Profiling of sunk cost industries by soft clustering techniques: Turkey case

M. Akif Arvas*

Department of Economics,
Van Yüzüncü Yıl University, Zeve Campus,
65080,Van, Turkey
E-mail: aarvas@yyu.edu.tr
*Corresponding author

Ahmet Selman Bozkır

Department of Computer Science,
Hacettepe University, Beytepe Campus,
06800, Ankara, Turkey
E-mail: selman@cs.hacettepe.edu.tr

Abstract: The aim of this study is to determine the convenience of Sutton's game-theoretic sunk cost theory by profiling Turkish manufacturing industries via fuzzy c-means (FCM) and expectation-maximisation (EM) clustering methods. The effects of sunk costs on market structure are analysed separately for exogenous and endogenous sunk cost industries in Turkish manufacturing sectors in the period 1992–2001. The results suggest that as market size increases, so does the level of advertising and R&D outlays leading to a concentrated market structure in endogenous sunk cost industries. On the other hand, in exogenous sunk cost industries, market structure becomes more competitive and number of new entrant firms rise as market size growths. According to the profiles constituted by FCM and EM methods, these findings are consistent with Sutton's predictions. Furthermore, the most important factors which discriminates the clusters are revealed by FCM method.

Keywords: Turkish manufacturing industries; sunk cost industries; clustering; fuzzy clustering; expectation-maximisation clustering.

Reference to this paper should be made as follows: Akif Arvas, M. and Selman Bozkır, A. (2013) 'Profiling of sunk cost industries by soft clustering techniques: Turkey case', *Int. J. Industrial and Systems Engineering*, Vol. 15, No. 3, pp.290–303.

Biographical notes: M. Akif Arvas is an Associate Professor of Economics at Yüzüncü Yıl University, Van, Turkey. He has studied the effects of R&D and Advertising expenditures on oligopolistic firms' behaviours and competition for his PhD dissertation in Turkey and Italy. His teaching and research interests include industrial organisation theory, mathematical economics and applied econometrics. He holds a PhD in Economics from Hacettepe University.

Ahmet Selman Bozkır received his BSc degree from the Computer Engineering Department of Eskisehir Osmangazi University in 2006. He received his MSc degree from the Computer Engineering Department of Hacettepe University in 2009. He is currently a Research Assistant in the Computer Engineering

Department of Hacettepe University and continuing his PhD studies at the same university. His areas of interest are data mining, data warehousing and web based application technologies.

This paper is a revised and expanded version of a paper entitled 'Profiling of sunk cost industries by various clustering techniques: Turkey case' presented at The Second International Fuzzy System Symposium, Hacettepe University, 17–18 November 2011.

1 Introduction

The structure-conduct-performance paradigm (SCP) in economics, which began with Bain (1956) asserts that there is a one-way chain of causation that runs from market structure (concentration rate of firms) to conduct (the pricing behaviour of firms), and then to performance (profitability). High concentration (few firms and less competition), which was argued, facilitated collusion and led to high profits. According to Bain, these high profits could be traced to the presence of certain 'barriers to entry' whose nature is taken to be *exogenously* given (not by firm's strategy but by industry conditions). Due to these high barriers, potential (new) firms or some small-sized firms will not able to find an opportunity to survive in an industry. Among these barriers, economies of scale, cost advantages and product differentiation, such as advertising intensity or research and development (R&D) intensity, are considered to be the most important ones, see, e.g., (Davies, 1989). In short, in Bain's approach, market structure determines firms' pricing behaviours and profits.

After Bain's and his followers' studies, the analysis of market structure has gained new momentum since the 1990s. One major factor behind this momentum has been the game-theoretical and mathematical researches of Sutton (1989, 1991, 2007). Sutton strictly criticises Bain's paradigm. According to Sutton, there is a reverse chain of causation according to which firms' behaviours and policies affect market structure (a competitive market or a collusive market). Further, in his theory, barriers to entry are *endogenously* determined. Sutton has developed a game-theoretic model on what determines behaviour of firms and market structure. In that framework, the level of sunk costs plays an important role.

What does a sunk cost mean? A sunk cost is both fixed and cannot be recovered; that is, it has no salvage value. Therefore it is an irrecoverable sunk cost and basically this part of sunk cost in Sutton's theory is treated as one of the most influential factors on entry decision and competition policies of firms. The framework of Sutton's theory relies upon the distinction between exogenous (Type 1) and endogenous sunk costs (Type 2).

Type 1 industries model rests on two-stage game. At the first stage, firms decide to enter a market and then incur the exogenous sunk costs. The most important exogenous sunk cost is the setup cost and this cost is outside of a firm's control. At the second stage of the game, firms compete over *prices*. But for those firms, the levels of R&D, innovation and advertising do not play an important role in both competition policy and in the formation of market structure. Thus, in these industries, there is a positive correlation between the number of firms and the market size. So, as the market size grows the level of concentration rate decreases. This means market structure becomes more competitive, since a growing market attracts many firms leading to an increase in the number of firms in the industry. What allows to firms to easily enter the market is the low level of sunk costs (low entry barriers). In addition, in these industries the level of R&D and advertising expenditures is expected to be low and unimportant.

