The Regional Economics Applications Laboratory (REAL) is a cooperative venture between the University of Illinois and the Federal Reserve Bank of Chicago focusing on the development and use of analytical models for urban and regional economic development. The purpose of the **Discussion Papers** is to circulate intermediate and final results of this research among readers within and outside REAL. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago, Federal Reserve Board of Governors or the University of Illinois. All requests and comments should be directed to Geoffrey J. D. Hewings, Director, Regional Economics Applications Laboratory, 607 South Matthews, Urbana, IL, 61801-3671, phone (217) 333-4740, FAX (217) 244-9339.
Web page: www.uiuc.edu/unit/real

TOWARD A QUANTITATIVE ANALYSIS OF INDUSTRIAL CLUSTERS I: FUZZY CLUSTERS VS. CRISP CLUSTERS

by

Chokri Dridi and Geoffrey J.D. Hewings

REAL 02-T-1 January, 2002

TOWARD A QUANTITATIVE ANALYSIS OF INDUSTRIAL CLUSTERS I: FUZZY CLUSTER VS. CRISP CLUSTERS

Chokri Dridi
Department of Agricultural and Consumer Economics, and Regional Economics Applications Laboratory (REAL), University of Illinois at Urbana-Champaign
cdridi@uiuc.edu

Geoffrey J.D. Hewings
Regional Economics Applications Laboratory (REAL), University of Illinois at Urbana-Champaign
hewings@uiuc.edu

ABSTRACT: This paper constitutes the first part of a two-paper series that makes use of fuzzy logic in an attempt to quantify input-output based cluster analysis. The objective of this paper is to examine the cluster structure of the sales and purchases profiles in input-output systems when the principle of 'excluded middle' is violated by the use of fuzzy set theory. The approach relies on results from the data analysis technique known as dual scaling (Dridi and Hewings, 2002a and 2002b). A crisp clustering algorithm known as Ward's algorithm (Ward, 1963) is used for comparison in an application with US input-output table.

JEL Classification: C45, C67, R15
Keywords: Classification, Fuzzy clusters, Fuzzy sets, Industrial clusters, Ward algorithm

Correspondence should be addressed to:

Chokri Dridi
Regional Economics Applications Laboratory
University of Illinois
607 S. Mathews #220
Urbana, IL 61801-3671 USA

Tel: (217) 244-7226 Fax: (217) 244-9339

TOWARD A QUANTITATIVE ANALYSIS OF INDUSTRIAL CLUSTERS I: FUZZY CLUSTER VS. CRISP CLUSTERS

1. INTRODUCTION

Economists and policy makers alike are often faced with data that require treatment to become useful for further uses such as analysis, pattern detection, interpretation, or comparison. For Bergman and Feser (1999), "industry cluster analysis can help exploit the growing wealth of regional economic data, provide a means of thinking effectively about industrial interdependence, and generate unique pictures of a regional economy that reveal more effective policy options." Popular methods to conduct cluster analysis include graph theory analysis (Campbell 1974), block decomposition analysis drawing on the theory of Markov chains (Hewings et al., 1997), value-chain-analysis used by Porter (1998a and 1998b), and input-output analysis (O'Huallachain 1984).

A particular set of data of interest to us is the input-output table of a country or a region; therefore, the methodology we are promoting is not concerned with the spatial dimension of the clusters, although we believe that the use of regional input-output tables should allow a thorough study of regional clusters. The nomenclature of input-output tables is based on an association structure that groups together sectors with similar activities that ultimately fall in one and only one of the following large categories; agriculture, industry, or services. However, this classification does not provide enough information on the functional and structural similarities between industrial profiles. The use of cluster analysis facilitates the identification of industries with similar sales or purchases profiles. The importance of clusters identification is not only of interest to policy makers, but as stated by Porter (1990), their existence influences the creation and development of firms through investment in infrastructure and R&D partnerships.

Nevertheless, traditional cluster detection and analysis are for the most part based on crisp sets, where each industry belongs to one and only one cluster; this restriction obviously limits the use of industrial clusters and yields an unrealistic description of the economy since industries in fact might have similarities not only with some but with all industries to varying degrees.

