of weights.

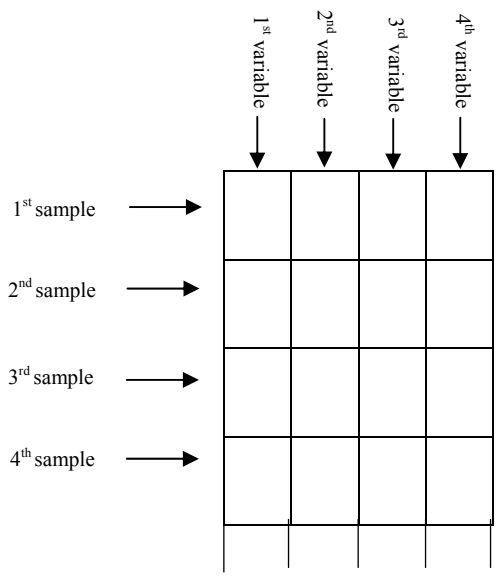


Figure 4-3: Data presentation

The variables might be the properties of an object, or a set of measurements measured at a specific time. The important thing is that every sample has the same set of variables. Thus, each column of the table holds all values for one variable. Some of the values may be missing, but the majority should be there.

Numeric and symbolic data can be handled by the toolbox, but only the former is utilized in the SOM algorithm. The numeric representation must be meaningful, for example 1,2 and 4 are corresponding to objects A,B and C should really mean that B is between A and C, and that the distance between B and A is smaller than the distance between B and C. Identification numbers, error codes, etc. rarely have such meaning, and they should be handled as symbolic data. In the toolbox, symbolic data can be inserted into string labels associated with each data sample. If the symbolic data are required in training the SOM, they can be converted into numerical variables using mapping or 1-of-n coding.

4.4.1. Initialization phase

There are two options, random and linear initialization, for initialization phase of the algorithm used in SOM toolbox.

4.4.1.1. Random initialization

SOM algorithms can be initialized using arbitrary values, then these arbitrary values are ordered by SOM algorithm in a few hundred initial steps, however this leads to increase in duration of the SOM process (Kohonen 2001).

4.4.1.2. Linear initialization

The ordered nodes at the initial state can provide some benefits throughout SOM process. The linear initialization method can be applied for providing the ordered nodes. In this method, firstly two eigenvectors of the autocorrelation matrix of X corresponding to the largest eigenvalues is determined, after that the eigenvectors extend across a to-dimensional linear subspace. A rectangular array, whose centroid coincides with that of the mean of the $X(t)$, is defined along this subspace, and this array can be used for the rectangular and hexagonal lattice. The initial weight values of the nodes are identified with the array points. And the uniform lattice spacing in the SOM is obtained with providing the relative numbers

of cells in the horizontal and vertical directions of the lattice proportional to the two largest eigenvalues (Kohonen 2001).

By ordering the nodes and approximating their density points to p(x), the learning rate at the beginning can be significantly smaller than unity, and the neighborhood function at the beginning is close to its final value, therefore the algorithm converges in a shorter time (Kohonen 2001).

4.4.2. Training phase

Two algorithms are available for training phase of the SOM in the SOM toolbox; namely sequential training and batch training algorithm phase.

4.4.2.1. Sequential training algorithm

In the sequential training algorithm, one sample vector x from the input data set is chosen randomly and the distances between it and all the weight vectors of the SOM are calculated by using some distance measure, such as Euclidean distance. At the end of that process, the winning node or Best-Matching Unit (BMU) is determined by comparing the distances between the sample vector and nodes. This process is repeated for each data point iteratively.

In the Toolbox, the process of distance computation is slightly more complicated because of two factors:

- Missing values: In the Toolbox, these are represented by the value not a number (NaN) in the vector or data matrix. Missing components are handled by simply excluding them from the distance calculation.
- Mask: each variable has an associated weighting factor, defined in the .mask field of the map and training struts. This is primarily used in the binary form for excluding certain variables from the BMU-finding process. However, the mask can get any values, so it can be used for weighting variables according to their importance.

After finding the BMU, the weight vectors of the SOM are updated so that the BMU is moved closer to the input vector in the input space. The weights of BMU's topological neighbors are also updated in order to preserve the typological structure of the map.

The training is usually performed in two phases. In the first phase, relatively large initial learning rate α_0 and neighborhood radius σ_0 are used. In the second phase learning rate and neighborhood radius are decreased with time.

4.4.2.2. Finding the winning node

Initially, the wining node should be identified. For that purpose, the distances between an input data and all the weight vectors of the SOM are computed. If the input data has M values and is denoted by $X=(x_1,x_2,\ldots,x_M)$, then each Kohonen layer node i will also have M weights and can be denoted by $W_i= (w_{i1} , w_{i2}, \ldots, w_{iM})$. The Euclidean distance between the data point and this node is calculated by using the following formula.

$$D_i = \sqrt{(x_1 - w_{i1})^2 + (x_2 - w_{i2})^2 + \ldots + (x_m - w_{im})^2} \qquad 4.1$$

After determination of Euclidean distance between the data points and all nodes, the minimum value of D_i is identified. During recall, the Kohonen node with the minimum D_i is determined as winner and has output of 1.0, while the other Kohonen nodes have an output of 0.0. Thus winning node is the closest to the input value and after assignment it represents the input value.

4.4.2.3. Conscience mechanism

One problem of the SOM is that most of the input data can be represented by a few nodes due to the effect of the initial randomization of the Kohonen node weights (Kiang and Kumar 2001). For example, assume the variables of the inputs were determined as latitude and longitude value, and randomly the nodes were selected as the center point of each continent, if all the input values were from the same continent, then the results of the analysis might be problematic. In brief, little information about the clusters of the input data would be found from the analysis in these cases due to the representation of all inputs made by one node. In addition, the activation of a neural unit is a function of the distance between the input pattern and neuron weights in competitive learning. During the learning phase, the activated neuron is the one learns the most from the input and it modifies its weights so it can be activated again if the same or a similar pattern is presented to the network. This means that the neurons that win the competition have a higher probability of wining again (Rizzo and Chella 2006).

The problem of too much winning by few nodes was solved by introducing the mechanism of a conscience. The conscience mechanism was firstly developed by Desieno (1988) by improving competitive learning algorithm. The goal of the conscience mechanism is to bring all processing elements available into the solution quickly, and to bias the competition process so that each processing element can win the competition with close to the 1/N probability desired for an optimal vector quantization (Desieno 1988).

The conscience mechanism adjusts the distance to encourage nodes that are not winning with an average frequency and to discourage node that are winning at an above average frequency. The conscience mechanism helps develop a uniform data representation in the SOM layer.

The other benefits which the conscience mechanism provides:

1. The nodes naturally represent approximately equal information about the input data.
2. Where the input space has sparse data, the representative Kohonen nodes compact the space.
3. Where the input space has high density, the representative Kohonen nodes spread out to allow finer discrimination (NeuralWare 2000).

During learning, it is necessary to keep track of the winning frequency of each node. This frequency is updated after the winning node is selected. The following formulas are used to calculate F_i.

For the winning node:

$$F_{inew} = F_{iold} + \beta(1.0 - F_{iold}) \qquad 4.2$$

For all other Kohonen nodes:

$$F_{inew} = F_{iold} + \beta(0.0 - F_{iold}) \qquad 4.3$$

β, a constant, is required for satisfying F_i does not reflect the random fluctuations in the data and it should be higher than 0 and very smaller 1 (Desieno 1988). For illustrating the process, following example is given. In this example, the number of nodes at initialization and β is given as 10 and 0.1 separately, then for each node i, F_i becomes 0.1 at the initial step of the process. Then after the first pass, the winning node would boost its new frequency of winning to F_i=0.19, while the other 9 nodes would have their frequency of winning drop to F_i= 0.09 and this process is repeated for every data point.

After conscience mechanism, the distance of the input vectors to the nodes were modified according to their winning frequencies, the following formula is used for determining the modified distance of the nodes.

$$D^* = D - \gamma * (N * F_i - 1) \qquad 4.4$$

where F_i is the frequency with which the node i has historically won, and N is the number of Kohonen nodes in the SOM layer and γ is a constant between zero and unity.

At initialization, $F_i = {}^{1}/_{N}$ and so initially D^* is equal to D. By considering the formula, it can be determined that when $F_i < {}^{1}/_{N}$, so D^* is decreased, hence chance of winning of corresponding node is increased, similarly when $F_i > {}^{1}/_{N}$, D^* is increased, this causes the decrease in the winning probability of the corresponding node. By comparing the D* of the all nodes, the winning node is determined. After determining the winning node, the node and its neighbor nodes are adjusted according to the input data.

4.4.2.4. Adjusting the weights of neighboring nodes

The formula for adjusting the weights in the winner's neighborhood is straightforward:

$$m_i(t+1) = m_i(t) + \propto (t) * h_{ci}(t)(x(t) - m_i(t)) \qquad 4.5$$

where t donates time. The x(t) is an input vector randomly drawn from the input data set at time t, $h_{ci}(t)$ the neighborhood kernel around the winner unit c and α(t) is the learning rate at time t, so that the neighborhood kernel and learning rate is changed with time. These two parameters control the learning of the algorithm.

The neighborhood width parameter determines to what extent the surrounding neurons are modified according to the input. The neighborhood kernel function is started with large value, equal to the size of the network itself suggested by Kohonen (1989) for minimizing the effect of the initial random weights assigned to the nodes (Kiang and Kumar 2001) and it shrinks to unity, in other words, at final step the winner nodes do not affect the weights of the its neighborhoods. Thus, the learning process is gradually shifting from an initial rough learning phase with a big influence area and fast-changing prototype vectors to a fine-tuning phase with small neighborhood radius and prototype vectors that adapt slowly to the samples (Kohonen 2001). This process is illustrated in Figure 4-4.

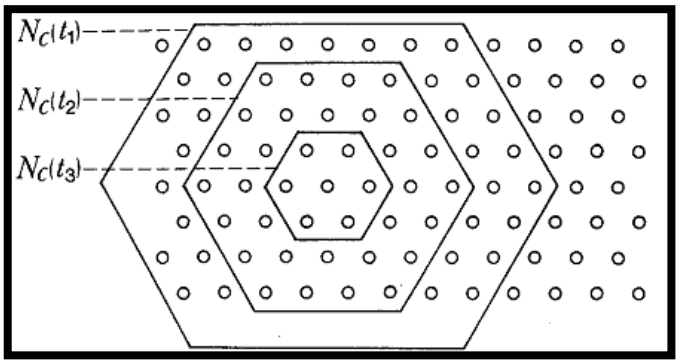


Figure 4-4: Examples of neighborhood kernel functions

The selection of the size of the neighborhood is required special attention, because if the neighborhood is too small to start with, the map will not be ordered globally. Instead various kinds of mosaic-like subdivisions of the map are seen, between which the ordering direction changes discontinuously (Kohonen 2001). There are four alternatives for neighborhood functions in the SOM toolbox. These are bubble ($h_{ci}(t) = 1(\sigma_t - d_{ci})$), Gaussian ($h_{ci}(t) = e^{-d_{ci}^2/2\sigma_t^2}$), cutgauss ($h_{ci}(t) = e^{-d_{ci}^2/2\sigma_t^2} 1(\sigma_t - d_{ci})$) and epanechikov ($h_{ci}(t) = max\,\{0.1 - (\sigma_t - d_{ci})^2\}$), where σ_t is the neighborhood radius at time t, d_{ci} is the distance between the map units c and i on the map grid and **1**(x) is the step function: **1**(x)=0 if x<0 and **1**(x)=1 if x$\geq$ 0 (Vesanto et. al., 2000).The following figure illustrates the shape these functions.

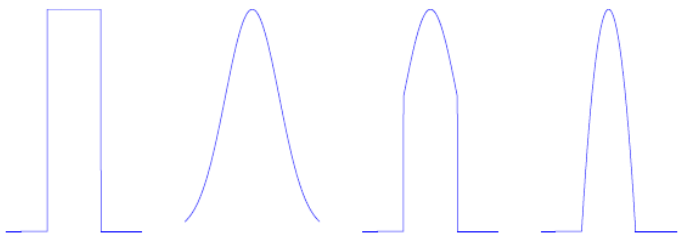


Figure 4-5: Different neighborhood functions. From left, bubble, Gaussian, cutgauss and epanechikov

The learning rate influences how size the neighbor nodes of the winning node are adjusted after each training step. And the purpose of the usage of learning rate is provided the convergence of the algorithm. The learning rate is started with reasonably high values, and it is suggested as close to unity by Kohonen (2001) and decreased monotonically. Three

learning rate function is available in the SOM toolbox, these are linear $\alpha(t) = \alpha_0(1 - T/t)$, power $\alpha(t) = \alpha_0(0.005/\alpha_0)^{t/T}$ and inverse $\alpha(t) = \alpha_0/(1 + 100t/T)$, where T is the training length and α_0 is the initial learning rate (Vesanto et al. 2000). Kohonen (2001) mentioned that an accurate time function is not important. However, especially for the large maps, the selection of an optimal learning rate may be crucial.

4.4.3. Batch training algorithm

The other algorithm can be used in the toolbox is batch training algorithm. The training steps of the algorithm are defined as follows,

1. For the initial reference vectors, take the first K training samples, where K is the number of reference vectors.
2. For each map unit i, collect a list of copies of all those training samples x whose nearest reference vector belongs to unit i.
3. Take for each new reference vector the mean over the union of the lists in N_i.
4. Repeat from 2 a few times (Kohonen 2001).

This algorithm is an iterative process. Firstly, the whole data set is presented to the map before any adjustments are made. In each training step, the data set is partitioned according to the Voronoi regions of the map weight vectors. After this, the new weight vectors are calculated as:

$$m_i(t+1) = \frac{\sum_{j=1}^{n} h_{ic}(t) x_j}{\sum_{j=1}^{n} h_{ic}(t)} \qquad 4.6$$

In this formula $c = argmin_k\{\|x_j - m_k\|\}$ is the index of the BMU of the data sample x_j. The new weight vector is a weighted average of the data samples, where the weight of each data sample is neighborhood function value h_{ic} (t) at its BMU c. As in the sequential training algorithm, missing values are simply ignored in calculating the weighted average (Vesanto and Alhoniemi 2000).

Alternatively, one can first calculate the sum of the vectors in each Voronoi set:

$s_i(t) = \sum_{j=1}^{nv_i} x_j$ where nv_i is the number of samples in the Voronoi set of unit i. then, the new values of the weight vectors can be calculated as:

$$m_i(t+1) = \frac{\sum_{j=1}^{m} h_{ij}(t) s_j(t)}{\sum_{j=1}^{m} nv_j h_{ij}(t)} \quad 4.7$$

where m is the number of the map units.

The learning rate parameter is not used in this algorithm unlike the sequential algorithm; therefore it has no convergence problems and yields more stable asymptotic values for the m_i than the sequential algorithm.

The size of the neighborhood set N_i can be similar to the sequential algorithms, and the radius of the neighborhood function is decreased throughout the process of the algorithm. At the last couple of iterations, N_i may contain the element i only, and the last steps of the algorithm are then equivalent with the K-means algorithm.

When the algorithms used in training of SOM are compared with each other, it can be concluded that the batch algorithm trains the data faster than the sequential algorithm due to the number of the updates performed during the process, specifically weights of neurons are updated only one time in batch training, whereas the sequential algorithm updates the weights of the neurons at each iterate step. On the other hand, the presentation of the clusters trained with batch training algorithm in the data space on maps is poor when compared to sequential (Kohonen 2001).

4.5. Determination of quality of the SOM

There are many parameters which adjust the outputs of SOM which can vary highly depending on the initial setting of parameters such as the number of output nodes, the decrease speed of neighborhood, and learning rate (Wang and Wang 2002), so the quality of the outputs should be examined in order to determine the parameters which give the most suitable solution for the data set. Two of the properties which most of the measures try to evaluate are topology preservation and vector quantization (Pölzlbauer 2004). For evaluating these properties, two measures are used in the SOM toolbox; namely quantization error and topographic error.

4.5.1. Quantization error

The quantization error is used for measuring the quality of vector quantization, thus this measure completely disregards map topology and alignment. The quantization error is

computed by determining the average distance of the sample vectors to the cluster centroids by which they are represented. The average quantization error can be calculated by using the following formula.

$$E = \frac{1}{N}\sum_{i=1}^{N} min\{\|x_i - m_c\|\} \qquad 4.8$$

where N is the total number of samples, x_i is the input data vector, and m_c is the best matching reference vector.

The number of weights and the neighborhood size determine the performance of the SOM, and therefore determines the quantization error. Intuitively, the quantization error decreases with the increasing number of weights. Sun (2000) suggests using small neighborhood size in order to reduce the quantization error and obtain good approximation of input distribution.

For any data set, the quantization error can be reduced by simply increasing the number of map nodes, because than the data samples are distributed more sparsely on the map (Pölzlbauer 2004). But with increasing the number of the map nodes, determining the clusters become more difficult because the number of the nodes are bigger than the number of the data points, so the data points are separated from each other. In other words, the projection quality decreases with increasing the vector quantization. Therefore, the determination of the best solution according to the quantization error should be made by considering the same number of the nodes.

4.5.2. Topographic error

Topographic error is used for calculating the quality of topology preservation. Topographic error is calculated as follows: for a data sample, the best matching unit is identified, and then the second best matching unit is determined for the same data points, if these two nodes are not adjacent on the map lattice, then this is considered as an error. This procedure is repeated for all of the data points. Finally, the total error is calculated and normalized to a range from 0 to 1 where 0 means perfect topology preservation. The data set is required for calculation of the topographic error, in other words the similarity distance is not enough for calculation of topographic error. Lastly, missing values is not an issue in calculation of topographic errors; in other words the topographic error can be calculated with missing values.

Topographic error is unreliable for small maps, results in very low values for maps that do not contain overly high numbers of nodes (Pölzlbauer 2004).

4.6. Parameters of SOM

There are a number of training parameters that need to be decided before training: map size (ie. the number of map units) and shape, neighborhood kernel function, neighborhood radius, learning rate and length of training. The user can freely specify all of these, however to minimize user effort, the toolbox also provides default values for them.

- The number of map units is $m=5*n^{0.54321}$, where n is the number of data samples.
- Map shape is rectangular sheet with hexagonal lattice. The ratio of side lengths corresponds to the ratio between two greatest eigenvalues of the covariance matrix of the data. The actual side lengths are then set so that their product us as close to the desired number of map units as possible.
- Neighborhood function is Gaussian $h_{ci}(t) = e^{-d_{ci}^2/2\sigma_t^2}$.
- The linear initialization along two greatest eigenvectors is tried, but if the eigenvectors can't be calculated, random initialization is used instead.
- Radius, as well as learning rate, is a monotonically decreasing function of time. The starting radius depends on map size, but the final radius is one. Learning rate starts from 0.5 and ends to (almost) zero.
- Training length is measured in epochs: one epoch corresponds to one pass through the data. The number of epochs is directly proportional to the ratio between number of map units and number of data samples *m/n.*

Although default values can be used for most parameters, SOM toolbox is most useful with high degree of interaction, as SOM toolbox is primarily intended as a tool for data understanding, it can never become a completely autonomous tool (Vesanto et al. 2000).

For strategic group analysis, SOM has a potential to provide an effective solution with its high capability of categorization of data as well as its outputs that help visualization of results. Its performance as a clustering tool will be compared with other methods in the next chapter, after discussing the fundamentals of fuzzy clustering.

CHAPTER 5

FUZZY CLUSTERING

The cluster analysis methods are divided into two groups, namely; classical cluster analysis and fuzzy cluster analysis. In classical (or hard) cluster analysis, each datum is assigned to exactly one cluster. Consequently, subsets provided at the end of the hard C-means analysis are non-empty and pairwise disjoint. Data points that are almost equally distant from two or more clusters cannot be presented by these methods adequately, in other words such data points are assigned to one of the clusters fully by a crisp partition method, although they should equally belong to all of them. However, this constraint is not valid for the fuzzy clustering approaches. In fuzzy clustering approach, data points can be assigned to more than one cluster and even with different degrees of membership to the different clusters (De Oliveira and Pedrycz 2007). Thus, this method can outperform hard clustering methods in many applications, especially when clusters are not well separated, the borders of the clusters are not sharp, and clusters overlap (Chuang et al. 1999).

Since the concept of fuzzy sets (Zadeh 1965) was introduced, fuzzy clustering has been widely discussed, studied, and applied in various areas, and different fuzzy clustering methods have been developed. One widely used algorithm which is the fuzzy C-means (FCM) algorithm, was firstly presented by Dunn (1974), Bezdek (1981) further developed the FCM clustering algorithm. Subsequent revisions came from Roubens (1982), Gath and Geva (1989), Gu and Dubuisson (1990), and Xie and Beni (1991), but the most commonly used algorithm is remained as Bezdek's FCM.

In this chapter, fuzzy C-means clustering algorithm will be explained, which will be further used for strategic group analysis of Turkish contractors. At the end of the chapter, a comparison will be made between the clustering methods introduced in this thesis; namely, traditional cluster analysis, self organizing maps and fuzzy clustering method.

5.1. Fuzzy C-means cluster (FCM) analysis

The FCM clustering has been applied to various disciplines such as climatic modeling, geologic modeling, suburban environment modeling, and soil-landscape modeling (Gorsevski et al. 2003). But, the application of the fuzzy clustering is very limited in business and economics.

FCM aims to minimize the fuzzy version of classical within groups' sum of squared error objective function according to the fuzziness exponent by using Picard iterations. The algorithm seeks to partition the data set into c subgroups or clusters on the basis of measured similarities among the vectors, so the number of the cluster is a prerequisite for starting the analysis. The Bezdek's objective function is shown in equation 5.1.

$$\text{Min } J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^m d^2(X_k, V,) \qquad 5.1$$

where c is the number of the clusters, n is the number of vectors (or data points), and m>1 is the fuzziness index. The quantity of d^2 (X_k,V_i) is the distance between observation X_k and the cluster centroid V_i. The choice of the distance measure is changed according to the variables (Hammah and Curran 1998), but the Euclidean distance is most commonly used.

According to the equation, two parameters affect the results of the algorithm, namely number of the clusters and m, degree of the fuzziness. The fuzziness of the memberships is controlled by m. m takes values higher than 1. The closer m is to 1, the crisper the membership values are. As the values of m become progressively higher, the resulting memberships become fuzzier (Hammah and Curran 1998). Pal and Bezdek (1995) advised that m should take the values between 1.5 and 2.5, and the number of clusters should be between 2 and $\sqrt{n}$.

In fuzzy clustering, firstly the initial centroids V_i (i=1,2,…,k) are chosen. After that, the membership matrix is computed by using the following formula;

$$u_{ij} = \frac{[d^2(x_j,V_i)]^{-\frac{1}{q-1}}}{\sum_{l=1}^{k}[d^2(x_j,V_i)]^{-\frac{1}{q-1}}} \quad \text{i=1,2,...,k j=1,2,...,n;} \qquad 5.2$$

Then the new centroids $\hat{V}_i$ (i=1,2,…,k) are computed as;

$$\hat{V}_i = \frac{\sum_{j=1}^{n} u_{ij}^q x_j}{\sum_{j=1}^{n} u_{ij}^q} \qquad 5.3$$

The degree of memberships is updated according to the membership matrix formula, according to the termination criterion, ε, the algorithm is stopped or iterated until it satisfies the criterion.

During the application of the algorithm, two constraints should be satisfied; $\sum_{j=1}^{n} u_{ij} > 0, i\in 1,...,c,$ and $j=1cuij=1, j\in 1,...,n.$

First constraint assures that no cluster is empty. This satisfies a requirement of the classical cluster analysis that no cluster is empty; in addition it avoids the trivial solution of minimization problem (De Oliveira and Pedrycz 2007). However, Krishnapuram and Keller (1993) illustrated a problem due to this constraint. For illustrating problem, the Figure 5-1 shows an example data set with two clusters. The FCM would produce very different membership values in cluster 1 for the points A and B, even though they are equally typical of this class. In addition, point A and point C may have equal membership values in cluster 1, even though point C is more typical than point A. In other words, because of this constraint, the membership of a point in a class depends on the membership of the point in all other classes.

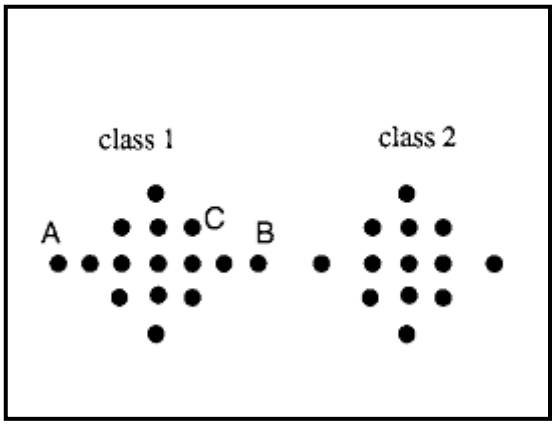


Figure 5-1: Example of a data set with two clusters (Krishnapuram and Keller 1993).

The second constraint is required for providing the sum of the membership degrees for each datum as one, so all data are included into the cluster partition equally. This is related to the requirement in classical clustering that partitions are formed exhaustively. Furthermore, by this condition, a normalization of the membership per datum is satisfied. Thus the membership degrees for a given datum can be assumed as the probabilities of its being a

member of the corresponding cluster (De Oliveira and Pedrycz 2007). However, this constraint also causes problems in determination of the membership degrees of each datum. For illustrating the problems, the data sets shown in the Figure 5-2 (Krishnapuram and Keller 1993). According to the first data set, two intersecting circular shell clusters are determined. In this case, point A is a good member of both clusters, whereas point B is a poor member. But, due to this constraint, both points are taken 0.5 membership degrees for both clusters (Krishnapuram and Keller 1993). The second data set is used for showing another problematic situation. Points A and B are two outlying data points, the distances of point A to the clusters are same, and thus it is assigned a membership degree of 0.5. Also, the same degrees of membership are assigned to the point B, even though this point is further away from both clusters and should be considered less typical. Consequently, point B receives fairly high membership degrees to both clusters (De Oliveira and Pedrycz 2007). In addition to these, the normalization of the memberships can further lead to undesired effects in the presence of noise and outliers. The fixed data point weight may result in high membership of these points to clusters, even though they are a large distance from bulk of data. Their membership values consequently affect the clustering results, since data point weight attracts cluster prototypes.

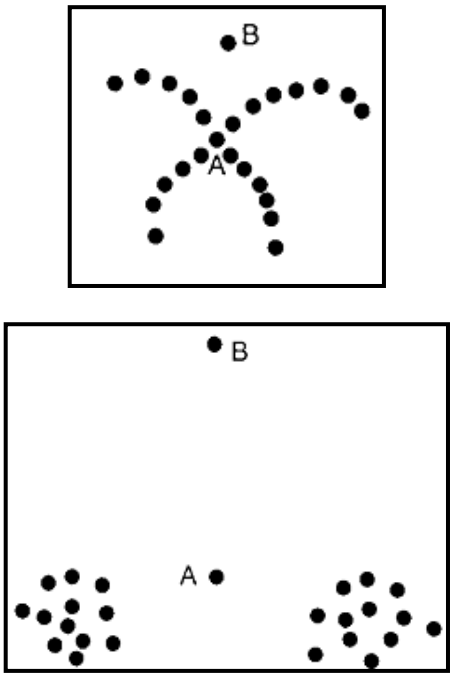


Figure 5-2: Example of two data sets with two clusters (Krishnapuram and Keller 1993)

Gath and Geva (1989) mentioned about three major difficulties during the application of the fuzzy clustering,

*The number of clusters cannot always be defined a priori, and one has to find a cluster validity criterion in order to determine the optimal number of clusters present in the data.

* The character and location of cluster centroids are not necessarily known priori, and initial guesses have to be made.

* There exist large variability in cluster shapes, variations in cluster densities, and variability in the number of data points in each cluster.

The FCM algorithm always needs a large number of memories to keep the data set and also requires a large amount of CPU time, especially for the large data sets. This may make it either unable to converge or lead to its convergence to an erroneous result in some conditions (Chuang et al. 1999).

5.2. Possibilistic C-means

The noise and outliers in the data can cause undesired effects on the output of the analysis due to the constraints of the FCM. Possibilistic c-means developed by Krishnapuram and Keller (1993) can be a remedy against undesireable normalization effects. In this algorithm, the normalization constraint is dropped. This requires a new objective function because dropping the normalization constraint leads to the mathematical problem that the objective function would reach its minimum for u_{ij} =0 for all data points. In other words, the data points are not assigned to the clusters. The modified objective function is formulated as;

$$j_m = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - u_{ij})^m \qquad 5.4$$

where η_i>0. The first term identical to the objective function of the FCM leads to a minimization of weighted distance. The second term suppresses the trivial solution since high memberships make the expression $(1 - u_{ij})^m$ become approximately zero. Thus the desire for assignments of data to clusters is expressed in the objective function J_m. The cluster specific constants η_i are used to balance the contrary objectives expressed in the two terms of J_m (Höppner et al. 1999). This value can be the same for all clusters if all clusters are expected to be similar.

As mentioned before, the Euclidean distance is widely used in application of fuzzy clustering; however this distance is only successful to identify the spherical clusters. Different methods have been proposed to overtake this problem, considering other distances between cluster centers and data points. Gustafson-Kessel (GK) algorithm and Gath-Geva (GG) algorithm are two of the methods for the remedy of this problem.

5.3. Gustafson-Kessel (GK) algorithm

In this algorithm developed by Gustafson and Kessel (1978), the standard FCM algorithm is extended by replacing Euclidean distance by a Mahalanobis distance induced by a positive definite, symmetric matrix in the FCM algorithm, so clusters of different geometrical shapes in one data set can be detected by an automatic adaption for each individual cluster. For a cluster i, the Mahalanobis distance is defined as;

$$D_{ijA}^2 = \left(x_j - \hat{V}_i\right)^T A_i\left(x_j - \hat{V}_i\right),\ 1 \leq \mathrm{i} \leq \mathrm{c},\ 1 \leq \mathrm{k} \leq \mathrm{N} \qquad 5.5$$

According to the Formula 5.5, each cluster is characterized by a symmetric and positive definite matrix A in addition to the cluster centers. This matrix is used as optimization variables in the c-means functional, thus allowing each cluster to adapt distance norm to the local topological structure of the data. The objective function can be minimized by selecting matrices with almost zero entries, in order to avoid this, matrix A must be constrained by a constant volume of clusters. The usual way of accomplishing this is constraining the determinant of A, $\det(A_i) = 1$. Thus; only the cluster shapes are variable now, but not the clusters' size. By changing the constraint value of determinant of each matrix A, the size of the clusters can be changed. However, the choice of the constants requires a priori knowledge about the clusters (Höppner et al. 1999).

Matrix A can be determined by using Lagrange multiplier method. The following expression is used for determining the matrix A;

$$A_i = [p_i \det(F_i)]^{1/n} F_i^{-1} \qquad 5.6$$

where F_i is the fuzzy covariance matrix of the ith cluster defined by:

$$F_i = \frac{\sum_{j=1}^{N} u_{ij}^m (x_j - \hat{V}_i)(x_j - \hat{V}_i)^T}{\sum_{j=1}^{n} u_{ij}^m} \qquad 5.7$$

However, the factor $\frac{1}{\sum_{j=1}^{n} u_{ij}^{m}}$ is not relevant for the result, because the matrices are scaled to the unit determinant (Höppner et al. 1999).

In comparison to the FCM algorithm, the performance of this algorithm for non-spherical clusters is better. However, this algorithm is more sensitive to initialization, because of larger research space (Bezdek 1999). Moreover, the computational cost of this algorithm is much higher due to matrix inversions. Because of these issues, Höppner et al. (1999) advised to initialize the cluster centers of the GK prototypes with the resulting prototypes of a FCM.

5.4. Gath-Geva (GG) algorithm

This algorithm developed by Gath and Geva (1989) is derived from a combination of FCM algorithm and the fuzzy maximum likelihood estimation (FMLE). An exponential distance measure based on FMLE is used instead of Euclidean distance in this algorithm; this distance measure is defined as;

$$D_{ij}(x_k, \hat{V}_i) = \frac{\sqrt{\det(F_i)}}{\propto_i} \exp\left(\frac{1}{2}(x_j - \hat{V}_i)^T F_i^{-1}(x_j - \hat{V}_i)\right) \qquad 5.8$$

where F_i is the fuzzy covariance matrix of the ith cluster and α_i is the priori probability of the ith cluster. According to the Formula 5.8 and Formula 5.9, the difference between the fuzzy covariance matrix of GK and GG algorithms is the weighting exponent, m. Instead of this, one is used in the formula, because the two weighted covariance matrices arise as generalizations of the classical covariance from two different concepts. αi is defined as;

$$\propto_i = \frac{1}{N}\sum_{j=1}^{N} u_{ij} \qquad 5.9$$

In contrast to FCM and GK algorithms, instead of an objective function, the GG algorithm is based on fuzzification of statistical estimators. Also, due to the exponential term in the distance formula, it is faster than the other algorithms. According to Höppner et al. (1999), this algorithm is suitable for clusters showing large variability of cluster shapes, densities and number of data points in each cluster. However, this algorithm is more likely to converge in a local minimum than the other algorithms. Also, the initialization affects the outputs of this algorithm more than other algorithms, in fact for the correct partition by the GG algorithm; the prototypes have to be initialized near the prototypes. For that purpose, GG algorithm should be started with the prototypes obtained from FCM or GK algorithms. Another problem of this algorithm is formation of floating point overflows, because of the

exponential function (Höppner et al. 1999). In other words, the distance which is relatively small according to the Euclidean distance can be determined as very large by the GG distances. Therefore, a modified exponential function that provides constant or only linearly increasing values when the arguments face an over-flow should be applied in this algorithm.

5.5. Cluster validation techniques

For evaluating the performance of the fuzzy clustering analysis methods, different validation techniques, namely fuzziness performance, modified partition entropy, partition index, separation index, Xie and Beni index are developed. All of these validation techniques examine different aspects of the clustering solutions, and none of them is perfect by itself.

5.5.1. Partition index

Partition index developed by Bensaid et al. (1996) accounts for properties of the fuzzy memberships and structure of the data and based on fuzzy compactness and separation. This measure aims to eliminate the some of the pitfalls of FCM algorithm. Partition index is defined as follows (Bensaid et al. 1996);

$$SC = \sum_{i=1}^{c} \frac{\sum_{k=1}^{n} \mu_{ik}^{m} \|x_k - v_i\|_A^2}{n_i \sum_{t=1}^{c} \|v_i - v_t\|_A^2} \qquad 5.10$$

where m is the fuzziness coefficient, $\|.\|_A$ is an inner product norm induced by matrix A, v_i is the centroid of the each cluster, and μ_{ik} is the fuzzy membership of object x_k belonging to class i, n is the number of the data points. The denominator part of the formula shows the separation of a fuzzy cluster (i) as the sum of the distances from its cluster center (v_i) to the centers of the other (c-1) clusters.

According to the 5.14, this index is the ratio of the sum of compactness ad separation of the clusters, in other words it is a sum of individual cluster validity measures normalized through division by the fuzzy cardinality of each cluster. In conclusion, a lower value of partition index indicates a better partition, and it is useful for comparing different partitions having equal number of clusters.

5.5.2. Fuzziness performance index

Fuzziness performance index (FPI) is derived from the partition coefficient proposed by Bezdek et al. (1984) for eliminating dependency of partition coefficient on number of clusters and the fuzziness index. More specifically, partition coefficient is decreased with increasing number of the clusters, so this validity index cannot be used directly, the knees of increase or decrease in the plot of the indices versus the number of clusters should be sought. In addition, this coefficient is sensitive to the fuzziness index, m (Halkidi et al. 2001). Partition coefficient is formulated as follows;

$$V_{PC} = \frac{1}{n} \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^2 \qquad 5.11$$

where n is the number of data points, and c is the number of the clusters. μ_{ik} is the membership of the ith data point to the kth cluster. From this equation, the V_{pc} ranges from $\frac{1}{c}$ to one. The closer the value of V_{pc} to $\frac{1}{c}$, the fuzzier the clustering is, whereas the closer the value of V_{pc} to the unity, the crisper the clustering is. The V_{pc} obtains its lowest value, in case of all membership values get value of $\frac{1}{c}$.

FPI estimation of the degree of fuzziness generated by a specified number of classes is formulated as follows (Roubens 1982);

$$FPI = 1 - \frac{cV_{PC} - 1}{c - 1} \qquad 5.12$$

The good clusters should not be very fuzzy, that is, less fuzziness shows better validity. In other words lower FPI shows a better solution. However, this validity index has some drawbacks. It eliminates some of the drawbacks of the partition coefficient, excluding the lack of direct connection to the geometry of the data (Dave 1996), because the data itself is not utilized in the calculations, whereas the memberships of the data points are used.

5.5.3. Modified partition entropy

Modified Partition Entropy (MPE) is derived from the partition entropy index for eliminating same drawbacks occurred in Partition coefficient. Partition entropy is defined as (Gan et al. 2007);

$$V_{PE} = -\frac{1}{n} \sum_{l=1}^{c} \sum_{i=1}^{n} \mu_{li} \log_a(\mu_{li}) \qquad 5.13$$

where a is the base algorithm, μ_{li} is the membership of a fuzzy c-partition. From this equation, the V_{PE} range in [0, $\log_a$ c]. The closer the value of V_{pc} to zero, the crisper the clustering is. This entropy is used to measure the fuzziness of the cluster partition only, which is similar to the partition coefficient.

This entropy has the same drawbacks as the partition coefficient; therefore the modified partition entropy estimation of the degree of disorganization created by a specified number of clusters is developed and defined as follows (Roubens 1982);

$$MPE = \frac{H}{\log_a c} \qquad 5.14$$

Like FPI, in this index the lowest value should be sought for determining the best solution for the data set.

5.5.4. Separation index

On the contrary of partition index (SC), the separation index uses a minimum-distance separation for partition validity (Bensaid et al. 1996). Separation index is defined as follow;

$$S(c) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^2\|x_j - v_i\|^2}{N \min_{i,k}\|v_k - v_i\|^2} \qquad 5.15$$

5.5.5. Xie and Beni index

The Xie-Beni index developed by Xie and Beni (1991) is also called the compact and separate fuzzy validity criterion. This index aims to identify an overall compact and separate fuzzy c-partition without assumptions as to the number of substructures inherent in the data. Xie and Beni index is defined as follows (Gan et al. 2007);

$$XB = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^2\|x_j - v_i\|^2}{N \min_{i,j}\|x_j - v_i\|} \qquad 5.16$$

where v_i is the centroid of the each cluster, μ_{ij} is the fuzzy membership of object x_j belonging to class i, n is the number of the data points.

From equation 5.15, the numerator shows the compactness within the clusters, whereas the denominator shows the separation between the clusters, so the smaller XB indicates a more compact and separate c-partition, in other words the smallest XB should be sought among the number of the clusters for determining the optimum number of the clusters.

5.6.A comparison of traditional cluster analysis, SOM and FCM

When it is compared with the other clustering algorithms, SOM is the one that has the greatest visualization capability. The outputs of the SOM can describe many properties of the data which can be described by using pages of reports provided by cluster analysis. In addition, the detailed information can be determined by using the SOM's outputs due to the easiness of interpretation of the visualized outputs. On the other hand, both of traditional cluster analysis and fuzzy C-means (FCM) have very limited visualization property. More detailed investigation of reports obtained as an output from these analyses is required for interpretation of findings.

Traditional clustering methods may involve a variety of algorithms but almost invariably build distinct self-contained clusters (Curry et al. 2001), whereas the neurons of the SOM are not mutually exclusive. This means that the final feature map, instead of showing several distinct clusters with differing characteristics, shows neighboring nodes which have many similar characteristics but differ perhaps on one or two, or in degree of intensity of characteristics (Curry et al. 2001), therefore if overlapping exists between the clusters, it can be determined through outputs of SOM. FCM evaluates the probability that the input belongs to each cluster; these probabilities can be used for gaining knowledge about the topography of the data structure. However, the membership of a data point must be summed to one in FCM, so the membership of a point in a class depends on the membership of the point in all other classes, thus indirectly on the total number of classes. Consequently, the fuzzy memberships of FCM cannot always represent the proper degrees of data points belonging to their clusters.

The traditional statistical methods can not be sufficient for analyzing the data containing many data points and large number of variables which describe these data points, however SOM method is considered as an effective method in dealing with high dimensional data. In addition, the FCM algorithm needs a large memory to keep the data set and a large amount of central processing unit (CPU) time to complete the analysis. This also decreases the detectability of the finer structures of the data, whereas due to the competitive learning, computer memory is not used by SOM algorithm intensively. This leads to effectively identification of the small sizes (Chuang et al. 1999).

The traditional cluster analysis methods are designed under strict assumptions of certain statistical distribution functions; however there is no need for making that kind of assumptions in application of SOM. For instance, continuous variables should satisfy normal

distribution assumption and categorical variables should satisfy multinomial distribution assumption in order to perform two-step cluster analysis effectively (Norusis 2004). Furthermore, the number of the clusters should be known at the inception of the K-means clustering method and FCM in order to start the analyses. Furthermore, the result of the FCM is highly affected from the selection of the fuzzy factor; therefore making assumptions about the fuzzy factor is a difficult and important process in FCM analysis. However, the number of the clusters is not a pre-request at the inception stage of SOM, and the correct number of clusters will be directly shown by the result itself. Although there is no need for assumption about the number of the clusters, in the application of SOM, there are also some parameters which affect the outputs of the SOM.

As mentioned before, the sorting ability of the traditional cluster analysis is an important problem for the reliability of the solutions. Whereas, the SOM can be a remedy for that problem, because The U-matrix does not give any results when there are no obvious clustering relations in the original space, thus, unreasonable arbitrary classification can be avoided (Zhang and Li 1993).

The prototype parameters and properties of the clusters of FCM can be greatly affected by the noise in data (Krishnapuram and Keller 1993). This may make it either unable to converge or cause it converge to an erroneous result in some conditions. Also, SOM algorithm is affected from the noise level and initial values (Chuang et al. 1999).

The applications of the methods to strategic group analysis are discussed in the next chapter.

CHAPTER 6

STRATEGIC GROUP ANALYSIS FOR TURKISH CONSTRUCTION COMPANIES

In order to carry out a strategic group analysis for a specific industry, first, a conceptual framework should be defined in which variables that reflect the strategic perspective of companies are identified. Second, the data about those variables should be analyzed by using an appropriate clustering method that has the capability to reveal the real strategic group structure within an industry. Within the context of this study, a framework, details of which will be discussed below is developed, necessary data is collected through a questionnaire survey and finally, the three clustering techniques, namely traditional cluster analysis, SOM and fuzzy clustering methods are applied to the data set to identify the strategic groups. The convergence between the strategic groups that are found by using each method are checked and consequently, by comparing and combining the outputs of these three clustering methods, strategic groups within the Turkish construction industry are identified. The performance differences between these strategic groups are revealed and results are interpreted.

In order to eliminate some of the shortcomings of strategic group analysis as mentioned in Chapter 2, following strategies are utilized;

1. A conceptual framework, which is thought to be a relevant model for the construction industry is defined. It is based on the work of Price and Newson (2003) and the strategic variables are determined by referring to previous work of Dikmen and Birgonul (2003) about strategic perspective of Turkish contractors.

2. In order to eliminate some of the problems of using cluster analysis, alternative clustering techniques are used and results are compared to find the best technique to minimize the risk of dependency of results on the selected technique.

6.1. The conceptual framework for strategic grouping

Thomas and Venkatraman (1988) mention that there is an absence of a generally accepted scheme for definition of strategic dimensions to be used for strategic group analysis. They distinguished between two types of studies that have defined strategy; those that viewed strategy in narrow terms (a single dimension such as size, location etc.) versus those that viewed strategy in relatively broader terms (a complex array of scope and resource deployment decisions). They argue that development of strategic groups using narrow conceptualization of strategy is unlikely to capture the complexity of the strategy construct and they suggest utilization of multidimensional constructs. A broader range of variables increases the probability of including relevant group-defining attributes. There are various researchers that used a broad perspective for defining strategic variables (Cool and Schendel 1987; Dess and Davis 1984; Fiegenbaum and Thomas 1990). However, defining a high number of variables from a broad perspective is not without criticisms. Barney and Hoskisson (1990) points out the potential for introducing noise into the data analysis process as all variables are usually given equal weight in generating clusters. Thus, if irrelevant factors are included into the analysis, the resulting clusters may not reflect the actual group structure. They suggest that some theoretical framework shall be chosen for choosing strategic variables and judging the quality of results. However, this theory has yet to be developed.

Thomas and Venkatraman (1988) suggested utilization of the following strategies for the selection of a conceptual framework:

1. It should match the key basis of competition in the market place.

2. It should bear a strong relationship to some of the common theoretical discussions on strategy types so that a cumulative perspective can be developed.

Thus, in this research, the findings of previous research studies about the competitive environment within the Turkish construction industry are used to identify the strategic dimensions for grouping. Moreover, the most common framework used in strategic grouping research, which is Porter's generic strategies is utilized to identify the variables about strategic position. The details about selection of the conceptual model and variables are given in the next section.

6.2. Selection of variables

Price and Newson (2003) mention about three dimensions of strategy that can be recognized in every real-life strategic problem situation:

1. Strategy content : product of strategy process
2. Strategy process : manner in which strategies come about
3. Strategy context: organizational and environmental circumstances under which strategy process and content are determined

In this research, it is also hypothesized that how strategies are formulated and the organizational context should be considered as well as the output, which is the competitive position (mode and scope) during strategic grouping.

6.2.1. Strategy content

The scope and mode of competition are the products of a strategy formulation process. An organization distinguishes itself by differentiating its offerings in some way. Basically there are two generic strategies: price differentiation (cost leadership as defined by Porter) and quality differentiation. Companies that utilize price differentiation strategy charge a lower price for their service/product and it seems as if it is the only way to survive in a market like construction industry where usually the lowest cost bidder is awarded the contract. Betts and Ofori (1992) argue that despite the competitive bidding system, price differentiation is not the only strategic approach in construction with increasing concern for quality and value for money. Although some argue that there are limited ways to differentiate services in the construction sector, there are several writers that suggest that construction offers considerable scope for differentiation (Hasegawa 1988; Tatum 1987). Kale and Arditi (2002) argue that contractors may compete on the basis of product and contracting service quality as well as product or process innovations. The level of competitive advantage gained from a particular mode of competition depends on priorities of clients, how they define project success and the level of sensitivity of the project success criteria to different factors. In this research, they are grouped into two categories: price and non-price factors. If a contractor is competing on the basis of price, the primary objective is to minimize costs. In the other category, which is quality differentiation, a contractor tries to maximize client satisfaction through high quality products and services, innovations, support services etc. In this category, company resources are organized in such a way that the client objectives are

identified and met successfully. Quality differentiation strategy does not mean that cost issues are not considered, however reducing cost is not the primary objective.

Companies should make strategic choices about the markets they serve and types of projects they undertake, which is about drawing the boundaries of an organization. Companies may have a diversification strategy or a focus strategy which determines their competitive scope. In this research, two dimensions are identified to group the companies with respect to their competitive scope: market-level and project-level diversification. Market-level diversification is about whether the company operates in construction market and/or sectors related to construction or unrelated markets as well. Also, amount of works in the domestic market and international markets are considered under the heading of market-level strategy. Project-level diversification has two dimensions: type of project and type of client. Whether a construction company carries out a single type of project or different projects is questioned as well as the variety of clients served.

6.2.2. Strategy process

During the strategic group analysis, "realized strategies" of firms are evaluated to identify possible similarities or differences between firms operating in the same market. A realized strategy has two components: deliberate strategy (plans realized) and emergent strategy (patterns developed in the absence of intentions) (Mintzberg 1987). Junnonen (1998) also stresses that realized strategy is composed of both intended and emergent actions which are sometimes formed, not only formulated. Thus, he argues that success of strategy process depends on both formalization of procedures and existence of the right climate for strategy formation. Dikmen and Birgonul (2004b) identified the following two mechanisms used during the strategy development process:

1. Formulation of strategies: A strategic planning system is required for formulation of strategies and realization of intended strategies. The effectiveness of this mechanism is mainly about existence of a system/structure that supports strategy formulation.

2. Emergence/formation of strategies: Strategies, other than intended ones, may emerge within the organization. Success of these strategies depends on the existence of the right climate and organizational culture that permits idea generation. The effectiveness of this mechanism is mainly about the power culture and management style of the organization rather than the existence of systems or structures. An appropriate environment within the

organization which fosters imagination and creative thinking is vital for emergence of successful strategies.

Based on the findings of a questionnaire answered by 60 respondents, Dikmen and Birgonul (2004b) found statistically significant differences between the competitiveness of Turkish companies that are grouped under different categories with respect to their strategic decision making characteristics. In this research, two strategic variables have been defined: systematic strategic planning and centralization of decision making. It is assumed that companies may be categorized considering whether they carry out systematic strategic planning and whether a collaborative environment exists within the company or not.

6.2.3. Strategy context

The resource-based view asserts the critical role of resources and capabilities of a firm as the basis for strategy and primary determinants of profitability. Under the title of strategy context, tangible (human resources and financial resources) and intangible resources (experience and relations) of companies are considered as well as the capabilities (managerial and technical capabilities). Dikmen and Birgonul's (2003) work on strategic perspective of Turkish contractors is referred for identification of major strengths and weaknesses. According to the results of the mentioned study, client relations and financial resources are the major determinants of competitive advantage in the Turkish construction industry.

Finally, based on strategy context, content and process, 13 variables are defined. These variables and how they are measured are depicted in Table 6-1.

Table 6-1: Variables used during strategic group analysis

Variable	Measure	Type
Differentiation strategy	Category 1. Price differentiation Category 2. Quality differentiation	Nominal
Diversification strategy	Category 1: Only construction/construction-related sectors Category 2: Diversified in sectors unrelated to construction	Nominal
Internationalization	Internationalization ratio (ratio of international workload to total workload)	Numeric
Type of projects (project type that has the highest percentage in total number of projects)	Category 1: Infrastructure Category 2: Housing + building Category 3: Industrial Category 4. Others	Nominal
Type of client (client type that has the highest percentage in total number of projects)	Category 1:Government Category 2:Private sector	Nominal
Strategic planning	Category 1: Systematic and regular strategic planning Category 2: No systematic approach for strategic planning	Nominal
Strategic decision-making	Category 1: Democratic and collaborative environment Category 2: Autocratic approach	Nominal
Relations with clients	Subjective rating using 1-5 scale	Ordinal
Human resources	Subjective rating using 1-5 scale	Ordinal
Managerial capability	Subjective rating using 1-5 scale	Ordinal
Technical capability	Subjective rating using 1-5 scale	Ordinal
Financial resources	Subjective rating using 1-5 scale	Ordinal
Experience	Subjective rating using 1-5 scale	Ordinal

6.2.4. Research methodology

The target population is selected as the members of the Turkish Contractors Association (TCA) which is an independent and non-profit professional organization that represents the

leading construction companies in Turkey. The business volume of its members encompasses nearly 70% of all domestic and 90% of all international contracting work done so far by the Turkish construction companies. From the beginning of the 1970s up to the present, member companies of TCA have completed over 3000 projects in 63 countries and their business volume abroad has reached approximately 65 Billion US Dollars (USD). Thus, the strategic group analysis carried out in this research covers only the medium-big contracting firms in Turkey. Small and local firms are excluded from the analysis.

A questionnaire form is designed and submitted to 136 members of TCA. In the questionnaire, each representative of the company is requested to give relevant information about the 13 strategic variables. The yearly turnover (average of the last 3 years) and age of the company are also questioned. Subjective reporting approach is used for performance assessment rather than collecting financial data. Dess and Robinson (1984) argue that subjective performance measures are most appropriate in examining relative performance within an industry. Each respondent evaluated his/her company's performance considering the previous 3 year period in terms of profitability, workload and other company objectives. The total number of returned questionnaires is 84 and the return rate is 0.62.

"Size" as defined in this study refers to the average turnover in the last 3 years. Thus, a high correlation is expected to exist between the performance rating and size, since one of the components of performance rating is "workload". Initially, size (measured in $/year considering the last 3 years) and age are not included in the cluster analysis by assuming that they do not directly reflect a firm's strategic perspective. However, correlations between the strategic variables, size and age should be calculated and if age and/or size are found to be strongly correlated with other variables, the list of variables to include in the analysis can be revised.