As a simple example of a Type 1 industry, assumes the total value of market demand is fixed at S (i.e., there is unit elasticity of demand). If all N firms in the market have the same constant marginal costs, and competition is Cournot (quantity competition), gross profit will be $\pi = S/N^2$. If each firm must incur a fixed cost, σ, upon entry, the free entry zero-profit condition ($\pi = \sigma$) gives the long run equilibrium number of firms: $N = [S/\sigma]^{1/2}$.

In contrast, an endogenous sunk cost is determined by business strategy. An endogenous sunk cost influences consumers' willingness to pay for firm's product sales. Thus, product differentiation (by R&D, innovation and advertising) plays a very important role. Sutton classifies advertising and R&D as the key endogenous sunk costs. In Type 2 industries model, there are three stages. At the first stage, firms take an entry decision and then, at the second stage, firms incur or determine R&D and advertising expenditures. At the final stage, firms compete over *product quality*, not over price. That means that a firm continuously has to make additional sunk expenditures in order to stay in the industry. As these expenditures increase, so does the level of entry barriers and this causes a few firms to survive and others to be obliged to exit from industry because of high entry barriers. On the other hand, greater product differentiation itself tends to decrease the toughness of price competition at small market sizes. Briefly, as market grows (an increase in the volume of sales), the level of advertising and R&D spending tends to increase. However the level of concentration does not converge to zero and remains high. This means that most of the industry sales will still be made by few incumbent firms which case describes the fundamental characteristic of Type 2 industry structure.

Type 2 industry theory assumes market demand is fixed at S, and allows consumer utility to depend not just on quantity consumed, but on the product of quantity and u, where u is an index of (perceived) product quality (by consumers). Sutton (1991) proposes the following relationship between endogenous sunk costs (denoted by E), and u: $E(u) = [\alpha / \gamma]\,[u^{\gamma} - 1]$ where $\gamma > 1$; a higher γ corresponds to more rapidly diminishing returns to investment; α reflects the unit cost of investing. Given strictly positive E, and noting that in equilibrium firms will invest until

$$\left.\frac{d\pi}{du}\right|_{u=\bar{u}} = \left.\frac{dE}{du}\right|_{u=\bar{u}} \tag{1}$$

the symmetric subgame perfect Cournot-Nash equilibrium with free entry $\left(\pi = \frac{S}{N^2} = E + \sigma\right)$ generates

$$N + N^{-1} - 2 = \frac{\gamma}{2}\left[1 - \left(\sigma - \frac{a}{\gamma}\right)\frac{N^2}{S}\right]. \tag{2}$$

In the limit, as $S \to \infty$, $[N + N^{-1}] \to [2 + (\gamma / 2)]$ which is a finite constant. Furthermore, $\mathrm{sgn}\{dN/dS\} = \mathrm{sgn}\{\sigma - (\alpha / \gamma)\}$, which confirms the possibility that concentration may actually rise with market size.

Current empirical analyses, which aim to investigate the effects of sunk costs on firms' choices and behaviours, hence on market structure, extensively use simplex method, panel data and recently stochastic frontier approaches. Instead of employing of these approaches, fuzzy c-means clustering (FCM) and EM (expectation-maximisation) methods are preferred to use in this study. According to what Peneder (1999) points out, cluster analysis represents a heuristic method for the exploration and identification of underlying patterns in data. Therefore, the goal of this study is to investigate and determine whether Sutton's predictions are robust for Turkish manufacturing industries via clustering approaches in data mining field.

As a descriptive data mining methodology, clustering is the finding of natural groups in whole dataset aiming to obtain maximum similarity among the cases in each cluster and maximum distances between different clusters. In literature, clustering methods are classified in different points of view. Yet, one classification approach divides all methods into two categories: *exclusive* and *overlapping* (Tan et al., 2005). In overlapping (non-exclusive) clustering, every case simultaneously falls into more than one cluster, whereas exclusive clustering assigns each case into only one cluster. In the case where the objects are decoupled well, exclusive clustering can be an ideal choice which builds a crisp classification of objects (Tan et al., 2005). But, as Tan et al. (2005) states that, objects cannot be segmented into well-separated clusters in most real world datasets, therefore there will be an indispensible arbitrariness in assignment of objects into particular clusters. As the dataset of this study consists of economical variables which are influenced by many complex real world factors, fuzzy clustering approach is mainly preferred due to its fuzzy modelling feature which is based on fuzzy set theory. However, as EM clustering technique is a robust method and does not prerequisite cluster count, it is employed for determining the optimal cluster count and comparison with fuzzy model.