Cluster identification producing crisp clusters are primarily useful for aggregation purposes and do not offer ways for a quantitative analysis of clusters. This paper is the first part of a two-paper series that departs from crisp clustering approaches to fuzzy clustering methods in an attempt to move toward a quantitative analysis of clusters. While this paper aims to further the investigation into cluster analysis, we make a departure from traditional cluster analysis by considering 'blurred overlapping' clusters through the adoption of *fuzzy* clusters. Fuzzy clusters, unlike crisp clusters where each element belong to one and only one cluster, allow for vagueness in the belongingness to a particular clusters. In addition to being more realistic, fuzzy cluster take into account the imprecision that often exists in the data, as well as the very real possibility that a sector may have close associations with more than one cluster. In a sense, fuzzy cluster analysis may resonate more formally with prior work that has come to be known as industrial complex analysis in which sets of industry associations were identified with the possibility of a sector belonging to more than one complex (see Czamanski and Ablas, 1979; Isard et al., 1959). In this regard, the fuzzy cluster approach provides the capability of exploring a more complex summary of structural interdependence in an economy, placing it between the mutually exclusive cluster systems on the one hand and the presentation of the complete set of linkages evident in an examination of the full input-output table. It does share more of the character of the Czamanski and Ablas (1979) approach; a sector involved with two or more clusters plays a different role in the economy than one that is entirely nested within just one cluster. The crisp versus fuzzy problem has a strong parallel in the primary versus secondary product issue in assigning firms to sectors (a problem that was addressed in the development of the make-use distinction in input-output accounting systems). The second paper explores applications of the Shapley value (Shapley, 1953) and other fuzzy measures and introduces the notion of a *lead* sector in a cluster.

The method employed here is based on results obtained from *dual scaling*, a term coined by Nishisato (1980). Dual scaling is considered a technique of multivariate descriptive analysis; it is not inferential and is similar to principal component analysis (PCA) for categorical data approach, except that in a single run it provides primal and dual solutions. Nishisato (1980 and 1994) offers a description of the dual scaling approach and Dridi and Hewings (2002b) showed how it can be used in the context of input-output models to decompose the internal structure of associations into open and closed loops of associations, leading to the identification of a finite number of stages of economic complexity. The results obtained from dual scaling are row weights and their dual, column weights. Those weights, among other things, can be used for cluster analysis. In O'Huallachain (1984), a reassessment of PCA was made where the method was applied to identify row clusters and column clusters, but the methodology was not used to assess the relationship between row-sectors and column-sectors as in Dridi and Hewings (2002a) and, furthermore, the clusters produced are crisp.

Without the need to revisit the dual scaling approach, in the next section a formal description of a popular hierarchical decomposition technique suggested by Ward (1963) is examined along with a fuzzy clustering approach. In section 3, for illustration, results from dual scaling are used for the 1990 US input-output tables to first identify crisp clusters using the

Ward's method and fuzzy clusters methodology and, secondly, to compare both methods and results. We conclude with a summary of results and further research possibilities.

2. Crisp and Fuzzy Clustering Methods

2.1. Ward's Algorithm for Crisp Clusters

Cluster identification is a relatively easy task if the data are represented by two variables, since a two-dimensional plot allows visual identification of all relevant clusters. However, this task becomes less straightforward for multi-dimensional representation of industries as is the case with input-output data decomposed into 10 dimensions using dual scaling (Dridi and Hewings, 2002a and 2002b). This decomposition obviously makes any visualization of the crisp clusters impossible and requires elaborate clustering techniques. A wide class of data partitioning has been based on hierarchical methods of agglomerative nesting; Ward (1963) introduced a very popular clustering method that we described in the present section.

At the beginning of the algorithm (step 0), each object (here, an industry) is considered as a separate cluster. The following iterative steps successively merge clusters with the smallest dissimilarity while leaving the others unmerged (step 1). Then, the dissimilarity between the new cluster and the rest of the clusters is computed (step 2) and step 1 is repeated until all industries are part of a single cluster (Kaufman and Rousseeuw, 1990).

The between cluster dissimilarity measure used by Ward is as follows:

$$d^2(R,Q) = \frac{2.\text{card}(\text{R}).\text{card}(\text{Q})}{\text{card}(R)+\text{card}(Q)}\left(\overline{x}(R)-\overline{x}(Q)\right)^2 \tag{1}$$

In the above expression *card*(.) is the number of elements (i.e. cardinality) in a given cluster, and $\overline{x}(.)$ is the centroid of the cluster. In a cluster with K variables (i.e. dimensions), the centroid of a cluster R is:

$$\overline{x}\left(R\right)=\left(\overline{x}_1(R),...,\overline{x}_k\left(R\right),...,\overline{x}_K(R)\right) \quad (2)$$

such that $\overline{x}_k\left(R\right)=\frac{1}{\text{card}\left(R\right)}\sum_{i\in R}x_{ik}$; where x_{ik} is the value for individual *i* along the dimension *k*, $\forall k\in K$.