Assumptions for the survey were;

- Each questionnaire sent to companies is designed to be answered by top executives.
- Respondents are medium-big size companies
- Each question should be answered by either comparing the respondent company with its competitors or by comparing the performance with market averages. This point has been made clear both in the questionnaires and during the interviews with the respondents.

6.2.4.1. Descriptive statistics of the respondents of the survey

- Statistics about Age: Age of the organization has been asked to relate company's age to its acquaintance to the markets it operates. Here in the questionnaire, age refers to the active years of the organization in construction business. The average age of the organizations is determined as 26.75, and standard deviation is calculated as 12.54. The oldest firm is determined as 63, and the youngest firm is determined as 3. The distribution of the companies according to their age is shown in Table 6-2.

Table 6-2: Distribution of the companies according to their ages

Age of the company in construction sector	**Frequency**	**%**
between 1-5	4	4.76
between 6-10	7	8.33
between 11-15	8	9.52
more than 15	65	77.39
Total	**84**	**100**

As seen from the above table, vast majority of the companies responded to the survey is older than 15 years (77.39%) in the construction business, and the data set is suitable for the target population which is medium-big size companies.

- Statistics about Turnover: "Size" is referred to average turnover in the last 3 years. According to the Table 6-3, the mean for yearly turnover is determined as 68.287 million $/year which is correlated with our assumption about the size of the companies, in addition to this, the median determined as 30 million $/year is also sufficient enough, however the minimum value 0.38 million $/year cause a deeper analysis, therefore the histogram of the total yearly turnover is illustrated in Figure 6-1. The standard deviation of yearly turnover is calculated as 93.672 million $/year, the reason of the largeness of this value can be explained by examining the maximum value and Figure 6-1. According to the Figure 6-1, seventeen companies is determined as having larger than 100 million $/year yearly turnover. Lastly, according to the Skewness and Kurtosis values which should be zero for normal distribution, it can be concluded that the size of the companies does not distribute normally. This leads to question the outputs obtained by the methods which assume that the variables show normal distribution like two-step cluster analysis.

Table 6-3: Descriptive statistics for yearly turnover

Descriptive Statistics	Value
Mean	68.278
Median	30.000
Mode	100.000
Std. Deviation	93.672
Variance	8774.494
Skewness	2.339
Std. Error of Skewness	0.263
Kurtosis	6.038
Std. Error of Kurtosis	0.520
Range	449.620
Minimum	0.380
Maximum	450.000

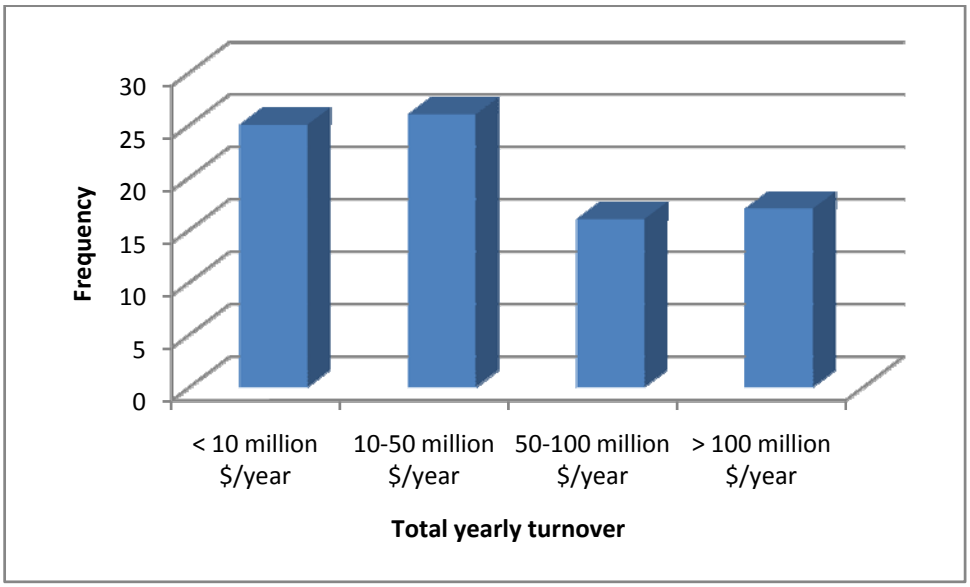


Figure 6-1: Distribution of respondents according to their total yearly turnover

6.3. Clustering techniques used for strategic group analysis in the literature

There are many studies on strategic groups in the management literature, however, almost all of them were conducted by using traditional cluster analysis methods, whereas this is a

criticism point of the existence of the strategic group concept, because cluster analysis' sorting ability is powerful enough that it will provide clusters even if no meaning groups are embedded in a sample (Ketchen and Shook 1996). In addition, Day et al. (1995) criticized the application of cluster analysis in strategic groups studies due to its ignorance of the casual models relating, its operational decisions with the attainment of an objective function developed by the managers for designing strategies. In addition, the companies do not aim only profit maximization; in fact, they are typically driven by multiple goals, however the multidimensionality of strategies cannot be captured by the traditional cluster analysis methods. Moreover, the boundaries of strategic groups are likely to be 'fuzzy' (Dranove et al. 1998), also Reger and Huff (1993) suggested that a strategic group might be best conceptualized as a core group of firms and secondary firms that share some of the attitudes of core members of different strategic groups; even, they proposed that the secondary firms in a strategic group perform better than the core members of this strategic group. McNamara et al. (2003) found out the supporting evidence for Reger and Huff's (1993) proposal about the performance difference between the secondary and core firms for commercial banks. Therefore, the membership of the firms is not enough to examine the structure of the market; in addition the typology of the market structure which cannot be determined by using traditional cluster analysis methods should be identified. Therefore, different clustering methods should be used in order to compare the outputs for identifying the most suitable solution for the data set. Kiang et al. (2001) concluded that the SOM can be used instead of cluster analysis especially for the skewed data. However, only two studies, in which SOM was used in the strategic group research, can be found in the literature: namely work by Serrano-Cinca (1998) and Curry et al. (2001). Also, there is only one research about strategic groups in which FCM was used by Hsu (2000).

Serrano-Cinca (1998) tried to determine the strategic groups of Spanish saving banks by using the published financial data for 1991. But, the purpose of carrying out SOM analysis in this study is to verify the findings from the cluster analysis, and the application process of SOM was not depicted in detail (Wang and Wang 2002). SOM was also used for finding out the strategic groups in the UK hotel industry by Curry et al. (2001). In contrast to the previous study, the published data was not used in this analysis, the responses of the hotel managers to the questions on their strategies were taken as data. The detailed information about the SOM was given, however information about the determination of the parameters of SOM was not explained in detail. After application of the SOM, K-means cluster analysis

was applied to the data set and strategic groups obtained from these two methods were compared with each other.

6.3.1. Analysis of the data set

The cluster analysis is affected from the structure of the data set; therefore the data set should be checked for some issues before the application of the cluster analyses, namely standardization of the variables and multi-linearity among the variables.

6.3.1.1. Standardization of the variables

As mentioned before, due to large scale difference between the variables, unsuitable clustering outputs can be obtained at the end of the analysis. The remedy of the high weights of variables with large ranges in cluster analysis can be standardization; however the standardization of the variables can cause elimination of meaningful differences among members, therefore descriptive analysis is performed to the data set in order to decide if the standardization is required, and the output of the descriptive analysis is shown in Table 6-4.

Table 6-4: Descriptive analysis of the variables

Variables	Minimum	Maximum	Mean	Std. Deviation	Variance
Type of client	0.000	1.000	0.703	0.327	0.107
Internationalization ratio	0.000	1.000	0.201	0.272	0.074
Type of project	1.000	3.000	1.600	0.730	0.533
Strategic planning	0.000	1.000	0.680	0.470	0.221
Strategic decision-making	0.000	1.000	0.740	0.442	0.196
Relations with clients	1.000	5.000	3.500	1.092	1.193
Human resources	1.000	5.000	3.580	0.984	0.969
Managerial capability	1.000	5.000	3.300	1.027	1.055
Technical capability	1.000	5.000	3.950	0.943	0.889
Financial resources	1.000	5.000	3.690	0.994	0.987
Experience	1.000	5.000	3.630	1.095	1.200
Differentiation strategy	0.000	1.000	0.790	0.413	0.170
Diversification strategy	0.000	1.000	0.550	0.501	0.251
Overall performance	1.000	5.000	3.450	1.080	1.166

According to Table 6-4, the variable which has the largest variance, 1.200, is determined as experience; variances of the other variables are close to this value. In addition, according to the means of the variables, minor differences are observed between the variables. For that reason, it is decided that the standardization is not required for the data set.

6.3.1.2. Multi-linearity among the variables

As mentioned before, high correlations between variables can cause flawed results by overweighting one or more underlying constructs. The remedy methods against the multi-linearity can be using Mahalanobis distance measure as similarity measure, or subjecting variables to factor analysis and using uncorrelated factor scores for each observation as the basis for cluster analysis (Ketchen and Shook 1996). The Mahalanobis distance measure cannot be used for all of the cluster analysis methods. Unfortunately, in application of cluster analysis with this distance measure, the variables are standardized automatically, whereas it was decided that no standardization is applied to the data set in this study. The other methods can cause exclusion of the factors which may provide unique, important information (Dillon et al. 1989), therefore before application of these methods, the data set is checked against multi-linearity by calculating Pearson correlations between the variables, and the correlation matrix is presented in Table 6-5. In this analysis, age and size variables are also included for determining existence of the correlation between these variables and remaining variables. But, age and size are not considered among the strategic variables for grouping. According to the Table 6-5, Age is determined as significantly correlated with only Size and Diversification strategy; in addition Size is significantly correlated with Technical capability and Age. Also, significant correlations are also observed between the Relations with client, Human resources, managerial capability, Technical capability, Financial resources, Experience and Differentiation strategy. As they are lower than the threshold of strong correlation, 0.70, it is assumed that no multi-linearity effect exists in the current data set. Therefore, no process is applied against multi-linearity in this study.

Table 6-5: Correlation matrix of the variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.Age	1.00															
2.Size	(**)0.44	1.00														
3. Type of Client	0.08	(*)0.32	1.00													
4.Internationalization	0.19	0.08	-0.17	1.00												
5. Types of Projects	0.17	-0.10	(**)-0.45	0.16	1.00											
6. Strategic Planning	(*)0.31	0.16	0.06	0.17	-0.06	1.00										
7.Strategic Decision Making	0.26	0.16	0.07	-0.11	0.05	(*)0.29	1.00									
8. Relations with Clients	0.26	0.18	0.02	0.11	-0.07	(**)0.55	(**)0.42	1.00								
9. Human Resources	(*)0.30	(*)0.29	-0.02	0.20	-0.03	(**)0.46	(**)0.38	(**)0.64	1.00							
10. Managerial Capability	0.27	0.17	-0.05	0.24	0.03	(**)0.55	(*)0.31	(**)0.53	(**)0.58	1.00						
11. Technical Capability	0.27	(**)0.37	0.02	0.07	-0.01	(**)0.46	(**)0.46	(**)0.56	(**)0.65	(**)0.64	1.00					
12. Financial Resources	0.22	0.15	0.09	-0.02	-0.05	(**)0.48	(**)0.44	(**)0.60	(**)0.61	(**)0.54	(**)0.61	1.00				
13. Experience	(*)0.29	(*)0.30	0.03	0.12	-0.03	(**)0.59	(**)0.52	(**)0.61	(**)0.65	(**)0.55	(**)0.64	(**)0.50	1.00			
14. Differentiation Strategy	0.18	0.02	0.05	0.18	-0.08	(**)0.45	(**)0.42	(**)0.62	(**)0.52	(**)0.47	(**)0.50	(**)0.48	(**)0.44	1.00		
15. Diversification Strategy	(**)0.40	0.18	-0.03	-0.10	0.16	0.04	0.11	0.02	0.13	0.27	0.26	0.15	0.18	-0.01	1.00	
16. Performance	(*)0.30	0.27	0.01	0.16	-0.00	(**)0.62	(**)0.48	(**)0.75	(**)0.75	(**)0.65	(**)0.76	(**)0.75	(**)0.74	(**)0.65	0.18	1.00

*. Correlation is significant at the 0.01 level (2-tailed).

**. Correlation is significant at the 0.001 level (2-tailed).

6.3.2. The application of traditional cluster analysis

The tool used for cluster analysis is SPSS 13.0 which offers three methods for cluster analysis; hierarchical cluster, K-means cluster and two-step cluster analysis, in addition to this S-plus is used for PAM procedure. As mentioned before, all of these methods have advantages and disadvantages over each other. For example, in hierarchical method, the number of clusters shall be determined after the cluster analysis, but for K-means method, the number of clusters should be known at the inception of the analysis. Hierarchical cluster analysis procedure makes only one pass through the data, therefore a poor early partition of data may persist throughout the analysis and lead to artificial results. Whereas, K-means cluster analysis can fix an early poor partition problem at later stages of the process by passing over the data more than once. The most important advantage of two-step cluster method is its ability to determine the number of the clusters automatically. Also, huge amount of data can be analyzed efficiently. However, accurate results can be obtained by using this method only if all variables are independent, continuous variables have normal distributions and categorical variables have multinomial distributions. In real life, these conditions can be provided rarely (Norusis 2004). Besides, the suitability of the clustering methods varies with changing types of data set. Specifically, K-means cluster analysis is more suitable for determining compact and hyper-spherical clusters, whereas centroid method reveal more accurate membership structure of the data set for irregularly shaped clusters. In this study, as it is impossible to determine the most appropriate clustering technique at the beginning due to non existence of information about the membership of the clusters, all of the available traditional cluster analysis methods are used and results are compared to obtain optimum solution.

The analysis is started with PAM procedure. After the application of PAM, the analysis is continued with all agglomerative hierarchical cluster methods. Lastly two-step and K-means cluster analysis methods are applied. At the end of the traditional cluster analysis methods, the outputs obtained from these methods are validated by using the validation techniques. In this way, the most suitable number of the clusters and best solution are determined.

6.3.2.1. The application of PAM procedure

The Euclidean distance is only available distance measure for all types of the methods, besides it is the most widely used method in the literature, therefore for satisfying the

stability between the methods, the Euclidean distance is decided to be used in all of the analyses.

Table 6-6: Number of clusters vs. silhouette coefficients

Number of cluster	Silhouette coefficients
2	0.30
3	0.24
4	0.13

PAM is not available in SPSS; therefore the cluster analysis by using this method is performed by S-PLUS. For application of this method, the number of the clusters should be known at the inception of the analysis; the analysis can be processed by determining the number of the clusters according to the silhouette coefficients obtained from different number of clusters. Table 6-6 shows the silhouette coefficients for solutions obtained for different number of clusters in this data. According to this table, the solution obtained for two clusters is the most suitable solution; however the silhouette coefficient value for two clusters solution is calculated below threshold recommended for a reasonable structure -0,51-, this indicates that the structure can be artificial, therefore additional methods of data analysis is required in order to verify these findings (Struyf et al. 1997).

6.3.2.2. The application of agglomerative hierarchical method

After application of PAM, all of the available agglomerative hierarchical cluster analysis methods are applied to data set: average linkage within groups, average linkage between groups, single linkage (nearest neighbor), complete linkage (furthest neighbor), centroid clustering, median clustering and Ward's method.

The dendrograms and inverse scree trees obtained as outputs of these analyses can be used for determining the most suitable number of the clusters for this data set. However, the number of the clusters cannot be obtained directly from these outputs; the researchers should determine the number of the clusters by examining these outputs subjectively, therefore the findings from same outputs can vary from case to case. The inverse scree tree is not an available output in the used software packages; therefore, Ms Excel 2007 is used for

illustration of the inverse scree trees. In addition, as mentioned before, inverse scree trees for median linkage, centroid linkage and average linkage within groups with combining clusters cannot be obtained.

The quality of the solutions derived from agglomerative methods can be evaluated by utilizing agglomeration coefficients of the methods. The threshold for this measure is recommended as 0.78 (Struyf et al. 1996).

6.3.2.2.1. The application of average linkage between groups

The inverse scree tree and dendrogram obtained at the end of the agglomerative analysis for average linkage between groups are shown in Figure 6-2 and Figure 6-3. According to the inverse scree tree, the sharpest increase is obtained for two clusters solution, the second sharpest increase is observed at the three clusters solution. According to the dendrogram, the newly formed cluster in three clusters solution has only two members. Consequently, it is decided that two clusters solution is the appropriate solution according to this analysis.

The agglomerative coefficient is determined as 0.74. According to 0.78, the obtained structure is good, but not good enough to pass the threshold for this study.

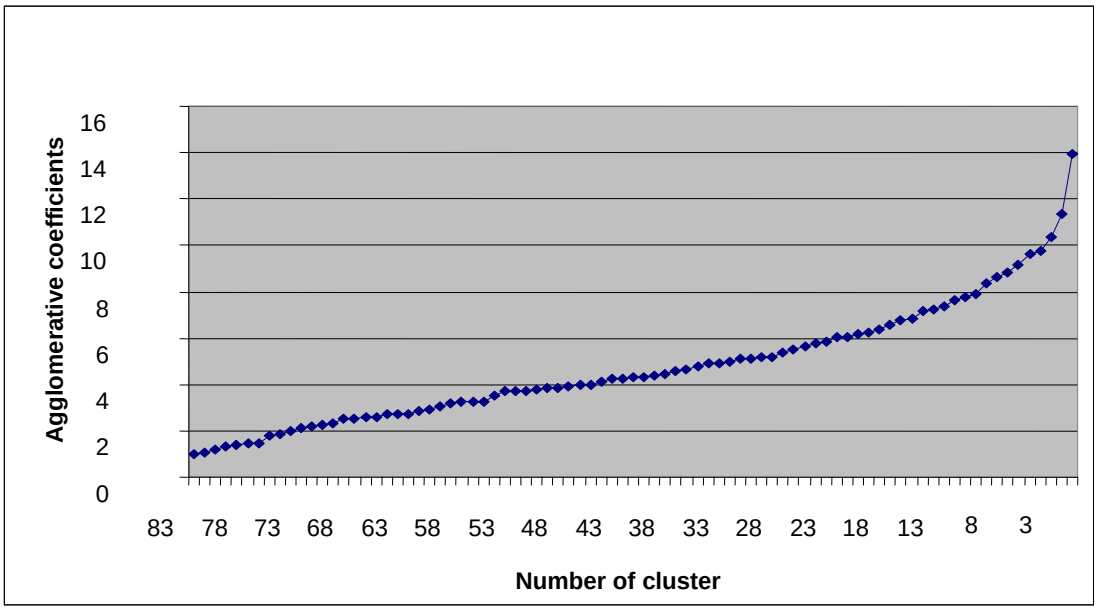


Figure 6-2: The inverse scree tree for average linkage between groups

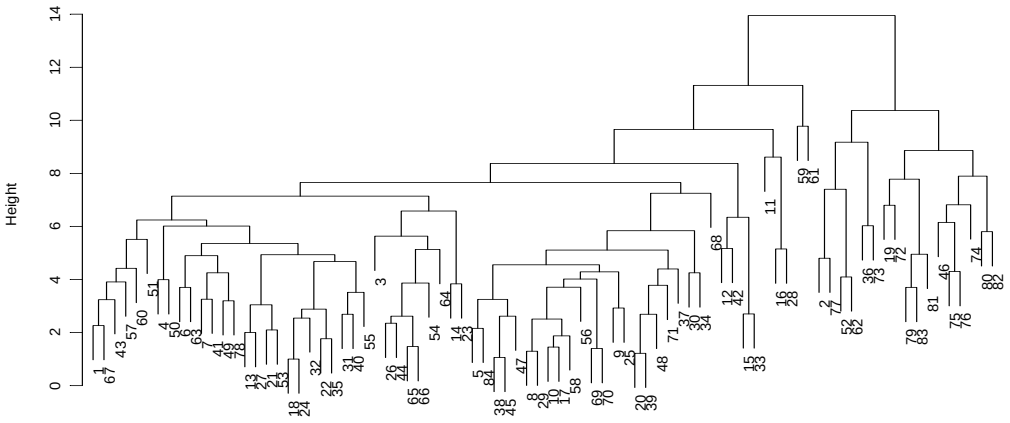


Figure 6-3: Dendrogram for average linkage between groups

6.3.2.2.2. The application of average linkage within groups

The inverse scree tree and dendrogram obtained at the end of the agglomerative analysis for average linkage within groups are shown in Figure 6-4 and Figure 6-5. As seen from the

Figure 6-4, the line does not increase continually, there are some declines in the line, and therefore, the conclusions obtained from inverse scree tree are questionable. In the dendrogram, for two clusters solution, it is observed that one of the clusters contains only five firms, the remaining seventy nine firms are comprised by the other cluster. This clustering structure is not preferred type of clustering structure. Consequently, according to the clustering structures obtained from the dendrogram, three clusters or four clusters solutions have more preferable cluster structures; therefore the three clusters solution is identified as appropriate.

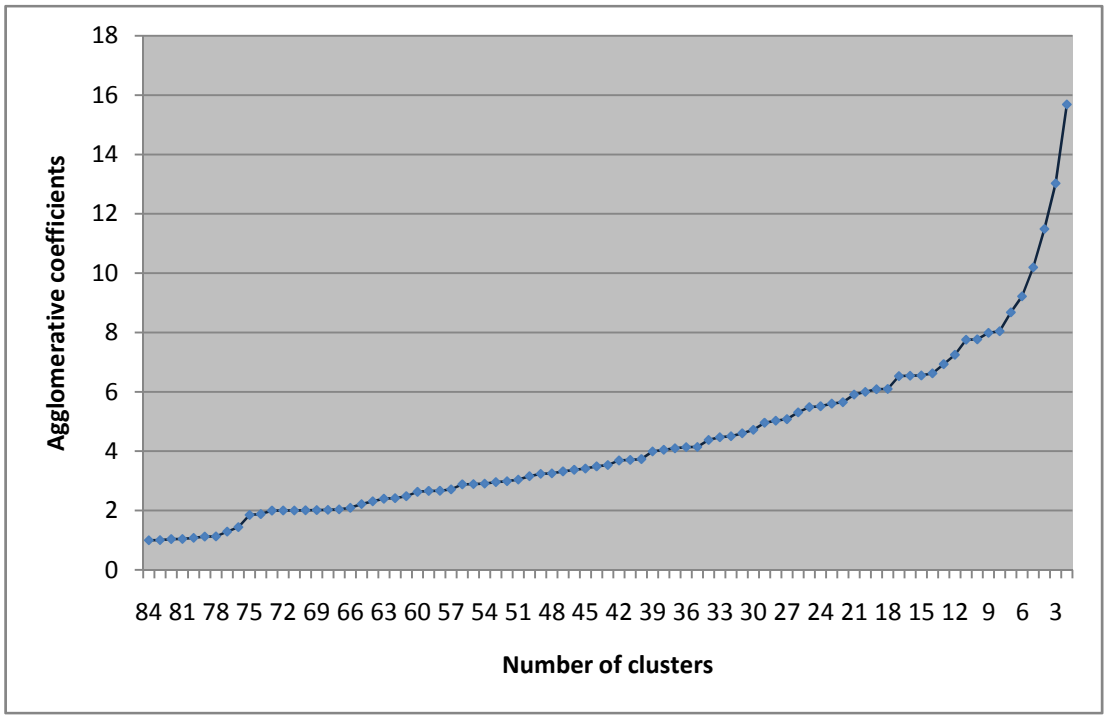


Figure 6-4: The inverse scree tree for agglomerative analysis (within groups)

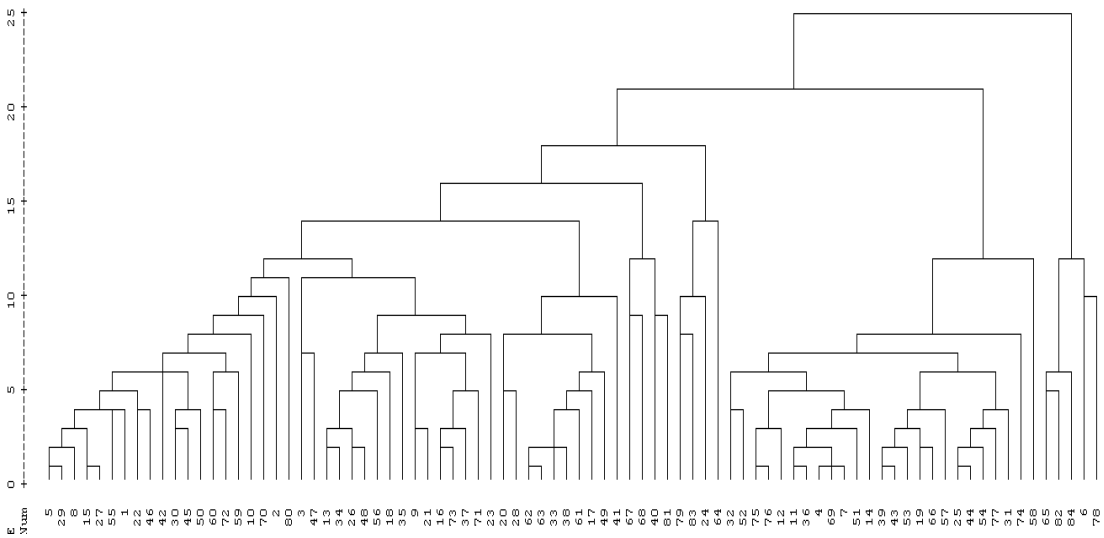


Figure 6-5: Dendrogram for average linkage within groups

6.3.2.2.3. The application of single linkage method

The inverse scree tree and dendrogram obtained at the end of the single linkage analysis are shown in Figure 6-6 and Figure 6-7. According to the inverse scree tree, the sharpest increase is obtained for four clusters solution, the second sharpest increase is observed at the two clusters solution. According to the dendrogram, in four cluster solution, two of the clusters contain only one firm, in addition one of the clusters contain four firms, whereas the remaining firms form the last cluster. As a result, the chaining, main problem of single linkage can be easily observed.

The agglomerative coefficient calculated as 0.56 shows that the solution obtained from this method is not competent to illustrate the clustering structure for this data set.

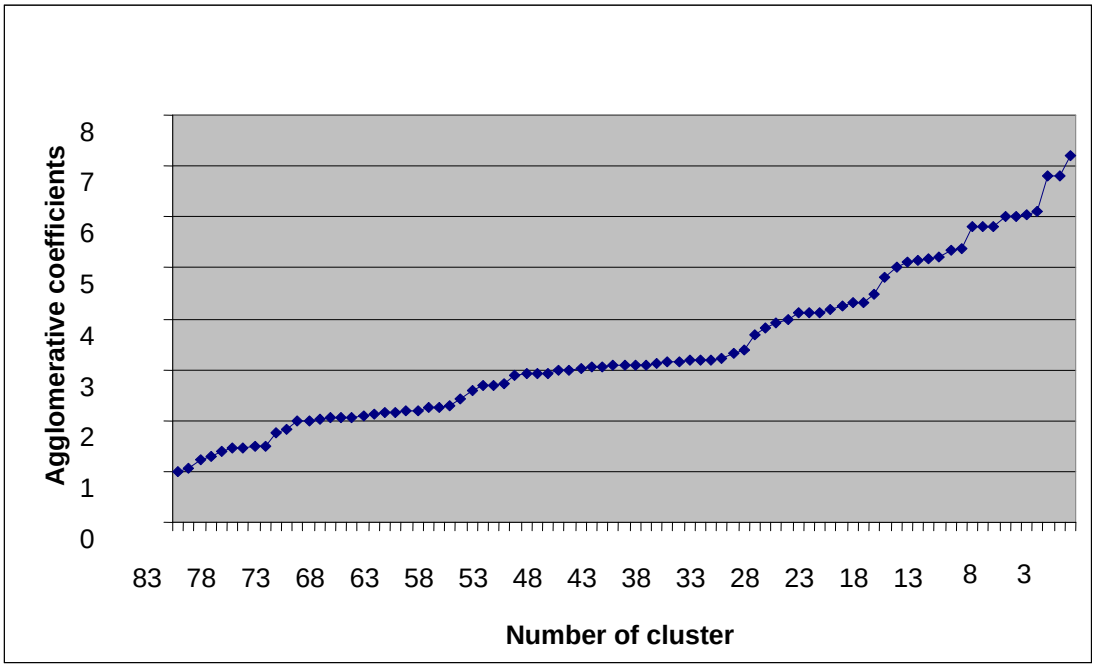


Figure 6-6: The inverse scree tree for single linkage

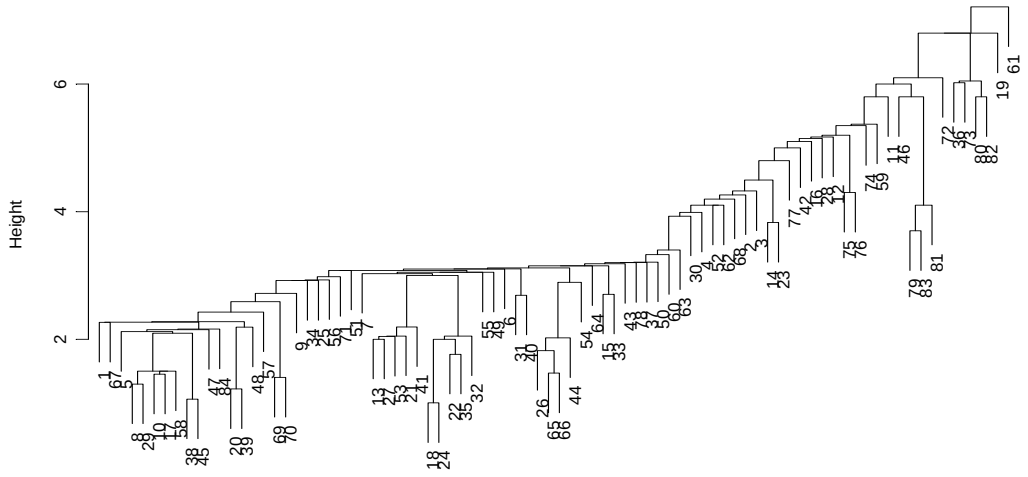


Figure 6-7: Dendrogram for single linkage

6.3.2.2.4. The application of complete linkage method

The inverse scree tree and dendrogram obtained at the end of the complete linkage analysis are shown at the following figures. According to the inverse scree tree, the sharpest increase is obtained for two clusters solution, the second sharpest increase is observed at the three clusters solution. According to the dendrogram, three clusters are distributed uniformly; this shows that three clusters solution is more suitable for this data set.

The agglomerative coefficient is determined as 0.87. This is higher than the defined threshold; in addition to this, the structure of the clusters based on number of data placed in is good enough to be accepted.

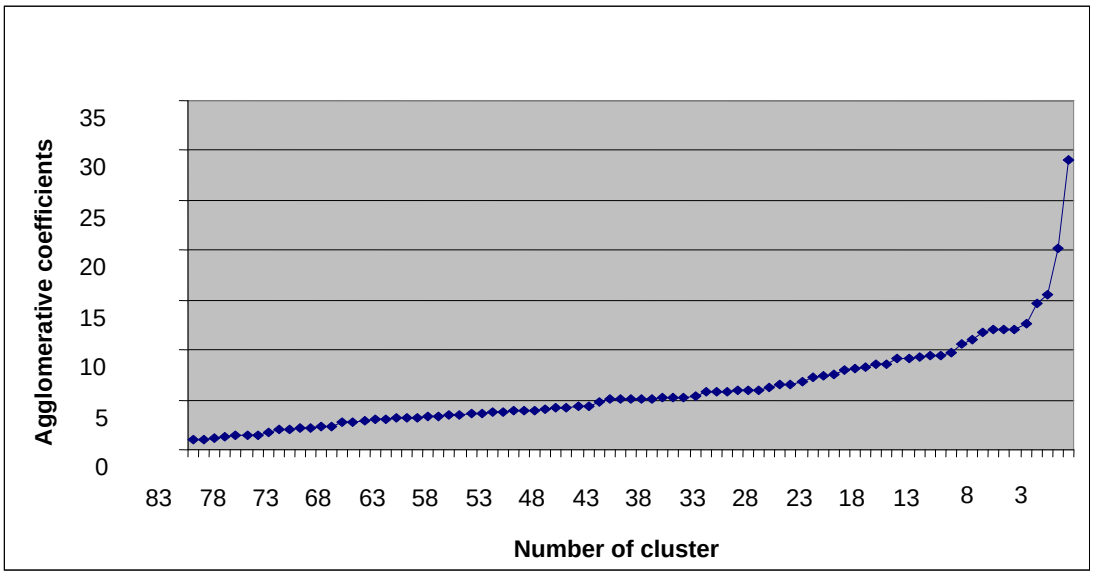


Figure 6-8: The inverse scree tree for complete linkage

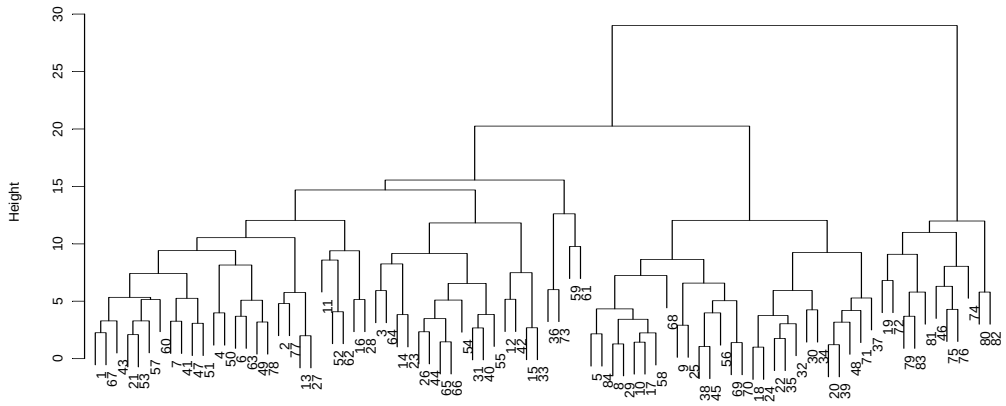


Figure 6-9: Dendrogram for complete linkage

6.3.2.2.5. The application of centroid linkage method

In the application of centroid linkage, the problem mentioned for the average linkage within groups is also valid; therefore the inverse scree tree of this method is not illustrated. Only dendrogram obtained from this analysis is shown in Figure 6-10. According to the dendrogram, two clusters solution is determined as the most suitable one according to the

merged heights and the clustering structure. In two clusters solution, the companies are distributed uniformly, whereas in three clusters solution, the newly formed cluster will contain only two firms which is not a preferable clustering type.

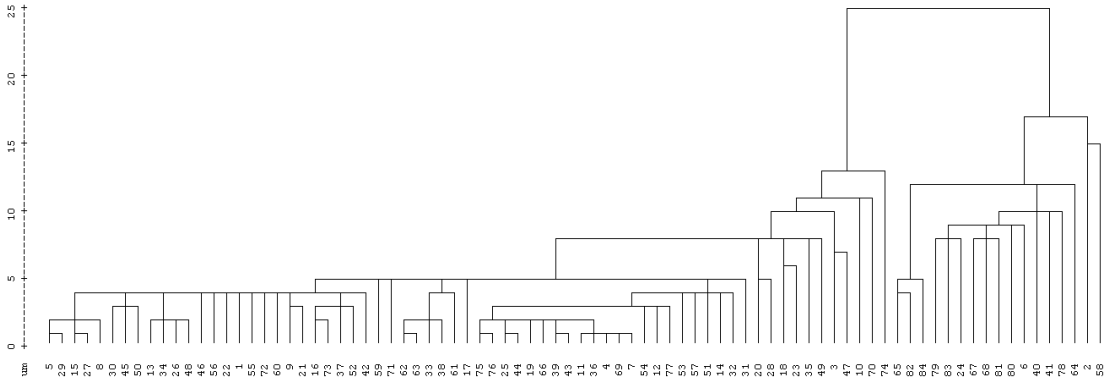


Figure 6-10: Dendrogram for centroid linkage

6.3.2.2.6. The application of median linkage

In the application of median linkage, the same problem mentioned for the average linkage within groups is also valid; therefore the inverse scree tree of this method is not illustrated. Only dendrogram obtained from this analysis is shown in Figure 6-11. According to the dendrogram, two clusters solution is determined as the appropriate solution according to the merged heights, but obtained clustering structure for two clusters is not preferable, because one of the cluster contains only two firms, whereas remaining all firms are comprised by the other cluster, unfortunately this situation is also valid for three or more number of clusters. Chaining problem as seen in single linkage method is also observed, therefore it should be expected that data points are assigned to the same clusters by these two cluster methods.

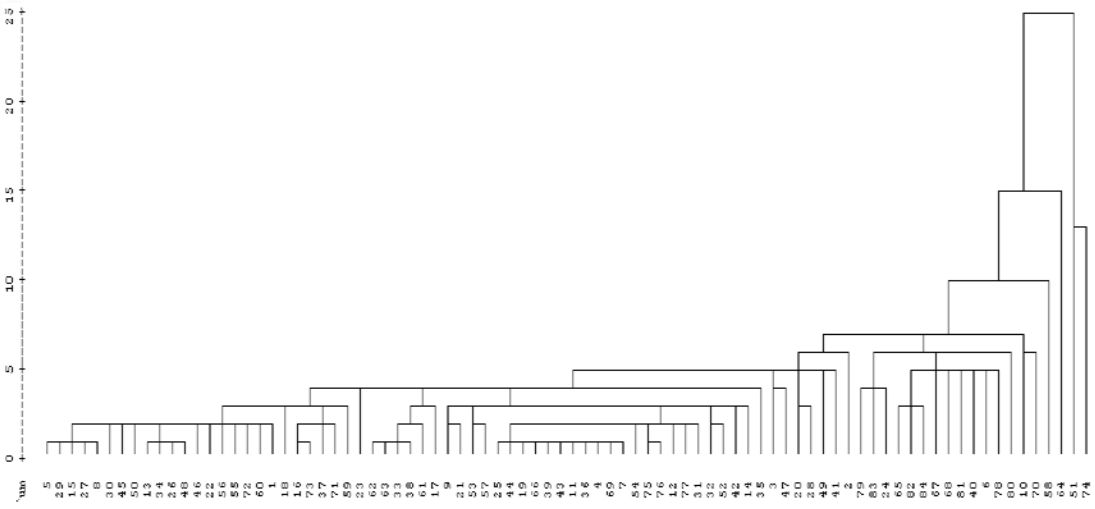


Figure 6-11: Dendrogram for median linkage

6.3.2.2.7. The application of Ward method

The inverse scree tree and dendrogram obtained at the end of the complete linkage analysis are shown at the following figures. According to the inverse scree tree, the sharpest increase is obtained for two clusters solution, but also the three clusters solutions should be considered as a valid alternative. According to the dendrogram, the sharpest merge is also observed at the last stage, however the height of the merge point of the clusters which form the first cluster at the last step is sufficient enough to define these two clusters are distinctive. In conclusion, three clusters solution is accepted as the appropriate solution for this data set.

The agglomerative coefficient (0.91) also illustrates that the great amount of clustering structure is revealed; also this is the highest value among agglomerative coefficients of all other methods. Besides, this method is advised by some authors, like Bacher (2002), Aldenderfer and Blashfield (1984).

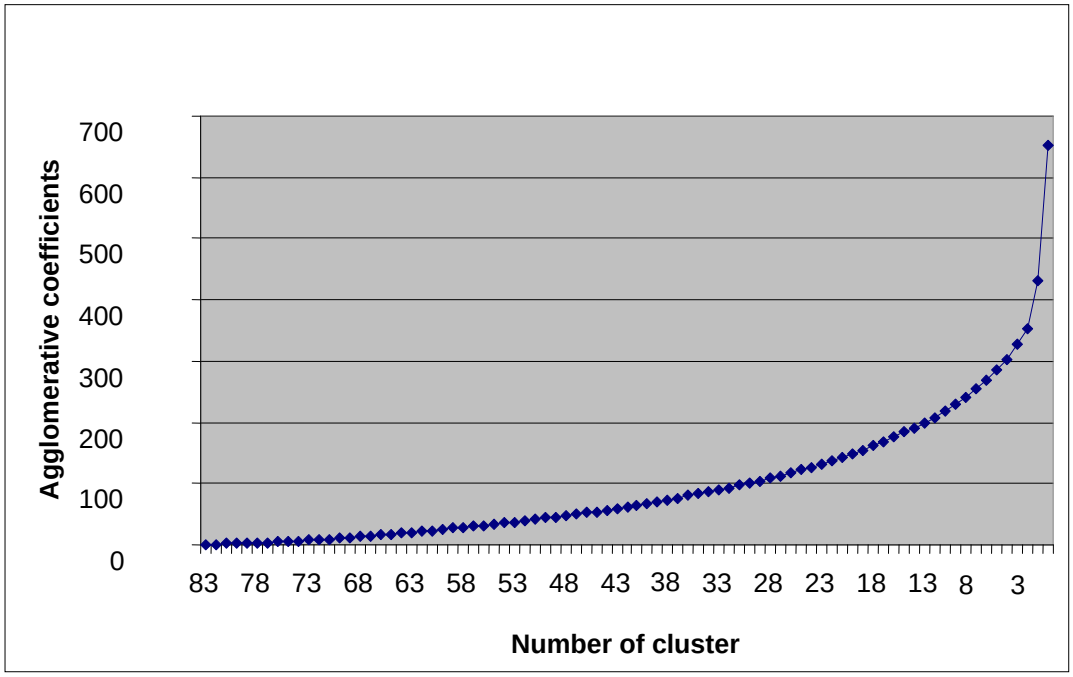


Figure 6-12: The inverse scree tree Ward's method

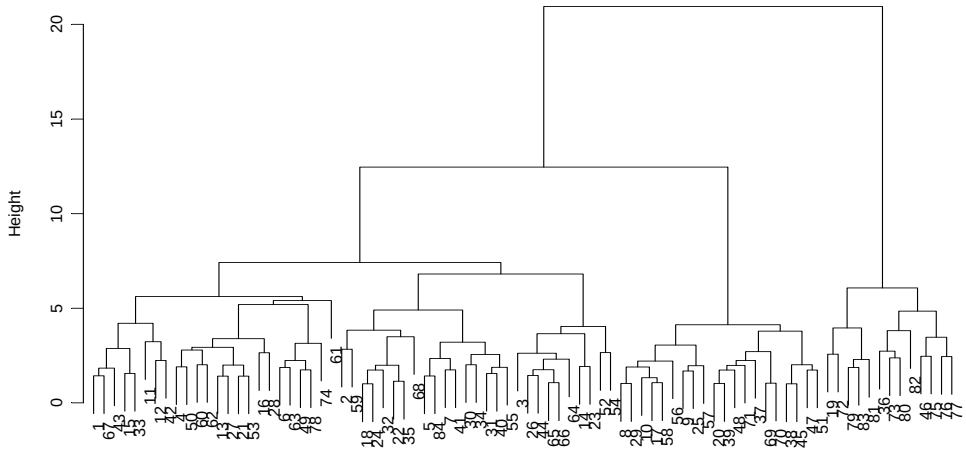


Figure 6-13: Dendrogram for Ward's method

6.3.2.3. Application of two-step cluster analysis

Two distance measures, log-likelihood and Euclidean distance are available for this analysis. The log-likelihood distance measure should be preferred when the data set is comprised of continuous and categorical variables. Therefore, the Euclidean distance measure is used in calculations. The stability between the methods is satisfied. The number of the clusters is identified by the analysis automatically according to the two separate criteria, Schwarz's Bayesian Criterion (BIC) and Akaike's Information Criterion (AIC). The following tables show the BIC and AIC process with changing the number of clusters. Smaller values of the BIC and AIC indicate better models, and in this situation, the "best" cluster solution has the smallest BIC. However, there are clustering problems in which the BIC and AIC will continue to decrease as the number of clusters increases, but the improvement in the cluster solution, as measured by the BIC and AIC Change, is not worth the increased complexity of the cluster model, as measured by the number of clusters, so the changes in BIC and changes in the distance measures should be used for determining the number of the clusters. A good solution will have a reasonably high ratio of BIC changes and ratio of distance measures. According to the Table 6-7 and Table 6-8, the appropriate number of the clusters is determined as three.

Table 6-7: Two-step cluster process with the BIC

# of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change (a)	Ratio of BIC Changes (b)	Ratio of Distance Measures (c)
1	1792.917			
2	1621.684	-171.233	1.000	1.639
3	1575.975	-45.709	0.267	2.368
4	1643.692	67.716	-0.395	1.085
5	1717.917	74.226	-0.433	1.059
6	1796.394	78.476	-0.458	1.112
7	1882.130	85.737	-0.501	1.088
8	1973.138	91.008	-0.531	1.135
9	2071.220	98.081	-0.573	1.099
10	2174.014	102.795	-0.600	1.056
11	2279.328	105.313	-0.615	1.066
12	2387.450	108.123	-0.631	1.132
13	2500.536	113.086	-0.660	1.028
14	2614.630	114.094	-0.666	1.117
15	2732.552	117.922	-0.689	1.054

Table 6-8: Two-step cluster process according to AIC

# of Clusters	Akaike's Information Criterion (AIC)	AIC Change (a)	Ratio of AIC Changes (b)	Ratio of Distance Measures (c)
1	1710.270			
2	1456.388	-253.881	1.000	1.639
3	1328.032	-128.356	0.506	2.368
4	1313.101	-14.931	0.059	1.085
5	1304.678	-8.422	0.033	1.059
6	1300.507	-4.171	0.016	1.112
7	1303.596	3.089	-0.012	1.088
8	1311.956	8.360	-0.033	1.135
9	1327.390	15.433	-0.061	1.099
10	1347.537	20.147	-0.079	1.056
11	1370.202	22.666	-0.089	1.066
12	1395.677	25.475	-0.100	1.132
13	1426.115	30.438	-0.120	1.028
14	1457.561	31.446	-0.124	1.117
15	1492.836	35.275	-0.139	1.054

6.3.2.4. Application of K-means cluster analysis

The pre-request of K-means cluster analysis is the number of the clusters. In other words the number of the clusters should be known or at least defined before the application of the analysis, and therefore for decreasing the number of the alternatives, the outputs obtained from the other cluster methods are examined. According to the outputs of the agglomerative cluster analyses, the most suitable solution is defined as two clusters solution, but according to the dendrograms of the complete and Ward's linkages, three clusters solution should also be taken into account as an alternative. In addition to this, the most suitable solution is determined as three clusters solution by the two-step cluster analysis for each criterion separately. According to these findings, two and three clusters should be selected as alternatives. Also, for determining the number of the clusters, cluster validity toolbox that has seven validity indexes, namely weighted inter-intra, silhouette, Davies-Bouldin, Calinski-Harabasz, Dunn, Krzanowski-Lai and Hartigan index are utilized. According to Krzanowski-Lai, Silhouette, Calinski-Harabasz and Dunn index, two clusters solution is the most suitable one, whereas Hartigan, Davies-Bouldin and weighted inter/intra indexes exhibit the most suitable solution for three clusters solution. These indexes verify the findings from the other cluster analysis methods.

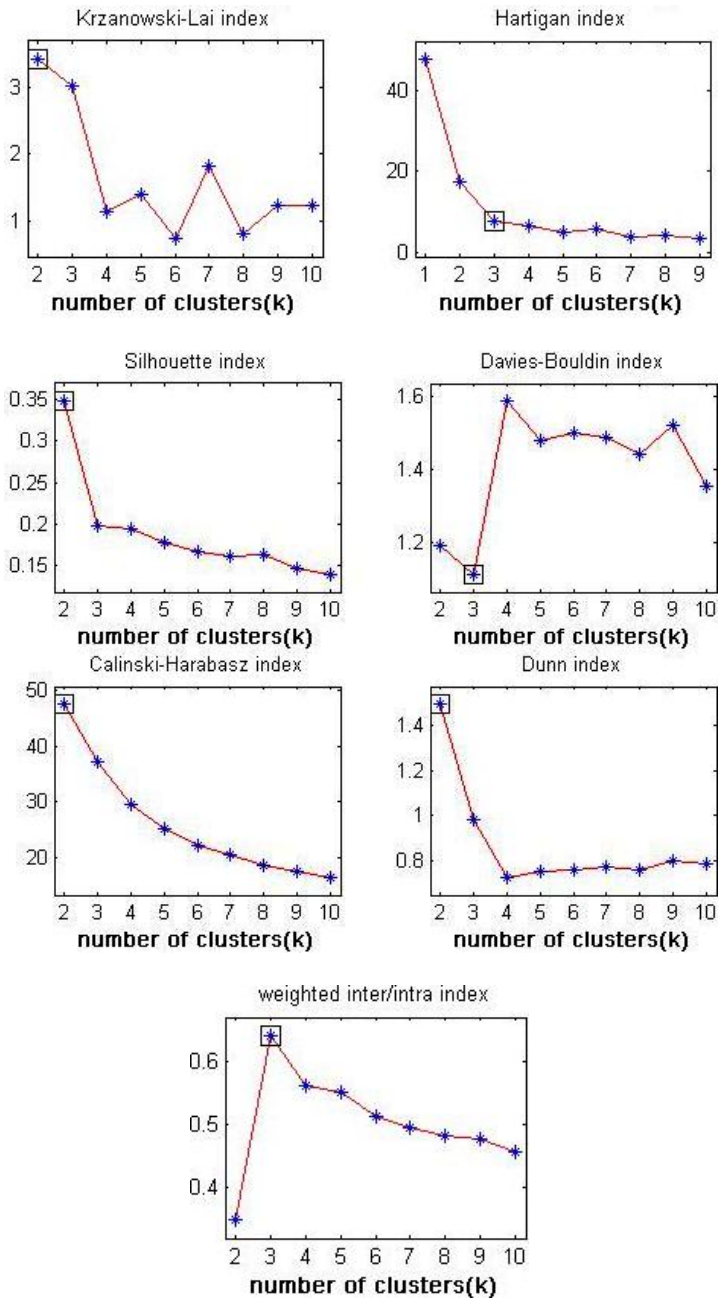


Figure 6-14: Validity indices for determining the number of the clusters

Since the important point in the determination of number of the clusters is that the new cluster can provide significant information that can be lost in the two clusters solution, the distance of the centers of the clusters shown in Table 6-9 is compared. By checking the distance between the centers of the clusters, the new cluster is determined as a required cluster for understanding the structure of the market properly.

Table 6-9: Distances between final cluster centers for two and three clusters

Cluster	1	2
1		3.843
2	3.843	

Cluster	1	2	3
1		2.147	5.491
2	2.147		3.475
3	5.491	3.475	

6.3.3. Determination of the number of the clusters

In conclusion, the most suitable alternatives regarding the number of clusters for this data set is determined as two or three according to the traditional cluster analysis methods. Specifically, according to two-step and K-means clusters analyses, three clusters solution is determined as the appropriate, whereas the most of the agglomerative cluster analysis methods reveal that two clusters solution is more suitable for this data set. Among the all agglomerative coefficients of agglomerative cluster analyses, the complete linkage and Ward's method have the highest value, whereas the single linkage method has the lowest value. Complete linkage and Ward's method also indicate that three clusters solution is a good alternative. Also, in the clustering methods in which two clusters solution is determined as appropriate, the distribution of the number of members within the cluster is not preferred; specifically one of two clusters obtained at the end of the single linkage method has only one member. Consequently, three clusters solution is decided as the most appropriate solution for this data set.

6.3.4. Determination of the strategic groups

In order to determine which clustering method fit the structure of the data set accurately, the stability of algorithms between each other and according to changes in population is tested and the number of the data points in a cluster is examined, specifically the methods which provide clusters that have only one or two members should not be used (Bacher 2002). Lastly, the external validity of the methods is examined by selecting the performance of the companies as external criteria in accordance with analysis of variances (ANOVA).

6.3.4.1. Stability of the cluster solutions

After determination of the number of the clusters, from the obtained cluster solutions, the most suitable one should be determined. For that purpose, firstly the stability between the clusters formed by using different algorithms is checked. For determining the stability between the clusters, rand indices of the clusters (as shown in Table 6-10) are calculated.