Fuzzy clustering is a very broad field and there are numerous studies in this subject. In literature, although fuzzy clustering has many examples in the engineering point of view such as grade estimation (Tütmez and Tercan, 2006), image segmentation (Noordam et al., 2002), soil clustering (Göktepe et al., 2005) and audio signal classification (Park, 2009), number of the studies related to sunk cost analysis and clustering is remarkably low. One of the first attempts is done by Peneder (1999). Peneder (1995) analysed industrial competitiveness via clustering techniques. In another study, he (Peneder, 1999) investigated and analysed the relations of intangible investments and human resource in new two taxonomies of manufacturing industries by hierarchical clustering.

This paper is organised as follows: employed clustering methods are explained in Section 2. In Section 3, data is introduced. Definition and details of data and how the FCM and EM clustering methods were applied is described in Section 4. Section 4 also covers the detailed results obtained by applying both clustering methods subjected to this paper. In Section 5, obtained cluster profiles are discussed and whether or not the Sutton's sunk cost theory is consistent with soft clustering techniques. The conclusion is made in Section 6.

2 Clustering

In 1965, Lutfi Zadeh (1965) demonstrated fuzzy set theory which allows a case to be assigned in a set with a degree of membership ranging from 0 to 1 and fuzzy logic which enables a statement to be true with a certainty between 0 and 1. Based on this theory, FCM clustering method, which aims to allow objects to be a part of more than one cluster, was first developed by Dunn (1973) and developed by Bezdek (1981). Liu and Xu (2008), describes FCM algorithm as a partitioning method which segments objects in clusters according to membership degrees. According to Höppner, FCM requires cluster count c as an input parameter and segments data into fuzzy clusters by providing typical prototypes for each of them (Höppner, 2002). Further, link between objects and cluster prototypes is expressed via a membership matrix. In essence, the goal of the FCM is to minimise the objective function given in (3):

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, 1 \le m \le \infty \tag{3}$$

where u_{ij} is the membership degree of x_i in the cluster j, m is a real number denoting the fuzziness coefficient greater than 1, x_i is the ith of d-dimensional data and c_j is the cluster centroid of cluster j. Further, fuzzy segmentation is done with the optimisation of the first equation by the update of membership u_{ij} and the cluster centroids c_j given in (4):

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-2}}}, \quad Cj = \frac{\sum_{i=1}^{N} u_{ij}^{m} \bullet x_i}{\sum_{i-1}^{N} u_{ij}^{m}} \tag{4}$$

Suppose that β denotes the stopping criteria ranging from 0 to 1 and k is the iteration count, whenever $max_{ij} \left\{ u_{ij}^{k+1} - u_{ij}^{k} \right\} < \beta$ than the minimisation process is stopped which means that local minimum of J_m is achieved. Further information about the algorithm of FCM can be found in (*A Tutorial on Clustering Algoritms, Fuzzy C-means Clustering,* 2011; Höppner, 2002).

As can be seen, FCM is an iterative algorithm and is being affected deeply by the fuzziness factor m. At this point, Tan et al. (2005) noted that if m is selected close to 1, FCM behaves like ordinary K-means algorithm. In contrast, increment of m results centroids to converge to the global centroid of all data objects. Thus, when m gets larger, clusters become fuzzier.

As stated in Tan et al. (2005), FCM has the advantage of describing the membership degree of each case to every cluster. However, as a shortcoming, FCM requires many experiments for a mature evaluation of clustering results. The main reason of this fact is the random selection of initial cluster centroids. Therefore, decision makers should be aware of this fact and run the algorithm for many times and get insight after all.

2.1 Brief review of EM clustering

EM clustering method is a kind of unsupervised soft clustering technique which is first introduced by Dempster et al. (1977). Conceptually, EM utilises mixture models which treat the data as a collection of cases from a mixture of different probability distributions

(Tan et al., 2005). These distributions define different clusters. Generally, EM algorithm tries to discover the parameters of the probability distribution that has the maximum likelihood of its attributes. EM consists of two iterative stages: e-step and m-step. In e-step, algorithm calculates the probability of each object belonging to each cluster. In m-step, the parameters of the probability distribution of each class are re-estimated and e-step is run again. When the distribution parameters converge to pre-defined value or number of iterations reaches the maximum limit, algorithm finishes building clustering model. In essence, every case belongs to every distribution (cluster) with some probability (Tan et al., 2005).

Mixture models constitute basis for EM framework. In Zhang et al. (2003), a deeper explanation of EM algorithm can be found. As Zhang et al. (2003) stated, let $\boldsymbol{x}$ be a d-dimensional random variable such that $\boldsymbol{x} = [x_1, x_2, x_3, \ldots, x_d]^T$ in finite mixture model having c-components then its probability density function can be expressed as:

$$p(x|\theta) = \sum_{n-1}^{c} \propto_n p(x|\theta_n) \tag{5}$$

conforming that $\propto_n \geq 0$ and $\sum_{n=1}^{c} \propto_n = 1$. Here θ_n stands for parameter of nth mixture model whereas $\theta = \{(\propto_n, \theta_n), n = 1,2,.., c\}$ denotes the parameter set of mixture models. At this point, EM plays the main role of parameter estimation of these finite mixture models.