In case the cluster *R* is obtained through a merger of clusters *A* and *B*, i.e. $\text{card}\left(R\right)=\text{card}(A)+\text{card}(B)$, its centroid is then:

$$\overline{x}\left(R\right)=\frac{\text{card}\left(A\right)}{\text{card}\left(R\right)}\overline{x}\left(A\right)+\frac{\text{card}\left(B\right)}{\text{card}\left(R\right)}\overline{x}\left(B\right) \quad (3)$$

2.2. Sectors Similarities and Fuzzy Classification [1]

Usually similarity studies and cluster analysis are used for aggregation purposes (Blin and Cohen, 1977), where each industry has to belong to a unique cluster: this should not exclude the possibility of overlap in clusters. In order to investigate clusters overlaps, we use fuzzy set theory where set distinctions are not crisp (see Miyamoto, 1990; Mirkin, 1996)

Over thirty-five years ago Zadeh (1965) introduced fuzzy sets and fuzzy logic, where he defines a fuzzy set as being "a class of objects with a continuum of grades of membership." Indeed, in fuzzy set theory, in contrast to Cantorian set theory, elements belong to a set according to a membership function that takes values in the range $[0,1]$ instead of taking a binary value from $\{0,1\}$, which means that Cantorian sets are particular cases of fuzzy sets with membership of 0 or 1.

<< insert figure 1 here >>

[1] This section draws on Dridi and Hewings (2002a)

In panel (a) of figure 1, assume four industries where an application of Ward's algorithm reveals a similarities between industries *A* and *B* form one hand and between industries *C* and *D* from the other, each pair of industries constitutes a distinct cluster (rectangle and oval). In panel (c) of the same figure an analysis conducted in line with Czamanski and Ablas (1979) might reveal that industries *A* and *B* are similar but also that industries *B* and *C* are similar too, while industry *D* is not similar to any of them. Therefore, industries *A* and *B* form a cluster (rectangle set) that overlaps with the cluster formed by industries *B* and *C* (oval set). In panel (b), using fuzzy sets, the analysis might reveal that while industries *A* and *C* belong fully to respectively the rectangular and oval clusters, industry *B* does not belong fully to either clusters but belongs to both of them with varying degrees while industry *D* does not belong to any of the two clusters.

Unlike the crisp clustering method, fuzzy clustering is concerned with pattern discoveries between sectors rather than their aggregations, hence, it is a discovery tool rather then a summarizing or simplification tool. The use of fuzzy set theory in social sciences in general and economics more particularly, helps avoiding certainties in conclusions based on data that might be inaccurate because of sample bias or collection errors or where interpretation may be less precise. Ragin (2000) suggests that the ability of fuzzy sets to explore and express a greater diversity in the data is useful in bringing closer data analysis and theoretical models in social sciences, since "fuzzy sets can be carefully tailored to fit theoretical concepts".

Formally, let *X* be a reference finite and countable space of points, where a generic point is denoted by *x*. A fuzzy set (or subset) $A \subseteq X$ is characterized by a real valued membership function $\mu_A(x)$ that associates with each point x a value from a real interval usually normalized to $[0,1]$. For notational clarity and to distinguish between fuzzy sets and crisp sets it is common to denote a fuzzy set *A* by $\underset{\sim}{A}$, however for ease of notation we will use, *A*, and specify whether is fuzzy or crisp when such specification is necessary. The fuzzy subset A of X is a set of ordered pairs $\left\{\left(x \mid \mu_A(x)\right); \forall x \in X\right\}$, where $\mu_A(x)$ is the grade or degree of membership of x in A.

$$\mu_A : X \rightarrow \left[0,1\right] \tag{4}$$

If we denote by A_k, $\forall k = 1,...,K$, all the subsets of the universal set X, then the following properties always hold:

$$\begin{cases} \mu_{A_k}(x) \in \left[0,1\right]; \forall x \in X, \forall k = 1,...K \\ \sum_{k=1}^{K} \mu_{A_k}(x) = 1 \end{cases} \tag{5}$$

One reason for the lack of use of fuzzy logic and fuzzy set theory in social sciences may be attributed to the problematic way the shape of the membership function is determined, an issue that has not received sufficient treatment in the literature. Indeed, although assuming an ad-hoc shape for the membership function - as is often done when dealing with probability distribution functions - might be convenient for theoretical studies, it constitutes a serious departure from the data's underlying patterns when used for applied work. An example of ad-hoc defined probability distribution is used in Jackson (1986) to estimate technical coefficients in input-output tables[2]. Various shapes for membership function were proposed: for the most common shapes used see Kaufmann (1975, p. 168-171), and Harris and Stocker (1998, p. 849-851). Kaufman and Rousseeuw (1990, Ch.4) advanced an algorithm, retained in the statistical package S-Plus 2000[3], to compute the membership values.