Table 6-10: Rand indices of the traditional cluster analysis methods

Cluster Analysis Methods		Two-step with BIC	Hierarchical							K-means with three clusters
			Aver. Link. (Between groups)	Aver. Link. (Within groups)	Single Linkage	Complete Linkage	Centroid Linkage	Median Linkage	Ward Linkage	
	Two-step with BIC	-								
Hierarchical	Aver. Link (Between groups).	0.614	-							
Hierarchical	Aver. Link. (Within groups)	0.798	0.549	-						
Hierarchical	Single Linkage	0.377	0.722	0.502	-					
Hierarchical	Complete Linkage	0.813	0.594	0.921	0.502	-				
Hierarchical	Centroid Linkage	0.631	0.980	0.544	0.703	0.588	-			
Hierarchical	Median Linkage	0.386	0.670	0.513	0.931	0.505	0.652	-		
Hierarchical	Ward's Linkage	0.762	0.644	0.684	0.387	0.711	0.645	0.394	-	
K-means with three clusters		0.804	0.675	0.788	0.454	0.847	0.668	0.466	0.757	-

The threshold for stability is determined as 0.7 which was recommended by Dreger (1986), Fraboni and Saltstone (1992). Unfortunately, the rand index a not available option in SPSS and S-plus, therefore the rand indexes between the algorithms are calculated by using the MS Excel 2007. According to the Table 6-10, two-step cluster with BIC analysis is stable with average linkage within groups, complete, Ward's linkage and K-means with three clusters. Average linkage between groups is stable with single linkage and centroid linkage, in addition average linkage between groups and centroid linkage has the highest rand index very close to the perfect fit. Average linkage within groups is stable with two-step with BIC, complete linkage , K-means with three clusters, also the rand index value between average linkage within groups and complete linkage is higher than 0.900. Single linkage is stable with average linkage between groups and median linkage, as mentioned before, the stability between the single and median linkage is expected due to the chaining problem observed in both of the method. For complete linkage, the stable between the two-step with BIC, average linkage within groups, Ward's linkage and K-means with three clusters is satisfied. Centroid linkage shows stability only with the average linkage between groups. Median linkage is stable with only single linkage. Ward's linkage and K-means with three clusters show stability with two-step with BIC, complete linkage and each other, in other words, the stability between these methods are satisfied, whereas this kind of stability is not existed between the other methods. In conclusion, only the clusters obtained from two-step with BIC, complete, Ward's linkage and K-means with three clusters are stable.

The population stability should also be checked by drawing sub-samples randomly; also the rand index can be used for evaluating stability. The number of drawn cases was determined as the nine, ten percent of the all data set, and a new data set which contains seventy five firms is formed. The analysis rerun for this data set and the rand indexes within population shown in Table 6-11 are calculated. According to the Table 6-11, two-step with BIC, Ward's linkage and K-means with three clusters show the highest stability within population, whereas centroid linkage and average linkage within groups show the worst stability within the population. In conclusion, the clusters obtained from two-step with BIC, Ward's linkage and K-means are determined as stable within population.

Table 6-11: Rand Indexes of population

The cluster method	Rand indexes
Two-step with BIC	1.00
Average Linkage (between groups)	0.94
Average Linkage (within group)	0.71
Single linkage	0.91
Complete linkage	0.88
Centroid linkage	0.70
Median linkage	0.91
Ward's method	0.98
K means with three clusters	1.00

Since the outputs of two-step cluster analysis varied with the changing order of the data, this analysis should be examined against the order of the data sets; therefore, the analysis is rerun for three more times with changing order for observing the changes occurred in the outputs. However, no difference or minor difference between the outputs of the analyses is observed.

6.3.4.2. The size of the clusters in cluster methods

Bacher (2002) advises to use the size of the clusters as an additional criterion in the cluster analysis and mentioned that the number of clusters should be as small as possible, also the solutions which contain clusters having small number of data points should be eliminated. Therefore, the Table 6-12 is prepared to see the number of data points in each cluster for three clusters solutions of all traditional cluster analysis methods. According to the Table 6-12, average linkage between groups method yields a cluster with two members, single linkage method yields two clusters with one member, centroid linkage method yields one cluster with two members, and the solution of median linkage method suggests one cluster with one member and one cluster with two members. The other methods yield clusters containing at least ten data points which is fair enough.

Table 6-12: The number of the data points in each cluster

	Cluster 1	Cluster 2	Cluster 3
two-step with BIC	24	39	21
Average Linkage (between groups)	17	65	2
Average Linkage (within group)	38	36	10
Single linkage	1	82	1
Complete linkage	27	46	11
Centroid linkage	2	67	15
Median linkage	2	81	1
Ward's method	20	51	13
K means with three clusters	34	36	14

According to the Rand indexes between traditional clustering methods, within population, the size of the clusters obtained from different traditional clustering methods, the solutions obtained from two-step with BIC, average linkage within groups, complete linkage, Ward's linkage and K-means methods are determined as more suitable for this data set, however according to the Table 6-11, the rand index of the average linkage within group is determined as one of the worst. Also the stability of this clustering method is not enough; therefore the solution obtained from this method is also eliminated. Finally some further analysis shall be carried out with the other four traditional cluster methods, namely two-step with BIC, complete linkage, Ward's linkage and K-means method.

6.3.4.3. External validity

Another criterion of validating the clusters obtained from different clustering method is the external validity. According to Bacher (2002), the clusters should correlate with external variables that are known to be correlated with the classification and that are not used for clustering. In addition, Thomas and Venkatraman (1988) advised to link the strategic groups to an external (criterion) variable, and two possibilities, namely performance measures and future strategic behavior, are suggested. Consequently, the external criterion is decided as performance of the companies, and for evaluating the external validity of the chosen traditional methods, ANOVA analysis is applied to these methods. The results obtained from ANOVA analysis shown in Table 6-13.

Table 6-13: ANOVA tables

Clustering Methods		Sum of Squares			F	Sig.
		Between Groups	Within Groups	Total		
Two-step Cluster Method		66.518	30.292	96.810	88.934	0.000
K-means Cluster Method		70.374	26.435	96.810	107.817	0.000
Hierarchical	Complete Linkage Method	63.148	33.661	96.810	75.978	0.000
	Ward Method	68.001	28.808	96.810	96.600	0.000

According to ANOVA results, it can be concluded that performance differences exist between clusters in all of the methods. By comparing the F values, it can be said that the clusters obtained by using K-means method satisfies the most heterogeneity between the clusters and homogenous within the clusters. Also, K-means method shows the highest stability within population. In conclusion, the solution of K-means method is decided as the most suitable solution for this data set.

6.3.5. Using SOM for strategic grouping

The tool used for SOM analysis is the SOM toolbox designed for Matlab. The clustering results of a SOM can vary highly depending on the initial setting of parameters such as map size, shape of the output neurons, training algorithm, shape of the map, the decrease speed of neighborhood and learning rate. The optimal map size is determined by the software automatically as 12x4. The neighborhood radius determines to what extent the surrounding neurons are modified according to the input data (Lansiluoto 2004). The neighborhood width function is initiated first with a large value, equal to half of the size of the network itself as suggested by Kohonen (2001) in order to minimize the effect of the initial random weights assigned to nodes. According to this, the neighborhood radius is determined as the half of the size of the network and decreased to unity with time according to different neighborhood functions. All neighborhood functions are tested for training, but they did not show any effect on the performance of analysis. The learning rate influences how the neighbor nodes of the winning node are adjusted after each training step (Lansiluoto 2004) and the purpose of the usage of learning rate is the convergence of the algorithm (Wang 2001). For large

maps, the selection of an optimal learning rate may be crucial (Kohonen 2001). In this analysis, since the provided map size is small, the determination of the learning rate is left to the software. The alternatives of remaining three parameters; shape of lattice, shape of the map and training algorithm, are changed at a time for determining the optimal solution. The quality of the results is evaluated according to three measures; quantization error, topographic error and average distortion measure, offered by the toolbox. As it can be seen in Table 6-14, the optimal solution is provided by using hexagonal shaped lattice, sheet shaped map and batch training algorithm.

Table 6-14: Quality measures of SOM

Shape of lattice	Shape of the map	Training Algorithm	Quantization Error	Topographic Error	Average Distortion Measure
Hexa	Sheet	Batch	1.558	0.012	16.852
Rectangular	Sheet	Batch	1.565	0.012	16.869
Hexa	Cyl	Batch	1.629	0.000	25.344
Rectangular	Cyl	Batch	1.661	0.024	24.800
Hexa	Toroid	Batch	1.643	0.012	26.039
Rectangular	Toroid	Batch	1.666	0.048	25.082
Hexa	Sheet	Sequential	1.659	0.000	16.684
Rectangular	Sheet	Sequential	1.646	0.012	15.808
Hexa	Cyl	Sequential	1.727	0.000	25.512
Rectangular	Cyl	Sequential	1.730	0.000	25.503
Hexa	Toroid	Sequential	1.747	0.071	27.751
Rectangular	Toroid	Sequential	1.727	0.000	25.030

The number of clusters in the SOM is usually decided by visual inspection of the map. The most widely used methods for visualizing the cluster structure of the SOM are distance matrix techniques, such as U-matrix. U-matrix developed by Ultsch and Siemon (1989) reflects distances between neighbouring map units. High values of the U-matrix indicate a cluster border; uniform areas of low values in U-matrix indicate clusters themselves. Another visualization method is to display the number of hits in each map unit. This visualization can also be used for clustering the SOM by using zero-hit units to indicate cluster borders. These two tools are used in the determination of the number of clusters. Figure 6-15 shows the median distance matrix and hit histogram. The median distance matrix can be described as the summary of the U-matrix. By examining these two figures, three clusters are identified as shown in Figure 6-15. The first cluster is determined by examining the hit histogram, where zero-hit units are observed at the top partition of the map. These zero-hit units separate the

first two rows from the other neurons, which lead to the formation of the first cluster. Since it has been observed that, the neuron in the third row has the biggest height, this neuron is re-examined, and it is found out that this increase originated due to neighboring neuron located at the bottom adjacent part. Due to this reason, this neuron is also considered as a part of the first cluster.

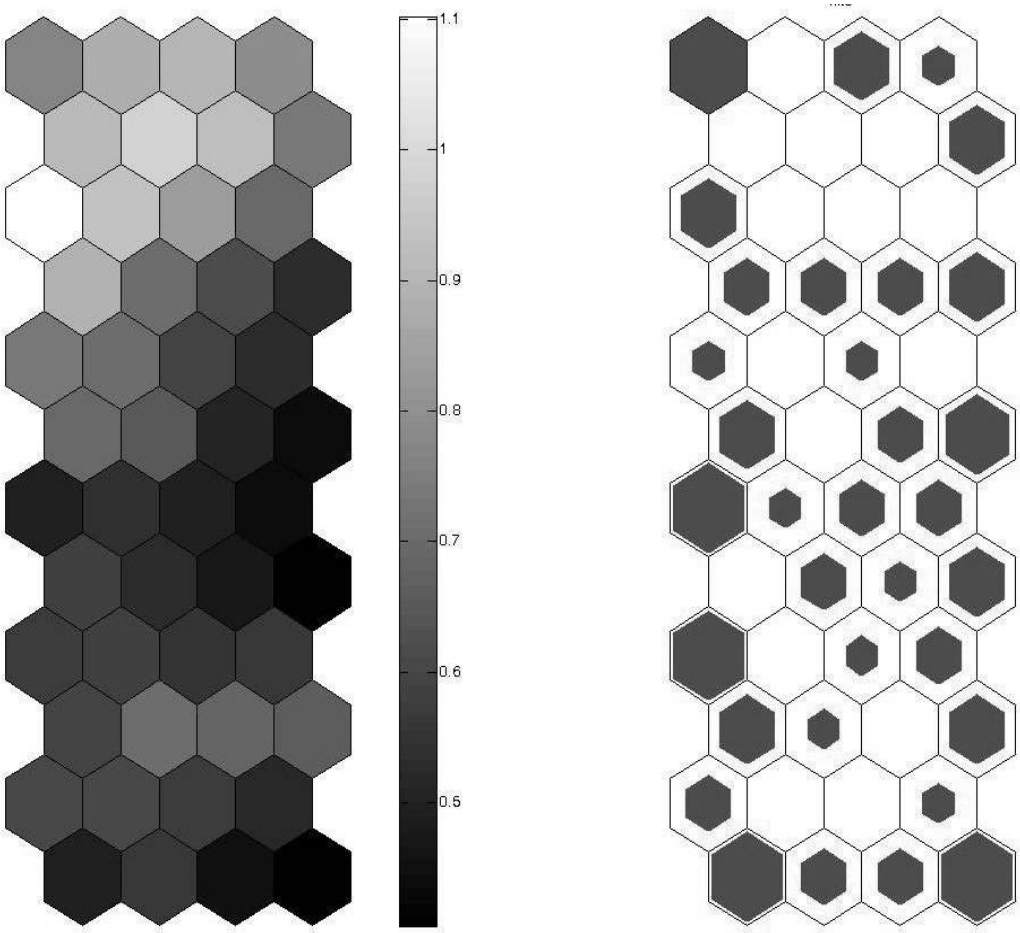


Figure 6-15: The median distance matrix and hit histogram

The zero-hit units can also be observed at the bottom portion of the map. Since, the extent of the boundary at the right hand side cannot be determined clearly, the median distance matrix has to be examined, in order to decide the number of clusters. As the neurons located at the right portion of the third row from the bottom display lighter colors compared to its neighboring neurons, it is decided that the border between the clusters should pass from this location.

In order to check the arbitrariness of the found clusters, the distance matrix is also considered. According to Figure 6-16, the hills can be seen at the boundaries of the clusters, this shows that the clustering relations in the original space really exist, thus, the found cluster is not arbitrary.

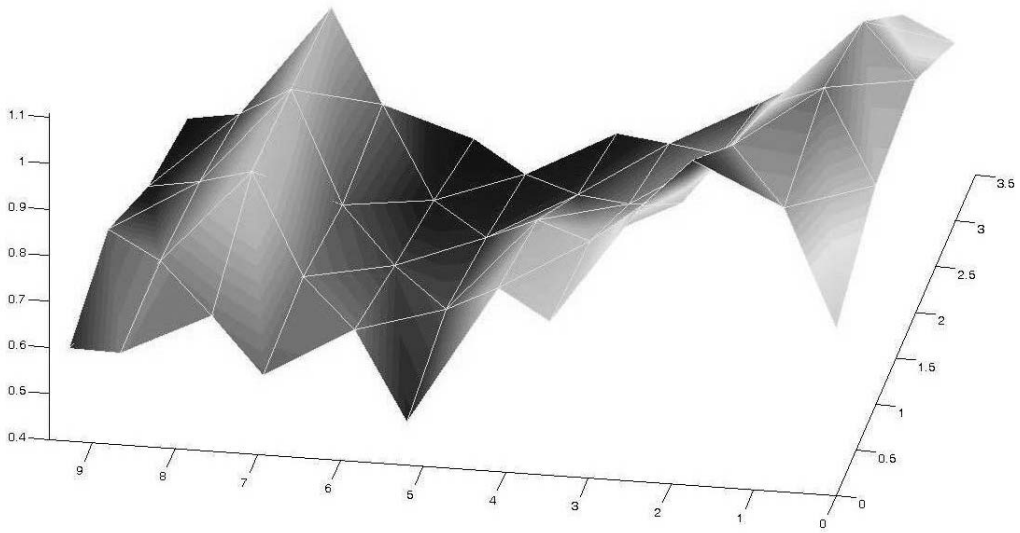


Figure 6-16: Distance matrix

In order to verify the findings, the method developed by Vesanto and Alhoniemi (2000) has been applied to the data set. In this method, clustering is carried out using a two-level approach, where the data is first clustered using SOM, and then, the prototypes of SOM is clustered. In the first level, a large set of prototypes-much larger than the expected number of clusters- is formed by using SOM. In the second level, these prototypes are taken as inputs. The number of clusters is determined by using the Davies-Bouldin index, a function of the ratio of the sum of within-cluster scatter to between-cluster separation. After deciding on the number of clusters, K-means cluster analysis is used and the borders of the clusters are determined.

By using Davies-Bouldin index graph, the number of clusters is determined as three, which are shown in Figure 2. The borders of the clusters as obtained by visual inspection and the borders obtained from this algorithm are approximately the same. The clusters, determined by using the two-level algorithm, are decided to be used in strategic group analysis. For determining existence of the performance difference between clusters, ANOVA has been performed. According to ANOVA results, statistically significant performance differences exist between the clusters (significance level= 1.42 e-23 << 0.01).

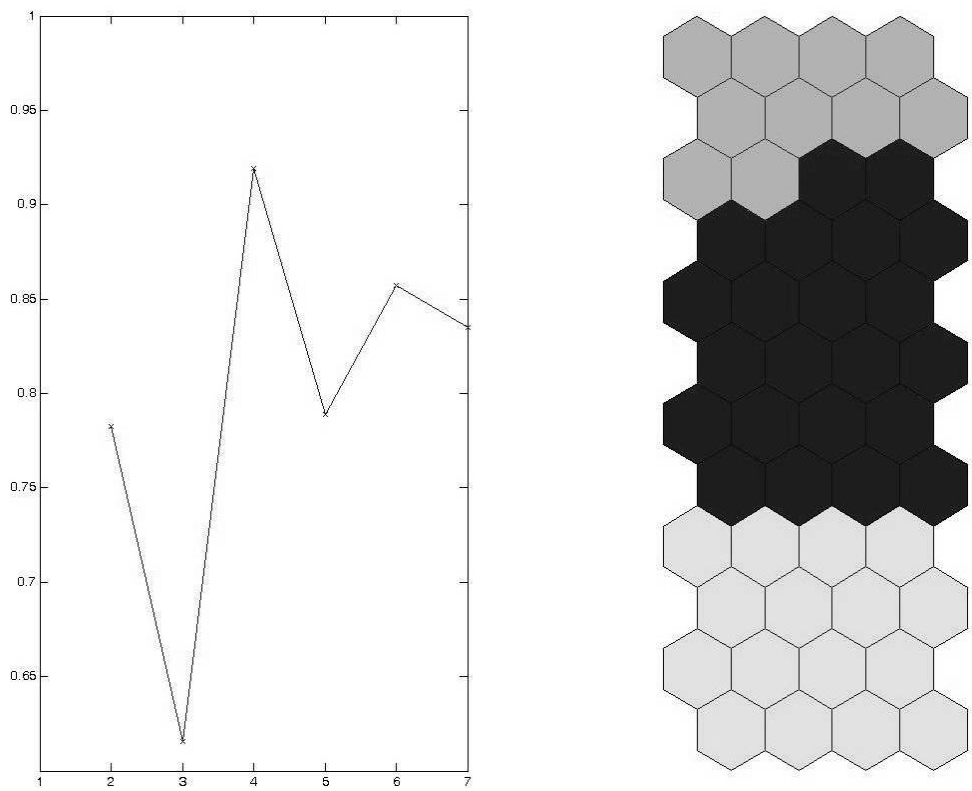


Figure 6-17: Davies-Bouldin index and borders of the clusters

6.3.6. Using FCM for strategic grouping

Clustering Toolbox developed for Matlab is used for the FCM analysis. As mentioned before; the results of FCM are affected from two parameters, number of the clusters and fuzziness index. For determining these parameters, the validity indices were calculated. Bensaid et al. (1996) clustered the validity indexes into three categories. Some methods measure partition validity, some validity measures deals with the evaluation of the properties

of the fuzzy membership, and others consist of validity measures that evaluate both properties of fuzzy memberships and structure of the data. However, none of them is perfect alone. Thus, different validity indices from different groups are used during the decision process. These validity indices are fuzziness performance index (FPI), modified partition entropy (MPE), partition index (SC), separation index (S), Xie and Beni's index (XB), Dunn's index (DI) and alternative Dunn index (ADI). The values of these indices are calculated for m equals to 1.5 and 2, respectively for checking if any difference exists in the general structure of the indices for different fuzziness parameters. The upper boundary of the number of the clusters is determined as $\sqrt{n}$, specifically as 9. The validity indices of the clusters are shown in Table 6-15.

Table 6-15: Validity indexes of FCM cluster analysis

For m=1.5

Validity Indexes	Number of Clusters							
	2	3	4	5	6	7	8	9
FPI	0.334	0.494	0.636	0.689	0.724	0.761	0.772	0.784
MPE	0.937	1.177	1.410	1.490	1.530	1.602	1.608	1.631
SC	0.343	0.256	0.290	0.289	0.256	0.273	0.266	0.280
S	0.004	0.005	0.005	0.005	0.004	0.005	0.005	0.005
XB	1.282	0.975	1.120	1.250	1.553	1.750	1.848	1.789
DI	0.298	0.182	0.197	0.211	0.186	0.263	0.224	0.224
ADI	0.043	0.008	0.000	0.007	0.004	0.001	0.001	0.002

For m=2

Validity Indexes	Number of Clusters							
	2	3	4	5	6	7	8	9
FPI	0.730	0.816	0.869	0.901	0.919	0.928	0.937	0.944
MPE	1.819	1.892	1.968	2.025	2.056	2.061	2.080	2.087
SC	0.539	0.372	0.378	0.409	0.417	0.372	0.374	0.377
S	0.006	0.007	0.007	0.007	0.007	0.006	0.006	0.006
XB	1.113	0.752	0.865	0.953	1.184	1.338	1.494	1.361
DI	0.232	0.189	0.189	0.197	0.224	0.211	0.211	0.186
ADI	0.089	0.009	0.003	0.002	0.002	0.001	0.000	0.016

SC, S, DI and ADI from the first category are used for determining the compactness and separation of the clusters. For a better solution, the clusters should be well compacted and

separated from each other. SC is the ratio of the sum of compactness and separation of the clusters, for that reason the lower value of SC indicates a better solution. Dunn's index is a ratio of within cluster and between cluster separations. According to these two indexes, the 3-clusters solution shows the best performance. FPI and MPE is involved into the second category of the validity indexes, namely they are used for evaluating the fuzziness of the solutions. The good clusters should not be very fuzzy and lower FPI and MPE show the better solutions. According to these measures, 2-clusters solution is optimum solution for this data set, and three-cluster solution is the second best solution. XB evaluates both properties of fuzzy membership and structure of the data. The smaller values of XB indicate a better clustering of the data set than larger values. According to the Table 6-15, the 3-cluster solution has the lowest XB value. 3-clusters solution is selected as the optimum solution for this data set.

For finding the optimum solution, the fuziness parameter (m) should also be determined, Odeh et al. (1992) advised that the best value of m for a given class is at maximum of the curve when plotting -[dj/dphi*sqrt(class)] vs m. Dj/dphi is the derivative of objective function value over fuzzy exponent. Dj/dphi values are calculated by using FuzMe 3.0 developed by Minasny and McBratney (2002). The Figure 6-18 is drawn for the number of the cluster equals to three, and according to this figure, the optimum solution is obtained when m equals to 1.7.

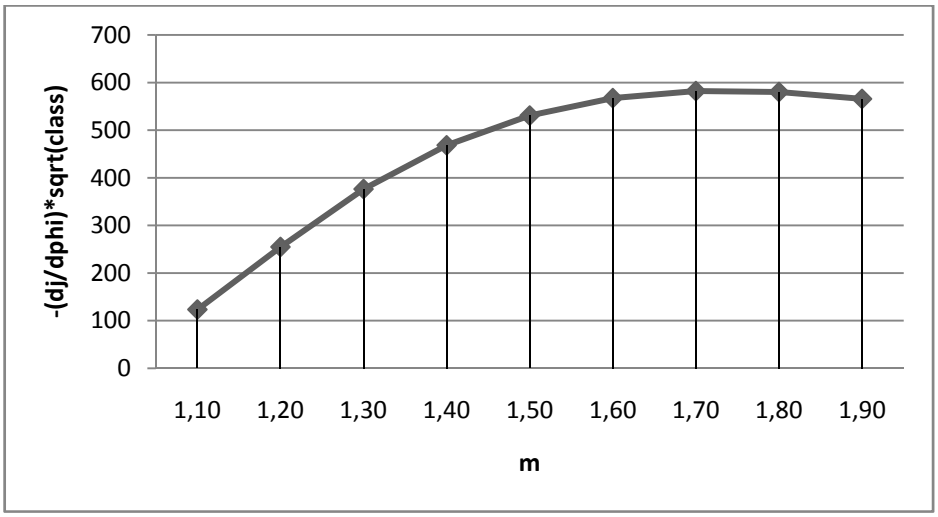


Figure 6-18: The graph of –(dj/dphi)*sqrt(class) vs m.

6.4. Discussion of the findings

In this study, three cluster analysis methods are applied to the data set separately. According to the different validation techniques applied to the methods, “three clusters solution” is determined as the most suitable solution by each cluster analysis. In the traditional cluster analysis, the degree of membership of the data points to a cluster can be determined by using the outputs; namely the distance between the centre of the clusters and the data points belonging to this cluster, however limited information about the typology of the data set can be provided. In SOM, the membership of the data points are not obtained directly, but the boundary of the clusters can be obtained by visual inspection of the U-matrix and zero units, and according these outputs, the membership for the data points can be identified and whether the clusters are overlapped or hard can be determined. Nevertheless the typology of the data set is obtained at the end of the analysis. The Figure 6-19 and Figure 6-16 shows the typology of the data set obtained by SOM analysis.

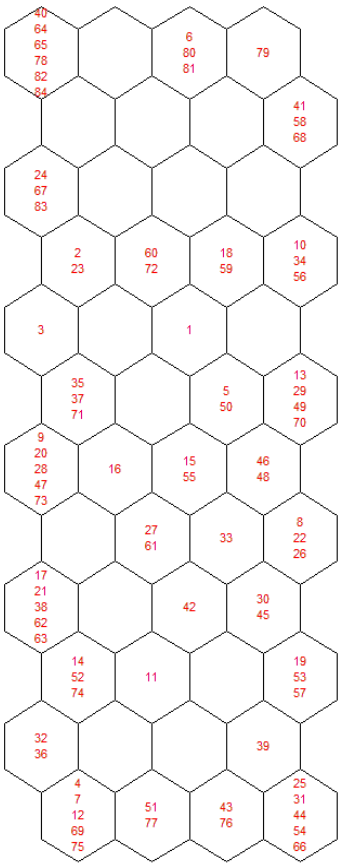


Figure 6-19: Typology of the data set

The FCM shows the degree of the memberships of the data points to all clusters. Table 6-16 shows the degree of the memberships of the all data points used in this analysis. By considering these degrees of membership, information about the typology of the data set can be obtained separately from the analysis. Although the information about the typology of the data points are obtained at the end of both SOM and FCM analyses, due to the visualization ability of SOM, the typology of the data points can be understood very easily as a result of this method.

Table 6-16: Membership of the data

ID	1st Cluster	2nd Cluster	3rd Cluster	ID	1st Cluster	2nd Cluster	3rd Cluster	ID	1st Cluster	2nd Cluster	3rd Cluster
1	0.118	0.719	0.164	29	0.029	0.800	0.172	57	0.033	0.385	0.582
2	0.334	0.465	0.201	30	0.031	0.404	0.565	58	0.419	0.396	0.185
3	0.304	0.472	0.225	31	0.030	0.240	0.730	59	0.150	0.696	0.154
4	0.015	0.132	0.853	32	0.028	0.209	0.763	60	0.197	0.639	0.164
5	0.044	0.819	0.137	33	0.033	0.562	0.405	61	0.113	0.565	0.322
6	0.824	0.130	0.047	34	0.117	0.671	0.212	62	0.020	0.359	0.621
7	0.021	0.144	0.836	35	0.121	0.575	0.304	63	0.044	0.532	0.425
8	0.026	0.710	0.264	36	0.014	0.148	0.838	64	0.827	0.118	0.055
9	0.077	0.469	0.455	37	0.062	0.664	0.274	65	0.888	0.075	0.037
10	0.203	0.593	0.204	38	0.037	0.427	0.535	66	0.018	0.177	0.805
11	0.012	0.159	0.829	39	0.014	0.143	0.843	67	0.706	0.210	0.084
12	0.023	0.155	0.822	40	0.850	0.102	0.049	68	0.508	0.375	0.117
13	0.069	0.737	0.194	41	0.402	0.445	0.153	69	0.017	0.136	0.846
14	0.027	0.287	0.686	42	0.034	0.451	0.515	70	0.139	0.517	0.344
15	0.062	0.687	0.252	43	0.015	0.133	0.852	71	0.099	0.667	0.235
16	0.028	0.559	0.414	44	0.017	0.181	0.803	72	0.240	0.648	0.112
17	0.080	0.489	0.432	45	0.035	0.438	0.527	73	0.048	0.515	0.438
18	0.191	0.643	0.166	46	0.038	0.705	0.257	74	0.089	0.406	0.505
19	0.013	0.210	0.777	47	0.120	0.507	0.373	75	0.025	0.151	0.824
20	0.092	0.536	0.372	48	0.026	0.767	0.206	76	0.022	0.149	0.828
21	0.044	0.369	0.587	49	0.090	0.658	0.252	77	0.024	0.177	0.800
22	0.042	0.577	0.381	50	0.059	0.666	0.275	78	0.832	0.116	0.052
23	0.306	0.494	0.200	51	0.022	0.164	0.815	79	0.714	0.209	0.077
24	0.589	0.285	0.126	52	0.035	0.303	0.662	80	0.741	0.191	0.069
25	0.019	0.168	0.813	53	0.027	0.329	0.644	81	0.815	0.132	0.053
26	0.035	0.624	0.341	54	0.021	0.189	0.790	82	0.847	0.102	0.051
27	0.042	0.629	0.329	55	0.082	0.675	0.243	83	0.805	0.143	0.052
28	0.088	0.582	0.330	56	0.097	0.735	0.169	84	0.854	0.098	0.048

In this study, according to the largeness of the membership degrees obtained, the data points are assigned to the clusters hardly. The purpose of this assignment is satisfying the

consistency between FCM, SOM and traditional cluster analysis methods in which the data points are assigned hardly. Thus, the rand indices between these techniques can be calculated. The rand indices between traditional cluster analysis and SOM, traditional cluster analysis and FCM and SOM and FCM are determined as 0.95, 0.94 and 0.96 respectively. This shows that the convergence between these three methods is achieved in this study. When the data set is scanned one by one, only four data points are determined to be clustered differently by SOM and traditional cluster analysis. Two of them are assigned to the first cluster by SOM, whereas these are assigned to the second cluster by traditional cluster analysis. Similarly, two of them that are assigned to the second cluster by SOM are assigned to the third cluster by the traditional cluster analysis. When the clusters obtained by SOM and FCM are considered, there are differences only in three data points. Two of these data points are assigned to the third cluster by SOM, whereas these are assigned to the second cluster by FCM, and one data point is differently assigned between the first and second cluster by these methods. Lastly, five data points are determined that assigned differently by traditional cluster analysis and FCM. Four of them are assigned to the second cluster by FCM, whereas these are assigned to the third cluster by traditional cluster analysis. Besides, one data point is differently assigned between the first and second cluster by these methods. Consequently, the difference is only between the adjacent clusters, not between the first and third clusters and vice versa, meaning that the discrepancy between findings is low. In addition, the data points assigned to different clusters by different methods are placed at the boundary of the clusters according to the typology of the data, and the height of the boundary where these points are placed is lower than the other parts of the boundaries, consequently it can be concluded that the clusters are overlapped.

Reger and Huff (1993) mentioned that the strategic groups should be conceptualized by degree of group membership and overlapping between strategic groups is necessary in order to map complex industries more realistically. In addition, according to Reger and Huff (1993) and McNamara et al. (2003), three types of firms characterize the industry: 1) core, 2) secondary, 3) solitary firms. Core firms are tightly adhere to the strategic groups which they belong, secondary firms are aligned loosely with the strategic group and adopted the specific strategies of a strategic group less consistently than core firms, and solitary firms form strategic groups on their owns. This study also indicates that the degree of the membership of the firms varies within a strategic group, in other words, all of the firms placed in a strategic group does not stick to the specific strategic posture of this strategic group equally.

The U-matrix and membership degrees found as a result of FCM can be used for identifying the core and secondary firms of these strategic groups.

In this study, it is concluded that the strategic groups in the Turkish construction industry tend to overlap. Some of the firms placed at the boundary of the strategic groups combine specific strategic postures of more than one strategic group. DeSarbo and Grewal (2008) developed the "hybrid strategic group" concept in order to explain the strategic posture of this kind of firms. They illustrated the competitive market structure by integrating hybrid strategic group concept in Figure 6-20. According to this figure, the overlapping between the pure strategic groups forms the hybrid strategic groups.

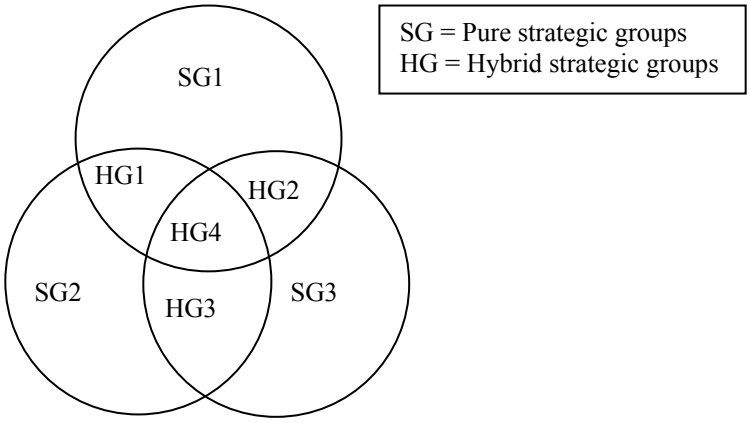


Figure 6-20: Pure and hybrid strategic groups (DeSarbo and Grewal 2008)

Consequently, by comparing and examining the clusters obtained by using SOM and FCM, the strategic group structure of the Turkish construction companies is illustrated in Figure 6-21 by combining the core, secondary firms concept and hybrid strategic groups.

This classification reveals no hybrid strategic group firms that combine strategic postures of all strategic groups. The six hybrid strategic group firms belong to strategic group 1 and 2, the eight hybrid strategic group firms belong to strategic group 2 and 3. However, no hybrid strategic group between strategic group 1 and 3 is identified, the reason could be explained by the argument that the strategic posture specific to these strategic groups are too distinct that no common features exist between these strategic groups. The thirteen pure strategic group firms belong to strategic group 1. The thirty-three pure strategic group firms belong to strategic group 2, but twenty-one of these firms from the secondary firms with in this group.

The twenty-four pure strategic group firms belong to strategic group 3, but seven of these firms from the secondary firms with in this group.

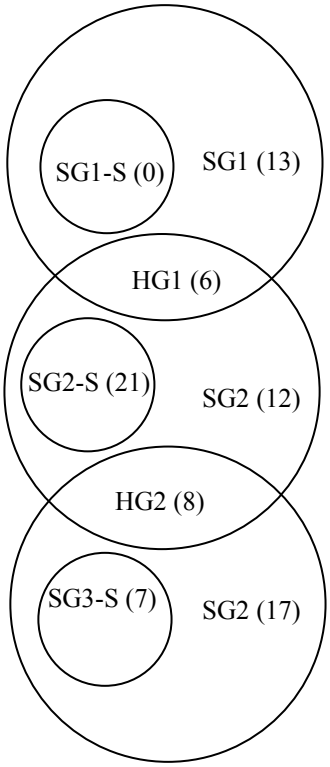


Figure 6-21: The strategic group structure of Turkish construction companies

In conclusion, the strategic group is a complex structure which is characterized by pure strategic groups consisted of core, secondary firms and hybrid strategic groups; therefore the traditional cluster analyses which form strict clusters and provide no information about the typology of the data, is not sufficient to understand this complex structure. DeSarbo and Grewal (2008) devised a modified version of NORMCLUS in order to identify traditional and/or hybrid strategic groups. However, this method is not capable of identifying the core and secondary firms in a pure strategic group whereas; SOM and FCM can be used for identifying the core and secondary firms and also hybrid strategic groups. Therefore, it can be concluded that they are competent to reveal this complex structure of an industry. It is believed that these two methods are appropriate alternatives in order to conduct strategic group analysis and using traditional clustering method only may lead to over-simplistic results.

6.5. Discussion of research findings

The variables that show highest contribution to the separation of the clusters show a rapid change on the borders of the clusters. Upon examination of Figure 6-22, the mobility barriers between the first and second strategic groups are determined as the strategic planning style and experience. For the second and third strategic groups, financial resources and managerial capability are found to be the most important mobility barriers. In addition, experience is determined as an important variable that affects the separation of second and third clusters. The contributions of factors such as internationalization ratio, type of projects and clients are found to be insignificant, since there is no uniform color distribution between the strategic groups. One can argue that strategy context is the major determinant of performance difference in different strategic groups. Groups specific findings are as follows:

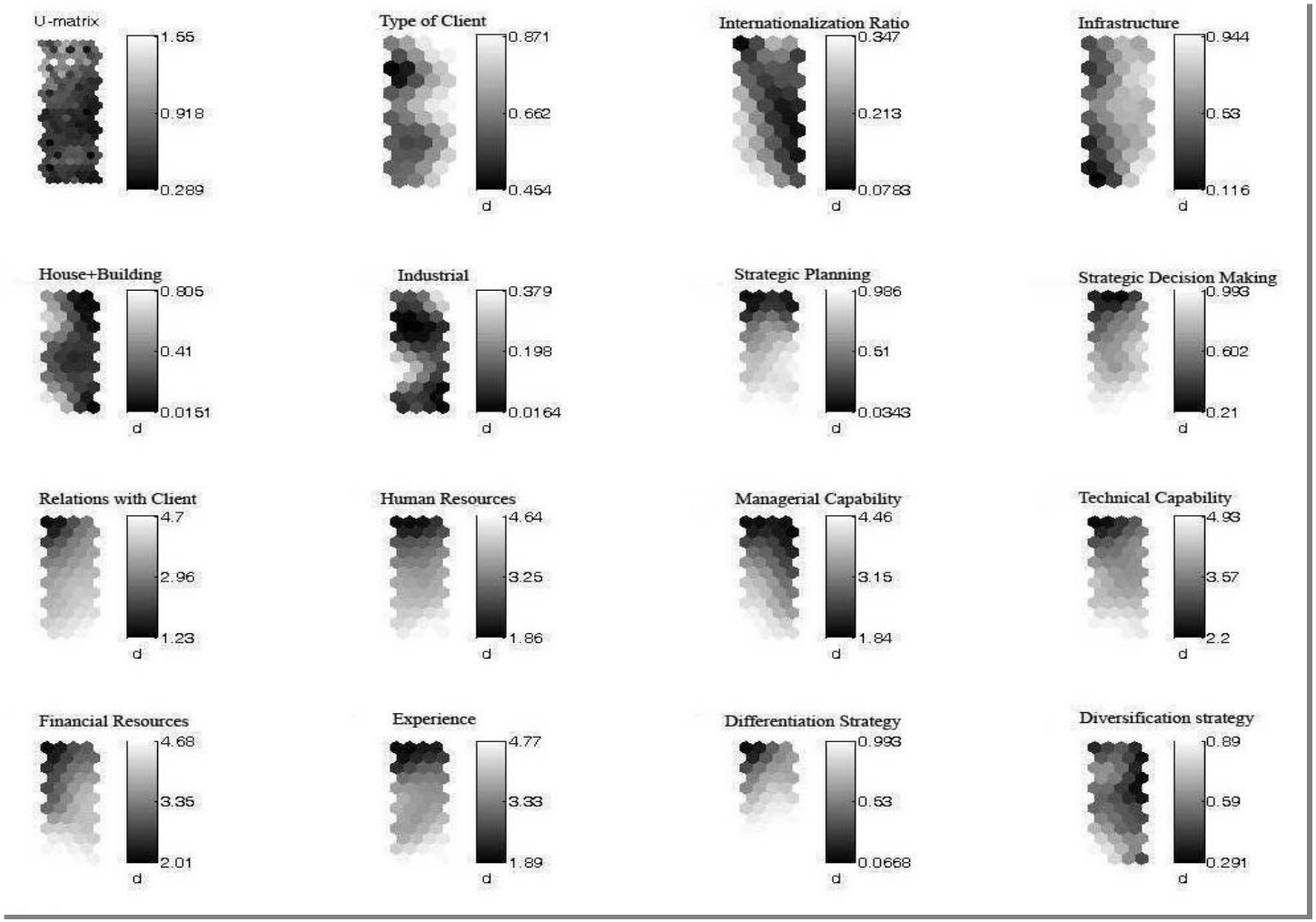


Figure 6-22: U- matrix and variables

Strategic Group 1: In this group, there exist 16 companies which have performance ratings in the range of 1-to-2 (on a scale of 1-5); only one company showing third level performance. This group comprises of comparatively small firms that mostly utilize a price differentiation and focus strategy. The basic difference from the other groups is that majority of firms in this group do not have a systematic and democratic strategy process. Also, they have weaknesses in terms of all sources of competitive advantage. They are especially weak in terms of client relations and experience.

Strategic Group 2: This group comprises of 36 companies. All of the strategy context variables have values in the range of 3-to-4, showing that their competitiveness is higher in all dimensions when compared with the firms in Strategic Group 1. Also, the percentage of firms having systematic strategic planning and democratic decision-making process is significantly higher. It may be argued that although majority of the firms in this group have non-price differentiation strategy, the level of resources and competencies that should backup this strategy are not very high. The performance of firms in this group is moderate and significantly higher than the performances in Strategic Group 1. The companies placed at the right side of the map (Figure 6-19) outperform the other companies in that group. The reason of this performance difference may originate from strong financial resources and democratic decision-making process.

Strategic Group 3: The number of firms in this group is 32. Majority of the firms in this group have a systematic strategic planning process and strategies are formulated in a democratic decision making environment. When compared to other clusters, the level of experience and client relations is significantly higher. The values of other strategic variables regarding the strategy context, which are financial resources, technical capability, human resources and managerial capability are also very high. All of the firms in this group utilize a non-price strategy, meaning that they differentiate themselves from others in terms of quality, innovation etc. Similarly, majority of the firms are diversified into sectors unrelated to construction. The performance ratings are in the range of 4-to-5 where, only one company showing average performance (3 out of 5) places itself on the border of second and third strategic groups. This high performance may result from the exceptional level of resources and capabilities which give them the opportunity to differentiate their services from others.

The observed relationship between group membership and firm performance can be taken as evidence of the existence of mobility barriers and the variables that contribute to the performance differences significantly can be considered as mobility barriers (McGee and

Thomas 1986). However, it is clear that the reasons of performance differences may be market-related or firm-specific factors as well as mobility barriers. If it is assumed that the mobility barrier is the major reason of performance differences between firms, "experience" constitutes the major mobility barrier in the Turkish construction industry. This can be considered as a reasonable finding because in many projects, the amount of similar works completed is a prequalification criterion.

6.6. Limitations of the study about performance implications of strategic group membership

There are certain shortcomings of this research. First, it cannot be argued that strategic group membership is the primary determinant of firm performance. There may be significant differences within a group due to firm level advantages such as existence of unique resources. Strategic group heterogeneity, as discussed by McNamara et al. (2003) is ignored in strategic group analysis. Although, some generalizations are made about strategic groups, it may not hold true for every firm in the same group. Second, the relation between strategic group membership and performance is studied at a single point in time. It is hard to make predictions about performance of groups in the future as strategic groups are dynamic. Also, group performance may change as a result of a change in the competitive environment. Third shortcoming is neglecting the role of macro environment in strategic group analysis. Researchers that support environmental determinism describe environment as the primary mechanism for explaining the performance of an organization. In strategic group analysis, it is hypothesized that environmental forces have the same impact on all firms and its implications are reflected in the performance ratings in the same manner. Also, the research has got some limitations due the measurement method used. Only single items are used to measure variables rather than constructs composed of a series of questions. Moreover, performance rating is a subjective rating that reflects the personal judgment of only one respondent.

Finally, since the data reflects the experiences of Turkish companies and the markets they operated in, the model may not be applicable to all construction companies, in other words the results of this study are specific for Turkish construction industry, thus cannot be generalized. However, the conceptual framework used in this study may further be applied to other countries to investigate the existence of strategic groups and possible performance differences between them.

CHAPTER 7

DIFFERENTIATION IN THE CONSTRUCTION INDUSTRY

According to the findings of the strategic group analysis, all of the companies in the strategic group that shows the highest performance among all strategic groups are found to follow differentiation (indicated as non-price differentiation in the conceptual framework depicted in Chapter 6) strategies rather than cost leadership (named as price differentiation in the conceptual framework). In the remaining parts of this thesis, "differentiation" is used instead of "non-price differentiation" having same meaning. As a result of strategic group analysis, it is concluded that differentiation provides a significant source of competitive advantage for the Turkish construction companies. Although the results demonstrate the success of differentiating companies, it is not clear how and in which parts of their value chain they differentiate their products/services. Thus, the content of differentiation is rather vague. Moreover, there are only a limited number of studies in the construction management literature that investigate the ways of differentiation. There are research studies about the drivers of differentiation that aim to develop generic conceptual frameworks. However, according to Porter (1985), the drivers of differentiation vary for each activity and may vary across industries for the same activity. Consequently, the current research was extended by establishing a conceptual model of differentiation and structural equation modeling was utilized to reveal the interrelations between drivers and modes of differentiation in the Turkish construction industry.

At this part of the study, firstly, a literature survey regarding the definition of differentiation and sustainability of competitive advantage based on differentiation strategies will be presented. Furthermore, a conceptual model developed for modeling the differentiation process within the Turkish construction industry will be explained. The structure and aims of the questionnaire study will be discussed and finally, survey findings will be presented as well as results of the statistical analysis.

7.1. Introduction:

According to Porter's generic competitive strategies' typology (1980; 1985) which has received the attention of many researchers and applied to various industries, the companies have basically two competitive strategies, namely, cost leadership and differentiation. Differentiation is defined as a position in which a competitor offers a product or service which has some unique characteristics for the customers. This uniqueness is perceived as valuable by the customers, so the companies can charge a premium price in the market that exceeds the required cost for providing the differentiation (Calori and Ardisson 1988). Many researches such as Porter (1985), Henderson (1981), Grant (1995) and Greenstein and Mazzeo (2006) mention the superiority of differentiation strategy for achieving and sustaining competitive advantage when compared to cost leadership. In fact, according to Mintzberg (1988), the only way for achieving higher performance can be established by differentiating from the rivals. The reason of this can be explained by the possibility to avoid price-based competition by means of differentiation. For a company that competes on lowest cost basis, there is only one possibility to maintain its position that is reducing the prices. However, for another firm that chooses the differentiation strategy, there are many ways to differentiate and maintain its competitive position (O'Brien and Al-Soufi 1993). Intensive competition may exist between the companies following cost leadership strategy, leading to a decrease in the profit margins of these companies whereas by following differentiation strategy, the companies can stay away from this kind of intensive competitive environment. Evans and Berman (1997) call this strategy as non-price based strategy. The price based and non-price based approach are compared in Figure 7-1.

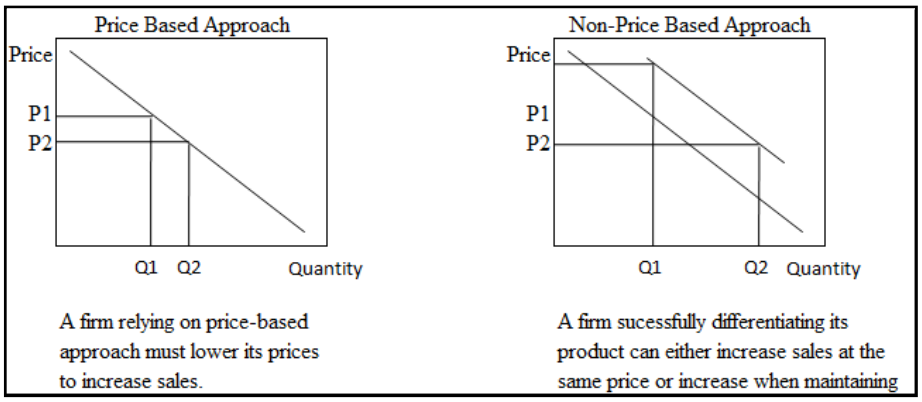


Figure 7-1: Price based vs non-price based strategies (Evans and Berman 1997)

According to Figure 7-1, an important outcome of differentiation is a reduction in consumer cross-price elasticity of demand, that is, the change in demand for one service relative to a change in the price of a competing service. The companies may offer more differentiated services for eliminating the elasticity of cross prices, by that way; the customers may be less affected from price increases. Therefore, the price/demand relationship can be modified by differentiating services/product leading to an increase in profitability (Fisher 1991). Finally, with the differentiation, the loyalty of the buyer to the company is strengthened, in other words, the customers of the differentiated companies may be more reluctant to switch their brands when compared with the costumers of the companies competing on cost leadership. Moreover, due to customer loyalty and brand identification, differentiation creates entry barriers for the new entrants into highly differentiated markets. Newcomers must spend large amounts on advertising and promotion to increase the brand awareness (Grant 1995). In addition to this, by differentiating their services/products, the companies can create mobility barriers, and reduce the power of buyers who feel they lack acceptable substitute products (Porter 1980).

Miller (1986) noted that there are at least two types of differentiation: innovative differentiation, and marketing and image differentiation. According to the first differentiation strategy, the companies try to create new products by means of higher quality, efficiency, design innovations, or new style. These companies should pay attention on research and development, and pioneering. Differentiation via product innovation often involves organizational renewal, reconfiguration of organizational resources and new technologies (Danneels 2002; Eisenhardt and Martin 2000; Miles and Snow 1978; Miller et al. 1984). The second strategy attempts to create a unique image for a product through marketing practices by offering attractive packages, good service, convenient locations and good product/service reliability. Therefore, these companies should spend large sums on advertising, sales force, and promotion for providing unique and positive messages, and unique communication activities for insulating themselves from future price competition (Boulding et al. 1994). However, Porter (1985) criticizes this perspective due to its limitation about the sources of differentiation; therefore he proposes that the potential of differentiation can arise anywhere in the value chain. Value chain is defined as a structured way of analyzing a business's constituents and its links to outside organizations by dividing the enterprise's activities into technologically and economically distinct value providers (Betts and Ofori 1994). The value chain consists of primary activities, such as inbound logistics, operations, outbound logistics, marketing and sales, and service; as well as support activities

such as firm infrastructure, human resource management, technology development, and procurement activities. Porter's view on differentiation covers all activities on the value chain as potential sources of innovation and differentiation.

Mathur (1988) attempts a new classification of generic strategies for adding value through effective differentiation. He introduces a distinction between merchandise and support differentiation. According to him, merchandise means a tangible or intangible offering which is being made available to the consumer. Support is provided along with the merchandise by advice, training or assistance. Either merchandise, support or both elements of the offering may be differentiated. Merchandise may be differentiated by content, image, or both. Similarly, support may be differentiated using expertise, personalization or both. MacMillan and McGrath (1997) also proposed a different approach called as consumption chain for discovering new ways for differentiation. They advised the companies should utilize their customer's entire experience with a product or service for improving their creative thinking, by that way, new points of differentiation can be continuously identified and successful differentiation strategies which have never been thought before can be developed. This approach is formed of two parts, namely "mapping the consumption chain" which captures the total experience of the customers about a product or a service and "analyzing your customer's experience" which shows managers how the most ordinary product or service can be differentiated in numerous ways by directed brainstorming about each step in the consumption chain.

Mintzberg and Quinn (1998) considered that the firms can differentiate themselves in six basic ways, namely price differentiation, image differentiation, support differentiation, quality differentiation, design differentiation, and un-differentiation strategy. The most important difference between the Porter (1985)'s and Mintzberg and Quinn (1998)'s approaches is about price differentiation. Porter (1980) argued that activities required for creating a differentiation situation is costly, in other words a firm who tries to create uniqueness should spend more money than the competitors for being better than them and he considers "cost leadership" as a separate category from differentiation. Dickson and Ginter (1987) believe that the price should be seen as a differentiation way for the firms and they define differentiation as the act of offering a product/service that is perceived to differ from its competitors' services/products on any physical and nonphysical characteristics including price. On the other hand, Warszawski (1996) excludes cost from differentiation possibilities, and he mentions about four possibilities of differentiation in project-based industries (such as construction), namely higher standard of product, higher quality of product, faster project

completion and more extensive service to clients. The disagreement between these two approaches could be explained by different perceptions of various authors' about the definition of differentiation. According to Porter (1985), differentiation can be established only by creating a uniqueness valued by the customers (this uniqueness cannot be the price of the product and service), whereas offering a product/service different than the competitors is seen enough for differentiation by Mintzberg and Quinn (1998), therefore offering a standard product with lower price is seen as a differentiation method.