As a consequence, EM can be counted to be similar to FCM. However, as Tan et al. (2005) states that EM algorithm is more general than FCM and K-means methods as many real world datasets are outcome of random processes and convenient with statistical models of mixture models. But, it may be slower on big datasets because of involving iterative probabilistic calculations at each step.

3 Data

Data subjected to this study is sourced from 75 Turkish four-digit manufacturing industries averaged over 1992–2000 periods provided by Turkish Statistical Institute (TÜİK). The data are based on information on plants. However, because of a methodological change in statistics, we end the data period in 2001 since survey data is classified according to Classification of Economic Activities in the European Community (NACE) Rev. 2. After 2001, the new data were organised by *establishment approach* which neither can be integrated nor be transformed to the previous dataset. Our set of variables consist of concentration rate (CR4), setup cost (σ), cost disadvantage ratio (CDR), advertising ratio (ADS) and R&D ratio (RDS). Detailed information about variables can be found in Table 1.

Table 1 Properties of attributes

Variable	*Type*	*Description*
CR4	Continuous	CR4 proxies market structure (Sutton, 1991, 2007; Robinson and Chiang, 1996; Giorgetti, 2003; Resende, 2009; Arvas, 2011) and it is a measure for the market share of four biggest firms sales in an industry. Therefore in a competitive market, CR4 is expected to be lower than that in a monopolistic market structure. The concentration measure most typically used is a logistic transformation of the concentration ratio log(CR4/(1–CR4)).
Setup cost (σ)	Continuous	σ is the minimum efficient scale (MES), the average value of the capital stock of those firms producing more than the median firm relative to industry sales (S), multiplied by the capital stock of the industry, *K*. That is, $\sigma = \left(\frac{mes}{s}\right) \times K$.
CDR	Continuous	Cost disadvantage ratio takes into account the costs that are caused by entry at the sub-optimal scale. It is measured as the ratio of average production quantity (total production/firm number) to MES (Arvas, 2011; Ilmakunnas, 2008; Yang and Kuo, 2007). A higher value implies that it is a relative disadvantage for smaller firms, such that the impact of CDR on the concentration ratio should be positive.
ADS, RDS	Continuous	ADS and RDS are measured as ratio of advertising and R&D spending to industry total sales. These variables are used to distinguish exogenous from endogenous sunk cost markets.

4 Methodology and applying clustering

The main objective of this study is to reveal the relationships of the five variables in sunk cost industries point of view and to cluster them in order to find similar/dissimilar industries for investigating and presenting the discriminative attributes. To achieve this goal, FCM algorithm is tried to employ first. However, FCM algorithm cannot detect optimal numbers of clusters and requires the cluster count to be an input. As a result of this fact, another powerful clustering technique named expectation-maximisation (EM) is used to learn the optimal cluster count.

Therefore, study is developed on two platforms. While FCM modelling is constituted on software R (The R Project for Statistical Computing, 2011), EM is applied on Microsoft SQL Server 2008 Analysis Services (White Paper: SQL Server 2008 Analysis Services Overview, 2011). Meanwhile, R is an open-source and free software package for statistics and has many plug-in modules for various types of statistical, mathematical and data mining related tasks. On the other hand, Microsoft SQL Server 2008 Analysis Services (MAS) is a commercial business intelligence software pack which provides various types of predictive and descriptive data mining algorithms.

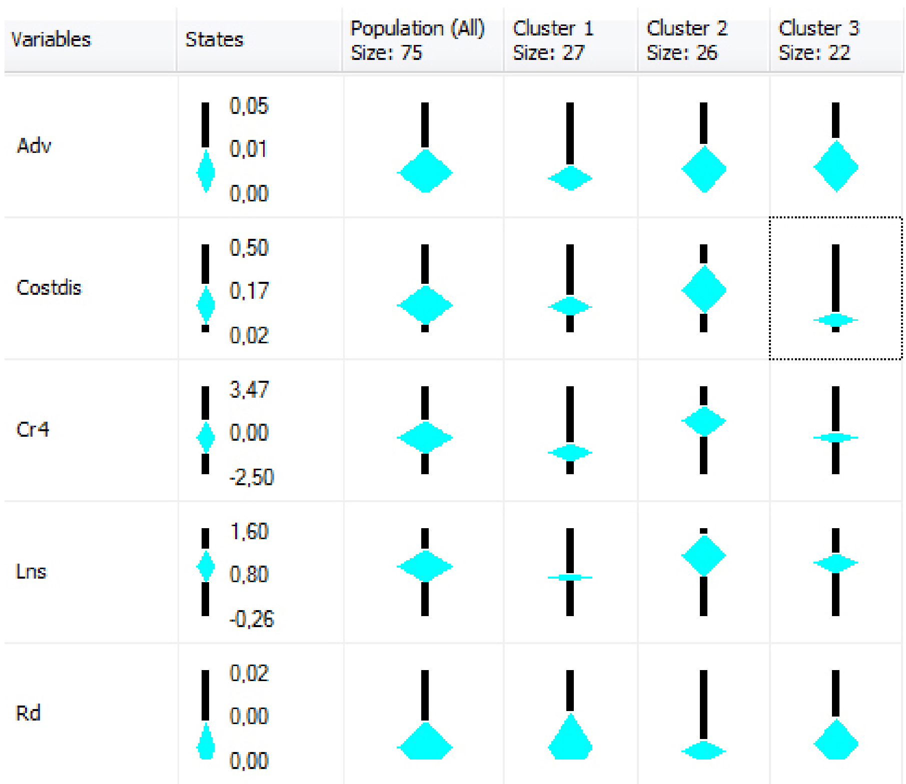
At the first stage, the dataset which contains 75 rows and five variables is converted from Excel to SQL Server database table by data import/export services of SQL Server.