The iterative algorithm proposed by Kaufman and Rousseeuw (1990) classifies r objects (here sectors) into k clusters based on the observation of s characteristics, variables, or

[2] This is not a criticism of Jackson's (1986) work since the underlying probability distributions for individual technical coefficients was not known and alternatives had to be proposed and evaluated in terms of their impacts on more macro economic analysis with an input-output model.
[3] A statistical package from Insightful Corp.

observations of a variable (here the dual scaling solutions, obtained by applying the method described in the appendix). With $d_{i,i'}$ the distance computed in equation (a.11) for the supply profiles and (a.12) for the demand profiles, for each sector *i* and cluster *k* there is a membership μ_{ik} value that solves the following program:

$$\begin{aligned}
&\min_{\mu_{ik}} \sum_{v=1}^{k} \frac{\sum_{i,i'}^{r} \mu_{iv}^2 \mu_{i'v}^2 d_{i,i'}}{2\sum_{i'=1}^{r} \mu_{i'v}^2} \\
&s.t. \\
&\mu_{iv} \geq 0 \qquad ;\forall i = 1,\ldots,r; \forall v = 1,\ldots,k \\
&\sum_{v=1}^{k} \mu_{iv} = 1 \qquad ;\forall i = 1,\ldots,r
\end{aligned} \tag{6}$$

The software package considered allows a maximum number of clusters $k = \frac{n}{2} - 1$, where in our case *n* is the number of industries (i.e. the size of the technical block). Aside from numerical considerations specific to the algorithm, the maximum number of clusters is set so as no sector represents a unique cluster and that the number of sectors is not equal across clusters. However, the optimal number of clusters has to minimize the objective value in (6). To assess the fuzziness of the resulting cluster, Dunn's partition coefficient is computed:

$$F_k = \sum_{i=1}^{r} \sum_{v=1}^{k} \frac{\mu_{iv}^2}{r} \tag{7}$$

$F_k \in \left[\frac{1}{k}, 1\right]$, for entirely fuzzy clustering $\mu_{iv} = \frac{1}{k}$ and $F_k = \frac{1}{k}$, and for entirely crisp sets $\mu_{iv} = 0$ or $\mu_{iv} = 1$ and $F_v = 1$. A normalized Dunn's coefficient taking values from $[0,1]$ is computed by:

$$F_k' = \frac{F_k - \frac{1}{k}}{1 - \frac{1}{k}} = \frac{kF_k - 1}{k - 1} \qquad (8)$$

Zadeh (1965) introduced elementary operations on fuzzy sets, such as equality, complementation, inclusion, intersection, and union. Consider the following two fuzzy subsets A_i and A_j of the universal set X, whose membership functions are $\mu_{A_i}(x)$ and $\mu_{A_j}(x)$.

equality: The subsets A_i and A_j are said to be equal if and only if their membership functions $\mu_{A_i}(x)$ and $\mu_{A_j}(x)$ are equal for all $x \in X$.

$$\mu_{A_i}(x) = \mu_{A_j}(x) \qquad ;\forall x \in X \qquad (9)$$

complementation: The complement of a fuzzy set A_i, denoted by A_i^c, has a membership function:

$$\mu_{A_i^c}(x) = 1 - \mu_{A_i}(x) \qquad ;\forall x \in X \qquad (10)$$

However, since fuzzy logic violates the excluded middle principle, the intersection of a fuzzy set and its complement does not produce necessarily the empty set.

inclusion or *containment*: The subset A_i is included or contained in A_j, and denoted by $A_i \subseteq A_j$ if and only if:

$$\mu_{A_i}(x) \le \mu_{A_j}(x) \qquad ;\forall x \in X \qquad (11)$$

intersection: The intersection of two fuzzy sets A_i and A_j is a fuzzy set $B = A_i \cap A_j$ whose membership function is defined as:

$$\mu_B(x) = \min\left\{\mu_{A_i}(x), \mu_{A_j}(x)\right\} \qquad ;\forall x \in X \qquad (12)$$

expression (12) can also be abbreviated by:

$$\mu_B = \mu_{A_i} \wedge \mu_{A_j} \tag{13}$$

union: The union of two fuzzy sets A_i and A_j is a fuzzy set $C = A_i \cup A_j$ whose membership function is defined as:

$$\mu_C(x) = \max\left\{\mu_{A_i}(x), \mu_{A_j}(x)\right\} \qquad ;\forall x \in X \tag{14}$$

expression (14) can also be abbreviated by:

$$\mu_C = \mu_{A_i} \vee \mu_{A_j} \tag{15}$$

α-cuts and α-level sets: For some analysis it might be necessary to exclude elements of a fuzzy set based on their membership value, the notion of *α-cuts* (Bojadziev and Bojadziev, 1997) can be used. An α-level fuzzy set *A*, defined as A_α, is:

$$A_\alpha = \left\{x \mid \mu_A(x) \geq \alpha; \forall \alpha \in [0,1], \forall x \in Y, Y \subseteq X\right\} \tag{16}$$

As mentioned earlier, the use of fuzzy logic allows for more flexibility in the treatment of industry clusters, the above fuzzy set operations will be used in the next section to discuss and extract information about clusters' structures of the sales and purchase profiles in input-output systems.

3. CLUSTERS IN THE 1990 US INPUT-OUTPUT TABLE

For illustration, both clustering methods described above are applied to a 33-by-33 input-output table[4] of the US for 1990 in current prices. The details of the aggregation are provided in table A.1 of the appendix.