Grant (1995) mentioned two ways to create differentiation; on the supply side, the firm should examine the activities that the firm performs and the resources that firm has, for figuring out the resources and capabilities to create uniqueness; on the demand side, by analyzing the customer demand, the companies can determine the potential for differentiation in a market, the willingness of customers to pay for differentiation, and the most promising positioning for the firms in terms of differentiation in relation to its competitors. However, achieving competitive advantage by differentiation is not enough for gaining benefits from this position continuously due to the competitive imitation. Wherever competitive success is observed, it will be imitated by competitors (Alchian 1950). This leads to a reduction in a firm's competitive advantage obtained by differentiation (Barney 1991; Dierickx and Cool 1989; Reed and DeFillippi 1990); therefore the companies should invest into strategies that resist competitive imitation for sustaining their returns from differentiation. According to Porter (1985), for the sustainability of a differentiation strategy, differentiation should depend on its sources and stem from multiple sources, rather than resting on a single factor such as product design. In addition, the companies should create switching costs, fixed costs incurred by the buyer when it changes supplier, during its differentiation process. Johnson and Scholes (2002) mentioned that the provided competitive advantage from differentiation can be sustained under two conditions, namely difficulties of imitation, and imperfect mobility. Due to the complexity of the competences which cannot be comprehended by the competitors, causal ambiguity of the competences that the competitors cannot understand the cause and effect relationship, and culturally embedment of the competencies in the organization, the imitation of differentiation may become difficult for the competitors. Imperfect mobility is the capabilities, resources and competences of an organization that could not be traded, like intangible assets such as brand image and reputation and co-specialization. Fisher (1991) advises that differentiation strategies developed by the companies should build on "isolating mechanisms". Isolating mechanisms are features of resources, like skills, knowledge, and capabilities that are tacit, unique,

invisible, complex, or path dependent that prevent other firms from obtaining and replicating them (Oliver 1997). Therefore the differentiation strategies built on "isolating mechanisms" provide long-term success by restricting competitive imitation. He classified the isolating mechanisms into two main groups, namely competitive isolating mechanisms, related to characteristics of the differentiating firm and its competitors; and customer-based isolating mechanisms, economic and physiological barriers related to consumers' brand selection decisions. Competitive isolating mechanism consists of information impactedness, response lags, economies of scale, producer learning, and channel crowding, and customer based isolating mechanism consists of buyer switching costs, buyer evaluation costs, and advertising crowding (Fisher 1991).

According to the typology of competitive systems developed by Boston Consulting Group (1981, 1985), there are two dimensions of differentiation: number of possible differentiation sources and size of the potential competitive advantage (Calori and Ardisson 1988). Construction sector can be considered as stalemate type of industry which has a few differentiation sources and a small potential competitive advantage, therefore the construction companies prefer competing on cost basis. Because, the technical regulations prepared by the governments determine the quality for a major part of the production, in addition due to the dominant power of the government as a major buyer, the price that a customer pays is decided according to financial and fiscal regulations (Pries and Janszen 1995). According to the Ball et al.'s (2000) study about medium-sized British construction companies, the contribution of the employed competitive methods to the performance of the companies is limited. Also, construction industry is a mature industry due to the availability of the industry standards and data (Sarshar et al. 1999). However, in the literature, there are many studies which show the positive influence of the differentiation strategy on the performance of the construction companies. For instance, Kale and Arditi (2002) determined considerable empirical support to the positive influence of differentiation on the firm performance for the US construction industry. In the same manner, Cheah et al. (2007) concluded that the differentiation strategies positively affect the firm performance as measured by revenue and profit growth for large local construction firms in China. Similarly, the previous part of the current research about strategic group analysis demonstrated the importance of differentiation for achieving success in the Turkish construction industry. In the forthcoming parts of this research, different sources of differentiation for construction companies will be defined and their contribution to level of differentiation will be investigated.

7.2. The ways that the companies can differentiate in the construction industry

Differentiation can be established, only if differentiation creates value to customers, otherwise the companies cannot ask for higher prices for their products/services that exceed the costs incurred by the firm (Porter 1980), therefore differentiation strategy fails. The companies trying to differentiate should understand what the customers want, how they choose the product/service, and what their motivations are for identifying opportunities of profitable differentiation (Grant 1995). Therefore, in this research, the key buying criteria of the clients are tried to be determined for identifying the ways that the construction companies can differentiate. Common evaluation criteria in the construction management literature include quality, cost and time criteria (Bennett 1983; Ward et al. 1991). Differentiation on cost basis means offering a standard product by charging lower price on it (Mintzberg and Quinn 1998). In the construction industry, intensive price competition usually occurs due to traditional competitive tendering based on only evaluation of cost and the companies try to reduce their costs by offering products under standard norms. In addition, in the construction industry, the awarding process of the contracts is performed before the construction; therefore whether the product will comply with the standards cannot be determined before the completion of the project. Consequently, in this study, productivity was decided as a source of differentiation instead of cost, because the only way to lower costs and duration without lowering quality standards is by means of productivity. Thus, the term productivity is used instead of cost. The term "differentiation" covers only non-price based strategies in the same line of thinking of Porter (1980; 1985).

Kale and Arditi (2003) added the innovation attributes of products/services as a differentiation way for the construction companies. Innovation may result in added value through cost decreases or high quality products. Harttman (2006) mentioned that the innovative ideas lead to the potential of differentiation in construction companies. Also, the importance of marketing and image for differentiation is mentioned by many authors, like Mintzberg and Quinn (1998), Miller (1986), Fombrun (1996). Cheah et al.(2007) mentioned reputation as one of the differentiation ways for Chinese construction companies.

Finally, the ways that a construction company can differentiate are defined in six categories: productivity, time/schedule, service quality, product quality, innovative solutions and positive corporate image. The reasons for this categorization will be discussed below. Moreover, satisfying all of these attributes requires different skills and resources (Porter 1980) which will be explained in the forthcoming sections.

7.2.1. Productivity

Productivity is associated with efficient use of the various factors of production like labor, equipment, and capital. The importance of productivity in construction industry as a competitive advantage is mentioned by many authors, like Arditi (1985), Lowe (1987), Kumaraswamy and Chan (1998). With increasing productivity, the companies can decrease total cost and duration of production while improving the quality (Kazaz and Ulubeyli 2007). According to the HSOM (1962) and DETR (1998), since the early 1960s, the construction industry has been continually criticized for its low productivity and poor quality (Fox et al. 2002). A construction company that can increase its productivity may have a strong competitive position when compared to its competitors due to savings from cost and time. Therefore, differentiation by higher productivity is defined as a possible strategic choice.

7.2.2. Time

Achieving on-schedule performance in construction projects, even delivering the constructed facilities ahead of schedule is one of the most important success factors in the construction industry. Some researchers, like Stalk (1988), Stalk and Hout (1990) mentioned about the benefits of the superior use of time. Laufer and Tucker (1987) point out that cost and time are received relatively more attention than quality. They further indicate that between cost and time, formal planning effort of most construction companies has been primarily focused on time planning, and to a lesser extent on resource allocation and its cash-flow implications. Due to gaining market' advantages of being first over competitors, project owners are increasingly placing greater demands on contractors to complete projects in expected time limits in today's fast-paced construction environment. Owners will most likely suffer the loss of expected profits as a direct result of delays in putting the facility or product into service. Consequently, contractors typically face liquidated damages for finishing late (Kog et al. 1999). However, the delays in construction projects are common. For instance, Morris and Hough (1987) evaluated the records of more than 4,000 projects between 1959 and 1986 and concluded that the success rate of projects is generally poor. He further emphasized that there are very few records showing completing the project ahead of the schedule. It was reported that 50-80% delays were experienced on 1627 World Bank sponsored projects between 1974 and 1988 (Chan and Kumaraswamy 2002). Therefore, due to the importance of time for the clients and frequent delays/poor performance in the sector, differentiating on time, can lead to competitive advantage to the construction companies.

7.2.3. Service Quality

According to Grönroos (1990), service is 'an activity or series of activities of more or less intangible nature that normally, but not necessarily, take place in interactions between the customer and service employees and/or physical resources or goods and/or systems of the service provider, which are provided as solutions to customer problems'. Lovelock (1991) defines it as 'a task, other than proactive selling that involves interactions with customers in person, by telecommunications, or by mail. It is designed, performed and communicated with two goals in mind: operational efficiency and customer satisfaction.' According to these definitions, the customer's evaluation of service is affected by any employee who has a contact with the customer, therefore provision of customer satisfaction is not commission of a department of the company, and in fact this should be embedded into the culture of the organization. Parasuraman et al. (1985) mentioned about three characteristics of the services; intangibility, heterogeneity, and inseparability. Firstly, most services are intangible; in other words, most services cannot be counted, measured, inventoried, tested, and verified, due to its intangibility. This leads difficulty for the companies to understand how customers perceive their services and evaluate service quality (Zeithaml 1981), therefore the companies tend to underestimate the benefits of the service quality. Second, the services are heterogeneous, namely the performance of services is not stable and varies from the producer to producer, customer to customer, and day to day. Finally, production and consumption of many services are inseparable, in other words service is simultaneously produced and consumed, and this leads to involvement of the costumers into the production phase, therefore the customer's input also determine the service quality. Siu et al. (2001) added perishability as fourth characteristic of service. Perishability means that services cannot be produced in advance, inventoried, and later made available to sale, in other words it cannot be stored. In other words, due to these properties of service, providing service with high quality is very difficult for the companies, but it is not impossible. For instance, Parasuraman et al. (1985) argue that there are five broad dimensions of service quality that are applicable to any service organization.

- Tangible: physical facilities, equipment, appearance of personnel, communication material.
- Reliability: ability to perform the promised service dependably and accurately.
- Responsiveness: the willingness to help customers and provide prompt service.
- Assurance: the knowledge and courtesy of employees and their ability to inspire trust and confidence.

• Empathy: the caring, individualized attention provided to customers

When the construction industry is concerned, service quality is also a hard-to-achieve but important topic. It is widely accepted that contractors should provide high quality services to the owners as well as a high quality product. Service quality contains the attributes of all kinds of services given by the contractor to the client within the context of his contractual obligations. The construction companies can differentiate by extending their services by advising the client in an early feasibility analysis, at various stages of design, in assistance with securing funding for the project, and in planning the operation and the maintenance of the project (Warszawski 1996). During each stage (feasibility, securing finance, design etc.), the communication quality, reliability, responsiveness, assurance and empathy of the contractor to the client may increase the service quality and finally, lead to higher client satisfaction.

7.2.4. Product Quality

Quality in product refers to achieving high quality, even beyond the requirements in the specifications of the constructed facility. Quality is one of the client's major concerns in the construction projects, however most of the construction companies confuse the differentiation with the concept of quality, whereas differentiation encompasses quality, even it is a much boarder concept (Porter 1985). In the past decade, the decline in construction quality has been reported, and this leads to decrease in customer satisfaction in the construction industry. The reasons of this decline in quality was explained by the implementation of new contracting techniques, inadequate pricing methods, lower productivity, insignificant technological growth, reduction in the industry's net worth, dependence on legal assistance , lack of co-operation among professional groups, disappearance of true general contractors, increasing dependence on project consultants, outmoded QA/QC programmes, outmoded safety programmes, and transference of professional liability (Yasamis et al. 2002). Besides, due to the combination of process fragmentation, product complexity, poor definition of quality attributes, and the one-off nature of many projects, achieving high level and consistent quality is very difficult for construction projects (Toakley and Marosszeky 2003). However, satisfying the quality requirements not only yields customer satisfaction in the current job, but also leads to enthusiasm about working with the company in additional works. Besides, a satisfied

customer perceives more value than originally anticipated. Achieving high quality can provide differentiation opportunities for the construction companies.

Mallon and Mulligan (1993) divided the quality into expected, or demanded quality, and attractive, or exciting quality. Expected quality relates to those elements of a product that the customer expects and that must be present to satisfy that customer, no additional amount of expected quality will increase customer satisfaction. Exciting or attractive quality is a product attribute that pleasantly surprises the customer. If this characteristic is not present, the customer's satisfaction is not reduced. If it is present and recognized however, the level of satisfaction is increased. In other words, a differentiated company should satisfy the exciting or attractive quality in their projects for creating a differentiated product from the other companies.

7.2.5. Innovative Solutions

Pedersen (1996) defines construction innovation as "the first use of a technology within a construction firm either in the process or in the product". Jones and Saad (2003) define innovation as whole sequence of events, which occurs over time, to find out a new idea that leads to enhanced performance. Both of these definitions agree on the "newness", which encompass two notions in most industrial researches, namely differentiation, and creation of competitive advantage (Jay Na and Ofori 2007).

Bowley (1960) classified the types of the innovations as product and process innovation. Product innovation is defined as: goods and services introduced to the market are either new or significantly improved with respect to fundamental characteristics. The innovations should be based on the results of new technological developments, new combinations of existing technology or utilization of other knowledge by the firm (Reichstein et al. 2005). "Process innovation" is an improvement in construction methods designed to accomplish usual construction operations or to improve the efficiency of a standard operation (Laborde and Sanvido 1994). Cannon and Hillebrandt (1990) offer four means of product differentiation (process innovation) in construction: by offering a range of project management methods; by extending their activities from construction into design; by extending activities into financial packaging; and by extending forward into commissioning and facilities management. In the construction management literature, the researches about the construction innovation are performed by focusing on the product innovation, whereas

the process innovation can also provide many gains for innovated companies (Slaughter 2000).

Innovation can provide various benefits to the innovator in his business. Previous studies have found that firm-level innovation increases a firm's export potential (Bleaney and Wakelin 2002), profits over long periods and even during recessions (Geroski 1993), credit ratings (Czarnitzki and Kraft 2004), chances of surviving in the market (Cefis and Marsili 2006) and market value (Toivanen et al. 2002). The ability to innovate can create possibilities for firms to gain competitive advantage over their industrial rivals.

In comparison to other sectors, according to the low levels of expenditure on activities related with innovation, construction is usually classified as a traditional and low-technology sector (Manseau and Seaden 2001). Nam and Tatum (1988) argue that there are five specific differences between the construction industry and manufacturing industry that limit the technological development of the construction industry. These differences are determined as immobility, complexity, durability, costliness and high degree of social responsibility. They cite these characteristics as creating the conditions for production processes that result in a 'locked system' in which innovation, or in their words, 'changing the status quo' becomes difficult. Reichstein et al. (2005) examine the liabilities that construction firms face in their innovative activities in comparison to other sectors, and determined six liabilities that inhibit innovative behavior, namely liability of projects, liability of immobility, liability of uncertain demand, liability of smallness, liability of separation, and liability of assembly. Consequently, liabilities of immobility and unexpected demand play a major role in shaping innovative behavior in the sector. Rosenfeld (1994) determined three major characteristics of the construction; capital intensiveness, legal responsibilities and great fragmentation as the greatest barriers to the innovation in construction. Due to these limitations and barriers, implementation of innovation in construction is a difficult process; therefore the companies which can realize the innovative solutions in their business process can create uniqueness in the construction industry where uniqueness leads to the differentiation from the other companies in the sector.

7.2.6. Corporate Image

Corporate image is "the mental picture of the company held by its audiences- what comes to mind when one sees or hears the corporate name or logo" (Gray and Balmer 1998). The various physical and behavioral attributes of the firm, such as business name, variety of

products/services, tradition, and ideology shape the corporate image, also the impression of quality communicated by each person interacting with the firm's clients has effect on the corporate image (Nguyen and Leblanc 2001). The corporate image has two principal components, namely the functional and the emotional component (Kennedy 1977). The functional component consists of tangible characteristics of the companies which can be easily measured. Psychological dimensions shaped by feelings and attitudes towards a firm constitute the emotional component. These feelings are derived from individual experiences with a firm and from the processing of information on the attributes that constitute functional indicators of image.

The importance of corporate image and reputation is recognized as critical corporate assets directly linked to competitive success of the companies (Gray and Balmer 1998). The construction companies also perceive the importance of the positive image in prequalification stage of the projects (Jennings and Holt 1998). The corporate image can be used as important a criterion that influences the customer's qualification process especially when the attributes of the companies are difficult to evaluate (Andreassen and Lindestad 1998). Also, the positive image has a direct effect on customer loyalty (Andreassen and Lindestad 1998; Nguyen and Leblanc 2001). Thus, using their positive image, the companies can eliminate the negative effects of rivalry, ask for a premium price for their products, and create additional leverage in negotiations with suppliers and creditors (Cheah et al. 2007). Moreover, the corporate image also affects the human resources of the companies; namely, the positive image of the companies can attract high quality personnel, and improve employee motivation (Christensen and Askegaard 2001). Nevertheless, since the benefits of the positive image are achievable in the long run, the construction companies who cannot perceive the importance of strategic thinking usually pay no attention to the creation of positive image.

The creation of a positive image requires much time, effort and cost. Firstly, a company should develop more than one image, because the construction projects involve many different parties, like clients, employees, subcontractors, suppliers and shareholders, and all of these parties have different experiences with and expectations from the company, therefore this leads to creation of various images of the company (Nguyen and Leblanc 2001). Although technological breakthroughs and unexpected achievements can improve it rapidly, conversely, it can be also destroyed by ignoring the needs and expectations of the various groups who interact with the firm (Dichter 1985; Herbig et al. 1994). In conclusion, the process of creating a corporate image is a difficult and risky process; but the companies

that manage to develop a positive image can differentiate themselves from the other companies.

As a conclusion, the different modes of differentiation for construction companies have been defined as productivity, time/schedule (rapid construction ability), service quality, product quality, innovative solutions and positive corporate image. It is hypothesized that all these differentiation strategies have a potential to be valued by the clients and may provide a unique competitive position for a construction company. However, the questions that should be answered are:

- What are the drivers for each mode of differentiation? In other words, how can the companies differentiate themselves from their rivals if they want to utilize a specific differentiation strategy?
- What are the sources of differentiation that result in competitive advantage?

The potential drivers of differentiation as identified within the context of this research are given in the next section.

7.3. Drivers of Differentiation

According to Porter (1980), the companies can differentiate themselves by creating uniqueness in their value chain. For construction companies, there are basically 2 levels of value chain, namely corporate level and project level. Moreover, in order to run their value chains effectively, different resources are needed at these levels. Besides, companies should integrate their value chains with those of others (partners, suppliers etc.) to create added value. Relations with other parties may be a major source of competitive advantage due to the high number of parties involved in the construction process and complexity of the supply chain in the construction industry. Within this context, four sets of drivers have been identified as: project management activities, corporate management activities, supporting resources and relations with other parties. It is hypothesized that companies can achieve competitive advantage by differentiating the tasks of project as well as corporate management. Furthermore, exceptional/differentiated resources and relations with other parties can create competitive advantage.

7.3.1. Project Management

It is hypothesized that a construction company may differentiate its value chain at the project level to create a unique position in the market. Project management activities have a potential to add value to the product/service given by the contractor and thus, constitute a source of differentiation for the contractor. Project management activities described below have been defined as potential drivers of differentiation in construction companies:

Schedule Management: A schedule is generally prepared as the job is being estimated. Its purpose at that time is primarily to determine the duration of construction so that an estimate of the general conditions or general requirements containing time-related costs can be prepared and included in the project's estimate (Levy 2002). Due to the nature of construction projects, there are many uncertainties about the project and external factors such as productivity of laborers, environmental, and weather conditions at the inception, therefore schedule planners have to make estimates of several parameters that may be the cause of a potential delay, and to complete the project on or ahead of schedule. Preparation of a successful schedule at the start of a project and managing it throughout the project can lead to differentiation advantage for the company.

Project Organization: It refers to an organization in which employees work temporarily together according to the project needs with a specific project responsibility during the life of the project to perform a specific task (Tosi et al. 2000), and after completion of the project, the employees of the organization will return to their own departments (Langford et al. 1995) or move into an another project. The coordination and integration of these temporary teams into an effective unit is the aim of the project organization, in other words an effective organization should determine lines of authority clearly and every member of the project will know what he or she is responsible for during the project (Lock 2007).

The project organization can affect the performance of individuals and companies, even according to the forms of organization, significant differences in the performance of the same people and companies can be observed. However, there is no one best way to organize a project due to the complex interaction between the environment, the task, the technology, human motivation and organizational structure (Shirazi et al. 1996), therefore each project organization should be designed according to the analyses of environment where the project

is undertaken, technology used during the project, and organizational structure. In conclusion, each project organization provides opportunities for creating uniqueness in their operations. Finally, a company that differentiates at project organization may develop competitive advantage over its rivals.

Contract Management: It refers to follow up contractual circumstances with other parties of the project by establishing contract administration systems, identifying the contractual circumstances, handling variations and additional work, and notifying the client variations (Andersson and Gunnarsson 2002). Many different parties like owners, contractors, subcontractors, suppliers are involved in a project; therefore the contractors should make different contracts with different parties for different purposes. Also, due to the uncertainties in the environment, risks and changes are frequently experienced during the project; therefore the contracts should provide incentives and safeguards to deal with the risks envisaged in advance and flexibility to deal with unforeseen circumstances (Turner 2004). Finally, poor contract management is determined as an important factor affecting the duration of projects, in other words it leads to delays in the projects (Frimpong et al. 2003; Mansfield et al. 1994; Odeh and Battaineh 2002). In conclusion, the contract management which can affect cost, time and quality of the projects is an important activity for a project team. A company that excels in contract management may create added value and differentiate itself from its rivals.

Resource Management: It refers to the activities about procurement of required amount of materials at the required quality, ensuring the availability of resources whenever required, and allocation of the resources to the identified tasks. The resources for any construction projects are manpower, materials, money and machines which are called as four M's of construction, without any of these resources the activities in the project cannot be performed, in other words, low resource coverage or long-lead time in resource acquisition can delay the project schedule by creating bottlenecks (Park 2005). Poor resource management is determined as an important factor causing delays in projects (Faridi and El-Sayegh 2006; Toor and Ogunlana 2008). Resource allocation is an important task of the resource management which affects the performance of the projects (Ashley et al. 1987; Pinto and Slevin 1988), since inadequate allocation of resources can lead to equipment idleness and higher consequent cost, especially when construction equipment is to be shared among a

number of project sites (Ammar and Mohieldin 2002). The resource allocation concerns two issues, namely resource leveling and resource scheduling. In resource leveling, fluctuations in resource usage over the project duration is tried to be minimized in order to avoid undesirable cyclic hiring and firing, and procurement of huge amount of material or equipment (Senouci et al. 2004). In resource scheduling, the project tasks are rescheduled in order to ensure the efficient utilization of limited available resources, meet the physical limits on resources and obtain a continuous usage of resources while keeping the unavoidable extension of the project to a minimum (Ammar and Mohieldin 2002; Hegazy 1999). A company can run resource management activities in a differentiated way when compared to its rivals and achieve a successful position.

Project Monitoring and Control: In a project, at the inception stage, a plan shows how the constructors will achieve the quality, time and budget restrictions that are demanded by the clients. However, due to the uncertainties about the project and nature of the work, the divergences from the plan usually occur, therefore the progress of the projects should be monitored during the project, and the differences between the actual work performance and the planned work performance should be measured. By this way, corrective actions can be taken by the project management team to achieve the preset objectives. The project monitoring should be performed at each stage of the project, and companies should respond quickly to changes, because time lags in such systems have been shown to degrade performance (Al-Jibouri 2003). Companies having effective monitoring and control systems can build a successful position for themselves as the system's effectiveness will directly affect the project success and result in additional value for the client.

Quality Management: It refers to an approach for achieving and sustaining high quality output in order to satisfy customer needs precisely and profitably. Quality is one of the major objectives of construction projects, in other words, the construction companies have to complete the projects according to the quality demanded by the clients in order to survive in the construction sector. However the construction sector is often criticized due to its poor performance on quality, because quality management is a difficult process in construction projects, since numerous parties with different objectives that are involved in a project have their impacts on quality, and the poor performance of any party can affect overall quality of the project. In addition, the construction industry is characterized by its non-standardization.

Since most products of construction industry are one-off type, standardized procedures are not valid in construction projects, therefore the concepts developed for manufacturing is not directly applicable to construction (Al-Momani 2000). Lastly, excessive changes in design are usually observed during the project due to errors and omissions in design, and changes in user needs or market demand throughout the project (Kanji and Wong 1998). However, the companies can impose the benefits of quality management such as improved product and services, reduced costs, more satisfied customers and employees, improved bottom-line financial performance, market share, and productivity (Lai and Cheng 2003; Walton 1988). With increasing importance of quality, different concepts, techniques and approaches are developed in the literature, such as quality assurance, quality control, total quality management, and quality function deployment.

- Quality control (QC): It is a body of routine activities to measure and control the quality of products for providing routine and consistent checks to ensure data integrity, correctness, and completeness, identifying and addressing errors and omissions, documenting and archiving inventory material and records about QC activities. QC systems include detailed product and performance specifications, a paperwork, and procedures control system, raw material and intermediate- stage product testing, reports about activities, performances of the members of activities and feedback of process information to appropriate personnel and suppliers (Dale 2003).

- Quality assurance (QA): It is a planned and systematic action necessary to provide adequate confidence that a product or service will satisfy given requirements for quality. QA focuses on how procedures can be adjusted to prevent the quality problems in order to satisfy a specific quality standard in every department of an organization. QA take a broader perspective according to the QC, because greater number of managers is involved in QA process, and it intends to increase awareness of the implications of quality throughout the workforce (Howard 2000).

- Total quality management (TQM): It is an integrated management philosophy involving every member of the organization for increasing customer satisfaction by continuously improving performance. TQM involves set of practices that focus on reducing rework, long-range thinking, increasing employee involvement and teamwork, process improvement, customer and supplier involvement, training and education of employee, cost effectiveness, defect- free work, improving motivation

of employees, team-based problem solving, and constant measurement of results (Burati Jr. et al. 1992; Powell 1995).

- Quality function deployment (QFD): It is a technique used for translating both spoken and unspoken customers' needs into appropriate technical parameters and prioritizing them in the design. Specific target parameters are determined according to these design parameters, and they are frequently checked against customers' needs throughout each stage of the product development process to ensure customers' satisfaction with the end product (Delgado-Hernandez 2007; Eldin and Hikle 2003). This approach is a broader and more integrated approach to deal with clients demands more systematically and defining what they want precisely to do it right in the first time. Due to benefits such as concentration on identification of client needs and expectations, planning, communication and concurrency, and uncertainty reduction, this approach also attracted the attention of construction industry (Dikmen et al. 2005b).

A company that can differentiate itself in managing quality based on the above discussed processes and using different techniques to maximize value for the client has a potential to achieve a strong competitive position.

Risk Management: It is a systematic approach for dealing with risk by establishing an appropriate context, setting goals and objectives, identifying and analyzing risks, influencing risk decision making, and monitoring and reviewing risks responses (Edwards and Bowen 1998). The construction industry and its clients are widely associated with a high degree of risk due to facing with a variety of situations involving many unknown, unexpected, frequently, undesirable and often unpredictable factors (Akintoye and MacLeod 1997). This leads risk management as an important part of decision making process in construction. It is evident that differentiating the risk management process, companies may create a strong competitive position due to better managed risks and a less vulnerable situation.

Project Cost Estimation: Most construction contracts are fixed price with penalties for cost overrun and delays. Usually contracts are awarded to the lowest cost bidder. Cost is a critical issue, because overestimated cost could result in loss of a contract whereas underestimated cost may lead to significant losses for the contractor. Despite the great

importance of cost estimation, it is a difficult and complex process because of the limited information in early stages of a project. Due to the limited information, the assumptions about uncertain items are made during the cost estimation process, in other words, cost estimation does not only consist of precise technical and analytical input but also the estimators subjective judgments and predictions about uncertain future (Leung et al. 2005). Furthermore, cost is an output of the performance of different parties involved in a project; therefore, integrated efforts of the various parties and their decisions regarding the design, technology and implementation of the project can have a significant impact on the overall project cost (Chan and Park 2005). Finally, due to the size and complexity of construction work, cost estimation process takes a long time to complete, however time spent on cost estimation is tried to be kept minimum by the clients and contractors. Differentiating in the cost estimation process has a potential to result in a differentiated competitive position for a contractor.

Environmental Management: Almost in all countries, environmental issues are receiving attention from governments, non-governmental institutions, commercial organizations in most of sectors of the economy and the general public (Ofori 1992). This leads to the broad and various legislations concerning the environmental issues (Pasquire 1999). As environmental pollution caused by construction activities is one of the major pollution source, this results in increased pressure on construction professionals for improving their environmental performance (Shen et al. 2005). However, the awareness of the construction companies about the environmental issues is limited and the construction activities are performed by several parties with different objectives, therefore none of them takes the responsibility for protecting the environment, even the participants see the responsibility as lying with others (Ofori 2000b). However, the activities related to environmental issues are a duty of the project management team due to the statues, regulations, codes, and general policies relating environmental issues about construction sector. Protection of the environment should be seen as a social responsibility by the companies. Also, the positive correlation between the positive social responsibility, including environmental issues and firm performance was mentioned in the literature (Capon et al. 1990; Klassen and McLaughlin 1996). Companies paying more attention to environmental issues and carry out environmental management activities effectively may build a differentiated position for themselves

Health and Safety Management: In the construction industry, the risk of fatality is five times more likely than in the manufacturing industry (Sawacha et al. 1999). Since certain types of the construction projects are of high risk, most of the accidents are concluded with death or major injury (Snashall 1990). Apart from loss of human lives and loss of morale, increase in indirect costs such as the cost of insurance, inspection, product, material, and legal costs are also observed. Furthermore, lost of working days due to classified incapability at work and industrial injuries are reported. Jannadi (1996) indicated that the most important factors affecting construction safety were: (1) maintaining safe working conditions; (2) establishing safety training; (3) cultivating good safety habits; (4) effective controlling of subcontractors; (5) maintaining a close supervision over workers; and (6) assignment of responsibility to all levels of management and workers. All of these factors are in the responsibility of the project management. In other words, with the strict health and safety management, the number of the accidents occurred in the projects and effects on project costs can be reduced. Companies having systematic health and safety assurance systems can outperform their rivals and increase their competitive advantage as a result of differentiation in this area.

7.3.2. Corporate management:

It is hypothesized that a construction company may differentiate its value chain at the corporate level to create a unique position in the market. Corporate management activities have a potential to add value to the product/service given by the contractor and thus, constitute a source of differentiation for the contractor. Corporate management activities described below have been defined as potential drivers of differentiation in construction companies:

Strategic Planning: It is a dynamic process of formulating and implementing decisions about an organization's future position and allocating its resources, such as its capital and employees according to the identified future opportunities for adapting to its ever-changing environment. The construction industry is a project-based industry; therefore the construction companies generally focus on a single and unique end product obtained at the end of a project. Consequently, they only focus on the planning and control of resources for execution of specific projects; this leads that the companies give more attention to project management rather than corporate management (Chinowsky and Meredith 2000). In

conclusion, the construction companies have traditionally neglect strategic planning and simply react to events; therefore they are affected from the economic recession and political instability severely (Langford et al. 1995). However, the construction industry is a highly dynamic sector due to the change pace of its operating environment, industry structures and product characteristics. Moreover, activities performed in the industry are affected from the pace of technological change in other sectors of the economy, uncertainty in the financial market, changing client demands due to the variations in taste, aspiration and purchasing power (Betts and Ofori 1992). Finally, environment in which the construction activities are performed contain more risk when compared to the other industries. In conclusion, strategic planning should be recognized as an important aspect of firms' overall activity which requires as much as routine operations (Betts and Ofori 1992). The positive link between the strategic planning and financial performance of the companies was mentioned in the literature, (Andersen 2000; Dikmen and Birgonul 2004a; Miller and Cardinal 1994; Pearce II et al. 1987). Companies that build strategic planning systems and use a strategic approach rather than an operational focus may differentiate their services and outperform their rivals.

Business Development: It is a strategic management activity for integrating business enterprises, creating innovative solutions to complex needs and requirements of the business environment, or thinking strategically about leading change (Rainey 2006). In business development process, a business is evaluated for determining its potential by imposing such tools as marketing, sales, information management and customer service. With globalization, technological changes, economic drivers, and social and environmental mandates, the global business environment becomes less predictable and more challenging. Moreover, clients, executives, and society expect and demand superior products, services, and operations with less waste, reduced impacts on health, safety, and the environment, therefore defining and improving the whole company is an important activity in order to achieve superior and sustainable performance that satisfies the challenges of the present and future expectations (Rainey 2006). Since the business development plans involve financial, legal and advertising skills, multi-discipline should work together for developing successful business development, in other words business development cannot be reduced to simple templates applicable to all or even most situations faced by real-world enterprises. The importance of business development as a part of the operations of the construction enterprise was also recognized (Betts and Ofori 1992). Hasegawa (1988) defined business development as one of the critical activities in order to accommodate the changes in the global construction

industry. Differentiating in business development activities may help companies to find niche markets, estimate market needs, recognize attractive project options in the global construction industry, matching company strengths/weaknesses with market threats/opportunities and thus, increase competitive advantage.

Human Resources Management: It refers to variety of activities, including deciding what staffing is needed, whether to use independent contractors or hire employees to fill these needs, recruiting and training the best employees, ensuring they are high performers, dealing with performance issues, and ensuring personnel and management practices conform to various regulations (McNamara 2002). Human beings are a basic resource of construction, and manpower cost forms the major part of total cost of the projects. In addition, the performance of the human beings is not predictable like the equipments and machineries. However, human resources management has been undervalued in the construction industry due to the high cost of human resource management, the fragmented nature of the industry, mobility of the workforce, the shallow management structure applied in the industry, subcontracting and the use of casual labor, the attitudes and education of construction managers (Langford et al. 1995). Human resources management is admitted as an important activity which affects the performance of organizations (Delaney and Huselid 1996; Huselid 1995). A construction company may use its effective human resources management system as a strategic weapon to create a differentiated position in the market.

Financial Management: It refers to the activities including set up of a total budgeting system, development of financial information reporting systems, internal auditing, and tracking of working capital requirements and rate of circulation (Cheah et al. 2007). With a strong financial management, a firm can track all its cost components more effectively and create greater accountability. Due to the characteristics of construction industry, the level of bankruptcy of construction companies is higher than the other sectors, while in many cases these companies can gain profit in the long term (Harris and McCaffer 2001). A sound financial management system is a must for surviving in the market especially when the market takes a downward turn. Moreover, as the needs of the market change in the direction in which contractors are no longer seen as constructors but also investors, securing finance and offering attractive financial packages for owners become a critical success factor. A company's financial management skills may help creation of differentiation advantage.

Professional Management: Most of the companies in the world are controlled by their founders, or by the founders' families (Burkart et al. 2003). As a result of this, these firms are characterized by their highly centralized decision making system, overdependence on one or two key individuals for firm's survival and growth and paternalistic atmosphere (Charan et al. 1980). However, as the companies expand and grow, in order to obtain the resources needed and manage the complex management processes to survive and maintain growth, the control of the firm should be ceded to professional managers (Gedajlovic et al. 2004). Also, the productivity of the professional managers is admitted as higher than the productivity of the family (Bhattacharya and Ravikumar 2005) due to the differences in management regime (Barth et al. 2005). Finally, since the family owners are cautious when making new investments and reluctant to raise loans or admit new investors (Agrawal and Nagarjan 1990; Gallo and Vilaseca 1996) due to the limited diversification of financial risk and a higher cost of capital (Demsetz and Lehn 1985), the family ownership companies may hesitate to make investments to differentiate themselves. In conclusion, it is hypothesized that professional managers' high capabilities and competences may help creation of competitive advantage using differentiation strategies.

Organizational Learning: It refers to the systematic promotion of a learning culture within an organization for adapting and developing organizational efficiency by improvement of their workforces' skills and capability individually and collectively (Dodgson 1993; Kululanga et al. 2001a). With the development of organizational learning behavior, the culture of the organizations is also changed, for instance, the workforces develop themselves from a "doing" workforce to 'thinking' workforce; in addition the organization start to search ways for better working continuously, this leads to continuous improvement in construction business process (Kululanga et al. 2001a). Also, the employees of the organizations gain benefits from organizational learning in form of development of skills and competencies (Chan et al. 2005). Finally, Dikmen et al. (2005a) demonstrate the direct and positive relationship between organizational learning and performance of the construction companies. However, for fully implication of organizational learning, the organizations should spend much effort and time. Therefore, organization learning is an important and difficult activity in corporate management for meeting the challenges of rapidly evolving business environment and survival of the organization in this environment.

Research and Development: It refers to a deliberate and managed process to improve the capacity and effectiveness of the construction industry to meet the national economic demand for building and civil engineering products, and to support sustained national economic and social development objectives (Ofori 2000a). Unfortunately, research and development (R&D) is always underestimated by the construction companies; however due to globalization and deregulation of markets, advanced technology is required to meet the major objectives of the projects (Raftery et al. 1998), therefore advanced technology become an important asset for the construction companies. Japanese construction companies can be a good example for enjoying the benefits of advanced technologies. According to Fraser and Zarkada-Fraser (2001), R&D expense of Japanese construction companies is much higher than the construction companies; even it is five times higher than the construction companies of United States. This leads to entry barriers that protect the home market against non-Japanese firms, and competitive advantage in world markets. In the rankings of the largest in terms of total construction revenue, five Japanese construction companies are listed in first twenty companies (ENR 2007). In conclusion, R&D is an important activity which could result in both formulation and implementation of a diversification strategy. Companies differentiating in research and development may secure a strong competitive position.

Tendering: In construction industry, most of the contracts are awarded by competitive tendering in the public sector such as central government departments, municipalities, and local authorities. This process is also valid for the clients from the private sector. Thus, most of the contracts are awarded after several contractors have submitted a tender (Harris and McCaffer 2001). In traditional method, the clients evaluate the tenders according to cost estimation; however the other factors such as quality plans, schedule plans and health and safety plans are also considered in the evaluation process. During a tendering process, the responsibilities and risks of the contractors are also determined. A company that has the ability to differentiate its tendering strategy has a potential to create competitive advantage in the market as it will be possible for the company to get more jobs and earn profits by determining a reasonable bid price that considers the costs, risks, opportunities and level of competition.

Claim Management: It refers to the process of employing and coordinating resources to progress a claim from identification and analysis through preparation, and presentation, to

negotiation and settlement (Ren et al. 2001). The aim of claim management is to ensure compensation of a detriment suffered by one party according to the contract in the execution of the contract (Kululanga et al. 2001b). Over the past three decades, the construction industry has experienced an increase in claims, liability exposures and disputes, along with an increasing difficulty in reaching reasonable settlements in an effective, economical and timely manner (Barrie and Paulson 1992). Also, due to the nature of the construction, the claims are observed during the projects very frequently. A company's claim management strategy significantly affects its profitability from a project and also, its long-term success. It is believed that companies that can carry out successful claim management tasks are more likely to succeed in the construction market and they can achieve competitive advantage by differentiating their claim management tasks from those of the others operating in the same market.

Knowledge Management: It is a process of acquiring, creating, sharing, utilizing, and storing knowledge from the internal and external business environments to achieve enhanced performance, increased value, competitive advantage, and return on investment, through the use of various tools, processes, methods and techniques (Huber 1991; Kamara et al. 2002). According to this definition, knowledge management is not only about collecting data, but also utilization of the collected data effectively within the organization for increasing the performance of the company. Kululanga and McCaffer (2001) argues that the knowledge is power, and learning rapidly and competently have become a permanent strategy for success in construction. However, due to the temporary organizations formed during the projects in the construction sector, capturing and transferring project knowledge is a difficult task for the companies. Consequently, this leads to the increased risk of 'reinventing the wheel', wasted activity, and poor project performance (Kamara et al. 2002). A construction company that has a good knowledge management system and learning ability may differentiate itself from the others and develop a competitive advantage.

7.3.3. Resources

Amit and Schoemaker (1993) defined resources as available factors owned and used by firms to develop and implement their strategies. According to the resource based view (Barney 1991; Wernerfelt 1984), the firms can be viewed as a collection of resources, skills and routines that can be used to achieve and sustain competitive advantage. This perspective

assumes that the construction companies can gain superior profits by differentiation through firms' set of resources (Leask and Parnell 2005) which are rare, unique, inimitable, durable, idiosyncratic, non-tradeable, intangible and non-substitutable (Oliver 1997). In other words, in order to develop sustainable competitive advantage, the companies should have differentiated resources.

The differences in resource selection and accumulation can be explained by within firm decision making and external strategic factors. Market imperfections defined as barriers to acquisition, imitation inhibit competitors' abilities to obtain or duplicate critical resources (Oliver 1997). Consequently, this leads to creation of firm-specific resources, and these unique resources help companies to gain differentiation advantage. Examples of unique resources include human resources, machinery and equipment, financial resources, experience and knowledge, and technological capabilities (Friedman 1984; Hofer and Schendel 1978; Mahoney and Pandian 1992; Wright et al. 2001).

Human Resources: Human resources denote the collection of people and all associated networks and structures within which they work together to meet the needs of the organization to acquire the necessary skills to conduct an effective and efficient business (Juran and Godfrey 1998). Human resources is recognized as an important resource rather than the practices and/or procedures used by the firm (Wright et al. 1994). Warszawski (1996) mentioned about human resources as probably the most important resource and the key to success in the construction industry. Human resources create sustainable competitive advantage due to their rare, inimitable, and non-substitutable nature (Wright and McMahan 1992). A firm's human resources practices also are considered as a source of competitive advantage (Lado and Wilson 1994). Human resource practices generally refer to a set of internally consistent practices adopted by firms to enhance the knowledge, skills, ability, and motivation of employees. As these practices support and develop the human resources and competencies, they add value to the firm (Wright et al. 1995). However, many organizations give relatively low priority on both human resources of the firm and human resource department, especially when faced financial difficulties, firstly the expenditures related with firm's people such as training, reward and headcounts are cut off (Barney and Wright 1998). Construction industry is a labor- intensive industry and human resources is a basic resource of construction. The shortage of skilled workers and experienced managers has reached dangerously low levels in today's market place and remains one of the major challenges

facing the industry (Levy 2002). In conclusion, high qualified and experienced human resource which is not available in all construction companies can lead to creation of uniqueness and a strong competitive position.

Machinery/ Equipment: Due to high wages and lack of skilled labors, production is mechanized in order to raise productivity by replacing labor with machines. In many developed countries, there has been a move towards a greater use of plant and machinery in building and civil engineering (Wells 2001). Also, the number and type of machinery and equipment owned by the company is a criterion in the selection of contractors for the clients as the machinery and equipment can be determinative for physical realization of some types of construction projects, such as dams, nuclear plants. Therefore, Friedman (1984) lists equipment and field resources as a primary strength for construction companies.

Experience and Knowledge: Knowledge is body of information, coupled with understanding and reasoning by expertise and skills acquired by a person through experience or education (Langford et al. 1995). The term knowledge when used in relation to a construction company usually encompasses features such as experience, concepts, values, beliefs, and ways of working that can be shared and communicated (Harris and McCaffer 2001). Today's work place is evolving from a skill based environment to knowledge based environment, even knowledge based tasks become as a central focus of organization operations (Chinowsky and Meredith 2000). As a result, knowledge has been treated as a key source of potential advantage for construction organizations (Bresnen et al. 2003; Kamara et al. 2002; Lima et al. 2003). However, due to high fragmented structure of the construction projects, knowledge does not exist as a totality. In addition, there is no single repository for project knowledge, which is fluid and constantly changing, due to the tight timeframe of construction projects, and the lack of sufficient resources and standard work processes for managing project knowledge, the demobilization of project team after the completion of a project, and the high staff turnover (Tan et al. 2007; Udeaja et al.). Knowledge can be explicit and tacit based on what circumstances it is considered and/or used. Explicit knowledge can be communicated externally and captured in formal models, rules and procedures (Vail 1999). Tacit knowledge is personal knowledge embedded in individual experience and involves intangible factors, such as personal beliefs, perspective, and the value system. Tacit knowledge is hard to articulate with formal language. Both types of

knowledge that accumulate within an organization as a result of previous experience can be a major source of competitive advantage for construction companies (Dikmen and Birgonul 2003).

Financial Resources: Financial resources refer to an organization's liquid assets such as cash, stock, bonds or investment capital that determine the company's financial capability to carry out projects. Strong financial structure lets the company to invest in high risk projects, in which the other companies cannot venture to invest, with prospects of higher returns. As a company's financial strength increases, its credibility and reputation also increases among its clients and suppliers (Warszawski 1996), therefore this persuades a client that a firm has sufficient stability and resources to undertake a contract (Fong and Choi 2000; Jennings and Holt 1998). Also, financial strength encourages the companies to make investments to put an innovative idea into practice due to the ability of the firm to take on the arising costs with respect to the risks of realizing the innovation (Hartmann 2006). On the other hand, lack of financial resources can cause construction delays or qualitative defects (Ghosh and Jintanapakanont 2004; Odeyinka and Yusif 1997), since the companies cannot continually secure surety bonds for ongoing project procurements, which in turn negatively affect a firm's image. Dikmen and Birgonul (2003) concluded that financial resources are the most critical resources and differentiated services for the Turkish construction companies through innovative project development.

Technology: It is a basic determinant of what and how the industry can build and where it is possible to construct. Technology can create competitive advantage by determining relative cost position or differentiation in its own right and/or through changing or influencing the other drivers of cost or differentiation (Porter 1985); therefore the available technology should be considered as an important factor in achieving differentiation strategies. Also, technological change affects the competition. It plays a major role in an industry's structural change, as well as in creating new industries (Porter 1985).

Nam and Tatum (1988) mentioned about five major characteristics of constructed products: immobility, complexity, durability, costliness and high degree of social responsibility which result in many limitations for construction technology. Therefore, the construction industry is traditionally defined as static and low technology industry. However, due to increasing technical complexity of constructed facilities and international competition, the importance

of advanced technologies in construction markets is increased. Technological advantages may lead to a strong position in the construction sector because the barriers to new entrants can be raised in large construction projects that require high-technology. Companies that invest in computing and engineering technology may use technological know-how as a strategic weapon and even create a market niche (Betts and Ofori 1992). A firm's proprietary technology not only strengthens the entry barriers, but also provides significant mobility barriers if this technology is incompatible with the existing culture or operations of rival firms (Oliver 1997). Consequently, the companies can differentiate themselves from their rivals by means of technological resources and know-how.

7.3.4. Relations

Construction is a project based industry, and all projects are carried on with the involvement of a number of parties such as client, contractor, subcontractors, suppliers, and government. The quality of a firm's relationship with these parties should be considered as a strategic asset that can have important implications on its operations and activities, as a result its economic performance (Johnson 1999; Ring and Van de Ven 1992). Especially, the impact of establishing and maintaining good relationships with the parties is gaining importance where the transactions involve high 'uncertainty' and high 'asset specificity' (Kale and Arditi 2001).

Developing long-term relationship with the parties leads to development of experience of working together and accumulation of information about how the companies conduct in performing the activities, therefore a conductive working environment resulting in effective communication flow and less chances of misunderstanding and disputes (Singh and Tiong 2006). Besides, the competitive positions of the companies can be improved upon by establishing good relationships with the related and supporting industries (Betts and Ofori 1992). The importance of long-term and high quality relationships as a source of competitive advantage was also mentioned by Dikmen and Birgonul (2003). Finally, the quality of any stage throughout the project is affected from the quality of the previous stages. Specifically, the quality of a project is directly related with the materials and equipment supplied by the suppliers, quality of the work packages performed by the subcontractors, and quality of the contributions of the clients and consultants. In other words, the contractors can produce high quality product if they can create and maintain close and long-term relationship among the parties involved in the process. However, the culture of construction industry depends on

negative relationships and attitudes of the companies towards each other due to traditional rivalry between the parties. These traditional relationships among different organizations or organizational units inhibit the companies to meet the increasing demands of clients regarding better-value-for-money, higher quality and shorter cycle times (Tah 2005). In conclusion, developing necessary relations within the market is a tedious task, but if a company has exceptional and unique relations, these may act as the major drivers of differentiation in the construction business.

Relationship with the suppliers: Rapidly changing competitive environment forces the companies to seek changes in the relationship with suppliers due to the value creation potential of the suppliers to create sustainable competitive advantage (Sheth and Sharma 1997). Consequently, the companies start to move from transaction oriented marketing strategies to relationship oriented marketing strategies. Long- term relationships with suppliers facilitate investment in equipment at lower manufacturing costs, and improve planning coordination and scheduling. Then, the suppliers get incentive to invest in technologies that can be used for several building projects. Due to the contractor's strict procedures and requirements, suppliers determine the most efficient ways for working with contractor (Singh and Tiong 2006). Finally, in order to develop a socially responsible business image, the construction companies should build good relationship with the suppliers (Petrovic-Lazarevic 2008).

Relationship with the subcontractors: Due to instability in demand and seasonal demand, the construction companies are suffering from the market ups and downs, however by subcontracting, the construction companies can be more flexible in responding to potential market ups and downs. In addition, the large construction companies do not prefer having large equipment parks and workforces for avoiding the idleness of these assets, leading to incapability of companies for carrying out the entire construction project with their own workforce and equipment (Kale and Arditi 2001) when needed. In addition, due to uncertainties and project complexity, the construction industry is risky; therefore the construction companies prefer subcontracting for sharing the risk in their projects (Anvuur and Kumaraswamy 2007). Consequently, subcontracting out work packages is very common practice in construction. However, some of problems occurred during the projects are originate from the performance of the subcontractors. Majid and McCaffer (1998) determine

the subcontractor delays as one of the main causes of project delays, and according to Love and Li (2000), damage created by subcontractors and poor workmanship are the primary reason of the defects. Due to the traditional relations throughout the project, a 'team spirit' between the subcontractors and contractors is not formed; this can lead to various barriers for productivity (Hsieh 1998). Moreover, the uncertainties related to a subcontractor's technical qualifications, timeliness, reliability and financial stability leads to risks (Akinci and Fischer 1998). The long-term and high quality relations with subcontractors can eliminate the uncertainties related with the subcontractor. Consequently, contractor performance is positively and strongly associated with establishing and maintaining long-term and high quality relationships with subcontractors (Kale and Arditi 2001).

Relationship with the clients: Due to the traditional relationships and contractual arrangements widely used in the construction industry, clients and contractors see themselves as adversaries leading to differences in values, goals and orientations of these parties (Bresnen and Marshall 2000). In other words, it impedes the emergence of 'team spirit' during the project. The other factor which affects the relationships of these parties is 'opportunistic' behavior. Due to the traditional tendering in which the contract is awarded according to the costs, the contractors are forced to give lowest bid at the expense of loss. Therefore, they try to impose changes in user needs and market demand modifying the client preferences, and errors and omissions in drawings and specifications for increasing their profit, consequently, the clients become suspicious of contractor claims and suggestions (Kadefors 2004). These factors lead to mistrust environment in the construction, in turn, poor project performances in terms of cost, time and quality. Developing high quality and long-term relationships affects the performance of the project positively; in addition the contractors can award contracts for extra projects from the same client without competing due to the preference of the clients to work with same contractors that have proven their competencies with successfully completed clients' previous projects. In other words, high quality relationships with clients can be considered as a critical resource on creating competitive advantages in construction companies (Dikmen and Birgonul 2003). Also, the clients can eliminate transaction costs regarding the contract awarding process.

Relationship with the competitors: According to Porter (1980), the competition in an industry, in turn ultimate profit potential of the market is driven by five competitive forces;

namely industry competitors, suppliers, buyers, substitutes, and potential entrants. New entrants to an industry who bring new capacity, the desire to gain market share, and often substantial resources leads to intensification of the competition, in turn decrease in profit margins. The companies can prevent new entrants by developing high entry barriers, in other words, entry barriers should be considered as collective asset by the companies for improving them by acting collectively, however due to mistrust between the competitors, they prefer acting only for their benefits rather than collective benefits. But, the companies can overcome this problem by good relationships with competitors. In addition, due to increasing competition based on globalization and technologically complexity of the projects, the companies try to join their efforts to achieve goals that each firm, acting alone, could not attain easily; in other words, gain competitive advantage in the market place. They can achieve this by establishing partnerships, joint-ventures and strategic alliances, therefore the companies can access to new technologies or markets, provide wider range of products/ services, provide economies of scale in joint research and/or production, access to knowledge beyond the firm's boundaries, share risks and access to complementary skills (Powell 1987). However, the establishment and maintenance of these compositions requires high quality relationship between the parties (Mohr and Spekman 1994). Especially in the construction industry, joint ventures are widely used and establishing successful joint ventures is one of the major critical success factors (Ozorhon 2007).

Relationship with internal stakeholders: The organizations are consisted of different departments possessing different responsibilities in order to achieve same goal, and good relationship with these departments satisfies effective communication flow within the organization and avoids misunderstanding between the internal groups. Even, strong ties within internal networks are required in order to motivate individual actors to transfer complex, sensitive information (Hansen 1999). This leads to increase in the productivity of the organization, which in turn affects the performance of the projects positively. Specifically, Bresnen and Marshall (2000) reported that emergence of problems due to the clear differences in objectives between the project teams and other internal departments which provides resources to the project teams. Especially, for the construction companies which carrying on projects worldwide, good relationships within internal stakeholders is an important determinative of the project success, therefore developing good internal relationships is considered as an important performance criteria in evaluating manager's performance (Dainty et al. 2003).