Then, to learn the optimal cluster count, EM algorithm is run on MAS by setting the CLUSTER_COUNT parameter as 0 to gain natural clusters. The other parameters of EM clustering on MAS are left as default values. In EM approach, there is no obligation to define the numbers of clusters. So, with this option optimal cluster count is detected as three.

As can be seen on Figure 1, 75 industries are partitioned in three clusters with 27, 26 and 22 members respectively. Because of having all attributes as continuous variables, mean and standard deviations can be seen on Figure 1. Coloured diamonds indicate the mean values and the vertical height of diamonds depict the value of standard deviations.

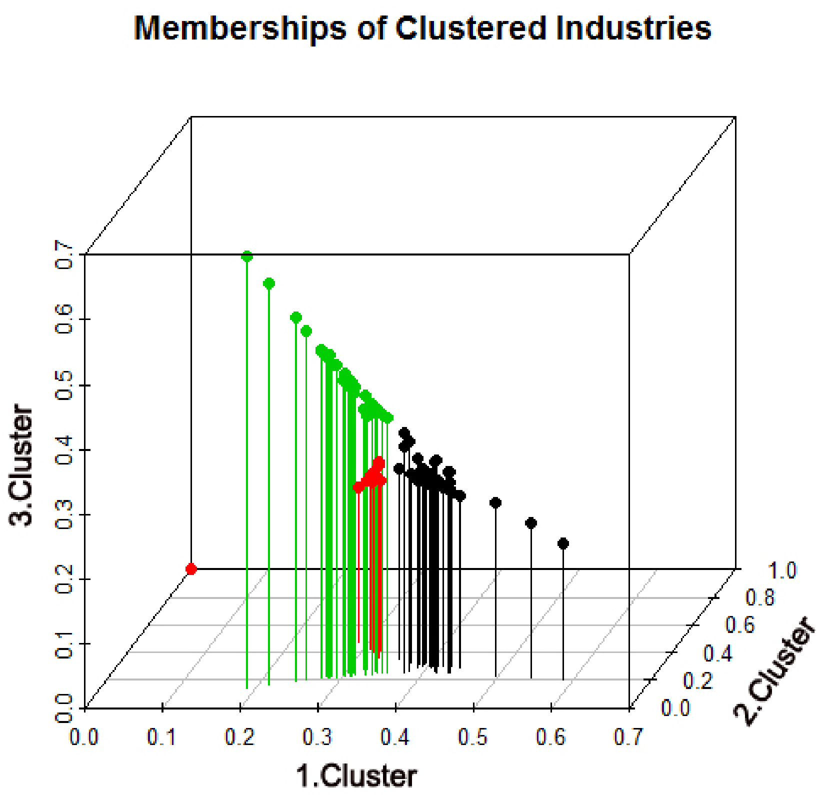
At the second stage, after getting optimal cluster count from EM algorithm, FCM is applied in R. For this aim, SQL Server data table is converted to comma separated value (CSV) format. As being rapidly growing software, R has many data mining related packages. For FCM, '*e1071*' package is employed. Cluster count is set as three. Fuzziness factor is set as 3, 4 and 5 respectively. Besides, maximum iteration count is set to 250. Remaining parameters are kept as default. With this configuration, FCM algorithm is executed over ten times. As expected, due to the random selection of cluster centroids at each startup, results varied slightly but cluster profiles generally remained consistent with Sutton's prediction. To gain insight, one clustering schema is chosen randomly in several iterations.