[4] Source: OECD Input-Output database

The application of Ward's algorithm to the US input-output data results in the dendogram in figure 2 for the sales profiles and in figure 3 for the purchases profiles. The hierarchical agglomeration reveals substantial differences between the two profiles.

<< insert figure 2 here >>

<< insert figure 3 here >>

If we limit the number of clusters to fifteen, the clusters' memberships of each industry are very different for the sales and purchases profiles in addition to being mostly concentrated in one or two clusters (tables 1 and A.2). The number of clusters is chosen to allow for comparisons with the application of fuzzy clusters that generates under the optimality criterion, fifteen clusters for the case at hand. In table 1, for the sales profile clusters number 1 and 5, cover about half the industries, while for the purchases profile clusters number 1, 2, 3, and 4 cover 20 out of 33 industries. The remaining industries do not belong to any of the clusters mentioned above, for the most each industry represents a separate cluster, indeed for both sales and purchases profiles clusters number 7 through 15 are represented by only one industry.

<< insert table 1 here >>

In figure 4, we provide plots that Kaufman and Rousseeuw (1990) call banner plot because of their resemblance to a waving flag. The banner plots contain the same information as the dendograms in figures 2 and 3, but are constructed to show the successive mergers from left to right. Each industry is represented by a horizontal line and the shorter that line the lower is the industry in the agglomerative hierarchy of figures 2 and 3. In figure 4, the fraction of the dark area to the total area of the graph, is called the *agglomerative coefficient* ($AC \in [0,1]$) and it is equal to 0.76 for the sales profiles and 0.48 for the purchases profiles. While the *AC* for the sales can be considered high, the purchases profile *AC* value is relatively low compared to a perfect clustering value of 1. The *AC* is an indicator of the strength of the obtained clustering structure. In the case of the US input-output data, the *AC* indices are very different and show a stronger clustering for the sales profile then for the purchases profile.

<< insert figure 4 here >>

A major shortcoming of crisp clusters seem to be the fact that many clusters are represented by one or two industries only while other group a large number of industries, there is an obvious lack of diversity in clusters' structure and a lack of density in many clusters. This criticism of crisp clusters might be attributed to the level of aggregation in the data, however fuzzy clustering techniques applied to the same data, as will be shown below, give clusters that are diverse and dense.

The application of the dual scaling technique in the appendix to the US 33-by-33 interindustry input-output flows, generates a maximum number of solutions of 32, which means that each industry can be plotted in a space of dimension 32, where each solution occupies a dimension. The 32 solutions for the weights for the sale profiles (rows) and the purchase profiles (columns).

Applying the fuzzy clustering algorithm in Kaufman and Rousseeuw (1990) to the sales and purchases profiles of the US input-output table provides the memberships to 15 fuzzy sets; the maximum number of sets was fixed by S-Plus. However, lowering the number of sets to 14 for example proved to yield non-optimal values in terms of the objective function in (6); the membership values are provided in tables A.3 and A.4 of the appendix. The overall fuzziness of the clustering can be assessed by Dunn's index in table 2.

<< insert table 2 here >>

For the US, Dunn's index suggests a similar fuzziness in both sales and purchase profiles. While the significance of the difference between the sales profiles Dunn's index and the purchases profiles Dunn's index cannot be statistically ascertained, their values suggest that an industry in the purchases or sales profile is equally likely to be part of more than one cluster that it is the case for a sale profile.

At the cost of loss of information, if we force each sector to belong to the set in which it holds the highest membership and look for the closest neighbor to that set; we can obtain a "forced" classification as shown in table 3.

<< insert table 3 here >>

Even though the forced classification is not precise since it is based only on the highest fuzzy membership that assigns each industry to a single cluster, table 3 shows that the cluster pattern obtained from the "forced" classification is very different from the results obtained using the Ward algorithm. Indeed, industries 1, 3, 5, 7, 11, 15, 17, and 27 in the sales profile are similar since they have the same closest crisp set and the closest neighbor to that set, this is not confirmed by the results from Ward's algorithm in table A.2, were only industries 1, 2, and 3 are similar as they belong to the same cluster. Tables 4 and 5 show that even by disregarding details regarding the membership values provided, the combination of just two clusters offers more diversity in describing the clusters in the economy. In table 4 for the sales profile, 6 industries have as second closest neighbor the row cluster number 8, however those industries cannot be considered similar since their first closest neighbors are different.

In tables A.3 and A.4, it is easy to check for the complementation property, for add-up $\sum_{i=1}^{4} \mu_{A_i}(x) = 1; \forall x \in X$, and for the absence of strict inclusion in all sets. Unlike the equality between fuzzy sets, the existence of inclusion raises issues of redundancy of the inner set or subset, that efficient algorithms are expected to avoid producing when dealing with fuzzy cluster analysis. If a set $A_j \subset A_i$ then only the cluster A_i should be considered and A_j is redundant and therefore should not emerge as a cluster and that is what the algorithm is doing. The intersection operation presents attractive features, such as the detection or a crude indicator of particular sectors in sales and purchases profiles. Because fuzzy memberships are defined on a continuum, the intersection and union operations on fuzzy sets allow assessing the industries' *global* importance in the whole economy and their *local* importance in particular clusters.