Relationship with the other parties: In addition to client, contractor, subcontractors, competitors, suppliers; many other parties are also involved throughout the construction process; such as governmental authorities, designers, consultants, and labor unions. Also, the relationships with these parties should be considered as an asset which affects the performance of the projects. For example, if a company decides to cut cost, and also reduces its labor force, labor unions may show their dissatisfaction by threatening to strike (Arthur 1992). The smooth relations with labor unions can minimize the likelihood of strikes, slowdowns and disputes. Similarly, relationships with governmental authorities may be very important in cases where the project success may significantly be affected from bureaucratic delays. Depending on the type of project delivery system and responsibilities between the parties, relations between the contractor, designer, construction manager and engineer may constitute one of the critical success factors for a project. Establishing and sustaining good relations with those parties may lead to further alliances in the long run, creating more job opportunities as well as short-term/project success.

Based on the above discussions, a conceptual framework has been developed to model the interrelations between the drivers and modes of differentiation adopted in the Turkish construction industry. A questionnaire has been designed to test the validity of the model, details of which are presented in the next chapter together with its statistical findings

CHAPTER 8

QUESTIONNAIRE FINDINGS ON DIFFERENTIATION STRATEGIES ADOPTED BY TURKISH CONTRACTORS

It is evident that there are a number of ways to differentiate and create a strong position in the construction industry. However, the dynamic relationship between the differentiation strategy and its drivers should be investigated in order to propose a strategic map for the contractors so that they can develop a relevant strategic perspective as well as an action plan. In order to investigate the complex structure that consists of drivers and modes of differentiation, a questionnaire study has been designed in which the respondents are requested to comment on the following issues:

- Modes of differentiation (or differentiation strategies/methods) that create a source of competitive advantage in the Turkish construction industry.
- The activities, resources or relations (or drivers of differentiation) that can be differentiated by the Turkish contractors to achieve competitive advantage.

In this chapter, findings of the questionnaire study will be discussed and statistically significant relations between the identified attributes will be examined to develop a complete model that explains the links between different modes and drivers of differentiation.

8.1 Questionnaire study

In this study, a questionnaire form composed of 54 questions was designed and posted on the internet. The respondents who identified themselves as either general managers, chairmen or heads of business development/ strategic planning divisions were informed about the questionnaire by sending e-mails. The major aim of the questionnaire was to reveal the interrelations between drivers and modes of differentiation in the Turkish construction industry. A sample of the questionnaire is provided in Appendix B. The survey is comprised of three parts. First part of the questionnaire aims to gain general information about the companies; namely age, size, area of expertise, total turnover, overseas turnover, diversity in

international markets. The aim of the second part of the questionnaire regarding variables under the heading of "project management", "corporate management", "resources", and "relationships" is evaluating the potential of the variables that lead to differentiation in gaining competitive advantage. Last part of the questionnaire concerns the potential of the modes of differentiation that create a source of competitive advantage. The 62 of the questionnaires returned back. The target section of the construction sector for this survey has been determined as medium-big sized companies whose total turnover has been calculated as 775.544 US$M in the last 3 years for the 62 respondents. The reason of this target selection is the limited resources of the small-sized construction firms which facing difficulties in differentiating themselves from their rivals (Kale and Arditi 2003), this leads to preference of cost leadership strategies in their business process. Subjective reporting approach is used for potential assessment to create competitive advantage based on differentiation. The respondents were asked to rank the listed differentiation modes and differentiation drivers using a Likert scale where 1 denotes least potential level and 5 denotes the highest potential level.

8.2. Descriptive statistics

The first step in the analysis of these data is to search for some descriptive statistics, so as to identify the general characteristics of the respondents which attended the survey. The following part discusses the descriptive statistics about the characteristics of the companies.

8.2.1. General information about the respondent companies

- **Number of years of operation in construction sector**

In the questionnaire, "age" refers to the active years of the organization in the construction business. The average age of the organizations was determined as 34.27, and standard deviation was calculated as 14.55. The oldest firm is determined as 70 years old, and the youngest firm was determined as 2 years old. The distribution of the companies in terms of their age is shown in Figure 8-1.

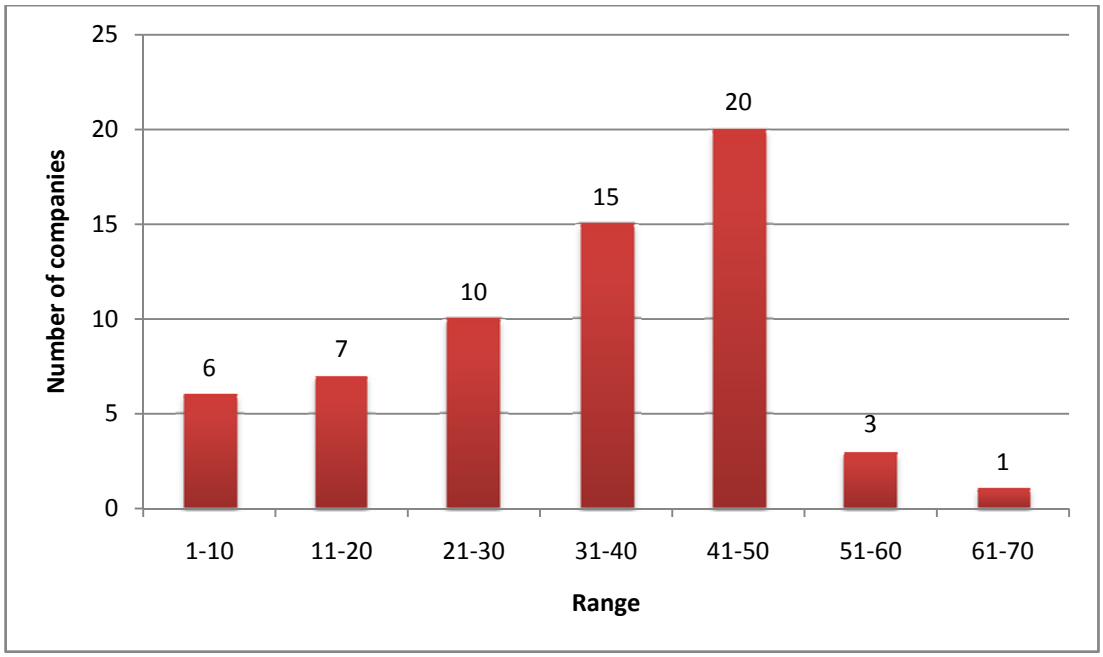


Figure 8-1: Distribution of companies in terms of age

• **Size of the companies**

In the questionnaire, "size" refers to the business volume of the companies when compared to those of other companies competing in the Turkish construction sector. The size is measured by a subjective 1-5 scale and respondents compared their sizes with respect to other firms that operate in the Turkish construction industry. According to Table C.1 (given in appendix C), the average size of the companies involved in this questionnaire is 4.23, in addition, the median and mode of the size of companies is 5, also the large (34; 54.84%) and medium-large (16; 25.81%) companies form the vast majority of the companies as seen in Figure 8-2. Therefore, the data set is suitable for investigating the behaviors of large and medium companies operating in the Turkish construction sector.

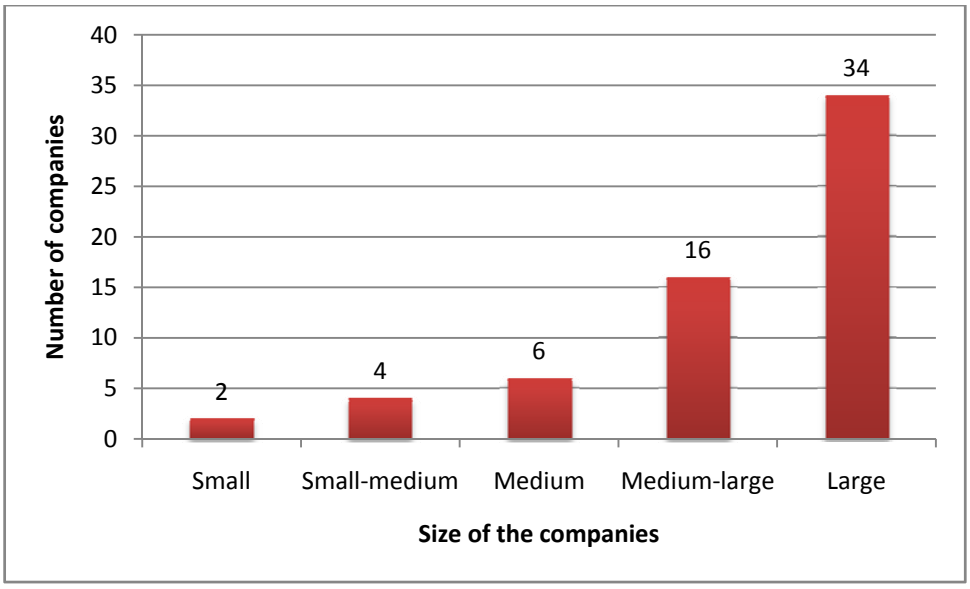


Figure 8-2: Distribution of number of companies in terms of size

• **Areas of expertise**

In the questionnaire, six fields of expertise; namely infrastructure, industrial facilities, housing, building, water structures (like dams, irrigation systems etc.), and transportation structures (highway, tunnel, bridge etc.) were identified. The respondents were asked to rank these fields of expertise according to total amount of projects performed in each field. According to Figure 8-3, only 6.50% of the companies performed construction in one field, 19.40% carried out projects in two fields, 24.20% carried out projects in three fields, 16.10% carried out projects in four fields, 12.90% carried out projects in five fields, and 21.00% carried out projects in all of the specified fields. The companies can be classified as "focus" and "diversified" according to the number of the areas they were involved, and the threshold for this classification was determined as three areas, therefore 31 companies were determined as "diversified" and the others were classified as "focus".

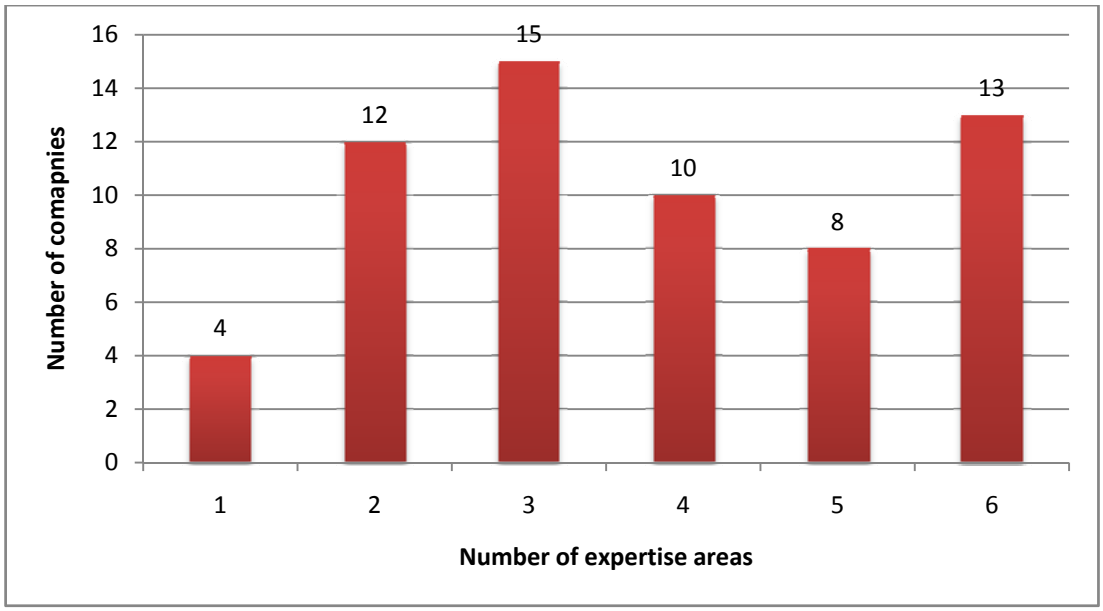


Figure 8-3: Distribution of number of companies in terms of number of expertise areas

Respondent companies have expertise mainly on transportation structures (27.40%), industrial facilities (22.60%), building (16.10%), housing (12.90%) and water structures (12.90%). Figure 8-4 shows the distribution of companies in terms of their expertise.

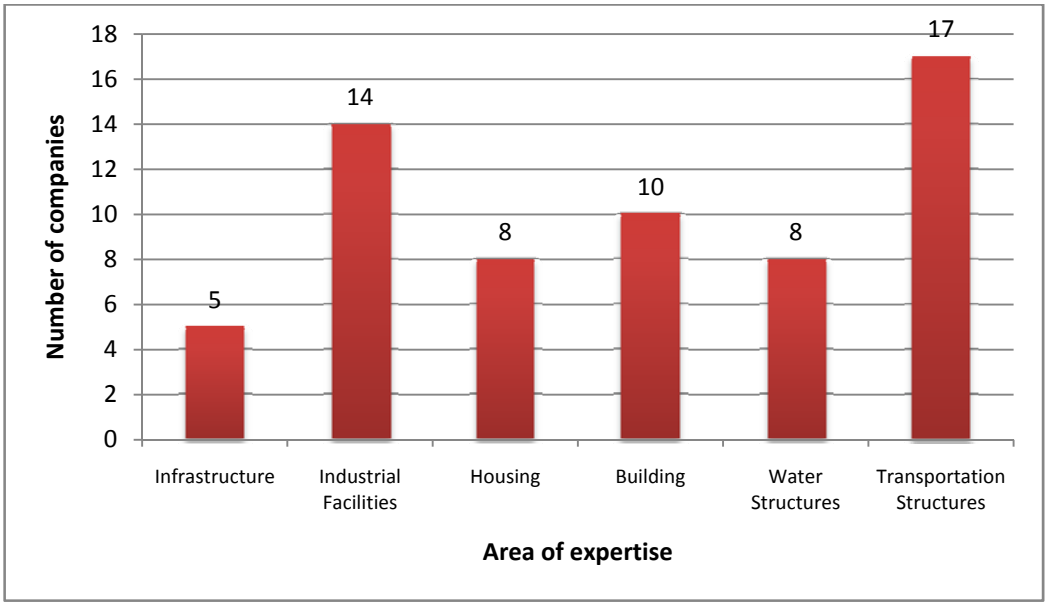


Figure 8-4: Distribution of companies in terms of expertise areas

• **Total Turnover**

Total turnover is referred to total turnover in the last 3 years in domestic and international construction projects. One of the respondents did not identify the total turnover and domestic turnover of his company; therefore it is excluded from the analysis. The average total turnover of the respondent companies is 775.544 US$M. The maximum value for total turnover is 10 Billion USD, and the minimum value of total turnover is 1.17 Million USD. Figure 8-5 shows the distribution of companies in terms of their total turnover.

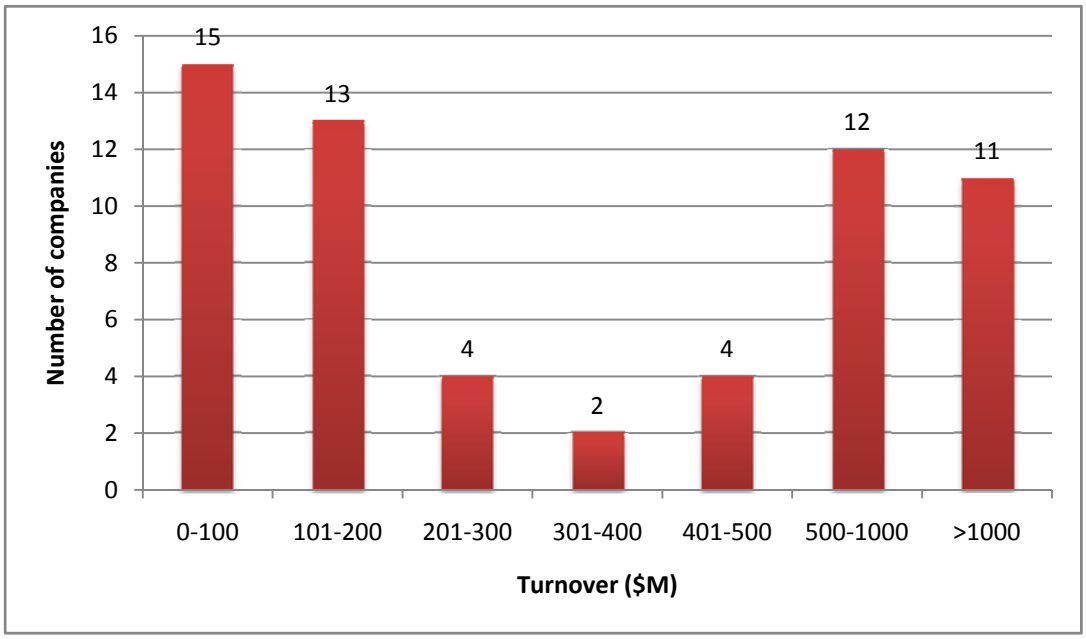


Figure 8-5: Distribution of companies in terms of total turnover

• **Overseas Turnover**

Overseas turnover is referred to total turnover in the last 3 years from only international construction projects. The average overseas turnover of the respondent companies is 542.510 US$M. The maximum value for overseas turnover is 10 Billion USD, and 9 companies perform projects only in Turkey. Figure 8-6 shows the distribution of companies in terms of their overseas turnover.

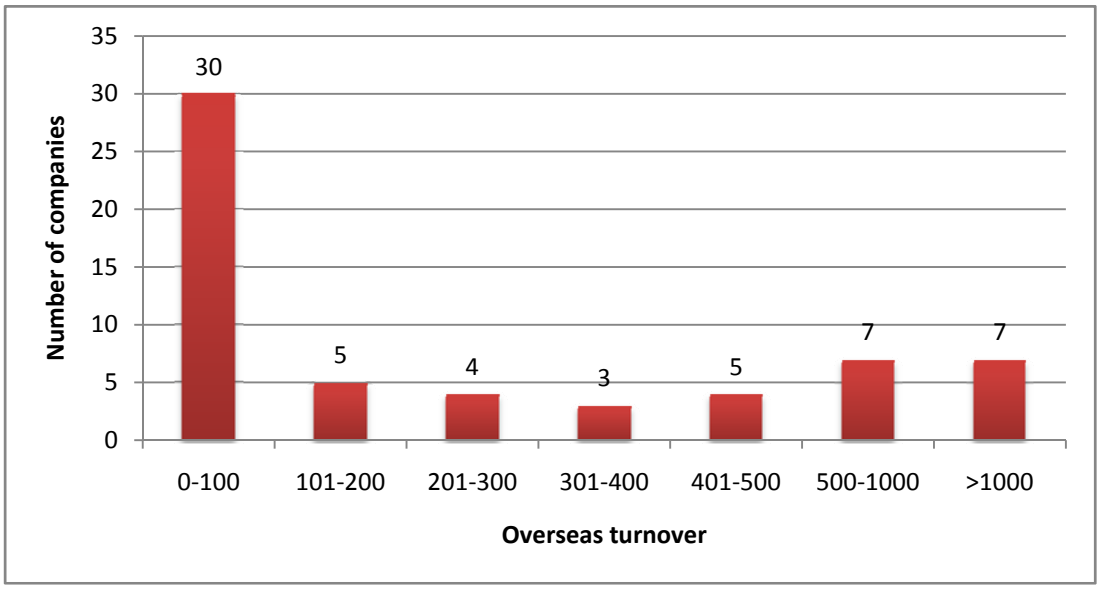


Figure 8-6: Distribution of companies in terms of overseas turnover

• **Diversity in international markets**

The respondents were asked to rank the first five countries according to the total size of the projects performed in these countries. The Turkish construction companies were determined as active in 38 countries worldwide. The following figure (Figure 8-7) illustrates the distribution of these countries. As expected, Turkey was determined as where the country in which the Turkish construction companies are most active. Libya, Russian Federation, Qatar, Saudi Arabia, and United Arab Emirates are the other countries where the Turkish construction companies perform projects actively. If these countries are classified according to their geographical locations, four groups can be determined; Turkey, Russian Federation and the countries separated from the Soviet Union, Arabic countries and others including United States of America, Ireland etc.

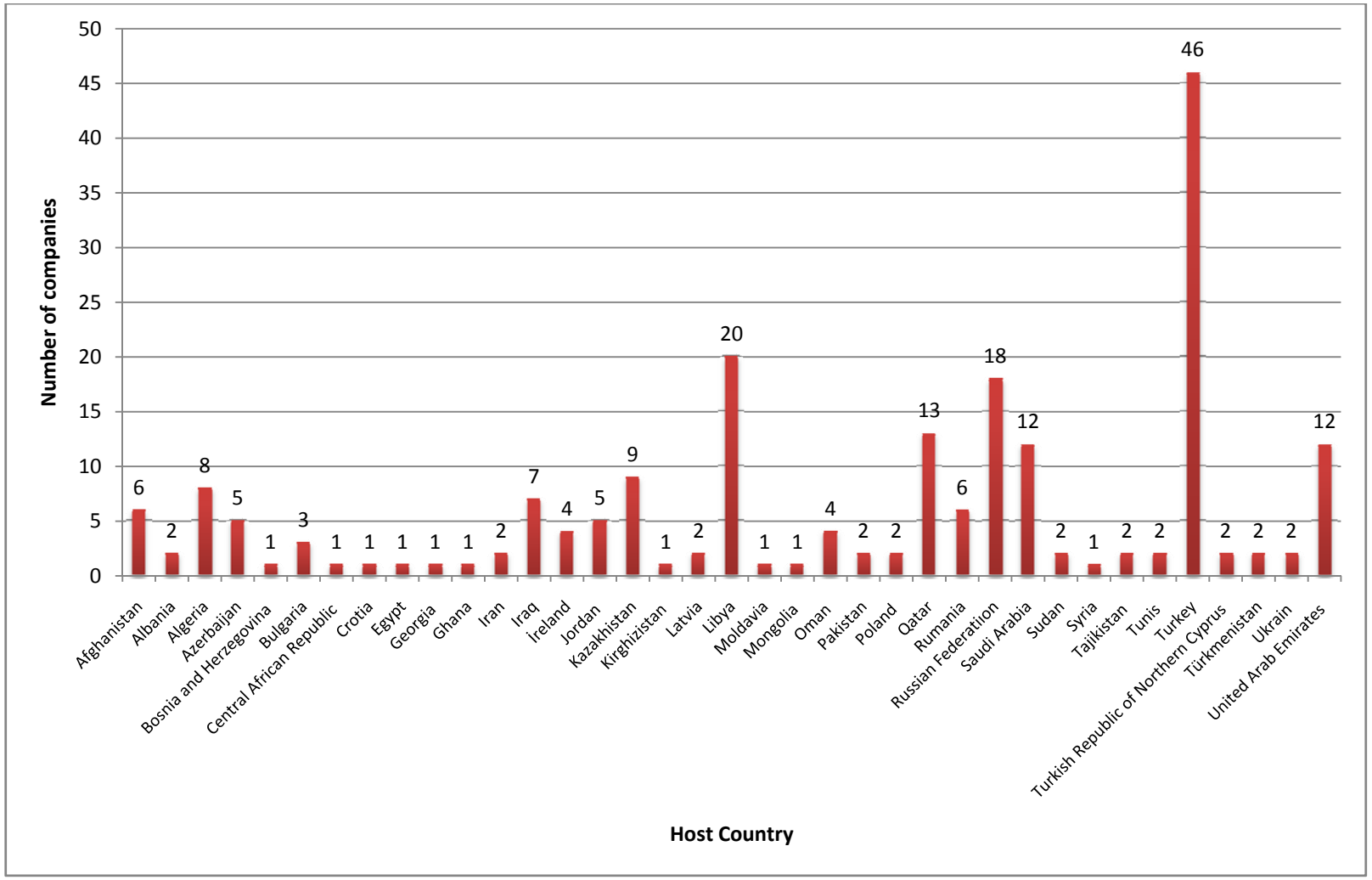


Figure 8-7: Distribution of companies in terms of countries where they perform projects

According to the results, % of companies that perform projects in one, two, three, four and five countries are 14.75%, 18.03%, 14.75%, 8.20% and 44.26%, respectively. In other words, approximately half of the companies perform projects in at least five countries, so respondent companies have a large diversity in international markets.

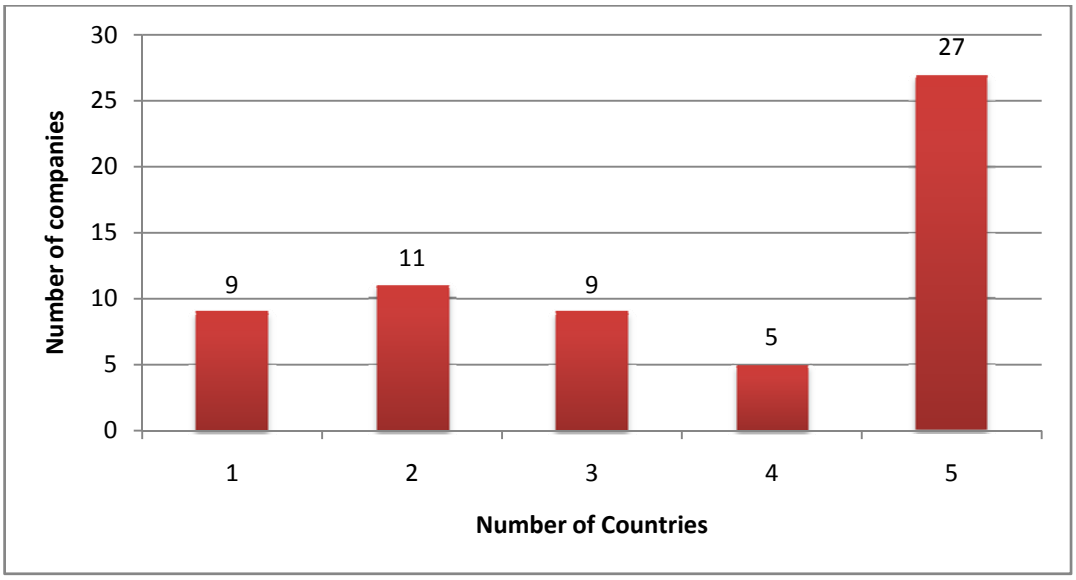


Figure 8-8: Distribution of companies in terms of number of countries

As a summary, results demonstrate that the respondent firms are large-size construction companies that operate in international markets and carry out various types of construction in Turkey as well as more than 30 different countries.

8.2.2. Overview of the modes of differentiation

As mentioned before, six modes/methods of differentiation were determined for the Turkish construction companies. The respondents were asked to comment on the potential of each mode of differentiation to create competitive advantage. In this study, according to the means, it is clear that the respondents placed relatively the same emphasis (all methods received an average rating above 4) to the identified differentiation methods. In other words, all of the identified differentiation modes are valid for the Turkish construction industry, the companies can differentiate themselves from the other companies by applying any one of these differentiation methods and they can be used to create competitive advantage. When

the means of the ratings are considered, productivity, time and innovative solutions have slightly higher ratings when compared with the other methods. This shows that the potential to create competitive advantage is the highest in the modes of differentiation.

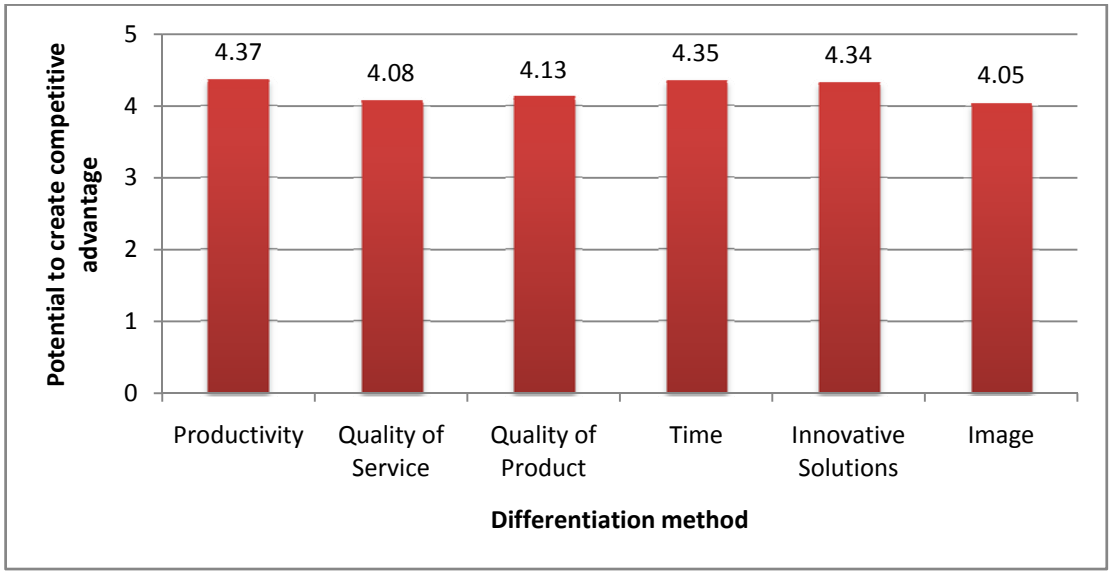


Figure 8-9: Ratings of differentiation methods

8.2.3. Statistical findings about drivers of differentiation

In this part of the questionnaire, the respondents were asked to evaluate the potential of achieving competitive advantage by using a differentiation strategy in the given tasks or resources. Statistical information on each driver of differentiation is given from Figure 8-10 to Figure 8-13.

Among the variables under the heading of project management; project cost estimation and project organization were determined as the most important variables. The ratings associated with all tasks of project management except environmental management and health and safety management were determined to be approximately equal, whereas environmental issues received the lowest rating among others.

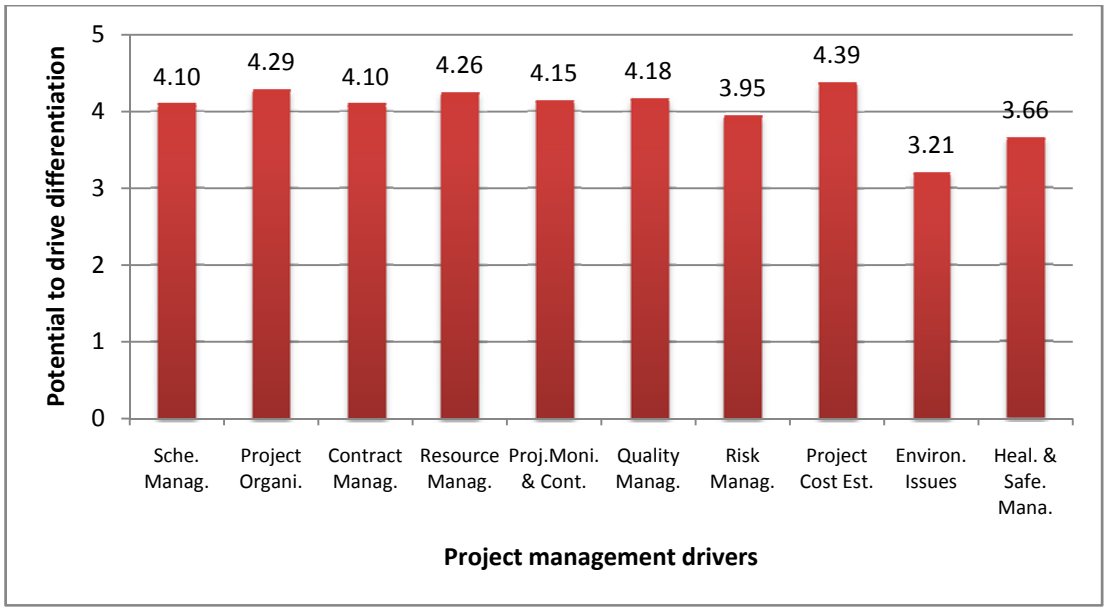


Figure 8-10 : Distribution of ratings related with project management

Strategic planning, organizational structure and professional management were determined as the most important variables under the heading of "corporate management"; whereas knowledge management and research and development were found to be the least important tasks that the companies can differentiate to achieve competitive advantage.

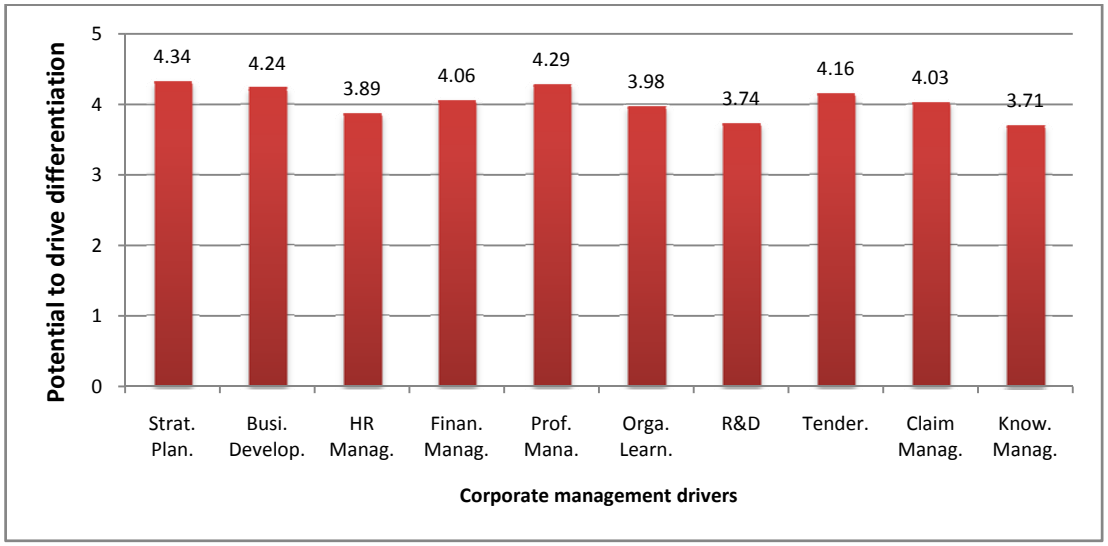


Figure 8-11: Distribution of ratings related with corporate management

The most significant resource that can be differentiated and used to create competitive advantage has been determined as "experience" among all resources succeeded by human resources. Machinery and equipment which can be easily obtained from the suppliers by all of the companies received the lowest rating from the respondents.

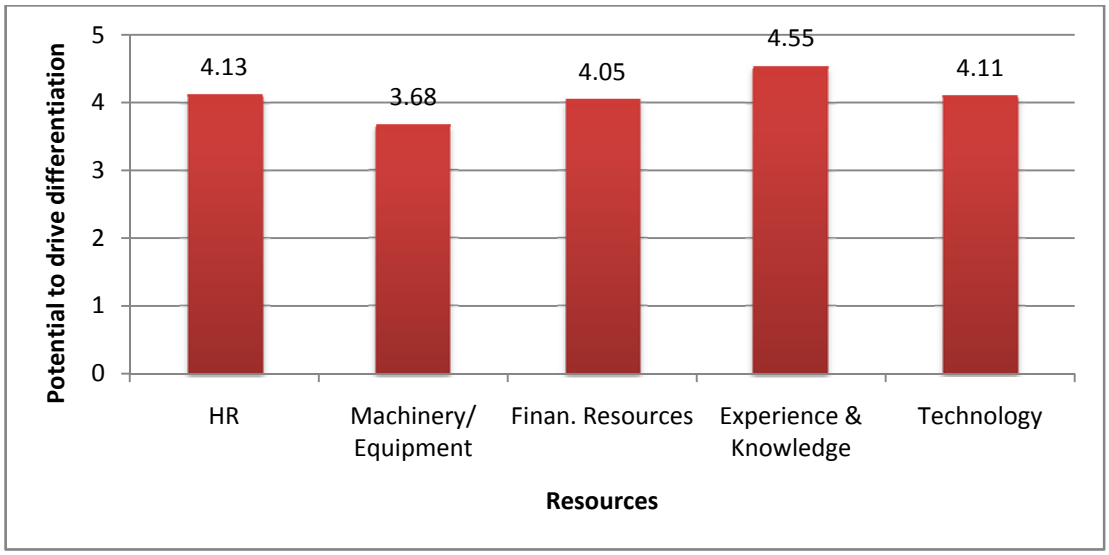


Figure 8-12: Distribution of ratings related with resources

Relations are found to receive lower scores from the respondents when compared to other categories. This may be interpreted as the possibility to differentiate in relations and potential to create competitive advantage by means of differentiated relations is lower with the exception of "relations with clients". According to the respondents, the "relations with the clients" is the most important variable, whereas the "relations with the competitors" is the lowest.

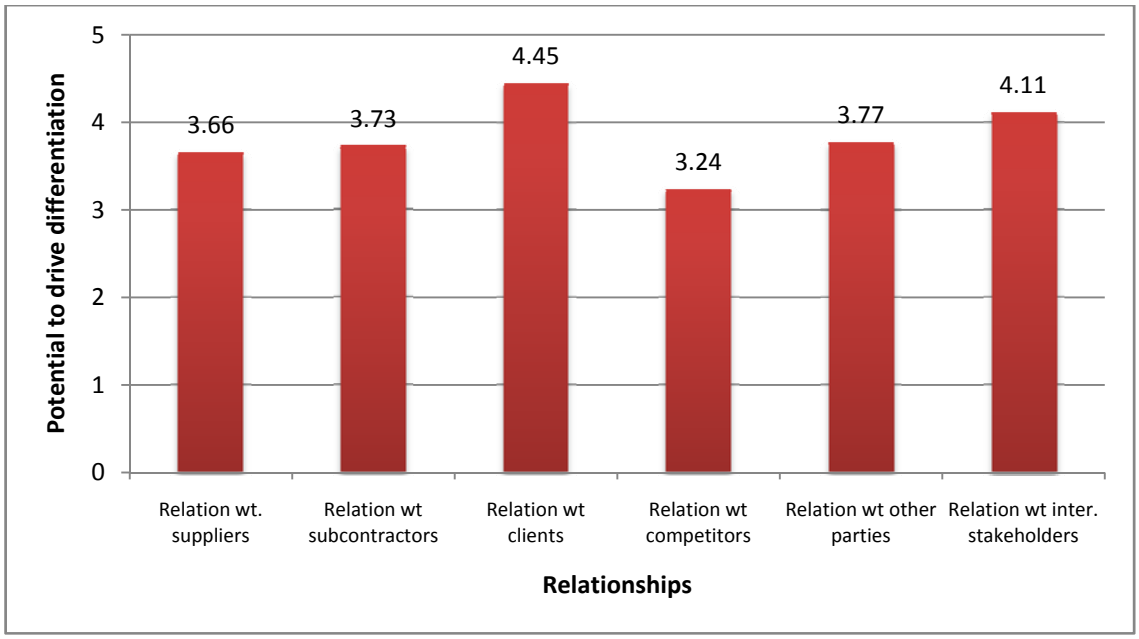


Figure 8-13: Distribution of ratings related with relations

8.2.4. Discussion of the relationships between companies profile, drivers and modes of differentiation

In this section, in the light of the general information about the respondent companies, the companies' profile was identified. Besides, the relationships between the differentiation drivers; namely project management, corporate management, resources and relationships, and differentiation modes and companies' profile were investigated by using crosstabs for nominal variables and correlation analysis for continuous variables by using SPSS. Firstly, the companies were categorized according to the number of the countries where they perform projects actively. The threshold for this categorization was decided as three countries, in other words the companies which perform projects in three or less than three countries are categorized as focus in terms of number of countries in which they perform projects, whereas other companies are categorized as diversified in terms of countries in which they perform projects. The crosstab analysis with Chi-square, contingency coefficient, Phi and Cramer's V and lambda statistic methods were conducted to investigate relationships. According to this analysis, the potential of differentiation drivers to create competitive advantage do not vary within varying number of the countries in which they perform projects. Besides, the variation in differentiation modes was also investigated by

crosstab analysis. As a result, the number of the countries was also concluded to be an insignificant factor that explains the choice of differentiation modes.

Also, the companies were classified according to the internationalization ratio calculated by dividing overseas turnover to the total turnover. The companies whose internationalization ratio is higher than 0.50 are classified as international companies, other companies are classified as domestic companies. The same analysis, performed for the categorization based on number of the countries, was re-performed for this categorization. Consequently, companies in the international category and companies in domestic category show no difference in perceiving the potential of the differentiation drivers to create competitive advantage. About the differentiation modes, improving the quality of service was perceived as more applicable to differentiate and having more potential to create competitive advantage for Turkish market (Chi-square = 0.043). On the other hand, creation of positive corporate image as a differentiation mode was more appreciated by the companies which prefer performing international projects (Chi-square= 0.086). The potential of other differentiation modes to create competitive advantage is perceived similarly by the companies in these categories.

Table 8-1: The crosstabs of quality of service and corporate image

Quality of service		3	4	5
International based	Count	4	18	11
	% within	12.1%	54.5%	33.3%
Turkey based	Count	4	7	18
	% within	13.8%	24.1%	62.1%

Corporate image		2	3	4	5
International based	Count	0	8	11	14
	% within	0.0%	24.2%	33.3%	42.4%
Turkey based	Count	2	5	16	6
	% within	6.9%	17.2%	55.2%	20.7%

As mentioned before, the companies were classified as “focus” and “diversified” according to the number of the areas they were involved. The same analysis explained above was performed for this classification as well. Consequently, the significant difference in perception of the potential of project management as a differentiation driver was observed (Chi-square = 0.027). According to Table 8-2, project management is conceived as a more

important differentiation driver among the companies which prefer focus strategies. When the differentiation modes are examined within this categorization, no significant difference was observed between these categories, therefore the number of the expertise areas was concluded as an insignificant factor in determination of the potential of differentiation modes.

Table 8-2: The crosstabs of project management

Project management		3	4	5
Focus	Count	6	7	18
	% within	19.4%	22.6%	58.1%
Diversified	Count	5	17	9
	% within	16.1%	54.8%	29.0%

Finally, the Pearson correlation analysis between age, size and differentiation drivers was carried out; as a result, no significant correlation was observed between age, size and differentiation drivers. Also, the correlation between age, size and differentiation modes was investigated. The correlation matrix of this analysis is shown in Table 8-3. Consequently, according to Table 8-3, no significant correlation between these variables was observed. Finally, it is concluded that age and size are not significant factors which affect the potential of differentiation modes to create a source of competitive advantage for the Turkish construction companies.

Table 8-3: Correlation analysis between age, size and differentiation drivers

	Age	Size
Age	1	0.561(**)
Size	0.561(**)	1
Productivity	-0.189	-0.093
Time	-0.109	-0.126
Innovative Solutions	-0.180	0.025
Quality of Product	-0.022	-0.029
Quality of Service	0.027	0.053
Image	0.099	0.121
**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).		

8.2.5. Discussion of descriptive statistics

The target population for this study was decided as medium-big companies, according to the total turnover (average= 775.544 US$M) and size of the companies (average= 4.23), it can be concluded that the data set is suitable for the target population. Besides, the average age of the respondents calculated as 34.27 reveals that the respondents companies are experienced enough. Also, Figure 8-8 demonstrates that the respondent companies have international experience and a large diversity in international markets.

When the mode of the differentiation is considered, it is apparent that all of the differentiation modes are conceived as applicable to differentiate from the other companies; albeit productivity, time and innovative solutions have slightly higher ratings. This implies that if a company wishes to differentiate to create competitive advantage, they should give equal importance to all these modes, however, especially, they should try to increase the productivity of the labor, equipment, and capital; achieve on-schedule performance in construction projects; and develop product and process innovations.

The following table shows the variables under the heading of "Project management", "Corporate management", "Resources", and "Relationships" have highest potential to drive the companies to achieve competitive advantage by using a differentiation strategy. According to the Table 8-4, experience and knowledge, relation with clients, strategic planning and project cost estimation were determined as the most important variables among thirty-one variables. In other words, the companies should concentrate on these variables in order to differentiate successfully. On the other hand, the least important variables were determined as environmental management and relationship with competitors.

Table 8-4: Variables having highest potential

Project Management	
Project Cost Estimation	4.39
Project Organization	4.29
Resource Management	4.26
Corporate Management	
Strategic Planning	4.34
Professional Management	4.29
Business Development	4.24
Tendering	4.16
Resources	
Experience & Knowledge	4.55
Human Resources	4.13
Technology	4.11
Relationships	
Relation with Clients	4.45
Relation with Internal Stakeholders	4.11

8.3. Structural equation modeling (SEM)

SEM is a collection of statistical techniques that allow a set of relations between one or more independent variables (IVs), either continuous or discrete, and one or more dependent variables (DVs), either continuous or discrete, to be examined. Both IVs and DVs can be either measured variables (directly observed), or latent variables (unobserved, not directly observed). SEM is also referred to as causal modeling, causal analysis, simultaneous equation modeling, and analysis of covariance structures, path analysis, or confirmatory factor analysis**.** The latter two are special types of SEM.

The companies manage their projects and corporate through performing many activities, also they utilize from different resources in their business processes, and establish relationships with many different parties throughout the construction processes; therefore the potential of these variables cannot be evaluated by the respondents directly; in other words these variables should be identified by multi factors. Besides, there are at least five dependent variables; namely project management, corporate management, resources, relationships, and differentiation modes, used in this study, and four of these dependent variables have

potential to affect each other. Whereas, the traditional statistical methods can deal with a limited number of variables and is not sufficient for analyzing this kind of complex structure. Consequently, in this study, SEM was decided to be used in the analysis of the interrelationship between drivers and modes of differentiation in the Turkish construction industry.

8.3.1. Components of SEM

8.3.1.1. Variables:

There are two major types of variables in structural equation modeling: observed variables and latent variables.

Observed Variables:

These variables are also called indicators and/or manifest variables. Observed variables have data that can be directly measured by a researcher, for example numeric responses to a rating scale item on a questionnaire. The independent and dependent observed variables are designated by X and Y, respectively. Observed variables are indicated by squares.

Latent Variables:

A latent variable; also called an F-type variable, or factor; is not directly observable or measured directly, rather they are observed or measure indirectly by inferring constructs based on what observed variables are selected to define the latent variable. Examples of the latent variables in psychology are self-concept and motivation; in sociology, powerlessness and anomie; in education, verbal ability and teacher expectancy; in economics, capitalism and social class; in management, performance of the companies and project planning effectiveness (Byrne 2006). The latent variables are indicated by ellipses or circles.

The latent variables are further distinguished into two types: exogenous and endogenous latent variables. Exogenous latent variables are synonymous with independent variables; they cause fluctuations in the values of other latent variables in the model. Changes in the values of exogenous variables are not explained by the model; rather, they are considered to be influenced by other factors external to it. The latent exogenous variables are labeled ksi (ξ). Endogenous latent variables are synonymous with the dependent variables, and are

influenced by the exogenous variables in the model, either directly or indirectly. Fluctuation in the values of endogenous variables is said to be explained by the model because all latent variables that influence them are included in the model specification. The latent endogenous variables are labeled eta (η). The following figure shows two latent variables and their relationship. In this figure, F1 is the exogenous latent variable, and F2 indicates endogenous latent variable since F1 influences F2.

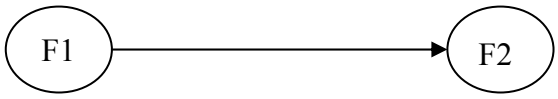


Figure 8-14: Exogenous and endogenous latent variables

The latent independent variables are measured by observed independent variables, traditionally denoted by X. the latent dependent variables are measured by observed dependent variables, traditionally denoted by Y. The latent variables are usually denoted by more than one observed variable, this approach provides benefits about the error assigned to the variables. In other words, if a single observed variable is used for inferring a latent variable, no measurement error will be associated with the measurement of the latent variable. Scilicet, it is assumed that the latent variable is perfectly measured by the single observed variables, which is typically not the case.

8.3.1.2. Errors:

E errors:

Every V variable that is predicted by other variables via a regression equation has associated with it an E, or error, variable. E variables are also called measurement error. The numbering of E variables is arbitrary, but by convention, the E number is matched to the V number. Thus, E7 is the error variable for V7. These errors are labeled as (ζ).

D errors:

Every factor that is predicted by other variables or factors has associated with it a D, or disturbance, variable. D variables are also called as residual error. The numbering of D variables is arbitrary, but by convention the numbers of F and D variables match. That is, D3 is the disturbance variable for F3. For the observed exogenous variables, these errors are

called delta (δ) and for the observed endogenous variables, these errors are called epsilon (ε).

8.3.1.3. Arrows:

For illustrating the influence between the observed and latent variables, the arrows can be used. Two types of arrows are used in the representation of SEM; single and double headed arrows. Single-headed arrows (⟶) represent the impact of one variable on another; and double-headed arrows (⟷) represent covariances or correlations between pairs of variables. These arrows are used for basic configurations which represent important components in the analytic process. The single-headed arrows from the latent to the observed variables are labeled lambda (λ). The single-headed arrows from the exogenous to the endogenous variables are labeled gamma (γ). The single-headed arrows from the endogenous variables to other endogenous variables are labeled beta (β). The double-headed arrows which represent the correlations or covariances among the exogenous variables are labeled phi (φ).

Structural equation models are most often represented graphically. Figure 8-15 shows a graphical representation of a structural equation model.

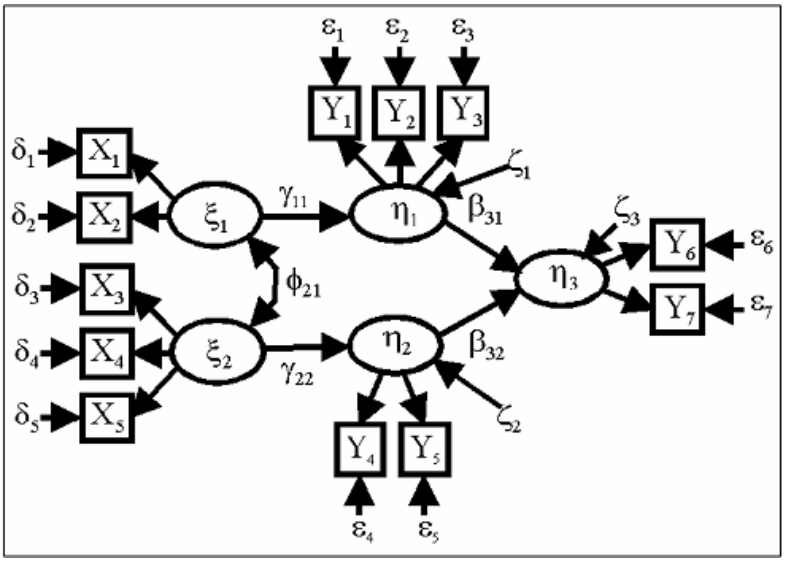


Figure 8-15: Graphical representation of a structural modeling

8.3.2. Strengths of SEM

The popularity of SEM has been increased in different fields; like psychology (Hershberger 2003), social science (Hayduk 1996), and biological sciences (Shipley 2000); due to advantages of this technique on traditional analysis methods (Tomarken and Waller 2005). Schumacker and Lomax (2004) mentioned about four major reasons for popularity of SEM. Firstly; the basic statistical methods can deal with a limited number of variables, which means that they are not capable of dealing with sophisticated theories developed. For example, in multiple regression, only one dependent variable can be used, however for understanding complex phenomena, the use of limited number of variables is not sufficient. In other words, multiple observed variables are required to better understand area of scientific inquiry. Therefore, SEM techniques are becoming preferred method for confirming (or disconfirming) theoretical models in a quantitative fashion. Secondly, in many disciplines, measurement error is one of the major issues; however measurement error and statistical analysis of data have been treated separately. While in SEM, the greater recognition is given to the validity and reliability of observed scores from measurement instruments. In addition, SEM techniques explicitly take measurement error into account when statistically analyzing data. Thirdly, more advanced models can be developed in SEM than the traditional analysis models, for example most first generation regression models such as linear regression, ANOVA can analyze only one layer of linkages between independent and dependent variables at a time, however more complicated variable relationships can be expressed in SEM, therefore a more complete picture of the entire model can be presented (Bullock et al. 1994). These advanced SEM models and techniques have provided many researchers with an increased capability to analyze sophisticated theoretical models of complex phenomena, thus requiring less reliance on basic statistical methods. In other words, the researchers can directly test the model of interest without using straw-man alternatives; set up to be easily refuted, required for evaluating researcher's theoretical hypothesis in using traditional analysis methods (Tomarken and Waller 2005). Finally, with the increasing popularity, different software programs which are user friendly and containing features similar to other Windows based software packages are developed for SEM analysis, such as LISREL, Amos, and EQS. Due to similarity of these software programs with the other Windows based software packages, learning process of this program becomes easier.