Figure 1 Discovered clusters of the EM clustering with distributions (see online version for colours)



In Figure 2, membership degrees are depicted in 3D diagram by utilising visualisation package (*scatterplot3d* library)

Figure 2 Membership values of objects clustered by FCM (see online version for colours)



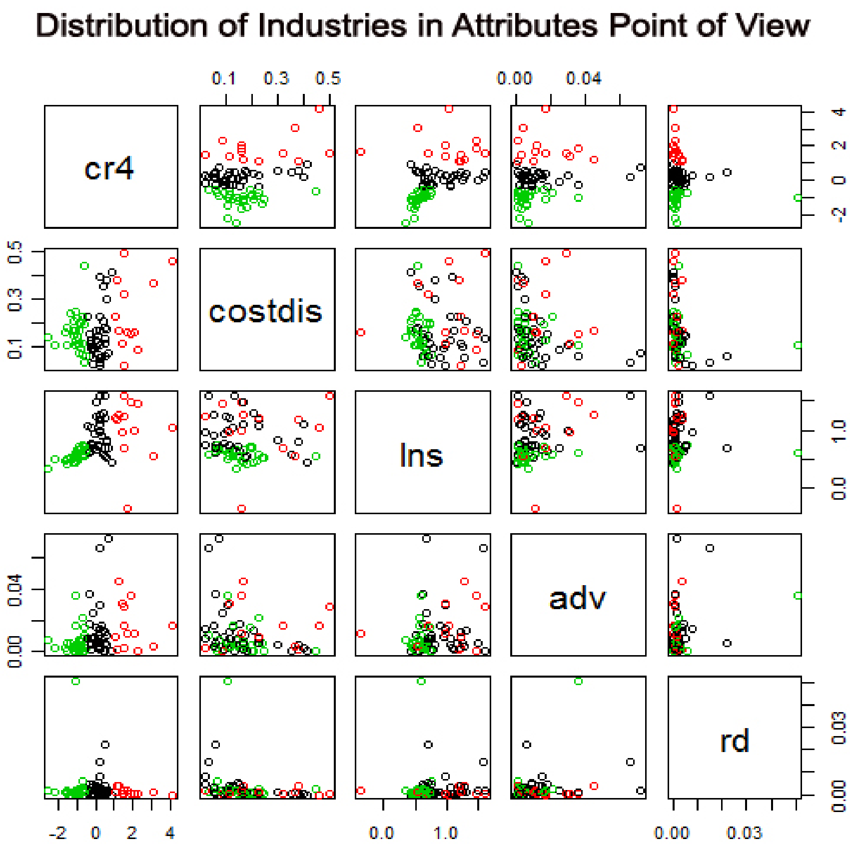
Three clusters are obtained by FCM with average of 0.445 at membership degrees. Discovered clusters have 33, 13 and 29 members respectively. Cluster centroids (mean values) are listed in Table 2.

Table 2 Centroids of cluster constituted by FCM

Variable	*Cluster 1*	*Cluster 2*	*Cluster 3*
CR4	0.1569351	1.4374431	–1.0125536
CDR (costdis)	0.1287300	0.1174372	0.1409288
Setup cost (lns)	0.7536169	0.9755752	0.6098400
Adv	0.0086474	0.0307446	0.0048858
R&D	0.0014739	0.0006050	0.0012260

If the mean values on Table 2 are considered, it can be seen that, Sutton's prediction is again consistent with FCM clustering schema like in EM. Cluster 1 and Cluster 2 industry characteristics reflect Type 1 and Type 2 markets as in EM. But in this case, only the numbers of members differ.

Figure 3 Attribute distribution by FCM (see online version for colours)



In Figure 3, *x-y* distribution graph of each attribute pair is plotted. In this grid style attribute paired graphic, clustered objects (industries in here) are being seen with three different colours. As can be seen on the distribution of clusters at x-y graphs, especially CR4 and costdis (CDR) attributes are more characteristic. In other words, clustered items are grouped more distinctive by CR4 and CDR features. The remaining features (setup cost, ADV and R&D) do not outcome a clear-cut clustered objects. According to the findings, it can be argued that CR4 and CDR attributes are major features which discriminate the clusters. This finding is tested in many try outs of FCM with different fuzziness factors and remained same during the study.

5 Discussion

Sutton's studies on sunk cost and market structure, both theoretically and empirically, has made a neat and robust contribution to profiling what factors lead to the formation of an industry's size distribution and structure. The robust predictions based on Sutton's 'lower bound to concentration' approach rely on separating exogenous and endogenous sunk costs that are determined by firms' strategies.

In this study, the analysis of exogenous and endogenous cost characters of Turkish manufacturing industries is taken into consideration. What distinguishes our study from others is the analytical method is employed here. The insufficient number of fuzzy and clustering based works on sunk costs and market structure in economics has encouraged us to make a contribution to the literature. According to our best knowledge, at least, we have not found yet any empirical research made in this field on Turkey. This is another motivation for us to focus on this issue within an interdisciplinary study.

Based on the theory, in an exogenous sunk cost industry, potential firms first incur the setup cost which is exogenously given by industry conditions, and then determine the quantity of their products to be supplied and, secondly, firms observe rivals' prices. The most discouraging entry barrier for a new firm deciding to enter to market is the setup cost of minimum efficient scale (MES) below which firm has to exit from the industry or market. Therefore, in case of the presence of low barriers, as a market becomes larger and profitable, then we should expect that the number of potential firms would go to infinity with decreasing concentration rate, thus leading to a competitive structure.