In figure 5, we provide a graphic representation of the intersection and union membership functions for the sales and purchases profiles for the US (from table A.4). The *intersection* operator indicates the degree of importance of a given industry to the overlap between clusters while the *union* operator indicates the importance of a given industry to all clusters. Figure 5.a reveals that for the sales profiles, industry 19, 'Ship building & Repairing', compared to the rest of the industries is more important to clusters overlap than it is important to all the industries in tandem (fig. 5.b), therefore we can say that industry 19 is *locally* important and *globally* unimportant. The rest of the industries seem to have close membership values in their belongingness to either of the intersection or the union of the fuzzy sets. It should be expected that industries having particularly low membership values for the fuzzy intersection of sets would have particularly high membership values for the fuzzy union of sets, this is because of the difference between the $\min\{.\}$ and the $\max\{.\}$ functions when the membership values for a given industry are not constant or close across clusters.

<< insert figure 5 here >>

For this particular dataset, it seems that there is alternation in the local and global importance of industries. It can be argued that odd numbered industries, which for the most are extraction or production of "raw" or "semi-finished" products, are locally important. Indeed a close look at the industry description (table A.1 in the appendix) shows that almost all locally important industries are followed by industries that produce a more "refined" or "finished" specific product, which are important to all clusters simultaneously. The pattern of alternation is of course tributary to the level of aggregation; for the case at hand the patterns ceases starting from industry 29 and the rest of the service industries.

It seems clear from the above comparison of fuzzy and crisp clusters that in addition to allowing the portrayal of relative importance of industries, the fuzzy clusters method allows for a better description of the existing clusters and the relation between clusters while the crisp clusters missed on detecting certain patterns of similarities between industries. By overlooking certain similarities between clusters, however weak they might be, crisp clustering ended up by putting most industries in one or two clusters (Table 1) and since an industry is forced to belong to one and only one cluster therefore the linkages between clusters are also overlooked. Although here we used only a 33-by-33 input-output table, fuzzy methods of cluster analysis can be safely used on large input-output data to unmask finer structures of similarities between industries and their relative importance, thing that crisp cluster analysis seems to fail to fully extract.

Unlike the binary value that the membership function in crisp clusters takes, continuous fuzzy membership allows for a larger number of operations and methods to extract information about the cluster and the industries composing it that goes beyond the simplistic observation that a cluster belongs or does not belong to a certain cluster. In addition, fuzzy clusters by their nature allow accounting for "spillovers" between industries, which for Porter (1998b) matter in the determination of the cluster boundaries. Porter (1998b) argues that "cluster boundaries should encompass all firms, industries, and institutions with strong linkages, whether vertical, horizontal, or institutional; those with weak or non-existent linkages can safely be left out". However, in this paper we showed that the *arbitrary* decision on what to include and what not to include in a cluster is no more a concern if low membership values are assigned to industries with weak similarities with other industries in a cluster.

4 Conclusion

Pattern identification and object classification are important and necessary steps to make a wealth of information useful for further analysis. Identifying similarities between industries is of importance to policy makers involved in regional development programs. Porter (1990) view clusters formation as a source of national advantages. Clusters formation and development can be achieved by various means such as the enlargement of existing clusters through the inclusion of new firms, private leadership, and technological leaps (Bekar and Lipsey, 2001). However, in most cases clusters emerge naturally, only then can government policies contribute to their growth, thus making the identification of clusters even more important for policy makers and private initiative.

Interest in industrial crisp cluster analysis based on input-output data decreased in the 1980s because of their limited ability to fully capture all the relationship that exist between industries. In using fuzzy cluster analysis, we hope to provide not only better ways to identify interindustry similarities but also to move toward a quantification of those clusters that goes beyond the binary notion of industries classification. Fuzzy sets theory, offers an alternative approach for cluster analysis not based on information concepts such as those proposed by Theil (1967) and Theil and Uribe (1967) but rather on the between-sector variance derived using dual scaling (Nishisato, 1980, 1994). While in this paper we relied on input-output data to derive fuzzy industrial clusters, any multivariate description of industries in terms of labor, investement and other variables can be used to accomplish that.

In this paper, we compared crisp cluster methods such as the agglomerative method proposed by Ward (1963) with fuzzy cluster methods and showed that crisp clusters are incapable to detect significant diversities between clusters of profiles that are functionally and structurally different, in addition, most industries were for the most part classified in a single cluster. We also showed how the membership values could be used for some elementary operations on cluster analysis.