Flexibility of SEM relative to many other statistical methods is also an advantage (Bollen 1989). SEM is flexible, because it allows to model relationships among multiple predictor and criterion variables, also latent variables are constructed, it provides model errors in

measurements for observed variables, and priori substantive/theoretical and measurement assumptions against empirical data are statistically tested (Chin 1998).

An advantage of the SEM approach is the easiness of the replication of the analysis by using given the matrices and information on the specific models. In other words, every outputs of SEM such as all the parameter estimates (structural paths and item loadings), overall model fit indices, R-square, and other analytical results can be reproduced according to the given the covariance matrix and details of the statistical model, therefore the reviewers can review the model and detect and possibly offer solutions to problems in the analysis, in addition they can perform further studies or reanalyzes with alternative models (Chin 1998).

Another important strength of SEM is the availability of global goodness of fit measures; therefore the models which involve a large number of linear equations can be evaluated. For the most alternative procedures that might be used in place of SEM to test such models, only separate "mini-tests" of model components that are conducted on an equation-by-equation basis are available, so the global model cannot be evaluated. In addition, different goodness of fit indices is developed for comparing the fit of alternative models that differ in complexity. In this regard, SEM supports the model comparison approach to data analysis (Tomarken and Waller 2005).

8.3.3. Limitations of SEM

Firstly, modeling technique of SEM is criticized, because the models are constructed for fitting the sample covariance instead of sample values themselves. This leads to elimination of much valuable information about the data such as means, skewness and kurtosis. Yet, sample covariance is only one of many characteristics of the data, and is not sufficient enough to understand the total structure of the data (Xie 2008).

The other limitation is the existence of equivalent models which are alternative models that suggest different relationships among latent constructs, in other words the different models which may be radically different can yield equivalent levels of fit (Shook et al. 2004). Therefore MacCallum et al. (1993) asserted that the existence of equivalent models could call into question the conclusions drawn from most SEM studies. MacCallum and Austin (2000) advised researchers to generate and evaluate the substantive meaningfulness of equivalent models in empirical studies. Therefore, examination of alternative models can provide some protection against a confirmation bias and provide support of a favored model.

Since the results of SEM are subject to sampling or selection effects with respect to at least three aspects of a study: individuals, measures, and occasions; the generalizability of findings is also a limitation of SEM. The researchers should specify the population of interest in their hypothesis, and acknowledge that the conclusions derived from the model cannot be generalized besides that population. The indicators which represent the latent variables should be chosen very carefully, because the nature of the construct can shift with the choice of indicators, which in turn can influence results and interpretation. The findings can show changes over time, therefore the effect of one variable on another cannot be identified as a single true effect, unless the variables do not change over any time period of interest. In conclusion, the findings about the study may be limited to the particular sample, variables and time frame exploited in the model, therefore the generalization of findings could lead to artificial conclusions (MacCallum and Austin 2000).

Another important limitation is the necessity of large sample sizes in order to maintain power and obtain stable parameter estimates and standard errors. Some methods are developed for parameter estimation of small sample size, after all estimates settle down at the smallest sample sizes, whereas standard errors are determined still according to larger number of samples (Bentler 2006). For example a chi-square test which is the most common fit measure is only recommended with moderate samples, in addition coefficient alpha, the most common measure of reliability, has several limitations; for instance coefficient alpha wrongly assumes that all items contribute equally to reliability (Shook et al. 2004).

8.3.4. Steps of SEM

In SEM applications, the researchers follow a logical sequence of five steps: model specification, model identification, model estimation, model testing, and model modification.

8.3.4.1. Model specification

Model specification involves using all of the available relevant theory, research, and information and developing a theoretical model. Thus, prior to any data collection or analysis, the researcher specifies a specific model that should be confirmed with variance and covariance data. In other words, available information is used to decide which variables to include in the theoretical model, and how these variables are related (Schumacker and Lomax 2004). For that purpose, the parameters are determined to be fixed or free. Fixed

parameters are not estimated from the data and are typically fixed at zero which is indicating no relationship between variables, whereas free parameters are estimated from the observed data and are believed by the investigator to be non zero. Asterisks are used for indicating the paths of free parameters. This step is stated as the hardest part of the SEM process (Cooley 1978).

8.3.4.2. Model identification

Identification refers to the amount of information available in the covariance matrix relative to the number of parameters that a researcher extracts from these data. The aim of the identification is satisfying the uniqueness of the parameters provided from the data. On the other hand, if the model cannot be identified, the different values for the parameters may satisfy the model. For illustrating the process, the process of determining the unique values of X and Y when the only available information is that X+Y=12, which is conceptually similar to the SEM, can be used. In this example, infinite number of solutions is available. Therefore, for determining a unique solution, the constraints, like setting X as a fix value, are required.

The parameters in a model can be specified either as a free parameter, a fixed parameter, or a constrained parameter. A free parameter is a parameter that is unknown and therefore needs to be estimated. A fixed parameter is a parameter that is not free, but is fixed to a specified value, typically 0 or 1. A constrained parameter is a parameter that is unknown, but is constrained to equal one or other parameters (Schumacker and Lomax 2004). In addition, model identification depends on the designation of parameters as fixed, free, or constrained due to its affect on the number of the parameters that required to be identified.

The necessary requirement for model identification is that the number of variance-covariance matrix observations equals or exceeds the number of the parameters estimated. According to this condition, the models are classified into three types; namely underidentified, just-identified or overidentified. Underidentification occurs when the number of variance-covariance matrix observations is not enough to obtain unique parameter estimates. Negative degree of freedom of a model refers to underidentification of the model; however it is not a necessary condition, in other words the model with positive degree of freedom can also be underidentified (Hershberger 2006). If a model is underidentified, then the parameter estimates are not to be trusted, however the model can become identified by adding constraints like the process given as an example.

Just-identification occurs when the number of variance-covariance matrix observations equals to the number of the parameters that leads to a unique solution, however, the just-identified model is not scientifically interesting because it has no degrees of freedom, in other words there is no way one can really test/confirm the veracity of this model, therefore the model can never be rejected (Byrne 2006). Overidentification occurs when there is more than one way of estimating a parameter (or parameters) because there is more than enough information in the model (Schumacker and Lomax 2004). The aim of SEM is to specify a model as overidentified, therefore the model has positive degrees of freedom that allows for testing/confirming the validity of the model.

8.3.4.3. Model estimation

The goal of estimation is to produce a $\sum(\theta)$ that converges upon the observed population covariance matrix, S, with the residual matrix (the difference between $\sum(\theta)$ and S) being minimized. When the elements in the matrix S minus the elements in the matrix $\sum(\theta)$ equal zero then $\chi^2 = 0$, that means a perfect model fit to the data. There are different methods available for estimation of the model, however, widely using models are maximum likelihood (ML), generalized least squares (GLS), least squares (LS), asymptotic distribution free (ADF), and robust.

Maximum likelihood (ML): is the most popular method among all estimation methods, because it provides unbiased, consistent estimates of population parameters, in addition admissible model solutions and stable parameter estimates can be obtained in case of small sample sizes (Meehan and Stuart 2007). ML methods are scale free, which means that the output of this method obtained by transformation of the scale of one or more of observed variables and output of untransformed output is related (Schumacker and Lomax 2004). However, ML assumes multivariate normality, and non-normality in the data can potentially lead to seriously misleading results. Nonetheless, ML is also determined quiet robust against the mild violation of the normality assumption (Chou and Bentler 1995; Curran et al. 1996; Muthen and Muthen 2002).

Generalized least squares (GLS): GLS is an adaptation of least squares to minimize the sum of the differences between observed and predicted covariances. It is probably the

second-most common estimation method after ML. GLS assumes multivariate normality like ML. The GLS estimation process requires much less computation than ML, however this should not be accepted as an advantage due to the powerful desktop computers with well-constructed SEM programs (Meehan and Stuart 2007). Olsson et al. (2000) evaluated the parameter estimates and overall fit of ML and GLS under different model conditions, including non-normality, and concluded that ML could provide more realistic indexes of overall fit and less biased parameter estimates for paths than did GLS under different misspecifications. In addition, ML performs better in small samples; therefore ML should generally be preferred with small sample sizes (Schermelleh-Engel et al. 2003).

Asymptotic distribution free (ADF): unlike the other methods, this method does not require multivariate normality assumption. Unfortunately, this method is data-intensive method, in other words, this methodology requires really large sample sizes to obtain model convergence and stable parameter estimates (Curran et al. 1996). However, Yuan and Bentler (1997) found out that ML is much less biased than ADF estimators for all distributions and sample sizes, but when the sample size is large ADF can compete with ML.

Robust: The strength of this method is that the normality is not requirement. In this method, the estimates provided from the methods are accepted, but the chi-square and standard errors are corrected to the non-normality situation. The chi-square test is corrected in the conceptual way described by Satorra and Bentler (1994). Also, robust standard errors developed by Bentler and Dijsktra (1985) are provided as an output of the robust analysis, and they are correct in large samples even if the distributional assumption regarding the variables is wrong (Bentler 2006). Although these robust statistics are computationally demanding, they have been shown to perform better than uncorrected statistics where the assumption of normality fails to hold and performs better than ADF (Chou et al. 1991; Hu et al. 1992). One important caveat regarding the use of robust statistics is that they can be computed only from raw data (Byrne 2006).

8.3.4.4. Model testing

After determination of the parameter estimates for a specified SEM model, the model should be examined against how well the data fit the data; in other words, how well the theoretical

model is supported by the obtained sample data should be determined. Different fit indices are developed for that purpose, namely chi-square, normed fit indices (NFI), non-normed fit indices (NNFI), comparative fit indices (CFI), Bollen fit indices (IFI), McDonald fit indices (MFI), goodness of fit indices (GFI) and root mean-square error of approximation (RMSEA).

Chi-square: it is calculated from the fitting function that is minimized by the estimation procedure used. Low and non-significant chi-square values are desired. There are two problems related to chi square as a fit index, firstly it has no upper, therefore the values of chi-square are not interpretable in standardized way. Second, it is very sensitive to sample size, in other words the chi-square values are affected from the sample size; for instance the same covariance matrix can provide different chi-square values with different sample sizes. Therefore, the chi-square statistic is divided by degree of freedom to reduce the sensitivity of this index to the sample size (Kline 1998).

Normed fit indices (NFI): The value of the NFI indicates the proportion in the improvement of the overall fit of the developed model to a null model. The NFI ranges from zero to one with higher values indicating better fit, however in small samples, it may not reach 1.0 when the model is correct (Bentler 2006). A rule of thumb for this index is that 0.95 is indicative of good fit relative to baseline model, while values greater than 0.90 may be interpreted as an acceptable fit (Schermelleh-Engel et al. 2003). The most important disadvantage of this index is that it is affected by sample size.

Non-normed fit indices (NNFI): Developed by Bentler and Bonnett (1980) by extending the work by Tucker and Lewis (1973) for the reliability of exploratory factor analysis models estimated by maximum likelihood, it is also known as Tucker-Lewis index (Bentler 2006). NNFI generally ranges from zero to one with higher values indicating better fit, since this index is not normed, values can fall outside the range 0-1. This index is also affected from the complexity of the model, in other words with increasing the complexity of the model, the values of this index decrease. The most important advantage of this index is that it is one of the fit indices less affected by sample size (Bentler 1990).

Comparative fit indices (CFI): It is developed by Bentler (1990) for avoiding the underestimation of fit in small samples. CFI is derived from the comparison between the hypothesized and independence models. The CFI ranges from zero to one with higher values indicating better fit. A rule of thumb for this index is that 0.95 is indicative of good fit relative to independence model, while values greater than 0.90 may be interpreted as an acceptable fit (Byrne 2006).

Bollen fit indices (IFI): It is developed to address the issues related to parsimony and sample size. IFI is computed basically by using the same procedure of NFI, except that degrees of freedom are considered (Byrne 2006)**.** A rule of thumb for this index is that 0.95 is indicative of good fit relative to independence model, while values greater than 0.90 may be interpreted as an acceptable fit (Meehan and Stuart 2007).

McDonald fit indices (MFI): It represents a normed measure of the centrality parameter that transforms the rescaled noncentrality parameter, which assesses model misfit. This index is similar to RMSEA, however this index does not provide for a fit degree of freedom interpretation (Byrne 2006). Values greater than 0.90 are mentioned as acceptable for MFI (Meehan and Stuart 2007).

Goodness of fit indices (GFI): It is developed by Jöreskog and Sörbom (1984) measures the relative amount of variances and covariances in the empirical covariance matrix S that is predicted by the model-implied covariance matrix, in other words it is the measure of how much better the model fits as compared to null model (Schermelleh-Engel et al. 2003). The GFI ranges from zero to one with higher values indicating better fit, but in some cases the values of GFI can fall outside the range 0-1. Values greater than 1 may be found with just identified models or overidentified models with almost perfect fit to the data; the reason of the negative values can be explained by the small sample size and poor fit of the model (Kline 1998). This means that the model fits worse than no model at all, this should not occur. A rule of thumb for this index is that 0.95 is indicative of good fit relative to baseline model, while values greater than 0.90 may be interpreted as an acceptable fit (Schermelleh-Engel et al. 2003). But, this index is not recommended due to being insufficiently and consistently sensitive to model misspecification and sample size (Byrne 2006).

Root mean-square error of approximation (RMSEA): It measures the degree of poor fit rather than measuring goodness of fit. The discrepancies between a candidate model and the covariance matrix from the population is measured, relative to the number of degrees of freedom in the model (Meehan and Stuart 2007), therefore this makes this index sensitive to the number of the estimated parameter in the model (Byrne 2006). MacCallum and Austin (2000) encouraged use of RMSEA for several reasons, firstly the RMSEA is adequately sensitive to model misspecification (Hu and Bentler 1998); and appropriate conclusions about model quality can be provided by utilizing interpretative guidelines (Hu and Bentler 1999; Hu and Bentler 1998); Finally and most importantly, a confidence interval is obtained as an output. RMSEA values smaller than 0.05 indicate a good fit, values between 0.05 and 0.08 indicate a an adequate fit, and values between 0.08 and 0.10 indicate a mediocre fit, whereas the values larger than 0.10 indicate that the model should not be accepted (Schermelleh-Engel et al. 2003). However, when the samples size is small, this index tends to over-reject true population models (Hu and Bentler 1999).

8.3.4.5. Model modification

When the covariance/variance matrix estimated by the model does not fit the sample covariance/variance matrix, then the model is modified and subsequently the new modified model is evaluated. The model can be modified by adding new pathways or removing the original ones. In addition, parameters can be changed from fix to free or from free to fix. Also, the different constraints can be applied to improve the model. Different procedures are available for determining how to modify the model, namely Lagrange Multiplier Index (LMtest) and the Wald test (Wtest). LMtest procedure is designed to test hypotheses on the statistical necessity of restrictions that exist in a model. The first type of restriction tested is whether the equality constraints that may have been imposed in a model are appropriate. The second type of restriction tested is whether fixed parameters, such as “missing" paths or covariances that are set to zero in the model, and hence would be better treated as free parameters and estimated in a future run. Univariate and multivariate LM statistics are produced to permit evaluation of the statistical necessity of these restrictions. Wtest is designed to determine whether sets of parameters that were treated as free in the model could be simultaneously set to zero without substantial loss in model fit. However, Wtests require quite large samples to be trustworthy (Bentler 2006).

8.3.5. Factors to consider when modeling

Sample Size

In structural equation modeling, for maintaining power and obtaining stable parameter estimates and standard errors, a larger sample size is required. The need for larger sample sizes is also due in part to the program requirements and multiple observed indicator variables used to define latent variables (Schumacker and Lomax 2004). As a result, the quality of results may occur from a given study with small samples will depend on the features of the model of interest, in other words, estimates settle down at the smallest sample sizes, whereas standard errors are determined still according to larger number of samples (Bentler 2006). There are different advices for sample sizes, for instance, Ding et al. (1995) found numerous studies that were in agreement that 100 to 150 subjects is the minimum satisfactory sample size, Boomsma (1982) recommended 400. Some sources advocate a minimum ratio of parameters estimated to the sample size, like 1:10, 1:20, as per multiple regressions (Bentler and Chou 1987; Stevens 1996). Unfortunately, the features of SEM also affect the minimum number of the sample size. For instance, the recommended sample size will vary with model complexity, because models with more variables and regression coefficients, correlations, and variances to estimate will require more data to produce stable and interpretable results (Meehan and Stuart 2007). The degree and kind of normality also affect the minimum number of sample; the data set is becoming alienated from the normality, the required sample size increase (Finch 1997; Muthen and Muthen 2002). In addition, the missing data also affect the minimum sample size, with increasing the number of the missing data, the required minimum number of sample size is increased approximately 18% (Muthen and Muthen 2002).

Normality

As mentioned before, most of the estimation methods assume normal distributions for continuous variables. Nonnormal data can occur because of scaling of variables or limited sampling of subjects, and the normality can on two levels; univariate and multivariate normality. Univariate normality concerns the distributions of the individual variables. The normality of a variable can be evaluated according to the standard error of skewness and kurtosis. The standard error of skewness is calculated by the ratio of skewness to its standard error, the values of standard error of skewness between -2 and 2 indicate the normality of the variable. A large positive value for skewness indicates a long right tail; an extreme negative

value indicates a long left tail. The standard error of kurtosis is calculated by the ratio of kurtosis to its standard error, the values of standard error of kurtosis between -2 and 2 indicate the normality of the variable. A large positive value for kurtosis indicates that the tails of the distribution are longer than those of a normal distribution; a negative value for kurtosis indicates shorter tails.

Missing Data

Little and Rubin (1987) mentioned that there are three kinds of missing data situations. Data missing completely at random occurs when the values of the missing scores are unrelated to why these data are missing, although this is probably a rare occurrence in applied settings. Data missing at random is when the values of the missing scores are related to why the data are missing but when the missing data values can be predicted by information that is available to the researcher. Non-ignorable data implies probabilistic information about the values that would have been observed. Different methods are available for handling these kinds of missing data; the following table summarizes some of these methods.

Table 8-5: Options for dealing with missing data (Schumacker and Lomax 2004)

Listwise	Delete subjects with missing data on any variable
Pairwise	Delete subjects with missing data on only the two variables used
Mean Substitution	Substitute the mean for missing values of a variable
Regression imputation	Substitute a predicted value for the missing value of a variable
Maximum likelihood	Find expected value based on maximum likelihood parameter estimation
Matching response pattern	Match variables with incomplete data to variables with complete data to determine a missing value

Listwise, pairwise, mean substitution, and regression imputation methods can be implemented practically, besides these methods can be found in most well-known statistical

program packages, however these methods are only appropriate when the amount of missing data is low, in addition these methods lead problems in analysis of the data; for instance, listwise deletion method can cause losing a large number of subjects, thus dramatically reducing the sample size. Pairwise deletion can compensate the reduction of the sample size by excluding data only when they are missing on the variables selected for analysis, however this could lead to different sample size for correlations and related statistical estimates, whereas maximum likelihood and multiple imputation outperform the other methods, as they maximize the amount of data used while producing unbiased parameter and standard error estimates (Allison 2003).

8.4. Research methodology:

In order to test the empirical validity of the methods and drivers of differentiation, data collected from a total of 62 questionnaires were analyzed using a software package called EQS 6.1, an SEM tool. The structure of the conceptual model that is composed of a number of direct and indirect interdependencies between the independent and dependent variables necessitated the usage of SEM.

In the first step, the method to be used for determining the parameters should be identified. For that purpose, the univariate kurtosis values shown in Table 8-6 were calculated for evaluating the normality of the variables. When variables demonstrate significant nonzero univariate kurtosis, it is certain that they will not be multivariately normally distributed. Kurtosis values for several variables (in particular, Project organization, Relationship with subcontractor, Technology, and Human resource management) bear serious considerations, however some of the kurtosis values can be considered as satisfactory. Therefore, for evaluating the normality, Mardia's normalized multivariate kurtosis was also considered. It should be roughly in the +3 to -3 range, if the data is multivariate normal. For each factor, the method used in model estimations was decided according to the kurtosis values of the variables forming that factor shown in Table 8-6 and Mardia's coefficients.

Table 8-6: Kurtosis values of the variables

Variables		Kurtosis
Project Management	Schedule Management	-0.850
	Project Organization	2.286
	Contract Management	-0.088
	Resource Management	-0.775
	Project Monitoring & Control	0.660
	Quality Management	-0.851
	Risk Management	-0.652
	Project Cost Estimation	1.102
	Environmental Management	-0.124
	Health and Safety Management	-0.863
Corporate Management	Strategic Planning	-0.624
	Business development	-0.752
	Human Resource Management	-1.170
	Financial Management	-0.893
	Instut. Stru.& Profess. Mana.	-0.798
	Organizational Learning	-0.580
	Research and Development	-0.045
	Tendering	-0.311
	Claim Management	-0.927
	Knowledge Management	1.149
Resources	Human Resources	-0.557
	Machinery/ Equipment	0.150
	Financial Resources	-0.823
	Experience& Knowledge	-0.429
	Technology	1.242
Relationships	Relationship with the suppliers	1.367
	Relationship with the subcontractors	2.302
	Relationship with the clients	-0.471
	Relationship with the competitors	-0.121
	Relationship with the other parties	0.395
	Relationship with internal stakeholders	0.046
Differentiation ways	Productivity	-0.696
	Service Quality	-0.211
	Product Quality	-0.215
	Time	-0.696
	Innovative Solutions	-0.775
	Positive Image	-0.439

8.4.1. Factor loading

As mentioned before, six variables are determined as methods/modes of differentiation according to the literature survey. In order to test whether these factors provide a good representation of the inter-correlations among indicators, a preliminary exploratory factor analysis was performed. According to Table 8-7, two factors were identified, first factor is

consisted of “service quality”, “product quality” and “positive image”, and second factor is consisted of “productivity”, “time”, and “innovative solutions”. According to variables involved, the first factor and second factor can be renamed as quality related differentiation, and productivity related differentiation, respectively.

Table 8-7 : Factor analysis of differentiation ways

Rotated Component Matrix		
	Component	
	1	2
Productivity	0.082	0.675
Service Quality	0.912	0.128
Product Quality	0.871	-0.010
Time	0.157	0.712
Innovative Solutions	0.047	0.794
Positive Image	0.480	0.149

The Mardia’s coefficients for these factors were calculated as -1.5840 and 0.6993 which are in the range of +3 to -3; therefore ML was decided to be used for determining the fit indices and significance of factor loadings. Three paths from the variables to the factor was decided as free, in addition the variances of errors were taken to be free parameters to be estimated, also variance of factor was also a free parameter, in other words there were seven free parameters to be estimated in these models, but a quick count to data involved, namely three variances of variables, and their covariance, indicated that there were six data points. Thus the models had more parameters than the data, and could not be uniquely estimated, in other words, the models were under-identified; therefore the constraints should be placed on these parameters in order to provide more data than effective parameters. For example, the residual variances of errors could be set equal to each other, so two parameters are eliminated, and then the models can be identified.

After determining the number of the factors, the reliability of these factors should be checked. Cronbach’s alpha coefficient, which measures the extent to which an item’s responses obtained at the same time are highly correlated with each other. It can be used for checking internal consistency and measures how consistently individuals respond to the

items within a scale. According to the EQS analysis results, Cronbach's alpha values, calculated as 0.675 for "productivity related differentiation", 0.690 for "quality related differentiation", were approximately equal to the threshold of 0.70 recommended by (Nunally 1978).

In addition, the fit indices of these factors were also examined for evaluating the unidimensionality, referring to the degree to which constituent variables represent one underlying latent variable of the factors. The results of the fit indices are represented in Table 8-9. In this study, the non-normed fit index (NNFI), the comparative fit index (CFI); Bollen's fit index (IFI), McDonald's fit index (MFI) and the root mean squared error of approximation (RMSEA) were used to evaluate the fit of the model. For the productivity related factor, all of the indices were determined above the mentioned thresholds for these indices (above 0.9), however the values of NNFI, CFI, IFI and MFI were calculated above one, which is the limit of these errors. The reason of this can be explained by sample size and degree of freedom. For the quality related factor, all of the indices except the NNFI were determined above the threshold values.

Another indicator of poor model fit is the presence of standard errors that are excessively large or small. For example, if a standard error approaches zero, the test statistic for its related parameter cannot be defined (Bentler 2006). Likewise, standard errors that are extremely large indicate parameters that cannot be determined. However, no definitive criterion is existed for "small" and "large" values of standard errors due to the dependence of the standard errors on units of the observed variables and the magnitude of the parameter estimate (Byrne 2006). On the other hand, the statistical significance of the parameter estimates can be improved by division of the parameter estimate by its standard error and it can be used for determining the non-significant parameters in the model. This evaluation is based on the 0.05 significance level. For both of the factors, all variables were determined as significant, consequently according to the fit indices and significance of the parameter estimates, the fitness of the factors for differentiation methods are quite satisfactory.

The same procedure applied for differentiation methods was applied for project management factor, firstly, according to the Mardia's coefficient, the method applied was determined. Since Mardia's coefficient was calculated as -3.895, ML-robust method was used in this factor. According to the Cronbach's alpha (0.798), the internal consistency was provided for this factor, too. However, the fit indices for this factor formed by using all variables determined in the conceptual model were determined under recommended perfect values,

therefore some modifications were required for improving these indices. For that purpose, the correlation matrix of the variables in terms of "project management" was examined. According to the Table C.5 (given in Appendix C), the "environmental management" and "health and safety management" variables were determined to be highly correlated (0.706). Because of emergence of "multicollinearity" between these variables, either one or the other should be excluded in the analysis and it makes little sense to include both in the same analysis because they are identical (Kline 1998). So by including one of these two variables at each time, analysis was repeated, as a result, the fit indices were improved, however they did not satisfy the recommended perfect values. According to Figure 8-10, these two variables were also highly correlated with "quality management"; therefore both of these two variables were excluded from the analysis. The analysis was repeated without these variables; however, as the Mardia's coefficient is - 1.5482, ML method was used in this analysis. At the end, all of the fit indices were satisfied with the recommended perfect values and all parameter estimates were determined at the 0.05 significance level.

Factor loading analysis was also conducted for corporate management. Firstly, for determining which method would be used in this analysis, Mardia's coefficient was examined, and according to the Mardia's coefficient that is equal to - 6.068, ML-robust method was used in this analysis. In the first round, claim management was determined as insignificant, and the fit indices of the model were under the recommended perfect values, therefore the claim management was excluded from the analysis. In the second analysis, all parameter estimates were determined as significant; however the fit indices were not good enough for satisfying the recommended thresholds. Therefore, the correlation matrix of the variables related with corporate management was examined. According to the Table C.7, the correlation between "organizational learning" and "professional management" was higher than the other correlations; therefore one of these variables was decided to be excluded from the analysis. In order to decide which variable should be eliminated from this factor loading, the analysis was repeated by excluding these variables separately. By comparing the determined fit indices of the models, the "organizational learning" variable was excluded from the analysis. Consequently, the last factor loading of corporate management was performed by the remaining eight variables by ML-robust method. The internal consistency was examined according to the Cronbach's alpha (0.740), it was determined as satisfactory according to the recommended values. In the end, all of the fit indices were satisfied with the recommended perfect values, as shown in Table 8-9, and all parameter estimates were determined at the 0.05 significance level.

The factor loading of resources were also performed according to the significance of parameter estimates and fit indices. But, firstly the method used in this analysis was determined as ML, as Mardia's coefficient was determined as - 0.523. For evaluating the "unidimensionality" of the factor loading, the fit indices and significance of the parameters were examined, and according to Table 8-8 and Table 8-9, the factor loading for the resources was determined as satisfactory. In addition, the internal consistency was also examined, and it was determined as satisfactory, so, no modification was performed for this model.

Finally, the factor loading of relationships were carried out. At the beginning of the analysis, six variables were available for this factor. However, "relationship with the suppliers" and "relationship with the subcontractors" were highly correlated (0.890). For avoiding "multicollinearity" in the analysis, one of the variables was eliminated by performing analyses for each variable. Consequently, according to the fit indices, "relationship with the subcontractors" was eliminated from the analysis. The analysis was re-started by determining the method used in this analysis, and according to Mardia's coefficient that is - 2.068, ML method was used. For evaluating the "unidimensionality" of the factor loading, the fit indices and significance of the parameters were examined, and according to Table 8-8 and Table 8-9, the factor loading for the resources was determined as satisfactory. Finally, the internal consistency was examined according to Cronbach's alpha that is (0.734) which is above the recommended threshold value, so the analysis was determined as satisfactory according to the required all conditions.

Table 8-8: Latent and continent variables of the model

Variables	Factor loading
Project Management	
Schedule Management	0.523
Project Organization	0.587
Contract Management	0.556
Resource Management	0.614
Project Monitoring/ Control	0.610
Quality Management	0.576
Risk Management	0.510
Project Cost Estimate	0.569
Environmental Management	0.528*
Health and Safety Management	0.550*
Corporate Management	
Strategic Planning	0.539
Business Development	0.697
Human Resources Management	0.58
Financial Management	0.441
Institutional Structure& Professional Management	0.475
Research and Development	0.63
Knowledge Management	0.403
Bidding	0.393
Organizational Learning	0.256*
Claim Management	0.043*
Resources	
Human Resource	0.414
Machine & Equipment	0.697
Financial Resources	0.500
Experience & Knowledge	0.434
Technology	0.681
Relationship	
Relationship with the suppliers	0.642
Relationship with the clients	0.544
Relationship with the competitors	0.684
Relationship with the other parties	0.779
Relationship with the internal stakeholders	0.309

*Eliminated variables due to "multicollinearity" and not significant at 0.05 level.

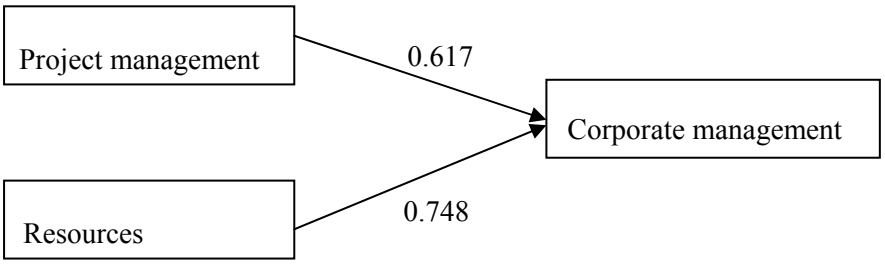
Table 8-9: Reliability values and fit indices for the constructs of the model

	Drivers of Differentiation				Dimensions of Differentiation	
	Project Management	Corporate Management	Resources	Relationships	Productivity related variables	Quality related variables
Cronbach's Alpha	0.786 (0.798)	0.740 (0.717)	0.688	0.734	0.675	0.69
NNFI	0.936 (0.535)	0.893 (0.584)	0.814	1.025	1.160	0.598
CFI	0.956 (0.649)	0.928 (0.686)	0.926	1	1	0.907
IFI	0.960 (0.676)	0.935 (0.712)	0.936	1.009	1.025	0.914
MFI	0.969 (0.687)	0.955 (0.736)	0.977	1.005	1.004	0.964
RMSEA	0.058 (0.150)	0.070 (0.135)	0.104	0	0	0.273
Chi square/dof	22.945/19= 1.208 (80.4782/34= 2.367)	24.658/19= 1.298 (72.0356/34= 2.119)	6.927/4= 1.732	3.442/4= 0.861	0.539/1= 0.539	2.214/1= 2.214

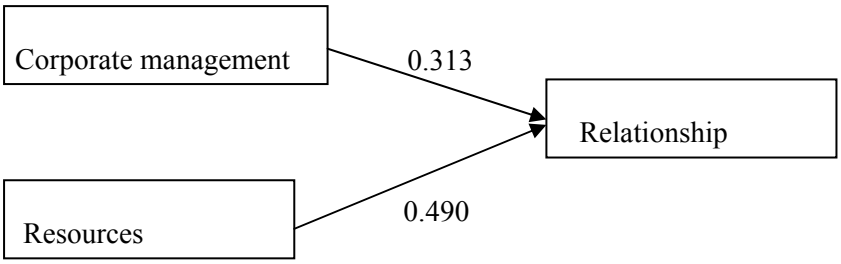
Numbers in parentheses belong to initial analysis before the model improvement.

8.4.2. Analysis of the structural equation model

In this model, due to the Mardia's coefficient that was calculated as - 21.199, and the small-sized sample, ML- robust methodology was used. In the first step, relationships among drivers of differentiation were analyzed. For that purpose, different hypothesized relationships for each driver were developed, according to significance of the parameter estimates, these hypothesized relationships were evaluated. At the end, the interrelationships for "corporate management" and "relationships" which are shown in Figure 8-16 were verified.



Corporate Management= 0.617*project management+ 0.748*resources+ 0.245D (R^2=0.940)



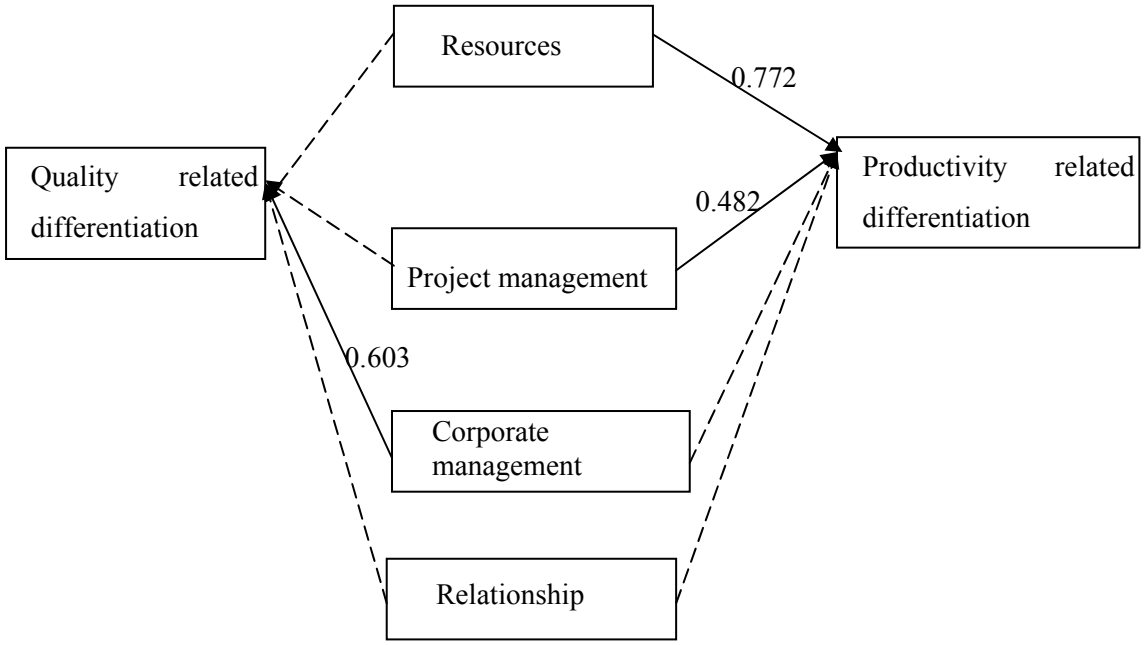
Relationship= 0.313*corporate management+ 0.490*resources+ 0.657D (R^2=0.568)

Figure 8-16: Interrelationships among the drivers of differentiation

According to this figure, it is determined that the "project management" and "resources" have direct and significant effects at the 0.05 significance level on 'corporate management". Also, "corporate management" and "resources" affect "relations" directly, in addition "project management" does not have a significant direct effect on relations, and however it has an indirect effect on relations through "corporate management".

In the second step of the structural model, the influences of each driving factor on "differentiation methods" were computed. Figure 8-17 shows the hypothesized relations

between the constructs of the model. The analysis suggests that "resources" is the main driver of "productivity related differentiation" with a path coefficient of 0.772. "Project management" has a moderate effect on "productivity related differentiation" (path coefficient: 0.482) according to the effect of the "resources". On the other hand, no significant influence of "corporate management" and "relationship" on "productivity related differentiation" was found. For "quality related differentiation", the influence of "corporate management" was determined as significant. On the other hand, "resources", "project management" and "relations" do not have significant direct effect on "quality related differentiation". However, "project management" and "resources" have an indirect effect on "quality related differentiation" through "corporate management". The most interesting finding is that "relationship" does not have any significant effect on differentiation ways. Links that are not statistically significant at 5% were eliminated from the model (shown in dashed lines in Figure 8-17).



Quality related differentiation= 0.603*corporate management + 0.698D (R^2 = 0.450)

Productivity related differentiation= 0.482*project management + 0.772*resources + 0.414D (R^2=0.829)

Figure 8-17: The structural equation model

Finally, the reliability values and fit indices for this model, as shown in Table 8-10 were examined. Cronbach's alpha, calculated as 0.888, shows that the internal consistency of the structural model is satisfied. Moreover, the $\chi 2$ to dof ratios, calculated as 1.094, is smaller than 3 as suggested by (Kline 1998). In addition, all of the fit indices except MFI are higher than the recommended perfect values of 0.90 demonstrate a good fit of the model to the data. Finally, the RMSEA value was examined and it was determined below the recommended perfect value of 0.10. As a conclusion, the model developed for this data set shows a satisfactory fit for the data. The overall model which shows the final factor loadings, marked next to light arrows, and path coefficients marked by the heavy arrows are presented in Figure 8-18.

Table 8-10: Fit indices for the model

Cronbach's alpha	0.888
NNFI	0.933
CFI	0.939
IFI	0.942
MFI	0.708
RMSEA	0.039
Chi square/dof	497.783/455=1.094

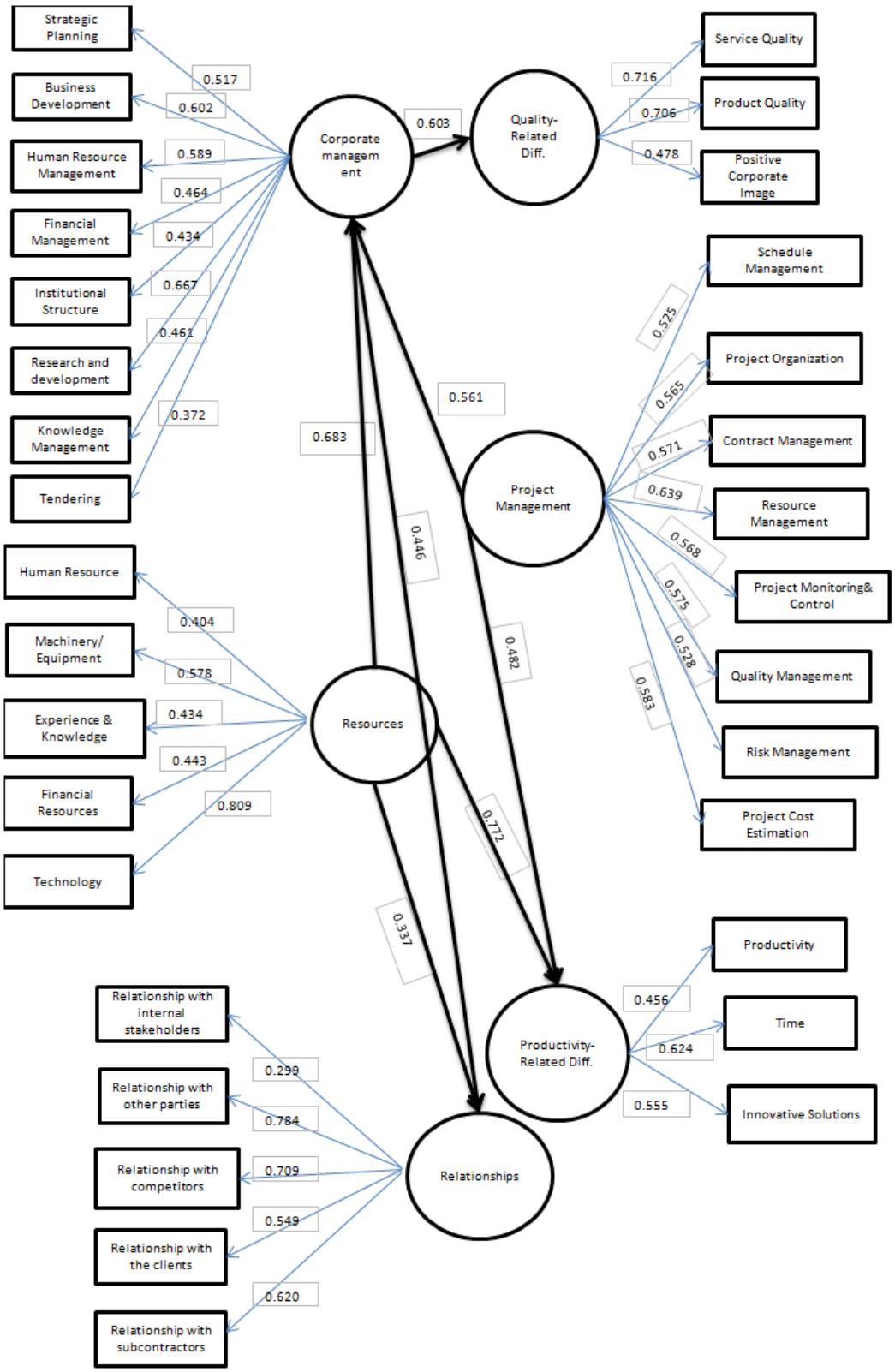


Figure 8-18: Overall structural model

8.5. Discussion of findings: Modes of differentiation

There are two choices that should be made by contractors in order to differentiate; first, selection of the mode of differentiation and second deciding on how to differentiate according to the selected mode. In this research, it is found that two modes of differentiation are possible in the Turkish construction industry. These modes were identified by exploratory factor analysis of six differentiation modes determined according to the buying criteria of the clients. They were termed as "quality related differentiation" and "productivity related differentiation" respectively. "Service quality", "product quality" and "positive image" were identified under the heading of "quality related differentiation"; on the other hand, "productivity", "time", and "innovative solutions" formed the "productivity related differentiation" mode. The first category is about creating more value to the customer by providing a differentiated service or product based on the highest quality, and creating a positive corporate image. The second category is about finishing the project on time, at the lowest cost by increasing productivity and providing innovative solutions to give the client a differentiated service to basically meet his expectations about cost and time. Schematic representation of findings is depicted in Figure 8-19 and Figure 8-20.

"Productivity", "time", and "innovative solutions" are determined as valid indicators of "productivity related differentiation". "Time" is observed to have the highest correlation with "product related differentiation" among other indicators. This is an expected result; since time is considered as one of the important success criteria in construction projects. Construction companies have to deliver the projects according to the certain milestones and before the final date of completion specified in contract documents. However, delays are commonly observed in construction projects. While some of the causes of delays can be prevented; thus, there is room for improving the speed of execution by advancing their business processes, capabilities and resources. Consequently, the construction companies can influence the completion date of their projects. The importance of time in differentiation was also mentioned by many authors, like Kale and Arditi (2003), Warszawski (1996), Calori and Ardisson (1988). Also, the findings show that the construction companies believe that they can differentiate by introducing product/process innovations to the market. The construction is usually classified as a traditional or low-technology sector (Manseau and Seaden 2001). The innovations such as construction materials and equipment that may occur in the construction industry are usually results of the efforts and researches of other industries; therefore the proprietary technology developed by the companies can lead to

uniqueness. Companies that can remove the barriers for innovation in the construction industry may create a major source of competitive advantage. Finally, due to the low productivity rates associated with the construction industry, companies that can achieve higher productivity rates may create an advantage due to this differentiated ability from their rivals. In conclusion, the construction companies can differentiate by achieving on-schedule performance in construction projects, even attempt to deliver constructed facilities ahead of the schedule, using the various factors of production like labor, equipment, and capital effectively, and developing product and process innovations that would be valued by their clients.

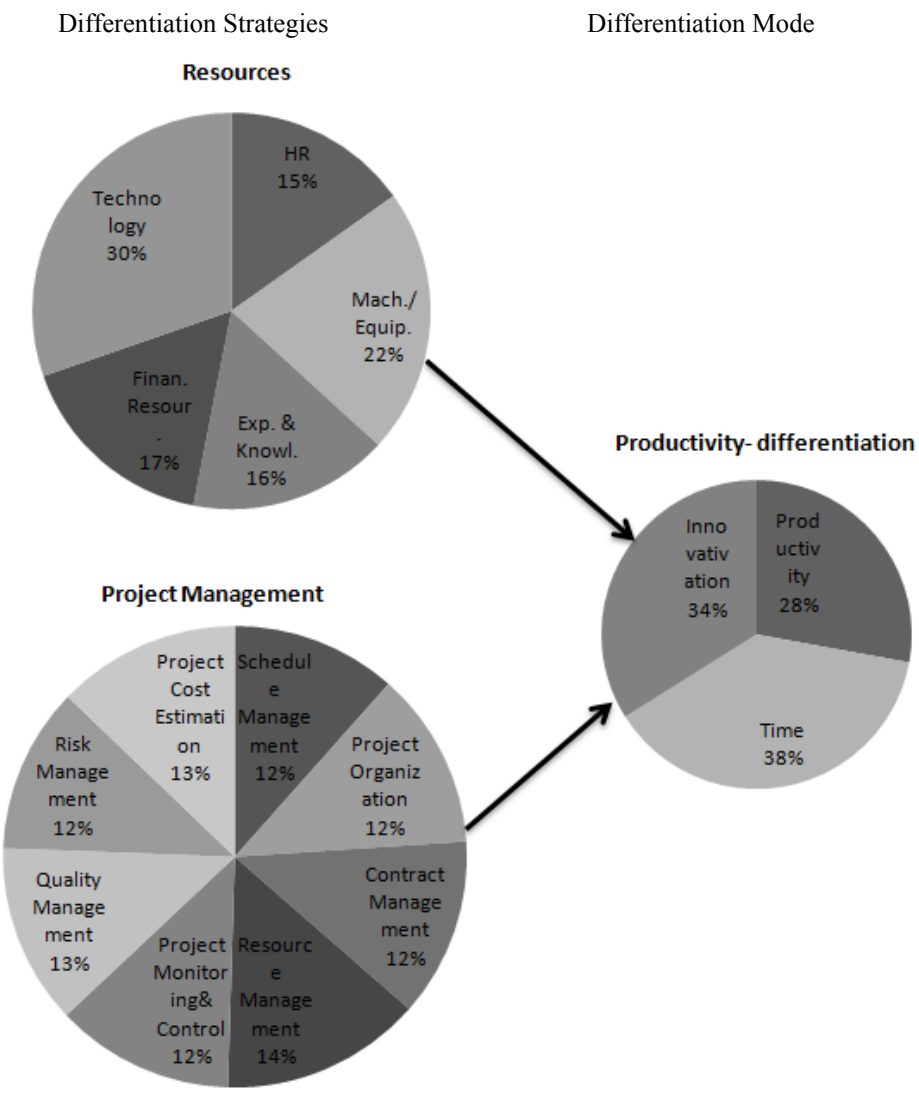


Figure 8-19: Strategies that support productivity-differentiation

The analysis suggests that "product quality", "service quality", and "positive corporate image" are all valid indicators that correspond to different and complementary dimensions of "quality related differentiation". Although all these indicators have high factor loadings, "product quality", and "service quality" are observed to show slightly higher correlation with "quality related differentiation". The quality is expected as a differentiation mode, since the construction companies can influence product quality by achieving high quality beyond the requirements in the specifications in constructed facility (Kale and Arditi 2003), also they can extend their services by influencing the quality of its contracting services through improving communications with clients and its consultants (Warszawski 1996; Yasamis et al. 2002). Besides, the companies can improve their corporate image by advertising, high quality sales force, and promotion for providing unique and positive messages, and unique communication activities (Boulding et al. 1994). It is interesting to note that the companies which try to differentiate based on quality should improve all of these dimensions; in other words, improving one of these dimensions does not guarantee the quality differentiation, the companies should consider all of these dimensions in order to have a competitive position based on quality differentiation. In conclusion, the construction companies can differentiate on quality by achieving high quality beyond the requirements in the specifications, providing high quality services to their clients, and strengthening image of the company in the eyes of the customers.

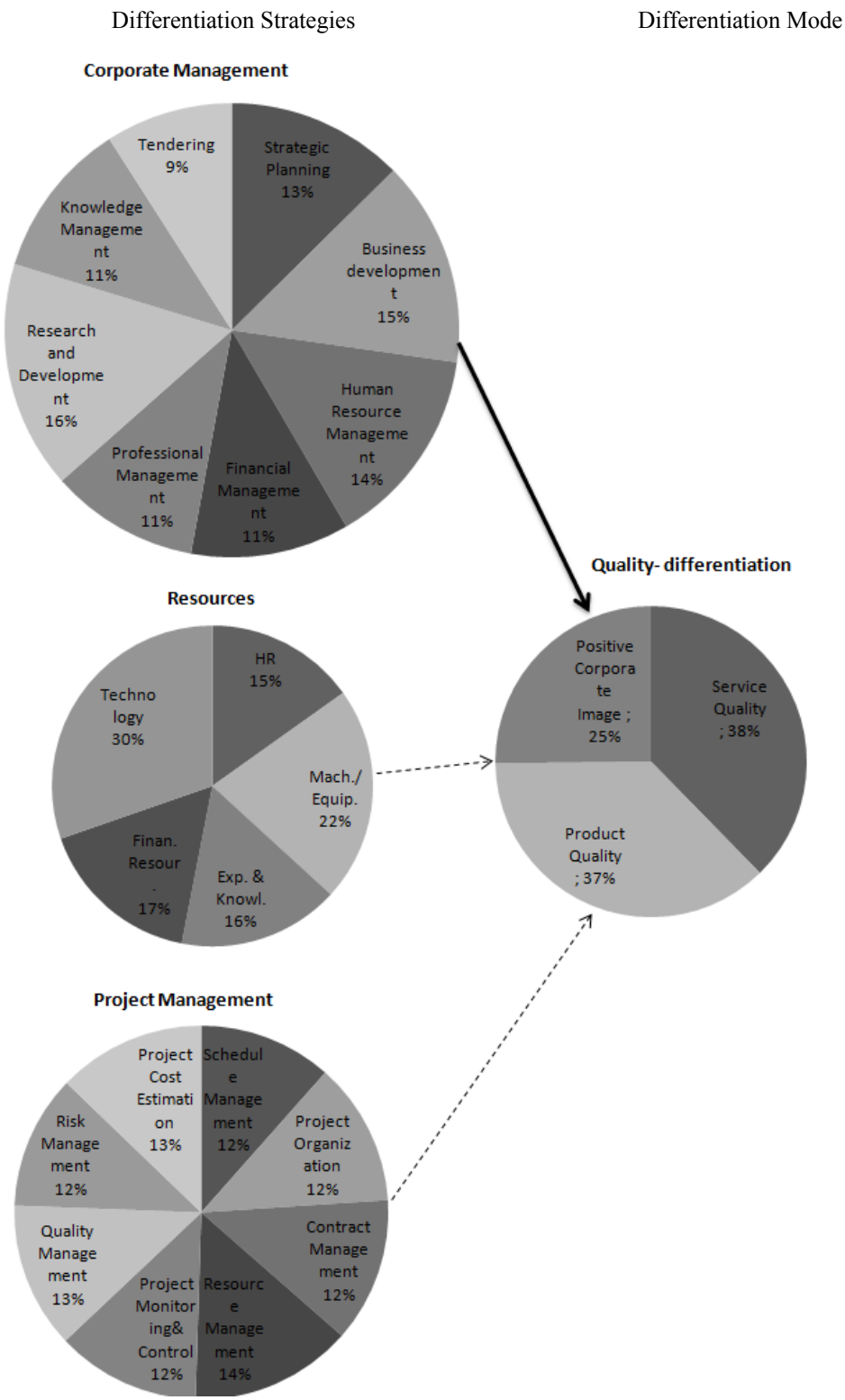


Figure 8-20: Strategies that support quality-differentiation

Another interesting finding of this research is that average of all differentiation modes was calculated above 4. Based on this, it can be concluded that the construction companies believe that the differentiation strategies are effective ways to create competitive advantage and all of the identified modes of differentiation are applicable to the Turkish construction industry.

8.6. Discussion of findings: About the relationships within the model

The drivers employed to achieve differentiation are important inasmuch as the selection of the differentiation modes. Statistical findings result in four constructs, which are proposed to be the drivers of "differentiation". In the measurement model, all constructs of the model were validated; indicators not having significant factor loadings were deleted, satisfactory values for reliability and fit indices were achieved. After validating all constructs within the model, the hypothesized relations between these constructs are tested.