But in an endogenous sunk cost industry, toughness of price competition leads to incumbent firms to involve in a non-price competition over which firms have to invest more in rising product quality. Since the quality is subject to product differentiation (both vertically and horizontally), in order to raise its level, firms need to incur additional sunk advertising and R&D expenditures independent of industry's conditions. So, ever-growing level of these sunk entry barriers both prevents potential entries and leads to some suboptimal incumbent firms to exit from industry. Thus, as firms exit from the market, then the number of rival firms shrinks and concentration rate rises. In such an industry configuration, we should expect that the level of advertising and R&D expenditures to rise.

According to EM clustering results, industries in cluster 1 have a Type 1 and a Type 2 characteristic in cluster 2 and 3. In cluster 1, as expected previously, the levels of advertising (Adv), cost disadvantage rate (Costdis) and concentration rate (CR4) are considerably lower than that in cluster 2 and 3. But the result with R&D is somewhat not clear. Theory asserts that in tough competition regime in Type 1 market, level of R&D does not necessarily increase because its level is endogenously given at the first stage and not a key strategic factor for competition. Some of these industries are slaughtering, preparing and preserving meat industry (3111 ISIC Rev 2 code); canning, preserving and processing of fruits and vegetables industry (3113); tobacco manufactures (3140); spinning, weaving and finishing textiles industry (3211); manufacture of drugs and medicines (3522); and manufacture of fabricated metal products, except machinery and equipment industries (381).

Controversially, in cluster 2, the level of R&D expenditures is lower than that what theory asserts. This implies that sunk costs created by research and development are not so important for Type 2 market. But the other endogenous key factor advertising seems to be strength on concentration. These industries are advertising intensive industries (Type 2A) within Type 2 category among which; canning, preserving and processing of fish, crustacean and similar foods industry (3114); sugar factories and refineries industry (3118); wine and malt industries (3132 and 3133); manufacture of paper and paper products group industries (341); manufacture of industrial chemicals, fertilisers, pesticides, synthetic resins, plastic materials and man-made fibres except glass industries (3511, 3512 and 3513); manufacture of engines and turbines (3821); manufacture of agricultural machinery and equipment (3822); manufacture of office, computing and accounting machinery (3825); manufacture of motorcycles and bicycles (3844); and manufacture of photographic and optical goods, watches and clocks (3851 and 3852) stand out.

In cluster 3, the level of two endogenous sunk factors fit well Sutton's predictions; both of whose levels are high and important in an expanding market. But in that cluster, Costdis variable seems to be lax implying that market conditions is sustainable for firms who operate with a suboptimal scale. In addition to this finding, market structure seems to be more competitive than cluster 2 is. Industries in this cluster are called advertising and R&D intensive industries (Type 2AR). In that classification, manufacture of dairy products (3112); manufacture of bakery products (3117); manufacture of furniture and fixtures (3320); manufacture of pottery, china, earthenware and glass products (3610 and 3620); basic metal industries (37); manufacture of electrical industry machinery, radio, television, communication equipment, electrical appliances, house wares and their apparatus (3831, 3832 and 3833); and ship building and repairing equipment industry (3842) are mentioned.

The similar results are obtained via FCM clustering method. The membership degrees in Table 2 point out that high/low values imply high/low level of costs or Cr4. Based on this clustering, cluster 1 and cluster 2 profile Type 1 and Type 2 characteristics. Degrees in cluster 3 (column 3) are not so clear for industry-type matching.

Our results may not be sufficient for concluding in detailed that the incumbent firms are cooperating or building entry barriers to exclude new entrants in the economy. Further studies should be made for Type 1 and Type 2 industries separately in the future.

6 Conclusions

However, increasing product quality and wide spreading product variety through innovation, product differentiation, research and development and advertisement are costless expenditures; they are basic requirements for manufacturing firms' competitiveness, modernisation and technology. In fact, what Sutton inspired from to develop his theory was this phenomenon.

With a weak national system of innovation and a negligible small share of R&D and advertising expenditures in GNP, Turkey is an interesting case of analysis since she has a relatively dynamic and productive manufacturing industry, the significance of which in long-term growth and international competitiveness is indisputable.

In this study, sunk cost industries are profiled by two well-known soft clustering algorithms: FCM and EM. As the profiling results present, Sutton's prediction is consistent with clustering results in both algorithms for Turkey case. Further, 'CR4' and 'costdis' attributes are revealed as the most important discriminators among the variables. According to findings, it is also seen that advertising expenditures are more effective than R&D in determining industrial clustering during the period taken into consideration.