Further research is still required to study the composition of a given industry cluster and what the role played by each industry; this compels us to consider clusters as coalitions with the application of cooperative game theory tools such as the Shapley value and to measure the entropy of fuzzy clusters. In the second paper in this series, the Shapley value and other fuzzy measures are introduced in the context of input-output based fuzzy cluster analysis.

REFERENCES

Bekar, C. and R. G. Lipsey (2001) Clusters and Economic Policies, paper presented at *Policies for the New Economy*, Montreal, June 26-27.

Bergman, E. M. and E. J. Feser (1999) *Industrial and Regional Clusters: Concepts and Comparative Applications*, The Web Book of Regional Science, Regional Research Institute, West Virginia University.

Blin, J.-M. and C. Cohen (1977) Technological Similarities and Aggregation in Input-output Systems: A Cluster-Analytic Approach, *Review of Economic Statistics* 59, 1: 82-91.

Bojadziev, G. and M. Bojadziev (1997): *Fuzzy Logic for Business, Finance, and Management*, World Scientific Publishing, Singapore.

Campbell, J. (1974) Selected Aspects of the Interindustry Structure of the State of Washington, 1967, *Economic Geography* 50, 1: 35-46.

Czamanski, S. and L.A.Q. Ablas. (1979) Identification of Industrial Clusters and Complexes: a Comparison of Methods and Findings, *Urban Studies*, 16: 61-80.

Dridi, C. and G. J.D. Hewings (2002a) Sectors Associations and Similarities in Input-Output Systems: An Application of Dual Scaling and Fuzzy Logic to Canada and the United States, Discussion Paper REAL 01-T-15, *Regional Economics Applications Laboratory, University of Illinois, Urbana* (forthcoming in *The Annals of Regional Science*).

Dridi, C. and G. J.D. Hewings (2002b) An Investigation of Industry Associations, Association Loops, and Economic Complexity: Application to Canada and the United States, *Economic Systems Research* 14, 3: 275-296.

Harris, J. W. and H. Stocker (1998) *Handbook of Mathematics and Computational Science*, Springer-Verlag, New York, Ch. 22.

Hewings, G. J.D., P.R. Israilevich, M. Sonis and G.R. Schindler (1997) Structural change in a metropolitan economy: the Chicago region, 1975-2010. In S. Bertuglia, S. Lombardo and P. Nijkamp (eds) Spatial Effects of Innovative Behaviour, Springer-Verlag, Heidelberg

Isard, W., E. W. Schooler, and T. Vietorisz (1959) *Industrial Complex Analysis and Regional Development*, MIT Press, Cambridge.

Jackson, R. W. (1986) The Full-Distribution Approach to Aggregate Representation in the Input-Output Modeling Framework, *Journal of Regional Science* 26, 3: 515-531.

Kaufman, L. and P. J. Rousseeuw (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, Ch. 4.

Kaufmann, A. (1975) *Introduction to the Theory of Fuzzy Subsets: Fundamental Theoretical Elements*, Vol. 1, Academic Press, New York.

Mirkin, B. (1996) *Mathematical Classification and Clustering*, Kluwer Academic Publishers, Dordrecht.

Miyamoto, S. (1990) *Fuzzy Sets in Information Retrieval and Cluster Analysis*, Kluwer Academic Publishers, Dordrecht, Ch. 1, 2, and 6.

Nishisato, S. (1980) *Analysis of Categorical Data: Dual Scaling and its Application*, University of Toronto Press, Toronto.

Nishisato, S. (1994) *Elements of dual scaling: An Introduction to Practical Data Analysis*, Lawrence Erlbaum Associates, New Jersey.

O'Huallachain, B. (1984) The Identification of Industrial Complexes, *Annals of the Association of American Geographer* 74, 3: 420-436.

Porter, M. E. (1998 a) Clusters and the New Economics of Competition, *Harvard Business Review* (November-December): 77-90.

Porter, M. E. (1998 b) *On Competition*, Harvard Business School Publishing, Boston, Massachusetts, Ch. 7.

Porter, M. E. 1990. *The Competitive Advantage of Nations*, The Free Press, New York.

Ragin, C. C. (2000) *Fuzzy-Sets Social Science*, The University of Chicago Press, Chicago.

Shapley, L. S. (1953) A Value for n-Person Games, in H. Kuhn (ed., 1997), Classics in Game Theory, Princeton University Press, Princeton, New Jersey, 69-79.

Theil, H. (1967) *Economics and Information Theory*, Rand McNally & Co., Chicago.

Theil, H. and P. Uribe (1967) The Information Approach to the Aggregation of Input-output Tables, *The Review of Economic Statistics* 49, 4: 451-462.

Ward, J. H. Jr (1963) Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association* 58, 301: 236-244.

Zadeh, L. A. (1965) Fuzzy Sets, *Information and Control* 8, 3: 338-353.