Project Management

The construction industry is a project-based industry; therefore the construction companies can achieve competitive advantage by differentiating themselves through the project management activities (Kezsbom and Edward 2001). In other words, the project management activities can furnish opportunities to drive differentiated products/services. As one of the most important factors of differentiation, "project management" extensively influences "productivity related differentiation" and influences "quality related differentiation" indirectly through "corporate management". Also, according to De Maio et al. (1994), "project management" plays an important role in new product development process.

If one relies on the high factor loadings of the variables that constitute the "project management" construct, the companies should give the highest priority to these activities in order to differentiate themselves based on productivity. Although all variables under heading of project management were determined significant within the model, "resource management" is observed to have the highest correlation among these variables. In other words, the companies should allocate the resources effectively and control the level of the resources throughout the project for securing differentiation. However, due to involvement of many different suppliers in a project, the coordination between these parties is a difficult

process and supply chain management is a critical success factor. It is believed that the traditional supply chain does not provide a basis for the true marketplace differentiation (Bovet and Martha 2000), therefore the companies should re-engineer their supply chains in order to achieve productivity and quality enhancements in their business (Kumaraswamy et al. 2006). Also, the companies can sustain their differentiation strategies based on supply chain management strategies, since the imitation of supply management strategies is difficult for the other companies due to the requirement of unique, experienced, and well-coordinated relationships between multiple parties in the supply chain (Mentzer 2004). Due to the low ratings of the variables, namely "health and safety management" and "environmental management" are excluded from the drivers of the differentiation. The low rating of "environmental management" could be explained by the argument that many managers view environmental management as compliance with environmental regulations (Klassen and McLaughlin 1996), since the environmental regulations are standardized, they may believe that the opportunities that lead to uniqueness is limited in "environmental management". This explanation could also be valid for "health and safety management".

Corporate Management

Although construction industry is a project-based industry, the corporate management activities constitute an important part of value chain of the business; therefore providing uniqueness in these activities is also expected to drive differentiation. As expected, the activities of "corporate management" extensively affect "quality based differentiation", however no statistically significant relation was observed between "corporate management" and "productivity related differentiation". This indicates that creation of quality culture is required in the construction industry in order to create and maintain quality performance, which in turn achieve differentiation advantage based on quality (Kanji and Wong 1998), therefore firstly the corporate management should espouse quality culture in their activities, in other words the managers should decide to improve the quality in their business, and reflect this purpose by their managerial decisions. Among the eight corporate management-related variables considered in this study, "research and development" ," business development" and "human resource management" are the most significant indicators based on their factor loadings. This finding demonstrates the importance of effectiveness/ uniqueness and differentiation in the value chain at the level of corporate management to provide a highest quality product/service to the client. In order to achieve a strong

competitive position based on high quality and differentiated services, the firms must be flexible enough to quickly shift production and organizational resources to meet changing markets and customer demands (Arthur 1992). Therefore, human resources must have the skill and training to perform a variety of different tasks; this demonstrates the importance of "human resource management". Research and development is an activity that is not carried out by majority of the contractors, thus offers a critical opportunity for firms to differentiate themselves from their rivals. Similarly, business development activity is of utmost importance as it directly deals with finding job opportunities all over the world and planning for the suitable future investments by considering the strengths and weaknesses of the company. Differentiating in this task has a potential to create competitive advantage based on high quality services/products.

Resources

The importance of the resources in order to gain competitive advantage was extensively emphasized in the literature (Amit and Schoemaker 1993; Barney 1991; Dierickx and Cool 1989; Wernerfelt 1984). Due to within company decision making, external strategic factors, and market imperfections, resources and capabilities cannot be assumed to be identical in each company (Oliver 1997). This leads to creation of firm-specific resources which are difficult to imitate and transfer and these unique resources drive to the differentiation ways of the companies. According to Mosakowski (1993), unique or specialized resources are required in order to realize differentiation strategies. In this study, as expected, "resources" are found to influence "productivity related differentiation" extensively, and "quality related differentiation" indirectly through "corporate management". It appears that the companies differentiate based on product should enhance resources in terms of technology (factor loading: 0.809), machinery and equipment (0.578), financial resources (0.443), experience and knowledge (0.434) and human resources (0.404). Although all these resources have moderate factor loadings, "technology" is observed to have the highest correlation with "resources". Porter (1985) also mentioned about the importance of technology for differentiation due to involvement of technology in achieving linkages among activities. Also, although the competitors can imitate the technology, relative advantage can be created and sustained due to the other complementary critical resources which make it difficult for competitors to utilize from the technology at same level (Keittinger et al. 1994). Especially, the role of information technologies in creating competitive advantage for firms through

differentiation has been mentioned in the literature (Bhatt and Grover 2005; Mata et al. 1995; Powell and Dent-Micallef 1997). Also, information technologies can enhance the productivity and service delivery by improving integration and facilitating communication in construction companies (Bjork 1999).

Relationships

Due to the labor intensive and project based structure in which a number of parties are involved, unique and long-term relationships based on trust and mutual dependence was expected to drive differentiation. Furthermore, strong and positive effect of high quality relationships with parties on the performance of the companies are mentioned by many authors (Betts and Ofori 1992; Hausman 2001; Kale and Arditi 2001). However, in this study it is surprising that data analysis did not provide any evidence of a significant direct or indirect relationship between "relationships" and "productivity based differentiation" as well as "relationship" and "quality based differentiation". This can be explained by the culture of construction industry which depends on negative relationships and attitudes of the companies towards each other due to traditional rivalry between the parties. Also, due to existence of ample number of suppliers and subcontractors, the companies engage in opportunistic behavior even though they are happy with existing suppliers and subcontractors, in addition, they fear of dependence on a smaller number set of suppliers and subcontractors by reducing the supplier and subcontracts choices (Sheth and Sharma 1997). In spite of no emergence of a link between the differentiation modes and relationship, according to the average ratings of the variables, it could be inferred that the construction companies also realize the importance of "relationship with the clients" (4.45). Fisher (1991) also emphasizes the importance of good buyer-seller relationships in order to sustain differentiation strategies.

As a conclusion, statistical findings can constitute a strong foundation on which the companies can build a differentiated position and choose the best mode of differentiation according to their strengths and weaknesses in their value chain. Likewise, they can try to develop necessary abilities necessary for achieving a competitive position based on a given mode of differentiation.

CHAPTER 9

CONCLUSION

The necessity of having a clear strategic perspective in order to achieve competitive advantage has been acknowledged in the literature. The selected strategies should be in accordance with objectives, competencies and competitive rules prevailing in the market. It is clear that a number of companies that compete in the same industry may have similar resources/competencies and develop similar strategic perspectives, therefore it is expected that these companies show similar performance, but this may not always hold true depending on the competitive rules prevailing in a market. Consequently, the relationship between the performance and strategies should be examined by analyzing the competitive structure of a market. The strategic group analysis is one of the methods for revealing the competitive structure. By conducting a strategic group analysis, the companies can figure out current strategic position of a firm within the competitive environment and formulate strategies in order to shift to a better performing cluster.

The first part of this study deals with the "strategic group analysis" of the major Turkish construction companies. In this part, it was aimed to carry out strategic group analysis by using different cluster analysis methods in order to identify clusters having different strategic positions. Then, the performance differences between the firms in different strategic groups were intended to be tested.

Initially, an extensive literature survey was conducted in order to develop a conceptual framework that fits to the characteristics of the construction industry. Consequently, a conceptual model was proposed based on the three dimensions of strategy; namely strategy content, strategy process and strategy context, as mentioned by Price and Newson (2003). A questionnaire survey which reflects the conceptual model developed for this study was administered to 136 medium-big Turkish construction companies. The total number of returned questionnaires was 84 leading to a return rate of 0.62. The respondent companies were the members of Turkish Contractors Association (TCA) that had an average age of 26.75 and had different levels of international experience. Average turnover of the respondent companies were determined as 68.287 million $/ year. The collected data were

analyzed by using traditional cluster analysis methods, self organizing map (SOM), and fuzzy-C means (FCM) in order to determine the most suitable cluster solution for this data set. Firstly, all available algorithms of traditional cluster analysis were performed. According to the inverse scree trees and dendograms obtained as outputs of hierarchical cluster analyses, and Schwarz's Bayesian Criterion (BIC) and Akaike's Information Criterion (AIC), the alternatives about the number of the clusters were determined as two and three. Finally, by comparing the distance of the centers of the clusters, the number of the clusters was identified as three. Then, the most suitable solution for this data set was determined by considering the stability between the methods as well as stability within the methods after applying small modifications and external criteria. The SOM analysis was carried out for this data set and by considering the U-matrix and Davies-Bouldin index, three strategic groups were identified. Finally, by using the FCM, three strategic groups were identified according to the different validity indices developed for this method. After all, the strategic groups of Turkish construction industry were identified by comparing and combining the findings of these methods.

The major findings of the first part of the study and contributions to literature can be listed as follows:

- Regarding the complex environment and conditions related to competitive strategies of the Turkish construction companies, based on an extensive literature survey, a conceptual model which reflects the general and industry specific strategic dimensions was proposed. This model can be used in order to identify the strategic groups in other sectors by modifying the industry specific strategic dimensions.
- At the end of the analyses, three strategic groups were identified for Turkish construction industry. First strategic group comprises of small firms that mostly utilize a price differentiation and focus strategy. The basic difference from the other groups is that majority of firms in this group do not have a systematic and democratic strategy process. Also, they have weaknesses in terms of all sources of competitive advantage. They are especially weak in terms of client relations and experience. Second strategic group comprises of moderate-sized companies that having higher systematic strategic planning and democratic decision-making process when compared with the firms in Strategic Group 1. According to the values of strategic context variables, it is concluded that the firms in this strategic group have higher scores than the firms in Strategic Group 1. It may be argued that although majority of the firms in this group have non-price differentiation strategy, the level

of resources and competencies that should backup this strategy are not very high. The last strategic group, Strategic Group 3, is determined to have the highest scores in terms of strategy context variables among all groups. Majority of the firms in this group have a systematic strategic planning process and strategies are formulated in a democratic decision making environment. When compared to other clusters, the level of experience and client relations is significantly higher. All of the firms in this group try to differentiate themselves from other companies. Similarly, majority of the firms are diversified into sectors unrelated to construction.

- According to the ANOVA results, a statistically significant performance difference existed between the strategic groups. The most successful strategic group whose performance ratings are in the range of 4-to-5 was determined as the Strategic Group 3. This high performance level may result from the exceptional level of resources and capabilities which give them the opportunity to differentiate their services from others. The performance of firms in the Strategic Group 2 is moderate and significantly higher than the performance in Strategic Group 1. SOM results show that the performance of the companies also varies within the strategic groups; specifically the companies placed at the right side of the map outperform the other companies in the Strategic Group 2. The reason of this performance difference may originate from strong financial resources and democratic decision-making process; also most of these companies prefer focus strategy.
- The mobility barriers between the first and second strategic groups are determined as the strategic planning style and experience. Therefore, the companies in the Strategic Group 1 firstly should change their strategic planning style, develop a more formal and systematic planning style in order to move to a better strategic position. Also, the companies can increase their experience level by hiring experienced managers, or they can establish partnerships with experienced companies. For the second and third strategic groups, financial resources and managerial capability are found to be the most important mobility barriers. The companies which try to improve their strategic position may enhance their financial management capabilities in order to use their limited financial resources effectively and try to establish partnerships with different parties (contractors, suppliers etc.) that can complement their weaknesses. Besides, they can improve managerial capabilities by hiring professional managers, consultants and management training. In addition, experience is determined as an important variable that affects the separation of second and third clusters.

- This analysis demonstrates that there are advantages and disadvantages of three clustering methods in identification of strategic groups. The most important strength of SOM is its greatest visualization capability. When compared with SOM, both traditional cluster analysis and FCM have very limited visualization properties. Traditional cluster analysis methods build distinct self-contained clusters. Specifically, the obtained clusters via the usage of these methods are not appropriate to reveal strategic groups in complex industries. Whereas, since SOM and FCM provides information about whether the clusters are overlapped or not, and typology of the data, these clustering methods appear to be good alternatives for identifying strategic groups in complex industries. The other important advantage of SOM is avoidance of unreasonable arbitrary classification, whereas traditional cluster analysis methods are criticized due to their sorting ability. At the end of the analysis, the effectiveness of the SOM and FCM was determined in mapping of the complex structure of industries such as construction.
- The strategic groups of Turkish construction companies were concluded by combining and comparing the findings of the three methods. Finally, it was concluded that the application of traditional cluster analysis methods which form strict clusters and provide no information about the typology of the data alone is not appropriate in the identification of strategic groups. The fuzzy classification and clustering methods which illustrate the typology of the data set can be appropriate alternatives of traditional cluster analysis methods. In order to reveal the strategic group structure of the Turkish construction industry, a new structure was proposed by combining the concept of core, secondary firms proposed by Reger and Huff (1993) and hybrid strategic groups proposed by DeSarbo and Grewal (2008). Furthermore, the strategic group structure of the Turkish construction companies is illustrated in Figure 6-21. According to this figure, no hybrid strategic groups that combine strategic postures of all strategic groups exists due to distinctness between strategic postures specific to Strategic Group 1 and 3. Also, no hybrid strategic group between Strategic Group 1 and 3 is identified. However, the six hybrid strategic group firms belong to Strategic Group 1 and 2; the eight hybrid strategic group firms belong to Strategic Group 2 and 3. Furthermore, Strategic Group 1 consists of only core firms; whereas other strategic groups are comprised of secondary and core firms.

The second part of the current study deals with the identification of the modes and drivers of differentiation that create competitive advantage in the Turkish construction industry. The inter-relationships between these drivers and modes are found out by using structural equation model (SEM).

Although the relationship between differentiation strategy and performance was investigated in many studies (Kale and Arditi 2003, Grant 1995, Porter 1981), and it was concluded that there is a positive effect of differentiation strategy on the performance of the companies, there are limited studies in the construction management literature that deal with the contents and context of differentiation together with its potential to create competitive advantage. Therefore, identification of modes of differentiation and a complete and valid list of determinants of differentiation was the main motivation in this study. According to the results of a detailed literature survey about differentiation, a conceptual framework was proposed to model the differentiation process in construction industry. A questionnaire form designed according to this conceptual framework was posted on web. 62 questionnaires were collected. The respondents were members of medium-big sized companies whose total turnover for last three years and average age has been calculated as 775.544 US$M and 34.27, respectively. The collected data were analyzed by using SEM to examine the interrelationships between the drivers of differentiation and their impact on differentiation modes to achieve competitive advantage. Firstly, factor loading was performed for each latent variable in order to determine the significance of the variables in formation of latent variables. According to the "multicollinearity" and "insignificance levels", some of the variables were eliminated. Finally, the validity of the proposed differentiation construct and its drivers were investigated through several tests. Hypotheses regarding the relations among determinants of differentiation were tested; and their influence on differentiation modes was analyzed.

The major findings of the second part of the study and contributions to literature can be listed as follows:

- Based on extensive literature survey, factors under the heading of "project management", "corporate management", "resources", "relationships" and the modes by which the companies can differentiate themselves were identified. By considering the inter-relationships among these factors, several hypotheses were developed. At the end, a valid conceptual model that demonstrates the relationships between the modes and drivers of differentiation was obtained.

- SEM has been selected as the most appropriate technique to analyze the complex relationships within the differentiation model. The first reason why SEM was used in this study is its capability to analyze advanced models that consist of multiple layers of linkages between dependent and independent variables, which was the case in differentiation process. The second reason is its ability to identify sophisticated variables comprising of multi factors, which was also necessary as the aim of the study was to reveal differentiation modes that may comprise of variables reflecting different aspects of differentiation.
- In this research, it is found out that two modes of differentiation are possible in the Turkish construction industry. These modes were identified by exploratory factor analysis of six differentiation modes determined at the start of the study. They were termed as "quality related differentiation" and "productivity related differentiation". "Service quality", "product quality" and "positive image" were identified under the heading of "quality related differentiation"; on the other hand, "productivity", "time", and "innovative solutions" formed the "productivity related differentiation" mode. The first category is about creating more value to the customer by providing a differentiated service or product based on the highest quality, and creating a positive corporate image. The second category is about finishing the project on time, at the lowest cost by increasing productivity and providing innovative solutions to the client to basically meet his expectations about cost and time.
- "Corporate management" influences "quality related differentiation" extensively, also "project management" and "resources" influence "quality related differentiation" indirectly through "corporate management". Concerning these findings, the construction companies can differentiate themselves based on quality only if they can embed the quality culture in their process. The improvements in project management are expected to lead to improvements in corporate management. Consequently, project management indirectly affects the quality related differentiation through corporate management.
- "Resources" influence "productivity related differentiation" extensively. In other words, the companies that try to differentiate themselves in terms of productivity should invest in unique or specialized resources, and they should utilize them effectively. Also, "project management" influences "productivity related differentiation" moderately, in other words, project management activities play an important role in productivity differentiation.

- According to the factor loadings of the variables that constitute the "project management" construct, "resource management" is observed to have the highest correlation among variables under the heading of "project management". In other words, the companies should allocate the resources effectively and control the level of the resources throughout the project for securing differentiation; therefore, the companies should re-engineer their supply chains in order to achieve productivity and quality enhancements in their business. Due to the low ratings of the variables, namely "health and safety management" and "environmental management" are excluded from the list of drivers of the differentiation. It can be concluded that the opportunities of differentiating "health and safety management" and "environmental management" are limited.
- "Research and development", "business development" and "human resources management" are the most significant indicators of corporate management based on their factor loadings. Human resources is an important indicator of the performance of the companies; however human resources of an organization should be flexible enough to meet changing markets and customer demands, especially in the companies adopting differentiation strategies. The flexibility of the human resources can be provided by continuous trainings in the organization, and comprehensive staff hiring processes. The employees should be encouraged to produce new ideas and share these ideas with their managers. Research and development is an activity that is not carried out by majority of the contractors, whereas it can provide opportunities in creating uniqueness in the market. This conclusion is also valid for business development activity.
- It appears that the companies that differentiate based on productivity should enhance "technology". Especially, the information technologies have a great potential in achieving competitive advantage for firms through differentiation. Also, information technologies can enhance the productivity and service delivery by improving integration and facilitating communication in construction companies (Bjork 1999). Therefore, the companies may establish an IT department in their organization, and develop robust conceptual models for IT-applications with an appropriate strategic perspective. Also, the culture of the organization should be changed in order to prevent employees' resistance against IT by training, participating and staff development.

- In spite of the fact that no link exists between the differentiation modes and relationship, the companies perceive the significance of "relationship with clients". Therefore, the companies should establish long-term relationships with clients by improving the trust, cooperation, teamwork and communications between the parties.

Recommendations for further work

The strategic groups of Turkish construction industry were identified for one time period. This study can be repeated in the future to understand the changes in market structure in time and investigate the dynamic structure of Turkish construction industry. Information about movements between groups can provide information about the nature and characteristics of mobility barriers. Also, the concept of dynamic strategic groups can be used to predict the future strategic movements of the companies.

In this study, the strategic dimensions were identified by conducting an extensive literature survey, and the respondents rate these strategic dimensions considering their companies using a Likert scale. The strategic groups can also be identified by using different data sources, like archival data and cognitive perspectives of managers. Therefore, by comparing the strategic groups obtained from different data sources, the obtained strategic groups at the end of this study can be justified; moreover the emergence of convergence between these methods can also provide evidence about the existence of the strategic groups.

The capability of fuzzy clustering in modeling of hybrid strategic groups was mentioned by DeSarbo and Grewal (2008). In this study, FCM was used in identification of strategic groups, whereas different fuzzy clustering are also available in the literature, and all of these methods have different approaches in identification of clusters, therefore different aspects of the data can be revealed by applying different fuzzy clustering methods. Furthermore, all of these methods have advantages and disadvantages over each other. Consequently, this study can be extended by carrying out strategic group analysis using different fuzzy cluster analysis methods.

Using the same parameters in this study, data from companies of other nationalities can also be collected in order to develop similar models and those models can be compared. When the differences among those models are observed, a more complete and reliable model that can be used for every company may be achieved.

REFERENCES

Acar, E., Koçak, I., Sey, Y., and Arditi, D. (2005). "Use of information and communication technologies by small and medium-sized enterprises (SMEs) in building construction." *Construction Management and Economics*, 23(7), 713-722.

Agrawal, A., and Nagarjan, N. J. (1990). "Corporate capital structure, agency costs, and ownership control: The case of all-equity firms." *Journal of Finance*, 45(4), 1325–1331.

Akinci, B., and Fischer, M. (1998). "Factors affecting contractors' risk of cost overburden." *Journal of Management in Engineering*, 14(1), 67-76.

Akintoye, A., and Skitmore, M. (1991). "Profitability of UK construction contractors." *Construction Management and Economics*, 9(4), 311-325.

Akintoye, A. S., and MacLeod, M. J. (1997). "Risk analysis and management in construction." *International Journal of Project Management*, 15(1), 31-38.

Al-Jibouri, S. H. (2003). "Monitoring systems and their effectiveness for project cost control in construction." *International Journal of Project Management*, 21(2), 145-154.

Al-Momani, A. H. (2000). "Examining service quality within construction processes." *Technovation*, 20(11), 643-651.

Alchian, A. A. (1950). "Uncertainty, evolution, and economic theory." *The Journal of Political Economy*, 58(3), 211-221.

Aldenderfer, M., and Blashfield, R. (1984). *Cluster Analysis*, Sage Publications, London.

Allison, P. D. (2003). "Missing data techniques for structural equation modeling." *Journal of Abnormal Psychology,* 112(4), 545-557.

Amel, D. F., and Rhoades, S. A. (1988). "Strategic groups in banking." *The Review of Economics and Statistics*, 70(4), 685-689.

Amit, R., and Schoemaker, P. J. H. (1993). "Strategic assets and organizational rent." *Strategic Management Journal*, 14(1), 33-46.

Ammar, M. A., and Mohieldin, Y. A. (2002). "Resource constrained project scheduling using simulation." *Construction Management and Economics*, 20(4), 323-330.

Andersen, T. J. (2000). "Strategic planning, autonomous actions and corporate performance." *Long Range Planning*, 33(2), 184-200.

Andersson, C., and Gunnarsson, P. (2002). "Contract Management: A Way of Increasing Profit in Construction Projects?" Chalmers University of Technology, Göteborg.

Andreassen, T. W., and Lindestad, B. (1998). "Customer loyalty and complex services: the impact of corporate image on quality, customer satisfaction and loyalty for customers with varying degrees of service expertise." *International Journal of Service Industry Management*, 9(1), 7-23.

Anvuur, A. M., and Kumaraswamy, M. M. (2007). "Conceptual model of partnering and alliancing." *Journal of Construction Engineering and Management*, 133(3), 225-234.

Arditi, D. (1985). "Construction productivity improvement." *Journal of Construction Engineering and Management*, 111(1), 1-14.

Arthur, J. B. (1992). "The link between business strategy and industrial relations systems in American Steel Minimills." *Industrial and Labor Relations Review*, 45(3), 488-506.

Ashley, D. B., Laurie, C. S., and Jaselskis, E. J. (1987). "Determinants of construction project success." *Project Management Journal*, 18(2), 69-79.

Bacher, J. (2002). "Cluster Analysis." *Script available from http://www. soziologie. wiso. uni-erlangen. de/koeln/script/script. pdf. Last Access November 2007*, 3.

Bacher, J., Wenzig, K., and Vogler, M. (2004). "SPSS two-step cluster: A first evaluation." In: *RC33 Sixth International Conference on Social Science Methodology*, Amsterdam.

Ball, M., Farshchi, M., and Grilli, M. (2000). "Competition and the persistence of profits in the UK construction industry." *Construction Management and Economics*, 18(7), 733-745.

Banfield, J. D., and Raftery, A. E. (1993). "Model-based Gaussian and non-Gaussian clustering." *Biometrics*, 49(3), 803-821.

Barney, J. (1991). "Firm resources and sustained competitive advantage." *Journal of Management*, 17(1), 99-120.

Barney, J. B., and Hoskisson, R. E. (1990). "Strategic groups: Untested assertions and research proposals." *Managerial and Decisions Economics*, 11(3), 187-198.

Barney, J. B., and Wright, P. M. (1998). "On becoming a strategic partner: The role of human resources in gaining competitive advantage." *Human Resource Management*, 37(1), 31-46.

Barrie, D. S., and Paulson, B. C. (1992). *Professional Construction Management: Including CM, Design-construct and General Contracting*, McGraw-Hill, New York.

Barth, E., Gulbrandsen, T., and Schønea, P. (2005). "Family ownership and productivity: The role of owner-management." *Journal of Corporate Finance*, 11(1-2), 107-127.

Bennett, J. (1983). "Project management in construction." *Construction Management and Economics*, 1(3), 183-197.

Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A., and Murtagh, R. F. (1996). "Validity-guided (Re)clustering with applications to image segmentation." *IEEE Transactions on Fuzzy Systems*, 4(2), 112-123.

Bensoussan, B., and Fleisher, C. (2000). "A FAROUT way to manage CI analysis." *Competitive Intelligence Magazine*, 3(2), 37-40.

Bentler, P. M. (1990). "Comparative fit indexes in structural models." *Psychological Bulletin*, 107, 238-246.

Bentler, P. M. (2006). *EQS 6 Structural Equation Program Manual*, Multivariate Software, Inc., Encino, CA.

Bentler, P. M., and Chou, C.-P. (1987). "Practical issues in structural modeling." *Sociological Methods Research*, 16(1), 78-117.

Bentler, P. M., and Dijkstra, T. (1985). "Efficient estimation via linearization in structural models." In: *Multivariate Analysis VI*, Krishnaiah, P. R. ed., Elsevier Science Publishers, Amsterdam: North-Holland, 9-42.

Betts, M., and Ofori, G. (1992). "Strategic planning for competitive advantage in construction." *Construction Management and Economics*, 10(6), 511-532.

Betts, M., and Ofori, G. (1994). "Strategic planning for competitive advantage in construction: The institutions." *Construction Management and Economics*, 12(3), 203-217.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, Norwell.

Bezdek, J. C. (1999). *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Kluwer Academic Publishers, New York.

Bezdek, J. C., Ehrlich, R., and Full, W. (1984). "FCM: Fuzzy c-means algorithm." *Computers and Geoscience*, 10(2-3), 191-203.

Bezdek, J. C., and Pal, N. R. (1998). "Some new indexes of cluster validity." *IEEE Trans. On Systems, Man., and Cybern.*, 28(3), 301-315.

Bhatt, G. D., and Grover, V. (2005). "Types of information technology capabilities and their role in competitive advantage: An empirical study." *Journal of Management Information Systems*, 22(2), 253-277.

Bhattacharya, U., and Ravikumar, B. (2005). "From Cronies to Professionals: The Evolution of Family Firms." In: *Capital Formation, Governance and Banking*, Klein, E. ed., Nova Science, Hauppauge, New York.

Bjork, B. C. (1999). "Information technology in construction: Domain definition and research issues." *International Journal of Computer Integrated Design and Construction*, 1(1), 3-16.

Blashfield, R. K., and Aldenderfer, M. S. (1978). "The literature on cluster analysis." *Multivariate Behavioral Research*, 13(3), 271-295.

Bleaney, M., and Wakelin, K. (2002). "Efficiency, innovation and exports." *Oxford Bulletin of Economics and Statistics*, 64(1), 3-15.

Bollen, K. A. (1989). *Structural equations with latent variables*, John Wiley & Sons New York.

Boomsma, A. (1982). "The robustness of LISREL against small sample size in factor analysis models." In: *Systems under indirect observation, Part 1*, Jöreskog, K. G. and Wold, H. eds., 149-173.

Boulding, W., Lee, E., and Staelin, R. (1994). "Mastering the mix: Do advertising, promotion, and sales force activities lead to differentiation?" *Journal of Marketing Research (JMR)*, 31(2), 159-172.

Bourassa, S. C., Hamelink, F., and Hoesli, M. (1997). "Defining residential submarkets: evidence from Sydney and Melbourne?" *Real Estate Research Unit Working Paper Series*, 3.

Bovet, D., and Martha, J. (2000). *Value Nets: Breaking the Supply Chain to Unlock Hidden Profits*, Wiley, New York.

Bowley, M. (1960). *Innovations in Building Materials: An Economic Study*, Gerald Duckworth & Co Ltd, London.

Bresnen, M., Edelman, L., Newell, S., Scarbrough, H., and Swan, J. (2003). "Social practices and the management of knowledge in project environments." *International Journal of Project Management*, 21(3), 157-166.

Bresnen, M., and Marshall, N. (2000). "Building partnerships: case studies of client-contractor collaboration in the UK construction industry." *Construction Management and Economics*, 18(7), 819-832.

Bullock, H. E., Harlow, L. L., and Mulaik, S. A. (1994). "Causation issues in structural equation modeling research." *Structural Equation Modeling*, 1(3), 253-267.

Burati Jr., J. L., Matthews, M. F., and Kalidindi, S. N. (1992). "Quality management organizations and techniques." *Journal of Construction Engineering and Management*, 118(1), 112-128.

Burkart, M., Panunzi, F., and Shleifer, A. (2003). "Family firms." *The Journal of Finance*, 58(5), 2167-2202.

Byrne, B. M. (2006). *Structural Equation Modeling With Eqs: Basic Concepts, Applications, and Programming*, Lawrence Erlbaum Associates, Mahwah, New Jersey.

Calinski, T., and Harabasz, J. (1974). "A dendrite method for cluster analysis." *Communications in Statistics*, 3(1), 1-27.

Calori, R., and Ardisson, J. M. (1988). "Differentiation strategies in 'stalemate industries'." *Strategic Management Journal*, 9(3), 255-269.

Cannon, J., and Hillebrant, P. M. (1990). "Diversification." In: *The Management of Construction Firms: Aspects of Theory*, Hillebrant, P. M. and Cannon, J. eds., Macmillan, London, 53-64.

Capon, N., Farley, J. U., and Hoenig, S. (1990). "Determinants of financial performance: A meta-analysis." *Management Science*, 36(10), 1143-1159.

Caves, R., and Porter, M. E. (1977). "From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition." *Quarterly Journal of Economics*, 91(2), 241-262.

Cefis, E., and Marsili, O. (2006). "Survivor: The role of innovation in firms' survival." *Research Policy*, 35(5), 626-641.

Chan, D. W. M., and Kumaraswamy, M. M. (2002). "Compressing construction durations: lessons learned from Hong Kong building projects." *International Journal of Project Management*, 20(1), 23-35.

Chan, P., Cooper, R., and Tzortzopoulos, P. (2005). "Organizational learning: conceptual challenges from a project perspective." *Construction Management and Economics*, 23(7), 747-756.

Chan, S. L., and Park, M. (2005). "Project cost estimation using principal component regression." *Construction Management and Economics*, 23(3), 295-304.

Charan, R., Hofer, C. W., and Mahon, J. F. (1980). "From entrepreneurial to professional management: A set of guidelines." *Journal of Small Business Management*, 18(1), 1-10.

Cheah, C. Y. J., Jian, K., and Chew, D. A. S. (2007). "Strategic analysis of large local construction firms in China." *Construction Management and Economics*, 25(1), 25-38.

Chin, W. W. (1998). "Commentary: issues and opinion on structural equation modeling." *MIS Quarterly*, 22(1), vii-xvi.

Chinowsky, P. S., and Meredith, J. E. (2000). "Strategic management in construction." *Journal of Construction Engineering and Management*, 126(1), 1-9.

Chinyio, E. A., Olomolaiye, P. O., Kometa, S. T., and Harris, F. C. (1998). "A needs-based methodology for classifying construction clients and selecting contractors." *Construction Management and Economics*, 16(1), 91-98.

Chiu, T., Fang, D. P., Chen, J., Wang, Y., and Jeris, C. (2001). "A robust and scalable clustering algorithm for mixed type attributes in large database environment." In: *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, San Francisco, California, 263-268.

Chou, C. P., and Bentler, P. M. (1995). "Estimates and tests in structural equation modeling." In: *Structural Equation Modeling: Concepts, Issues and Applications*, Hoyle, R. H. ed., Sage, Thousand Oaks, CA, 37-55.

Chou, C. P., Bentler, P. M., and Satorra, A. (1991). "Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: a Monte Carlo study." *British Journal of Mathematical and Statistical Psychology*, 44(2), 347-357.

Christensen, L. T., and Askegaard, S. (2001). "Corporate identity and corporate image revisited." *European Journal of Marketing*, 35(3/4), 292-315.

Chuang, K., Chiu, M., Lin, C., and Chen, J. (1999). "Model free functional MRI analysis using Kohonen clustering neural network and fuzzy C-means." *IEEE Transactions on Medical Imaging*, 18(12), 1117-1128.

Claver, E., Molina, J. F., and Tari, J. J. (2003). "Strategic groups and firm performance: The case of Spanish house-building firms." *Construction Management and Economics*, 21(4), 369-377.

Cool, K., and Dierickx, I. (1993). "Rivalry, strategic groups and firm profitability." *Strategic Management Journal*, 14(1).

Cool, K. O., and Schendel, D. (1987). "Strategic group formation and performance: The case of the U.S. pharmaceutical industry." *Management Science*, 33(9), 1102-1124.

Cooley, W. W. (1978). "Explanatory Observational Studies." *Educational Researcher*, 7(9), 9-15.

Cormack, R. M. (1971). "A review of classification." *Journal of the Royal Statistical Society. Series A (General)*, 134(3), 321-367.

Curran, P. J., West, S. G., and Finch, J. F. (1996). "The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis." *Psychological Methods*, 1(1), 16-29.

Curry, B., Davies, F., Evans, P., and Moutinho, L. (2001). "The Kohonen self-organizing map: An application to the study of strategic groups in the UK hotel industry." *Expert Systems*, 18(1), 19-31.

Czarnitzki, D., and Kraft, K. (2004). "Innovation indicators and corporate credit ratings: evidence from German firms." *Economics Letters*, 82(3), 377-384.

Dainty, A. R. J., Mi, C., and Moore, D. R. (2003). "Redefining performance measures for construction project managers: An empirical evaluation." *Construction Management and Economics*, 21(2), 209-218.

Dale, B. G. (2003). *Managing Quality*, Blackwell Publishing, Malden, MA.

Danneels, E. (2002). "The dynamics of product innovation and firm competences." *Strategic Management Journal*, 23(12), 1095-1121.

Dave, R. N. (1996). "Validating fuzzy partitions obtained through c-shells clustering." *Pattern Recognition Letters*, 17(6), 613-623.

Davies, D., and Bouldin, D. (1979). "A cluster separation measure." *IEEE Trans. Pattern Anal. Mach. Intell.*, 1(2), 224-227.

Day, D. L., Lewin, A. Y., and Li, H. (1995). "Strategic leaders or strategic groups: A longitudinal data envelopment analysis of the US brewing industry." *European Journal of Operational Research*, 80(3), 619-638.

De Oliveira, J. V., and Pedrycz, W. (2007). *Advances in Fuzzy Clustering and Its Applications*, Wiley, West Sussex.

Deboeck, G. J., and Kohonen, T. K. (1998). *Visual Explorations in Finance*, Springer-Verlag New York, Secaucus, NJ, USA.

Delaney, J. T., and Huselid, M. A. (1996). "The impact of human resource management practices on perceptions of organizational performance." *The Academy of Management Journal*, 39(4), 949-969.

Delgado-Hernandez, D. J. (2007). "Quality function deployment in construction." *Construction Management and Economics*, 25(6), 597-609.

Demsetz, H., and Lehn, K. (1985). "The structure of corporate ownership: Causes and consequences." *The Journal of Political Economy* 93(6), 1155–1177.

DeSarbo, W. S., and Grewal, R. (2008). "Hybrid strategic groups." *Strategic Management Journal*, 29, 293-317.

Desieno, D. (1988). "Adding a conscience mechanism to competitive learning." In: *Proceedings IEEE International Conference on Neural Networks*, IEEE, San Diego, California, 117-124.

Dess, G. G., and Davis, P. (1984). "Porter's (1980) generic strategies as determinants of strategic group membership and organizational performance." *The Academy of Management Journal*, 27(3), 467-488.

Dess, G. G., and Robinson, R. B. (1984). "Measuring organizational performance in the absence of objective measures: The case of the privately-held firm and conglomerate business unit." Strategic Management Journal, 5(3), 265-273.

De Maio, A., Verganti, R., and Corso, M. (1994). "A multi-project management framework for new product development." European Journal of Operational Research, 78(2), 178-191.

Dichter, E. (1985). "What's in an image." *Journal of Consumer Marketing*, 2(1), 75-81.

Dickson, P. R., and Ginter, J. L. (1987). "Market segmentation, product differentiation, and marketing strategy." *Journal of Marketing*, 51(2), 1-10.

Dierickx, I., and Cool, K. (1989). "Asset stock accumulation and sustainability of competitive advantage." *Management Science*, 35(12), 1504-1511.

Dikmen, I., and Birgonul, M. T. (2003). "Strategic perspective of Turkish construction companies." *Journal of Management in Engineering*, 19(1), 33-40.

Dikmen, I., and Birgonul, M. T. (2004a). "Impact of Strategic Planning on Competitive Advantage." In: *Proceedings of the ASCE Specialty Conference on Leadership and Management in Construction*, Hilton Head, South Carolina, USA, 201-211.

Dikmen, I., and Birgonul, M. T. (2004b). "Neural network model to support international market entry decisions." *Journal of Construction Engineering and Management*, 130(1), 59-66.

Dikmen, I., Birgonul, M. T., and Ataoglu, T. (2005a). "Chapter 10. Empirical investigation of organisational learning ability as a performance driver in construction." In: *Knowledge Management in the Construction Industry: A Socio-Technical Perspective*, Kazi, A. S. ed., Idea Group Publishing Inc., USA.

Dikmen, I., Birgonul, M. T., and Kiziltas, S. (2005b). "Strategic use of quality function deployment (QFD) in the construction industry." *Building and Environment*, 40(2), 245-255.

Dillon, W. R., Mulani, N., and Frederick, D. G. (1989). "On the use of the component scores in the presence of group structure." *Journal of Consumer Research*, 16(1), 106-112.

Ding, L., Velicer, W. F., and Harlow, L. L. (1995). "Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices." *Structural Equation Modeling*, 2(2), 119-144.

Dodgson, M. (1993). "Organizational learning: a review of some literatures." *Organization Studies*, 14(3), 375-394.

Dranove, D., Peteraf, M., and Shanley, M. (1998). "Do strategic groups exist? An economic framework for analysis." *Strategic Management Journal*, 19(11), 1029-1044.

Dreger, R. M. (1986). "Microcomputer programs for the rand index of cluster similarity." *Educational and Psychological Measurement*, 46(3), 655.

Dunn, J. C. (1974). "Well separated clusters and optimal fuzzy partitions." *Journal of Cybernetics*, 4(3), 95-104.

Edelbrock, C. (1979). "Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody." *Multivariate Behavioral Research*, 14(3), 367-384.

Edwards, P. J., and Bowen, P. A. (1998). "Risk and risk management in construction: a review and future directions for research." *Engineering Construction and Architectural Management*, 5(4), 339-349.

Eisenhardt, K. M., and Martin, J. A. (2000). "Dynamics capabilities: What are they?" *Strategic Management Journal*, 21(10/11), 1105-1121.

El-hamdouchi, A., and Willett, P. (1989). "Comparison of hierarchic agglomerative clustering methods for document retrieval." *The Computer Journal*, 32(3), 220-227.

El-Mashaleh, M., O'Brien, W. J., and Minchin, R. E. (2006). "Firm performance and information technology utilization in the construction industry." *Journal of Construction Engineering and Management*, 132(5), 499-507.

Eldin, N., and Hikle, V. (2003). "Pilot study of quality function deployment in construction projects." *Journal of Construction Engineering and Management*, 129(3), 314.

ENR. (2007). "The top 225 global contractors." *Engineering News Record*, 259(7).

Evans, J. R., and Berman, B. (1997). *Marketing*, 7th Ed., Prentice Hall Inc., New Jersey.

Everitt, B. (1974). *Cluster Analysis*, Wiley, New York.

Everitt, B. (1980). "Cluster analysis." *Quality and Quantity*, 14(1), 75-100.

Everitt, B., Landau, S., and Leese, M. (2001). *Cluster Analysis*, 4th Ed., Oxford University Press, London : Arnold ; New York.

Flavian, C., and Polo, Y. (1997). "Strategic group analysis (SGA) as a tool for strategic marketing." *European Journal of Marketing*, 33(5/6), 548-569.

Faridi, A. S., and El-Sayegh, S. M. (2006). "Significant factors causing delay in the UAE construction industry." *Construction Management and Economics*, 24(11), 1167-1176.

Fiegenbaum, A., McGee, J., and Thomas, H. (1988). "Exploring the linkage between strategic groups and competitive strategy." *International Studies of Management and Organizations*, 18(1), 6-25.

Fiegenbaum, A., and Thomas, H. (1990). "Strategic groups and performance: The U.S. insurance industry, 1970-84." *Strategic Management Journal*, 11(3), 197-215.

Fiegenbaum, A., and Thomas, H. (1995). "Strategic groups as reference groups: Theory, modeling and empirical examination of industry and competitive strategy." *Strategic Management Journal*, 16(6), 461-476.

Finch, J. F. (1997). "Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models." *Structural Equation Modeling*, 4(2), 87-107.

Fisher, R. J. (1991). "Durable differentiation strategies for services." *The Journal of Services Marketing*, 5(1), 19-28.

Fombrun, C. J. (1996). *Reputation: Realizing Value from the Corporate Image*, Harvard Business School Press, Boston, MA.

Fong, P. S. W., and Choi, S. K. Y. (2000). "Final contractor selection using the analytical hierarchy process." *Construction Management and Economics*, 18(5), 547-557.

Fox, S., Marsh, L., and Cockerham, G. (2002). "How building design imperatives constrain construction productivity and quality." *Engineering Construction and Architectural Management*, 9(5-6), 378-387.

Fraboni, M., and Saltstone, R. (1992). "The WAIS-R Number-of-Factors quandary: A cluster analytic approach to construct validation." *Educational and Psychological Measurement*, 52(3), 603-613.

Fraley, C., and Raftery, A. E. (1998). "How many clusters? Which clustering method? Answers via model-based cluster analysis." *The Computer Journal*, 41(8), 578-588.

Fraser, C., and Zarkada-Fraser, A. (2001). "The philosophy, structure and objectives of research and development in Japan." *Construction Management and Economics*, 19(8), 831-840.

Frazier, G. L., and Howell, R. D. (1983). "Business definition and performance." *Journal of Marketing*, 47(2), 56-67.

Friedman, W. (1984). *Construction Marketing and Strategic Planning*, McGraw-Hill Companies, New York.

Frimpong, Y., Oluwoye, J., and Crawford, L. (2003). "Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study." *International Journal of Project Management*, 21(5), 321-326.

Gallo, M. A., and Vilaseca, A. (1996). "Finance in family business." *Family Business Review*, 9(4), 387-401.

Gan, G., Ma, C., and Wu, J. (2007). *Data Clustering theory, Algorithms, and Applications*, SIAM, Philadelphia.

Gath, I., and Geva, A. B. (1989). "Unsupervised optimal fuzzy clustering." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 773-780.

Gedajlovic, E., Lubatkin, M. H., and Schulze, W. S. (2004). "Crossing the threshold from founder management to professional management: A governance perspective." *Journal of Management Studies* 41(5), 899-912.

Geroski, P. (1993). "The profitability of innovating firms." *Rand Journal of Economics*, 24(2), 198-211.

Ghosh, S., and Jintanapakanont, J. (2004). "Identifying and assessing the critical risk factors in an underground rail project in Thailand: a factor analysis approach." *International Journal of Project Management*, 22(8), 633-643.

Gordon, A. D. (1999). *Classification*, Chapman&Hall/CRC, Florida.

Gorsevski, P. V., Gessler, P. E., and Jankowski, P. (2003). "Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard." *Journal of Geographical Systems*, 5(3), 223-251.

Gower, J. (1971). "A general coefficient of similarity and some of its properties." *Biometrics*, 27(4), 857-874.

Grant, R. M. (1995). *Contemporary Strategy Analysis: Concepts, Techniques, Applications*, 2nd Ed., Blackwell Publishing, Oxford.

Gray, E. R., and Balmer, J. M. T. (1998). "Managing corporate image and corporate reputation." *Long Range Planning*, 31(5), 695-702.

Greenstein, S., and Mazzeo, M. (2006). "The role of differentiation strategy in local telecommunication entry and market evolution: 1999-2002." *Journal of Industrial Economics*, 54(3), 323-350.

Groth, R. (1998). *Data Mining: A Hands-on Approach for Business Professionals*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Grönroos, C. (1990). *Service Management and Marketing: Managing the Moments of Truth in Service Competition*, Lexington Books.

Gu, T., and Dubuisson, B. (1990). "Similarity of classes and fuzzy clustering." *Fuzzy Sets and Systems*, 34(2), 213-221.

Gustafson, D. E., and Kessel, W. C. (1978). "Fuzzy clustering with a fuzzy covariance matrix." *IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, 17(1).

Günter, S., and Bunker, H. (2003). "Validation indices for graph clustering." *Pattern Recognition Letters*, 24(1), 1107-1113.

Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (1995). *Multivariate Data Analysis with Readings*, Prentice Hall, New York.

Hair, J. R. (1998). *Multivariate Data Analysis*, Prentice-Hall, London.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001). "On clustering validation techniques." *Journal of Intelligent Information Systems*, 17(2), 107-145.

Hammah, R. E., and Curran, J. H. (1998). "Fuzzy cluster algorithm for the automatic identification of joint sets." *Int. J. Rock Mech. Min. Sci.*, 35(7), 889-905.

Hampson, K., and Tatum, C. B. (1997). "Technology strategy and competitive performance in bridge construction." *Journal of Construction Engineering and Management*, 123(2), 153-161.

Hansen, M. T. (1999). "The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits." *Administrative Science Quarterly*, 44(1), 82-85.

Harrigan, K. R. (1985). "An application of clustering for strategic group analysis." *Strategic Management Journal*, 6(1), 55-73.

Harris, F., and McCaffer, R. (2001). *Modern Construction Management*, Blackwell Publishing, Oxford.

Hartigan, J. A. (1975). *Clustering Algorithms*, John Wiley and Sons, New York.

Hartmann, A. (2006). "The context of innovation management in construction firms." *Construction Management and Economics*, 24(6), 567-578.

Hasegawa, F. (1988). *Built by Japan: Competitive Strategies of the Japanese Construction Industry*, John Wiley and Sons, New York.

Hatten, K., and Hatten, M. L. (1987). "Strategic groups, asymmetrical mobility barriers and contestability." *Strategic Management Journal*, 8(4), 329-342.

Hausman, A. (2001). "Variations in relationship strength and its impact on performance and satisfaction in business relationships." *Journal of Business and Industrial Marketing*, 16(7), 600-616.

Hawes, J. M., and Crittenden, W. F. (1984). "A taxonomy of competitive retailing strategies." *Strategic Management Journal*, 5(3), 275-287.

Hegazy, T. (1999). "Optimization of resource allocation and leveling using genetic algorithms." *Journal of Construction Engineering and Management*, 125(3), 167-175.

Henderson, B. D. (1981). *The Concept of Strategy*, The Boston Consulting Group, Boston, MA.

Herbig, P., Milewicz, J., and Golden, J. (1994). "A model of reputation building and destruction." *Journal of Business Research*, 31(1), 23-31.

Hershberger, S. L. (2006). "The problem of equivalent structural models." In: *Structural Equation Modeling*, Hancock, G. R. and Mueller, R. O., eds., IAP, Greenwich, Connecticut.

Hodgkinson, G. P. (1997). "The cognitive analysis of competitive structures: A review and critique." *Human Relations*, 50(6), 625-654.

Hofer, C. W., and Schendel, D. (1978). *Strategy Formulation: Analytical Concepts*, West Pub. Co., St. Paul.

Holt, G. D. (1997). "Classifying construction contractors." *Building Research and Information*, 25(6), 374-382.

Houthoofd, N., and Heene, A. (1997). "Strategic groups as subsets of strategic scope groups in the Belgian brewing industry." *Strategic Management Journal*, 18(8), 653-666.

Howard, H. (2000). *Managing Quality*, Scitech Educational, Broadstairs, Kent, UK.

Höppner, F., Klawonn, F., Kruse, R., and Runkler, T. (1999). *Fuzzy Cluster Analysis*, John Wiley and Sons, Chichester, England.

Hsieh, T. Y. (1998). "Impact of subcontracting on site productivity: Lessons learned in Taiwan." *Journal of Construction Engineering and Management*, 124(2), 91-100.

Hsu, T. (2000). "An application of fuzzy clustering in group positioning analysis." *Proceedings National Science Council of the Republic of China*, 10(2), 157-167.

Hu, L., and Bentler, P. M. (1999). "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." *Structural Equation Modeling*, 6(1), 1-55.

Hu, L., Bentler, P. M., and Kano, Y. (1992). "Can test statistics in covariance structure analysis be trusted?" *Psychological Bulletin*, 112(2), 351-362.

Hu, L. T., and Bentler, P. M. (1998). "Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification." *Psychological Methods*, 3(4), 424-453.

Huber, G. P. (1991). "Organizational learning: The contributing processes and the literatures." *Organization Science*, 2(1), 88-115.

Hunt, M. (1972). "Competition in major home appliance industry 1960-1970," Unpublished Doctoral Dissertation, Harvard University, Cambridge, MA.

Huselid, M. A. (1995). "The impact of human resource management practices on turnover, productivity, and corporate financial performance." *The Academy of Management Journal*, 38(3), 635-672.

Insightful. (2007). *S-PLUS® 8 Guide to Statistics, Volume 2*, Insightful Corporation, Seattle.

Janadi, M. O. (1996). "Factors affecting the safety of the construction industry." *Building Research and Information*, 24(2), 108-111.

Jardine, N., and Sibson, R. (1968). "A model for taxonomy." *Mathematical Biosciences*, 2, 465-482.

Jay Na, L., and Ofori, G. (2007). "Classification of innovation for strategic decision making in construction businesses." *Construction Management and Economics*, 25(9), 963-978.

Jennings, M., and Betts, M. (1996). "Competitive strategies for quantity surveying practices: The importance of information technology." *Engineering, Construction and Architectural Management*, 3(3), 163-186.

Jennings, P., and Holt, G. D. (1998). "Prequalification and multi-criteria selection: A measure of contractors' opinions." *Construction Management and Economics*, 16(6), 651-660.

Johnson, G., and Scholes, K. (2002). *Exploring Corporate Strategy, Text and Cases*, 6th Ed., Prentice Hall, Harlow.

Johnson, J. L. (1999). "Strategic integration in industrial distribution channels: managing the interfirm relationship as a strategic asset." *Journal of the Academy of Marketing Science*, 27(1), 4-18.

Jones, M., and Saad, M. (2003). *Managing Innovation in Construction*, Thomas Telford, London.

Junnonen, J. M. (1998). "Strategy formation in construction firms." *Engineering, Construction and Architectural Management*, 5(2), 107-114.

Juran, J. M., and Godfrey, A. B. (1998). *Juran's Quality Handbook*, McGraw-Hill Professional Publishing, Blacklick, OH, USA.

Kadefors, A. (2004). "Trust in project relationships: Inside the black box." *International Journal of Project Management*, 22(3), 175-182.

Kale, S., and Arditi, D. (2001). "General contractors' relationships with subcontractors: A strategic asset." *Construction Management and Economics*, 19(5), 541-549.

Kale, S., and Arditi, D. (2002). "Competitive positioning in United States construction industry." *Journal of Construction Engineering and Management*, 128(3), 238-247.

Kale, S., and Arditi, D. (2003). "Differentiation, conformity, and construction firm performance." *Management in Engineering*, 19(2), 52-59.

Kamara, J. M., Augenbroe, G., Anumba, C. J., and Carrillo, P. M. (2002). "Knowledge management in the architecture, engineering and construction industry." *Construction Innovation*, 2(1), 53-67.

Kanji, G. K., and Wong, A. (1998). "Quality culture in the construction industry." *Total Quality Management*, 9(4), 133-140.

Kaski, S., Honkela, T., and Kohonen, T. (1998). "Websom: Self-organizing maps of document collections." *Neurocomputing*, 21(1), 101-117.

Kaufman, L., and Rousseeuw, P. (1987). "Clustering by means of medoids." In: *Statistical Data Analysis Based on the L1-Norm and Related Methods*, Dodge, Y., ed., Birkhauser, Amsterdam.