Due to the requirement of several iterations and inconsistencies caused by randomness in generation of centroids in FCM, performed more consistent and well defined cluster discoveries in the data compared to FCM. Another advantage of EM clustering schema is that it does not require initial cluster number and hence produces natural clusters. Therefore, natural clustering techniques are suggested to researchers especially for providing both consistency and easy to use environment. However, if there is an insight or background knowledge about the profiles hidden in data, the algorithms which require cluster count parameter (e.g., FCM, *K*-means) can be employed by changing this parameter but this approach requires several iterations and deep analysis of revealed clusters.

Separation of exogenous and endogenous sunk cost industries of Turkey and discovering the individual natural cluster properties of them is being planned as a future work of this study.

References

A Tutorial on Clustering Algorithms, Fuzzy C-means Clustering (2011) available at http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html (accessed on 24 September 2011).

Arvas, M.A. (2011) 'Effects of price competition on non-price competition and market structure in endogenous sunk cost industries: Turkish manufacturing industry', PhD thesis, Hacettepe University, Ankara.

Bain, J.S. (1956) *Barriers to New Competition*, Harvard University Press, Cambridge.

Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York.

Davies, S. (1989) *Economics of Industrial Organization*, Longman, Burnt Mill, Harlow.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society*, Vol. 39, No. 1, pp.1–38.

Dunn, J.C. (1973) 'A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters', *Journal of Cybernetics*, Vol. 3, No. 3, pp.32–57.

Giorgetti, M.L. (2003) 'Lower bound estimation-quantile regression and simplex method: an application to Italian manufacturing sectors', *The Journal of Industrial Economics*, Vol. 51, No. 1, pp.113–120.

Göktepe, A.B., Altun, S. and Sezer, A. (2005) 'Soil clustering by fuzzy c-means algorithm', *Advances in Engineering Software*, Vol. 36, No. 10, pp.691–698.

Höppner, F. (2002) 'Speeding up fuzzy c-means: using a hierarchical data organization to control the precision of membership calculation', *Fuzzy Sets and Systems*, Vol. 128, No. 3, pp.365–376.

Ilmakunnas, P. (2008) 'Lower bounds of concentration in a small open economy', *Review of Industrial Organization*, Vol. 32, No. 1, pp.19–33.

Liu, J. and Xu, M. (2008) 'Kernelized fuzzy attribute C-means clustering algorithm', *Fuzzy Sets and Systems*, Vol. 159, No. 18, pp.2428–2445.

Noordam, J.C., van der Broek, W.H.A.M. and Buydens, L.M.C. (2002) 'Multivariate image segmentation with cluster size insensitive fuzzy c-means', *Chemometrics and Intelligent Laboratory Systems*, Vol. 64, No. 1, pp.65–78.

Park, D. (2009) 'Classification of audio signals using fuzzy c-means with divergence-based kernel', *Pattern Recognition Letters*, Vol. 30, No. 9, pp.794–798.

Peneder, M. (1995) 'Cluster techniques as a method to analyse industrial competitiveness', WIFO Working Papers 80, WIFO.

Peneder, M. (1999) 'Intangible investment and human resources. The new WIFO taxonomy of manufacturing industries', WIFO Working Papers 114 WIFO.

Resende, M. (2009) 'Concentration and market size: lower bound estimates for the Brazilian industry', *Economics Bulletin*, Vol. 29, No. 3, pp.2468–2474.

Robinson, W.T. and Chiang, J. (1996) 'Are Sutton's predictions robust? Emprical insights into advertising, R&D, and concentration', *The Journal of Industrial Economics*, Vol. 44, No. 4, pp.389–408.

Sutton, J. (1989) 'Endogenous sunk costs and the structure of advertising intensive industries', *European Economic Review*, Vol. 33, Nos. 2–3, pp.335–344.

Sutton, J. (1991) *Sunk Costs and Market Structure*, The MIT Press, Cambridge.

Sutton, J. (2007) 'Market structure: theory and evidence', *Handbook of Industrial Organization*, Elsevier, Vol. 3.

Tan, P.N., Steinbach, M. and Kumar, V. (2005) *Introduction to Data Mining*, Addison-Wesley.

The R Project for Statistical Computing available at http://www.r-project.org/ (accessed on 21 October 2011).

Tütmez, B. and Tercan, A.E. (2006) 'Use of fuzzy modeling approach in grade estimation', Madencilik, Vol. 45, No. 2, pp.39–47.

White Paper: SQL Server 2008 Analysis Services Overview, available at http://www.microsoft.com/sqlserver/2008/en/us/wp-sql-2008-analysis-services.aspx (accessed on 21 October 2011).

Yang, C.H. and Kuo, C.C. (2007) 'Lower bounds of concentration in Taiwan's manufacturing industries: do exports matter?', *Applied Economics*, Vol. 39, No. 16, pp.2389–2401.

Zadeh, L.A. (1965) 'Fuzzy sets', *Information and Control*, Vol. 8, No. 3, pp.338–353.

Zhang, B., Zhang, C. and Yi, X. (2003) 'Competitive EM algorithm for finite mixture models', *Pattern Recognition*, Vol. 37, No. 1, pp.131–144.