Kaufman, L., and Rousseeuw, P. J. (1990). *Finding Groups in Data*, Wiley, New Jersey.

Kaufman, R. L. (1985). "Issues in multivariate cluster analysis: Some simulation results." *Sociological Methods and Research*, 13(4), 467-486.

Kauko, T. (2003). "On current neural network applications involving spatial modeling of property prices." *Journal of Housing and the Built Environment*, 18(2), 159-181.

Kazaz, A., and Ulubeyli, S. (2007). "Drivers of productivity among construction workers: A study in a developing country." *Building and Environment*, 42(5), 2132-2140.

Keittinger, W., Grover, V., Guha, S., and Segars, A. (1994). "Strategic information systems revisited: A study in the sustainability and performance." *MIS Quarterly*, 18(1), 31-58.

Kennedy, S. (1977). "Nurturing corporate images." *European Journal of Marketing*, 11(3), 120-164.

Ketchen, D. J., and Shook, C. L. (1996). "The application of cluster analysis in strategic management research: An analysis and critique." *Strategic Management Journal*, 17(6), 441-458.

Kezsbom, D. S., and Edward, K. A. (2001). *The New Dynamic Project Management: Winning Through the Competitive Advantage*, Wiley, New York.

Kiang, M. Y., and Kumar, A. (2001). "An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications." *Information Systems Research*, 12(2), 177-194.

King, B. (1967). "Step-wise clustering procedures." *Journal of the American Statistical Association*, 62(317), 86-101.

Kiran, K. B., Vijaya, B. N., Vishnuvardhan, R. M., and Giridharan, N. (2007). "DNA fingerprinting and phylogenetic analysis of WNIN rat strain and its obese mutants using microsatellite markers." *Biochemical Genetics*, 15(1), 77-91.

Klassen, R. D., and McLaughlin, C. P. (1996). "The impact of environmental management on firm performance." *Management Science*, 42(8), 1199-1214.

Kline, B. R. (1998). *Principles and Practice of Structural Equation Modeling*, The Guilford Press, New York.

Kog, Y. C., Chua, D. K. H., Loh, P. K., and Jaselskis, E. J. (1999). "Key determinants for construction schedule performance." *International Journal of Project Management*, 17(6), 351-359.

Kohonen, T. (1989). *Self-Organization and Associative Memory*, Springer-Verlag, New York, USA.

Kohonen, T. (2001). *Self Organizing Maps*, Springer, New York, USA.

Krishnapuram, R., and Keller, J. (1993). "A possibilistic approach to clustering." *IEEE Transactions on Fuzzy Systems*, 1(2), 98-110.

Krzanowski, W. J., and Lai, Y. T. (1988). "A criterion for determining the number of groups in a data set using sum-of-squares clustering." *Biometrics*, 44(1), 23-34.

Kululanga, G. K., Edum-Fotwe, F. T., and McCaffer, R. (2001a). "Measuring construction contractors' organizational learning." *Building Research and Information*, 29(1), 21-29.

Kululanga, G. K., Kuotcha, W., McCaffer, R., and Edum-Fotwe, F. (2001b). "Construction contractors' claim process framework." *Journal of Construction Engineering and Management*, 127(4), 309-314.

Kululanga, G. K., and McCaffer, R. (2001). "Measuring knowledge management for construction organizations." *Engineering Construction and Architectural Management*, 8(5/6), 346-354.

Kumaraswamy, M. M., and Chan, D. W. M. (1998). "Contributors to construction delays." *Construction Management and Economics*, 16(1), 17-29.

Kumaraswamy, M. M., and Dissanayaka, S. M. (2000). "ISO 9000 and beyond: From a Hong Kong construction perspective." *Construction Management and Economics*, 18(7), 783-796.

Kumaraswamy, M. M., Palaneeswaran, E., Rahman, M. M., Ugwu, O. O., and Ng, S. T. (2006). "Synergising R&D initiatives for e-enhancing management support systems." *Automation in Construction*, 15(6), 681-692.

Laborde, M., and Sanvido, V. (1994). "Introducing new process technologies into construction companies." *Journal of Construction Engineering and Management*, 120(3), 488-508.

Lado, A. A., and Wilson, M. C. (1994). "Human resource systems and sustained competitive advantage: A competency-based perspective." *Academy of Management Review*, 19(4), 699-727.

Lai, K.-h., and Cheng, T. C. E. (2003). "Initiatives and outcomes of quality management implementation across industries." *Omega*, 31(2), 141-154.

Langford, D., Hancock, G. R., Fellows, R., and Gale, A. W. (1995). *Human Resources Management in Construction*, Longman, Harlow, Essex.

Lansiluoto, A. (2004). *Economic and Competitive Environment Analysis in the Formulation of Strategy, a Decision-Oriented Study Utilizing Self-Organizing Maps*, Publications of the Turku school of economics and business administration, Turku.

Lanyon, S. M. (1985). "Detecting internal inconsistencies in distance data." *Systematic Zoology*, 34(4), 397-403.

Laufer, A., and Tucker, R. L. (1987). "Is construction project planning really doing its job? A critical examination of focus, role and process." *Construction Management and Economics*, 5(3), 243-266.

Leask, G. (2004). "Strategic groups and the resource based view: Natural complements enhancing our understanding of the competitive process." Ashton University, UK, Working Paper.

Leask, G. (2005). "Use of strategy frameworks to analyse competitive dynamics in the pharmaceutical industry." *Journal of Medical Marketing*, 5(3), 209-218.

Leask, G., and Parnell, J. A. (2005). "Integrating strategic groups and the resource based perspective: understanding the competitive process." *European Management Journal*, 23(4), 458-470.

Leung, M.-Y., Olomolaiye, P., Chong, A., and Lam, C. C. (2005). "Impacts of stress on estimation performance in Hong Kong." *Construction Management and Economics*, 23(9), 891-903.

Levy, S. M. (2002). *Project Management in Construction*, McGraw-Hill Professional, New York, USA.

Lewis, P., and Thomas, H. (1990). "The linkage between strategy, strategic groups, and performance in the U.K. retail grocery industry." *Strategic Management Journal*, 11(5), 385-397.

Lima, C., Stephens, J., and Bohms, M. (2003). "The bcXML: Supporting e-commerce and knowledge management in the construction industry." *Itcon Journal*, 8(Special Issue), 293-308.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, John Wiley, New York.

Lock, D. (2007). *Project Management*, 9th Ed., Gower Publishing Ltd., Aldershot.

Love, P. E. D., and Li, H. (2000). "Quantifying the causes and costs of rework in construction." *Construction Management and Economics*, 18(4), 479-490.

Lovelock, C. H. (1991). *Services marketing*, Prentice Hall, Upper Saddle River, NJ.

Lowe, J. G. (1987). "The measurement of productivity in the construction industry." *Construction Management and Economics*, 5(2), 101-113.

MacCallum, R. C., and Austin, J. T. (2000). "Applications of structural equation modeling in psychological research." *Annual Review of Psychology*, 51(1), 201.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., and Fabrigar, L. R. (1993). "The problem of equivalent models in applications of covariance structure analysis." *Psychological Bulletin*, 114(1), 185-199.

MacMillan, I. C., and McGrath, R. G. (1997). "Discovering new points of differentiation." *Harvard Business Review*, 75(4), 133-145.

Macnaughton-Smith, P., Williams, W. T., Dale, M. B., and Mockett, L. G. (1964). "Dissimilarity analysis: A new technique of hierarchical sub-division." *Nature*, 202(4936), 1034-1035.

Mahoney, J. T., and Pandian, J. R. (1992). "The resource-based view within the conversation of strategic management." *Strategic Management Journal*, 13(5), 363-380.

Majid, M. Z. A., and McCaffer, R. (1998). "Factors of non-excusable delays that influence contractors' performance." *Journal of Management in Engineering*, 14(3), 42-49.

Mallon, J. C., and Mulligan, D. E. (1993). "Quality function deployment- a system for meeting customers' needs." *Journal of Construction Engineering and Management*, 119(3), 516-531.

Manseau, A., and Seaden, G. (2001). *Innovation in Construction: An International Review of Public Policies*, Spon Press, London.

Mansfield, N. R., Ugwu, O. O., and Doran, T. (1994). "Causes of delay and cost overruns in Nigerian construction projects." *International Journal of Project Management*, 12(4), 254-260.

Marriott, F. H. C. (1971). "Practical Problems in a Method of Cluster Analysis." *Biometrics*, 27(3), 501-514.

Mascarenhas, B. (1989). "Strategic group dynamics." *The Academy of Management Journal*, 32(2), 333-352.

Mascarenhas, B., and Aaker, D. A. (1989). "Mobility barriers and strategic groups." *Strategic Management Journal*, 10(5), 475-485.

Mata, F. J., Fuerst, W. L., and Barney, J. B. (1995). "Information technology and sustained competitive advantage: A resource-based analysis." *MIS Quarterly*, 19(4), 487-505.

Mathur, S. S. (1988). "How firms compete: A new classification of generic strategies." *Journal of General Management*, 14(1), 30-57.

McGee, J., and Thomas, H. (1986). "Strategic groups: Theory, research, and taxonomy." *Strategic Management Journal*, 7(2), 141-160.

McGee, J., and Thomas, H. (1992). "Strategic groups and intra-industry competition." *International Review of Strategic Management*, 3, 77-98.

McNamara, C. (2002). *Field Guide to Leadership and Supervision for Nonprofit Staff*, Authenticity Consulting, LLC.

McNamara, G., Deephouse, D. L., and Luce, R. A. (2003). "Competitive positioning within and across a strategic group structure: The performance of core, secondary, and solitary firms." *Strategic Management Journal*, 24(2), 161-181.

Meehan, J. C., and Stuart, G. L. (2007). "Using structural equation modeling with forensic samples." *Criminal Justice and Behavior*, 34(12), 1560-1587.

Mehra, A., and Floyd, S. W. (1998). "Product market heterogeneity, resource imitability and strategic group formation." *Journal of Management*, 24(4), 511-531.

Mentzer, J. T. (2004). *Fundamentals of Supply Chain Management: Twelve Drivers of Competitive Advantage*, Sage Publications Inc., Thousand Oaks, California.

Miles, R. E., and Snow, C. C. (1978). *Organizational Structure, Strategy, and Process*, McGraw-Hill, New York.

Miller, C. C., and Cardinal, L. B. (1994). "Strategic planning and firm performance: A synthesis of more than two decades of research." *The Academy of Management Journal*, 37(6), 1649-1665.

Miller, D. (1986). "Configurations of strategy and structure: Towards a synthesis." *Strategic Management Journal*, 7(3), 233-249.

Miller, D., Friesen, P. H., and Mintzberg, H. (1984). *Organizations: A Quantum View*, Prentice-Hall, Englewood-Cliffs, N.J.

Milligan, G. W. (1980). "An examination of the effect of six types of error perturbation on fifteen clustering algorithms." *Psychometrika*, 45(3), 325-342.

Milligan, G. W., and Copper, M. C. (1988). "A study of standardization of variables in cluster analysis." *Journal of Classification*, 5(1), 181-204.

Milligan, G. W., and Isaac, P. D. (1980). "The validation of four ultrametric clustering algorithms." *Pattern Recognition*, 12(2), 41-50.

Mintzberg, H. (1987). "Five Ps for strategy." *California Management Review*, 30(1), 11-24.

Mintzberg, H. (1988). "Generic strategies: Toward a comprehensive framework." In: *Advances in Strategic Management*, Shrivista, P., Huff, A., and Dutton, J., eds., JAI press, Greenwich.

Mintzberg, H., and Quinn, J. B. (1998). *Readings in the Strategy Process*, Prentice Hall, Upper Saddle River, NJ.

Mirkin, B. G. (1996). *Mathematical Classification and Clustering*, Kluwer Academic Publishing, Dordrecht.

Mohamed, S. (1999). "Empirical investigation of construction safety management activities and performance in Australia." *Safety Science*, 33(3), 129-142.

Mohr, J., and Spekman, R. (1994). "Characteristics of partnership success: Partnership attributes, communication behavior, and conflict resolution techniques." *Strategic Management Journal*, 15(2), 135-152.

Morris, P. W. G., and Hough, G. H. (1987). *The Anatomy of Major Projects: A Study of the Reality of Project Management*, Wiley, Oxford, UK.

Mosakowski, E. (1993). "A resource-based perspective on the dynamic strategy-performance relationship: An empirical examination of the focus and differentiation strategies in entrepreneurial firms." *Journal of Management*, 19(4), 819-839.

Muthen, L. K., and Muthen, B. O. (2002). "Teacher's corner: How to use a Monte Carlo study to decide on sample size and determine power." *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599 - 620.

Nam, C. H., and Tatum, C. B. (1988). "Major characteristics of constructed products and resulting limitations of construction technology." *Construction Management and Economics*, 6(2), 133-147.

Nath, D., and Gruca, T. S. (1997). "Convergence across alternative methods for forming strategic groups." *Strategic Management Journal*, 18(9), 745-760.

NeuralWare, I. (2000). *Neural Computing, NeuralWorks Professional II/PLUS*, NeuralWare, Carnegie, Pennsylvania.

Ng, R. T., and Han, J. (1994). "Efficient and effective clustering methods for spatial data mining." In: *Proceedings of the 20th International Conference on Very Large Data Bases*, San Francisco, CA, USA, 144-155.

Nguyen, N., and Leblanc, G. (2001). "Corporate image and corporate reputation in customers' retention decisions in services." *Journal of Retailing and Consumer Services*, 8(4), 227-236.

Norusis, M. (2004). *SPSS 13.0 Statistical Procedures Companion*, Prentice Hall, Upper Saddle-River, N.J.

Nunally, J. (1978). *Psychometric Theory*, McGraw-Hill, New York.

O'Brien, M. J., and Al-Soufi, A. (1993). "Electronic data interchange and the structure of the UK construction industry." *Construction Management and Economics*, 11(6), 443-453.

Odeh, A. M., and Battaineh, H. T. (2002). "Causes of construction delay: Traditional contracts." *International Journal of Project Management*, 20(1), 67-73.

Odeyinka, H. A., and Yusif, A. (1997). "The causes and effects of construction delays on completion cost of housing projects in Nigeria." *Journal of Financial Management of Property and Construction*, 2(3), 31-44.

Ofori, G. (1992). "The environment: The fourth construction project objective?" *Construction Management and Economics*, 10(5), 369-395.

Ofori, G. (2000a). "Globalization and construction industry development: Research opportunities." *Construction Management and Economics*, 18(3), 257-262.

Ofori, G. (2000b). "Greening the construction supply chain in Singapore." *European Journal of Purchasing and Supply Management*, 6(3-4), 195-206.

Oja, M., Kaski, S., and Kohonen, T. (2003). "Bibliography of self-organizing map (SOM) papers: 1998-2001 Addendum." *Neural Computing Surveys*, 3, 1-156.

Oliver, C. (1997). "Sustainable competitive advantage: Combining institutional and resource-based views." *Strategic Management Journal*, 18(9), 697-713.

Olsson, U. H., Foss, T., Troye, S. V., and Howell, R. D. (2000). "The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality." *Structural Equation Modeling*, 7(4), 557-595.

Osborne, J. D., Stubbart, C. I., and Ramaprasad, A. (2001). "Strategic groups and competitive enactment: A study of dynamic relationships between mental modes and performance." *Strategic Management Journal*, 22(5), 435-454.

Ozorhon, B. (2007). "Modeling the performance of international construction joint ventures," Unpublished Doctoral Dissertation, Middle East Technical University, Ankara.

Pal, N. R., and Bezdek, J. C. (1995). "On cluster validity for the fuzzy C-means model." *IEEE Transactions on Fuzzy Systems*, 3(3), 370-379.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. (1985). "A conceptual model of service quality and its implications for future research." *Journal of Marketing*, 49(4), 41-50.

Park, M. (2005). "Model-based dynamic resource management for construction projects." *Automation in Construction*, 14(5), 585-598.

Pasquire, C. (1999). "The implications of environmental issues on UK construction management." *Engineering Construction and Architectural Management*, 6(3), 276-286.

Pearce II, J. A., Robbins, D. K., and Robinson, R. B., Jr. (1987). "The impact of grand strategy and planning formality on financial performance." *Strategic Management Journal*, 8(2), 125-134.

Pedersen, D. O. (1996). "The economics of innovation in construction. Economic management of innovation, productivity and quality in construction." In: *CIB W 55 Building Economics 7th International Symposium*, Katavic, M. ed., Zagreb , Croatia, 158-184.

Peña, J. M., Lozano, J. A., and Larrañaga, P. (1999). "An empirical comparison of four initialization methods for the K-Means algorithm." *Pattern Recognition Letters*, 20(10), 1027-1040.

Peteraf, M., and Shanley, M. (1997). "Getting to know you: A theory of strategic group identity." *Strategic Management Journal*, 18(Summer special issue), 165-186.

Petrovic-Lazarevic, S. (2008). "The development of corporate social responsibility in the Australian construction industry." *Construction Management and Economics*, 26(2), 93-101.

Pinto, J. K., and Slevin, D. P. (1988). "Critical success factors across the project life cycle." *Project Management Journal*, 19(3), 67-75.

Platts, W. S. (1980). "A plea for fishery habitat classification." *Fisheries*, 5(1), 2-6.

Porac, J. F., Thomas, H., and Baden-Fuller, C. (1989). "Competitive groups as cognitive communities: The case of Scottish knitwear manufactures." *Journal of Management Studies*, 26(4), 397-416.

Porter, M. E. (1979). "The structure within industries and companies' performance." *Review of Economics and Statistics*, 61(2), 214-227.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, Free Press, New York.

Porter, M. E. (1985). *Competitive Advantage*, Free Press, New York.

Powell, T. C. (1995). "Total quality management as competitive advantage: A review and empirical study." *Strategic Management Journal*, 16(1), 15-37.

Powell, T. C., and Dent-Micallef, A. (1997). "Information technology as competitive advantage: The role of human, business, and technology resources." *Strategic Management Journal*, 18(5), 375-405.

Powell, W. (1987). "Hybrid organizational arrangements: New form of transitional development." *California Management Review*, 30(1), 295-336.

Pölzlbauer, G. (2004). "Survey and comparison of quality measures for self-organizing maps." In: *Proceedings of the Fifth Workshop on Data Analysis (WDA-04)*, Elfa Academic Press, Sliezsky dom, Vysok´e Tatry, Slovakia, 67-82.

Price, A. D. F., and Newson, E. (2003). "Strategic management: Consideration of paradoxes, processes, and associated concepts as applied to construction." *Journal of Management in Engineering*, 19(4), 183-192.

Pries, F., and Janszen, F. (1995). "Innovation in the construction industry: The dominant role of the environment." *Construction Management and Economics*, 13(1), 43-51.

Raftery, J., Pasadilla, B., Chiang, Y. H., Hui, E. C. M., and Tang, B. S. (1998). "Globalization and construction industry development: Implications of recent developments in the construction sector in Asia." *Construction Management and Economics*, 16(6), 729-737.

Rainey, D. L. (2006). *Sustainable Business Development*, Cambridge University Press, New York.

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

Reed, R., and DeFillippi, R. J. (1990). "Causal ambiguity, barriers to imitation, and sustainable competitive advantage." *Academy of Management Review*, 15(1), 88-102.

Reger, R. K., and Huff, A. S. (1993). "Strategic groups: A cognitive perspective." *Strategic Management Journal*, 14(2), 103-124.

Reichstein, T., Salter, A. J., and Gann, D. M. (2005). "Last among equals: A comparison of innovation in construction, services and manufacturing in the UK." *Construction Management and Economics*, 23(6), 631-644.

Ren, Z., Anumba, C. J., and Ugwu, O. O. (2001). "Construction claims management: Towards an agent-based approach." *Engineering Construction and Architectural Management*, 8(3), 185-197.

Ring, P., and Van de Ven, A. (1992). "Structuring cooperative relationships." *Strategic Management Journal*, 13(7), 483-498.

Rizzo, R., and Chella, A. (2006). "A comparison between habituation and conscience mechanism in self-organizing maps." *IEEE Transactions on Neural Networks*, 17(3), 807-810.

Rosenfeld, Y. (1994). "Innovative construction methods." *Construction Management and Economics*, 12(6), 521-541.

Roubens, M. (1982). "Fuzzy clustering algorithms and their cluster validity." *European Journal of Operational Research*, 10(3), 294-301.

Rumelt, R. P. (1984). "Towards a strategy theory of the firm." In: *Competitive Strategic Management*, Lamb, R. B., ed., Prentice-Hall, Englewood Cliffs, NJ, 556-570.

Sarshar, M., Finnemore, M., Haigh, R., and Goulding, J. (1999). "Spice: is a capability maturity model applicable in the construction industry?" In: *Eighth International Conference on Durability of Building Materials and Components, 8 dbmc*, Vancouver; Canada, 2836-2843.

Satorra, A., and Bentler, P. M. (1994). "Corrections to test statistics and standard errors in covariance structure analysis." In: *Latent Variables Analysis: Applications for*

Developmental Research, von Eye, A. and Clogg, C. C., eds., Sage, Thousand Oaks, CA, 399-419.

Sawacha, E., Naoum, S., and Fong, D. (1999). "Factors affecting safety performance on construction sites." *International Journal of Project Management*, 17(5), 309-315.

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). "Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures." *Methods of Psychological Research Online*, 8(2), 23-74.

Schumacker, R. E., and Lomax, R. G. (2004). *A Beginner's Guide to Structural Equation Modeling*, Lawrence Erlbaum Associates, New Jersey.

Seber, G. A. F. (2004). *Multivariate Observations*, Wiley, New York, Chichester.

Selim, S. Z., and Ismail, M. (1984). "K-means-type algorithms: A generalized convergence theorem and characterization of local optimality." *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(1), 81-86.

Senouci, A. B., Eldin, N. N., and Asce, M. (2004). "Use of genetic algorithms in resource scheduling of construction projects." *Journal of Construction Engineering and Management*, 130, 869.

Serrano-Cinca, C. (1998). "From financial information to strategic groups: A self-organizing neural network approach." *Journal of Forecasting*, 17(1), 415-428.

Shen, L.-Y., Lu, W.-S., Yao, H., and Wu, D.-H. (2005). "A computer-based scoring method for measuring the environmental performance of construction activities." *Automation in Construction*, 14(3), 297-309.

Sheth, J. N., and Sharma, A. (1997). "Supplier relationships: Emerging issues and challenges." *Industrial Marketing Management*, 26(2), 91-100.

Shirazi, B., Langford, D. A., and Rowlinson, S. M. (1996). "Organizational structures in the construction industry." *Construction Management and Economics*, 14(3), 199-212.

Shook, C. L., Ketchen, D. J., Hult, G. T. M., and Kacmar, K. M. (2004). "Research notes and commentaries: An assessment of the use of structural equation modeling in strategic management research." *Strategic Management Journal*, 25(4), 397-404.

Singh, D., and Tiong, L. K. (2006). "Contractor selection criteria: Investigation of opinions of Singapore construction practitioners." *Journal of Construction Engineering and Management*, 132(9), 998-1008.

Siu, G. K. W., Bridge, A., and Skitmore, M. (2001). "Assessing the service quality of building maintenance providers: mechanical and engineering services." *Construction Management and Economics*, 19(7), 719-726.

Slaughter, E. S. (2000). "Implementation of construction innovations." *Building Research and Information*, 28(1), 2-17.

Snashall, D. (1990). "Safety and health in the construction industry." *British Medical Journal*, 301(6752), 563-564.

Sokal, R. R., and Michener, C. D. (1958). "A statistical method for evaluating systematic relationships." *University of Kansas Science Bulletin*, 38(1), 1409-1438.

SPSS, I. (2001). "The SPSS two-step cluster component." White Paper Technical Report.

Stalk, G. (1988). "Time--The next source of competitive advantage." *Harvard Business Review*, 66(4), 41-51.

Stalk, G., and Hout, T. M. (1990). *Competing Against Time: How Time-based Competition is Reshaping Global Markets*, Free Press, New York.

Stevens, J. (1996). *Applied Multivariate Statistics for the Social Science*, Laurence Erlbaum Associates, Mahway, NJ.

Struyf, A., Hubert, M., and Rousseeuw, P. (1996). "Clustering in an object-oriented environment." *Journal of Statistical Software*, 1(4), 1-30.

Struyf, A., Hubert, M., and Rousseeuw, P. J. (1997). "Integrating robust clustering techniques in S-PLUS." *Computational Statistics and Data Analysis*, 26(1), 17-37.

Sun, Y. (2000). "On quantization error of self-organizing map network." *Neurocomputing*, 34(1-4), 169-193.

Tah, J. H. M. (2005). "Towards an agent-based construction supply network modeling and simulation platform." *Automation in Construction*, 14(3), 353-359.

Tan, H. C., Carrillo, P. M., Anumba, C. J., Asce, M., Bouchlaghem, N. D., Kamara, J. M., and Udeaja, C. E. (2007). "Development of a methodology for live capture and reuse of project knowledge in construction." *Journal of Management in Engineering*, 23(1), 18-26.

Tang, M. J., and Thomas, H. (1992). "The concept of strategic groups: Theoretical construct or analytical convenience." *Managerial and Decision Economics*, 13(4), 323-329.

Tatum, C. B. (1987). "Business planning for design and construction firms." *Journal of Management in Engineering*, 3(2), 117-126.

Thomas, H., and Venkatraman, N. (1988). "Research on strategic groups: Progress and prognosis." *Journal of Management Studies*, 25(6), 537-555.

Timm, N. H. (2002). *Applied Multivariate Analysis*, Springer, New York.

Toakley, A. R., and Marosszeky, M. (2003). "Towards total project quality - A review of research needs." *Engineering, Construction and Architectural Management*, 10(3), 219-228.

Toivanen, O., Stoneman, P., and Bosworth, D. (2002). "Innovation and the market value of UK firms, 1989-1995." *Oxford Bulletin of Economics and Statistics*, 64(1), 39-61.

Tomarken, A. J., and Waller, N. G. (2005). "Structural equation modeling: Strengths, limitations, and misconceptions." *Annu. Rev. Clin. Psychol.*, 1(1), 31-65.

Toor, S.-U.-R., and Ogunlana, S. O. (2008). "Problems causing delays in major construction projects in Thailand." *Construction Management and Economics*, 26(4), 395-408.

Tosi, H. L., Mero, N. P., and Rizzo, J. R. (2000). *Managing Organizational Behavior*, 4th Ed., Blackwell Publishers, Massachusetts.

Turner, J. R. (2004). "Farsighted project contract management: Incomplete in its entirety." *Construction Management and Economics*, 22(1), 75-83.

Tyron, R. C. (1939). *Cluster Analysis*, Edward Brothers, Ann Arbor.

Udeaja, C. E., Kamara, J. M., Carrillo, P. M., Anumba, C. J., Bouchlaghem, N., and Tan, H. C. "A web-based prototype for live capture and reuse of construction project knowledge." *Automation in Construction*, In Press, Corrected Proof.

Vail, E. F. (1999). "Knowledge mapping: Getting started with knowledge management." *Information Systems Management*, 16(4), 1-8.

Van der Laan, M., Pollard, K., and Bryan, J. (2003). "A new partitioning around medoids algorithm." *Journal of Statistical Computation and Simulation*, 73(8), 575-584.

Vesanto, J., and Alhoniemi, E. (2000). "Clustering of the self-organizing map." *IEEE Transactions on Neural Networks*, 11(3), 586-600.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (1999). "Self-organizing map in Matlab: the SOM toolbox." In: *Proceedings of the Matlab DSP Conference*, Espoo, Finland, 16-17.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). *SOM Toolbox for Matlab 5*, Helsinki University of Technology.

Walton, M. (1988). *The Deming Management Method*, Perigee, New York.

Wang, S. (2001). "Cluster analysis using a validated self-organizing method: Cases of problem identification." *International Journal of Intelligent Systems in Accounting, Finance & Management*, 10, 127-138.

Wang, S., and Wang, H. (2002). "Knowledge discovery through self-organizing maps: Data visualization and query processing." *Knowledge and Information Systems*, 4(1), 31-45.

Ward, S. C., Curtis, B., and Chapman, C. B. (1991). "Objectives and performance in construction projects." *Construction Management and Economics*, 9(4), 343.

Warszawski, A. (1996). "Strategic planning in construction companies." *Journal of Construction Engineering and Management*, 122(2), 133-140.

Wells, J. (2001). "The construction industry in the twenty-first century: Its image, employment prospects and skill requirement." International Labor Organization, Geneva.

Wernerfelt, B. (1984). "A resource-based view of the firm." *Strategic Management Journal*, 5(2), 171-180.

Wiggins, R. R., and Ruefli, T. W. (1995). "Necessary conditions for the predictive validity of strategic groups: Analysis without reliance on clustering techniques." *The Academy of Management Journal*, 38(6), 1635-1656.

Williams, W. T. (1971). "Principles of Clustering." *Annual Reviews in Ecology and Systematics*, 2(1), 303-326.

Wright, P. M., Dunford, B. B., and Snell, S. A. (2001). "Human resources and the resource based view of the firm." *Journal of Management*, 27(6), 701-721.

Wright, P. M., and McMahan, G. C. (1992). "Theoretical perspectives for strategic human resource management." *Journal of Management*, 18(2), 295-320.

Wright, P. M., McMahan, G. C., and McWilliams, A. (1994). "Human resources and sustained competitive advantage: A resource-based perspective." *The International Journal of Human Resource Management*, 5(2), 301-326.

Wright, P. M., Smart, D. L., and McMahan, G. C. (1995). "Matches between human resources and strategy among NCAA basketball teams." *The Academy of Management Journal*, 38(4), 1052-1074.

Xie, X. L., and Beni, G. (1991). "A validity measure for fuzzy clustering." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 841-847.

Xie, Y. (2008). "Structural equation modeling: where are advantage?" <http://www.yihui.name/en/read.php/33.htm> (May 20, 2008).

Xu, R., and Wunsch, D. (2005). "Survey of clustering algorithms." *IEEE Transactions on Neural Networks*, 16(3), 645-678.

Yasamis, F., Arditi, D., and Mohammadi, J. (2002). "Assessing contractor quality performance." *Construction Management and Economics*, 20(3), 211 - 223.

Yuan, K.-H., and Bentler, P. M. (1997). "Improving parameter tests in covariance structure analysis." *Computational Statistics and Data Analysis*, 26(2), 177-198.

Zadeh, L. A. (1965). "Fuzzy sets." *Information and Control*, 8(3), 338-353.

Zeithaml, V. A. (1981). "How customer evaluation processes differ between goods and services." In: *Marketing of Services Conference*, American Marketing Association, Chicago, 186-190.

Zhang, T., Ramaksrishnan, R., and Livny, M. (1996). "BIRCH: An efficient data clustering method for very large databases." In: *Proceedings of the 1996 ACM SIGMOD international conference on management of data*, International Conference on Management of Data, Montreal, Canada, 103 – 114.

Zhang, X., and Li, Y. (1993). "Self-organizing map as a new method for clustering and data analysis." In: *Proceedings of 1993 International Joint Conference on Neural Networks*, IJCNN, Nagoya, 2448-2451.

Zúñiga-Vicente, J., de la Fuente-Sabaté, J. M., and Suárez González, I. (2004). "Dynamics of the strategic group membership- performance linkage in rapidly changing environments." *Journal of Business Research*, 57(12), 1378-1390.

Zupa, J. (1982). *Clustering of Large Data Sets*, Research Studies Press, New York.

APPENDIX A

A SAMPLE OF THE QUESTIONNAIRE

Unless otherwise stated, please use the average of last 3 years' figures while answering the following questions.

1. Please define the age of your organization : _____ years

2. By considering only the construction projects completed by your company within the last three years, please indicate your company's average yearly turnover value: _____ $/yr

3. By considering all the construction projects completed by your company so far, please indicate which type of projects form the highest percentage in your company's portfolio:

▪ Infrastructure (sewerages, pipe lines, city infrastructure etc.) :	☐
Housing and building construction (houses, touristic facilities, hospitals, universities, governmental buildings etc.) :	☐
Industrial construction (factories, refineries, powerhouses, etc.) :	☐
Others (roads, tunnels, bridges, dams, sprinkler systems etc.)	☐

4. By considering all the construction projects completed by your company so far, please indicate the ratio of international numeric workload to total workload.

: _____ %

5. By considering all the construction projects completed by your company so far, please indicate your major client which has the highest percentage in total number of projects.

☐ Government ☐ Private sector

6. Is there any activity related with systematic and regular strategic planning? Please choose the appropriate answer:

☐ Yes ☐ No

7. In your company are strategic decisions decided by democratic and collaborative ways or autocratic approach. In the light of the definitions supplied below, please indicate the strategic decision-making approach

Democracy: The thoughts and ideas of the employees working at any level of the organization are taken into account in decision concerning the organization or project. The power inside the organization is not held by a few executives but shared between the levels. Multi-disciplinary teamwork orientation, creativity and innovative actions are fostered in the organizations.

Autocracy: Company/Project is managed by a person/people who holds/hold the powering his/their hands. The decisions concerning the organization/project reflect only

the thoughts/decisions of the top level executives. Empowerment of employees in any departmental level is never observed inside the organization.

☐ Democratic approach ☐Autocratic approach

8. By considering all the projects that your company involved in so far, please define the strength of relations your company develops with its client organizations and customers:

☐ Very Poor ☐ Poor ☐Poor/Strong ☐Strong ☐Very Strong

9. Please indicate the level of investment and opportunities provided for employees in your organization (like providing job related trainings, proposing seminars, congresses, fairs, etc.):

☐Very Low ☐Low ☐Medium ☐ High ☐Very High

10. Please indicate the managerial capability of your organization on the following scale

☐ Very Poor ☐ Poor ☐Poor/Strong ☐Strong ☐Very Strong

11. Please define the technical capability of your organization on the following scale:

☐ Very Poor ☐ Poor ☐Poor/Strong ☐Strong ☐Very Strong

12. Please define the financial resources of your organization on the following scale:
☐ Very Poor ☐ Poor ☐Poor/Strong ☐Strong ☐Very Strong

13. Please indicate to what extend your company experience provide competitive advantage:

☐ Very Poor ☐ Poor ☐Poor/Strong ☐Strong ☐Very Strong

14. Please indicate the mostly utilized generic strategies of your organization:

☐ Cost leadership strategy ☐ Quality/Service differentiation strategy

15. Does your company operate in the sectors <u>related</u> with the construction business? (for example, production of construction materials, operation and maintenance services)

☐ Yes ☐ No

16. Does your company operate in the sectors <u>unrelated</u> with the construction business? (like textile, finance, tourism, etc.)

☐ Yes ☐ No

17. By considering the other companies involved in the construction industry, please define the level of performance for your organization:

☐Very Low ☐Low ☐Medium ☐High ☐ Very High

APPENDIX B

A SAMPLE OF THE DIFFERENTIATION QUESTIONNAIRE

The concept of the questionnaire

According to Porter (1980), the companies can adopt two modes of competition for creating and sustaining competitive advantage.

1. Cost leadership: implies that a firm emphasizes low cost relative to its competitors. Such an approach calls for a strong emphasis on cost reductions by adopting tight cost and overhead control, avoiding marginal customer accounts minimizing cost across the departments, and conducting operations and activities in an efficient manner.
2. Differentiation: implies that a firm offers something unique and unmatched by its competitors, and valued by the industry, which enables the firm to command higher prices than industry average. Such an approach calls for differentiating different aspects of the business such as the products or services offered, the technology used, the delivery system offered, the marketing approach adopted, and a wide range of other aspects, depending on a particular industry's characteristics.

According to the research about "strategic group analysis in Turkish construction companies", one of the most important factors which determined the performance difference between the companies was determined as mode of the strategy. In conclusion, the companies adopting differentiation strategy shows higher performance than the other companies.

The major objective of this research is determination of how the companies can differentiate themselves from the other companies, and which activities placed in the value chain of the construction companies have effect on creating differentiation.

1. Define the age of your organization in construction sector: **……..** Years
2. By considering only the construction projects completed by your company within the last three years including international and national projects, please indicate your company's total turnover value: **…….** $
3. By considering only the construction projects completed by your company within the last three years, please indicate your company's total international turnover value: **…….** $
4. By considering the other construction companies in Turkish construction sector, please indicate the size of your company.

☐ Small ☐ Small-medium ☐Medium ☐Medium-large ☐Large

5. According to the total size of the projects performed by your company, indicate first five countries in order.

I. ……..
II. ……..
III. ……..
IV. ……..
V. ……..

6. By considering all the construction projects completed by your company so far, please rank the type of projects in your company's portfolio:

Type of the project	**Rank**
Infrastructure (sewerages, pipe lines, city infrastructure etc.)	
Industrial Facilities (factories, refineries, powerhouses, etc.)	
Housing	
Building (touristic facilities, hospitals, universities, etc.)	
Water Structures (dams, irrigation systems, etc.)	
Transportation Structures (roads, tunnels, bridges, airports, etc.)	
Others (please indicate with rank)	

7. The construction companies can differentiate themselves from the other companies in which activities of the project management. Please indicate the potential of these activities for creating a competitive advantage by differentiating according to the scale.

Activities	Very Low	Low	Medium	High	Very High
Schedule Management					
Project Organization					
Contract Management					
Resource Management					
Project Monitoring & Control					
Quality Management					
Risk Management					
Project Cost Estimation					
Environmental Management					
Health and Safety Mana.					

8. The construction companies can differentiate themselves from the other companies in which activities of the corporate management. Please indicate the potential of these activities for creating a competitive advantage by differentiating according to the scale.

Activities	Very Low	Low	Medium	High	Very High
Strategic Management					
Work Development					
Human Resource Management					
Financial Management					
Institutional Structure & Professional Management					
Organizational Learning					
Research and Development					
Tendering					
Claim Management					
Knowledge Management					

9. The construction companies can differentiate themselves from the other companies in which resources. Please indicate the potential of these activities for creating a competitive advantage by differentiating according to the scale.

Resources	Very Low	Low	Medium	High	Very High
Human Resource					
Machinery and Equipment					
Financial Resources					
Experience and Knowledge					
Technology					

10. The construction companies can differentiate themselves from the other companies in which relations. Please indicate the potential of these activities for creating a competitive advantage by differentiating according to the scale.

Relations	Very Low	Low	Medium	High	Very High
Relationship with the suppliers					
Relationship with the subcontractors					
Relationship with the clients					
Relationship with the competitors					
Relationship with the internal stakeholders					
Relationship with the other parties					

11. Please indicate the potential of the differentiation ways for your company according to the scale.

Relations	Very Low	Low	Medium	High	Very High
Differentiation by increasing the productivity					
Differentiation by increasing the service quality					
Differentiation by increasing the product quality					
Differentiation by shortening duration of the project or completing the project on time					
Differentiation by creating innovative solutions					
Differentiation by creating positive image					

APPENDIX C

DESCRIPTIVE STATISTICS

Table C.1: Statistics for general information about the companies

Statistics	*Age (Years)*	*Size*	*Total Turnover ($M)*	*Overseas Turnover ($M)*	*Number of countries(1-5 scale)*	*Number of areas (1-5 scale)*
Mean	34.27	4.23	775.54	542.51	3.49	3.73
Standard Error	1.85	0.14	181.21	173.92	0.20	0.20
Median	35.00	5.00	265.00	105.00	4.00	3.50
Mode	30.00	5.00	1000.00	0.00	5.00	3.00
Standard Deviation	14.55	1.08	1415.32	1358.38	1.56	1.59
Sample Variance	211.75	1.16	2003133.65	1845203.87	2.42	2.53
Kurtosis	-0.28	1.41	30.44	40.30	-1.45	-1.17
Skewness	-0.38	-1.44	4.93	5.91	-0.38	0.09
Range	68.50	4.00	9998.83	10000.00	4.00	5.00
Minimum	1.50	1.00	1.17	0.00	1.00	1.00
Maximum	70.00	5.00	10000.00	10000.00	5.00	6.00
Count	62.00	62.00	61.00	61.00	61.00	62.00

Table C.2: Statistics for project management

	Statistics	Schedule Mana.	Project Organization	Contract Mana.	Resource Mana.	Project Monitoring & Control	Quality Mana.	Risk Mana.	Project Cost Estimation	Environmental Management	Health and Safety Mana.
Statistic	N	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00
	Range	3.00	3.00	2.00	2.00	3.00	2.00	2.00	3.00	3.00	3.00
	Minimum	2.00	2.00	3.00	3.00	2.00	3.00	3.00	2.00	2.00	2.00
	Maximum	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	Mean	4.10	4.29	4.10	4.26	4.15	4.18	3.95	4.39	3.21	3.66
	Std. Deviation	0.94	0.69	0.59	0.68	0.67	0.69	0.66	0.69	0.83	0.90
	Variance	0.88	0.47	0.35	0.46	0.45	0.48	0.44	0.47	0.69	0.82
	Skewness	-0.57	-1.08	-0.02	-0.36	-0.51	-0.25	0.05	-0.99	0.47	0.05
	Kurtosis	-0.85	2.29	-0.09	-0.78	0.61	-0.85	-0.65	1.10	-0.12	-0.86
Std. Error	Mean	0.12	0.09	0.08	0.09	0.09	0.09	0.08	0.09	0.11	0.11
	Skewness	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
	Kurtosis	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60

Table C.3: Statistics for corporate management

	Statistics	Strategic Planning	Work Development	Human Resource Man.	Financial Management	Instut. Stru.& Profess.Mana.	Organizational Learning	Research and Development	Tendering	Claim Management	Knowledge Management
Statistic	N	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00	62.00
	Range	2.00	2.00	2.00	2.00	2.00	3.00	4.00	2.00	3.00	4.00
	Minimum	3.00	3.00	3.00	3.00	3.00	2.00	1.00	3.00	2.00	1.00
	Maximum	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	Mean	4.34	4.24	3.89	4.06	4.29	3.98	3.74	4.16	4.03	3.71
	Std. Deviation	0.63	0.67	0.75	0.70	0.69	0.82	1.05	0.61	0.81	0.82
	Variance	0.39	0.45	0.56	0.49	0.47	0.67	1.11	0.37	0.66	0.67
	Skewness	-0.39	-0.32	0.19	-0.09	-0.45	-0.34	-0.67	-0.08	-0.25	-0.71
	Kurtosis	-0.62	-0.75	-1.17	-0.89	-0.80	-0.58	-0.05	-0.31	-0.93	1.15
Std. Error	Mean	0.08	0.09	0.10	0.09	0.09	0.10	0.13	0.08	0.10	0.10
	Skewness	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
	Kurtosis	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60

Table C.4: Statistics for resources

	Statistics	Human Resources	Machinery/ Equipment	Financial Resources	Experience& Knowledge	Image	Technology
Statistic	N	62.00	62.00	62.00	62.00	62.00	62.00
	Range	3.00	3.00	2.00	2.00	4.00	3.00
	Minimum	2.00	2.00	3.00	3.00	1.00	2.00
	Maximum	5.00	5.00	5.00	5.00	5.00	5.00
	Mean	4.13	3.68	4.05	4.55	3.92	4.11
	Std. Deviation	0.78	0.78	0.69	0.56	0.95	0.70
	Variance	0.61	0.62	0.47	0.32	0.90	0.50
	Skewness	-0.45	-0.62	-0.06	-0.77	-0.56	-0.75
	Kurtosis	-0.56	0.15	-0.82	-0.43	0.01	1.24
Std. Error	Mean	0.10	0.10	0.09	0.07	0.12	0.09
	Skewness	0.30	0.30	0.30	0.30	0.30	0.30
	Kurtosis	0.60	0.60	0.60	0.60	0.60	0.60

Table C.5: Statistics for relations

	Statistics	Relationship with the suppliers	Relationship with the subcontractors	Relationship with the clients	Relationship with the competitors	Relationship with internal stakeholders	Relationship with the other parties
Statistic	N	62.00	62.00	62.00	62.00	62.00	62.00
	Range	4.00	4.00	2.00	4.00	2.00	4.00
	Minimum	1.00	1.00	3.00	1.00	3.00	1.00
	Maximum	5.00	5.00	5.00	5.00	5.00	5.00
	Mean	3.66	3.73	4.45	3.24	4.11	3.77
	Std. Deviation	0.87	0.83	0.62	0.84	0.58	0.88
	Variance	0.75	0.69	0.38	0.71	0.33	0.77
	Skewness	-0.67	-1.02	-0.67	-0.15	0.01	-0.44
	Kurtosis	1.37	2.30	-0.47	-0.12	0.05	0.40
Std. Error	Mean	0.11	0.11	0.08	0.11	0.07	0.11
	Skewness	0.30	0.30	0.30	0.30	0.30	0.30
	Kurtosis	0.60	0.60	0.60	0.60	0.60	0.60

Table C.6: Statistics for Differentiation methods

	Statistics	Productivity	Service Quality	Product Quality	Time	Innovative Solutions	Positive Image
Statistic	N	62.00	62.00	62.00	62.00	62.00	62.00
	Range	2.00	3.00	3.00	2.00	2.00	3.00
	Minimum	3.00	2.00	2.00	3.00	3.00	2.00
	Maximum	5.00	5.00	5.00	5.00	5.00	5.00
	Mean	4.37	4.08	4.13	4.36	4.34	4.05
	Std. Deviation	0.58	0.86	0.91	0.68	0.70	0.82
	Variance	0.34	0.73	0.84	0.46	0.49	0.67
	Skewness	-0.26	-0.65	-0.80	-0.58	-0.58	-0.46
	Kurtosis	-0.70	-0.21	-0.22	-0.70	-0.78	-0.44
Std. Error	Mean	0.08	0.11	0.12	0.09	0.09	0.10
	Skewness	0.30	0.30	0.30	0.31	0.30	0.30
	Kurtosis	0.60	0.60	0.60	0.60	0.60	0.60

Table C.7: Correlation matrices for the determinant constructs of the model

	Schedule Management	Project Organization	Contract Management	Resource Management	Project Monitoring& Control	Quality Management	Risk Management	Project Cost Estimation	Environmental Management
Schedule Management	1								
Project Organization	0.389(**) 0.001	1							
Contract Management	0.249(*) 0.026	0.252(*) 0.024	1						
Resource Management	0.375(**) 0.001	0.366(**) 0.002	0.346(**) 0.003	1					
Project Monitoring& Control	0.237(*) 0.032	0.474(**) 0.000	0.293(*) 0.010	0.348(**) 0.003	1				
Quality Management	0.227(*) 0.038	0.305(**) 0.008	0.318(**) 0.006	0.322(**) 0.005	0.473(**) 0.000	1			
Risk Management	0.219(*) 0.044	0.283(*) 0.013	0.346(**) 0.003	0.284(*) 0.013	0.199 0.060	0.449(**) 0.000	1		
Project Cost Estimation	0.400(**) 0.001	0.245(*) 0.028	0.431(**) 0.000	0.382(**) 0.001	0.337(**) 0.004	0.199 0.061	0.294(*) 0.010	1	
Environmental Management	0.205 0.055	-0.022 0.432	0.224(*) 0.040	0.164 0.101	0.120 0.176	0.476(**) 0.000	0.434(**) 0.000	0.028 0.415	1
Health and Safety Management	0.214(*) 0.048	0.108 0.201	0.215(*) 0.047	0.011 0.465	0.243(*) 0.028	0.492(**) 0.000	0.355(**) 0.002	0.162 0.104	0.706(**) 0.000

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

Table C.7: Correlation matrices for the determinant constructs of the model (continued)

	Strategic Planning	Work development	Human Resource Management	Financial Management	Instut. Stru.& Profess.Mana.	Organizational Learning	Research and Development	Tendering	Claim Management
Strategic Planning	1								
Work development	0.310(**) 0.007	1							
Human Resource Management	0.363(**) 0.002	0.317(**) 0.006	1						
Financial Management	0.287(*) 0.012	0.246(*) 0.027	0.390(**) 0.001	1					
Inst. Structure& Prof. Mana.	0.263(*) 0.019	0.415(**) 0.000	0.192 0.067	0.165 0.099	1				
Organizational Learning	0.330(**) 0.004	0.007 0.478	0.344(**) 0.003	0.030 0.407	.503(**) 0.000	1			
Research and Development	0.333(**) 0.004	0.484(**) 0.000	0.440(**) 0.000	0.179 0.082	.241(*) 0.030	.223(*) 0.041	1		
Tendering	0.243(*) 0.029	0.468(**) 0.000	0.149 0.124	0.208 0.053	0.161 0.105	-0.028 0.415	0.143 0.133	1	
Claim Management	0.043 0.371	0.016 0.452	0.006 0.481	-0.149 0.124	-0.076 0.278	0.100 0.220	0.068 0.301	.357(**) 0.002	1
Knowledge Management	0.227(*) 0.038	0.220(*) 0.043	0.267(*) 0.018	0.263(*) 0.019	.299(**) 0.009	.238(*) 0.032	.292(*) 0.011	-0.069 0.296	-0.085 0.256

Table C.7: Correlation matrices for the determinant constructs of the model (continued)

	Relationship with the suppliers	Relationship with the subcontractors	Relationship with the clients	Relationship with the competitors	Relationship with the other parties
Relationship with the suppliers	1				
Relationship with the subcontractors	0.890(**) 0.000	1			
Relationship with the clients	0.381(**) 0.001	0.371(**) 0.001	1		
Relationship with the competitors	0.405(**) 0.001	0.236(*) 0.032	0.353(**) 0.002	1	
Relationship with the other parties	0.502(**) 0.000	0.430(**) 0.000	0.433(**) 0.000	0.541(**) 0.000	1
Relationship with internal stakeholders	0.242(*) 0.029	0.168 0.096	0.085 0.257	0.314(**) 0.006	0.181 0.079

Table C.7: Correlation matrices for the determinant constructs of the model (continued)

	Human Resources	Machinery/ Equipment	Financial Resources	Experience& Knowledge	Image	Technology
Human Resources	1					
Machinery/ Equipment	0.257(*) 0.022	1				
Financial Resources	0.233(*) 0.034	0.394(**) 0.001	1			
Experience& Knowledge	0.322(**) 0.005	0.184 0.076	0.269(*) 0.017	1		
Image	0.037 0.389	0.097 0.227	0.182 0.078	0.115 0.186	1	
Technology	0.242(*) 0.029	0.512(**) 0.000	0.259(*) 0.021	0.338(**) 0.004	0.260(*) 0.021	1

Table C.7: Correlation matrices for the determinant constructs of the model (continued)

	Productivity	Service Quality	Product Quality	Time	Innovative Solutions	Positive Image
Productivity	1					
Service Quality	0.137 0.144	1				
Product Quality	0.156 0.113	0.700(**) 0.000	1			
Time	0.243(*) 0.028	0.317(**) 0.006	0.031 0.407	1		
Innovative Solutions	0.332(**) 0.004	0.091 0.242	0.110 0.198	.363(**) 0.002	1	
Positive Image	0.031 0.406	.322(**) 0.005	0.167 0.097	0.116 0.185	0.171 0.092	1

APPENDIX D

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name	: Cenk Budayan
Nationality	: Turkish
Date and Place of Birth	: 09.03.1980, Ankara
Marital Status	: Single
Phone	: +90 312 210 54 02
Fax	: +90 312 210 54 01
Email	: budayan@metu.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
BS	METU Civil Engineering	2003
High School	Milli Piyango Anatolian High School, Ankara	1998

WORK EXPERIENCE

Year	Place	Enrollment
2004- Present	METU Department of Civil Engineering	Research Assistant
2002 July	Enis A.S.	Intern Engineering Student
2001 July	Limak A.S.	Intern Engineering Student

PUBLICATIONS

1. Budayan C., Dikmen I. and Birgonul M. T., (2007), "Strategic group analysis by using self organizing maps", Proceedings of ARCOM 2007 (Association of Researchers in Construction Management), Belfast, UK, 3-5 September, Vol.1, pp.223-232.

2. Budayan, C., Dikmen, İ. ve Birgönül, M.T. (2007), "Türk İnşaat Sektöründe Stratejik Grup Analizi", 4. İnşaat Yönetimi Kongresi, 30-31 Ekim, İstanbul, s.139-148.

3. Dikmen, I., Birgonul, M. T. and Budayan C. "Strategic group analysis in the construction industry", ASCE Journal of Construction Engineering and Management, submitted for publication.

AWARDS

1. The 'Rod Howes' Commemorative Award, with the research paper entitled "Strategic Group Analysis by Using Self Organizing Maps" presented in the 23rd Annual ARCOM Conference, Belfast, 3-5 September (2007) (Co-authors: Irem Dikmen and M. Talat Birgonul)

2. 2004-2005 Academic year METU graduate courses performance award