Table 1: Crisp clusters cardinality

Cluster	Profiles	
	Sales	Purchases
1	**6**	**5**
2	**2**	**7**
3	**2**	**4**
4	**3**	**4**
5	**10**	**2**
6	1	**2**
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
Total	*33*	*33*

Table 2: Dunn's index of fuzziness

	US	
	Sales Profiles	***Purchases Profiles***
Objective	2.8103	2.4839
Dunn's Coeff.	0.4261	0.4429
Dunn's Norm.	0.3851	0.4031

Table 3: "Forced" classification of US fuzzy clusters

	Sales Profile		Purchases Profile	
Ind.	**Closest Crisp Set**	**Neighbor**	**Closest Crisp Set**	**Neighbor**
1	1	8	1	2
2	2	14	2	15
3	1	8	1	15
4	3	8	3	15
5	1	8	4	1
6	4	8	5	11
7	1	8	1	7
8	5	14	6	15
9	1	14	1	2
10	6	8	7	5
11	1	8	4	2
12	7	8	8	2
13	1	7	9	8
14	8	6	9	8
15	1	8	4	2
16	9	10	10	13
17	1	8	4	11
18	10	8	11	13
19	1	14	4	2
20	11	7	12	8
21	1	6	1	13
22	12	10	13	11
23	1	6	1	11
24	13	8	14	2
25	1	14	1	2
26	14	15	4	2
27	1	8	1	2
28	15	14	15	2
29	1	14	1	15
30	1	14	1	2
31	1	15	1	2
32	1	14	1	2
33	1	14	1	2

Table 4: Cluster combinations cardinality for industries in the sales profile: a forced classification

Closest Crisp Set	**Neighbor Set**						***Total***
	R6	**R7**	**R8**	**R10**	**R14**	**R15**	
R1	**2**	1	**8**		**7**	1	***19***
R2					1		*1*
R3			1				*1*
R4			1				*1*
R5					1		*1*
R6			1				*1*
R7			1				*1*
R8	1						*1*
R9				1			*1*
R10			1				*1*
R11		1					*1*
R12				1			*1*
R13			1				*1*
R14						1	*1*
R15					1		*1*
Total	***3***	***2***	***14***	***2***	***10***	***2***	*33*

Table 5: Cluster combinations cardinality for industries in the purchases profile: a forced classification

Closest Crisp Set	**Neighbor Set**								***Total***
	C1	**C2**	**C5**	**C7**	**C8**	**C11**	**C13**	**C15**	
C1		**8**		1		1	1	**2**	**13**
C2								1	1
C3								1	1
C4	1	**4**				1			**6**
C5						1			1
C6								1	1
C7			1						1
C8		1							1
C9					**2**				**2**
C10							1		1
C11							1		1
C12					1				1
C13						1			1
C14		1							1
C15		1							1
Total	*1*	***15***	*1*	*1*	***3***	***4***	***3***	***5***	*33*

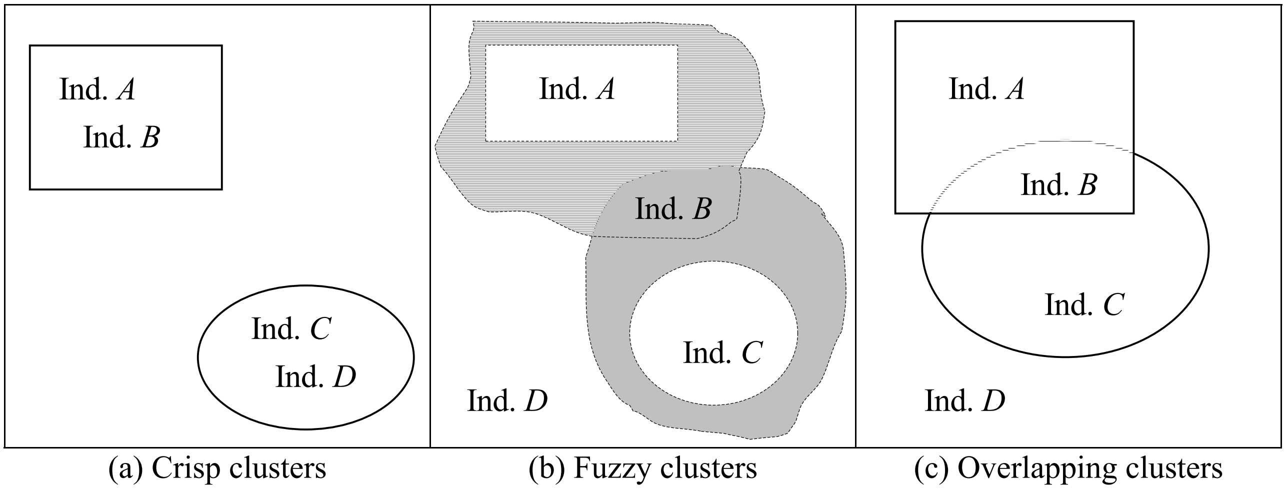


Figure 1: Crisp, fuzzy, and overlapping classifications

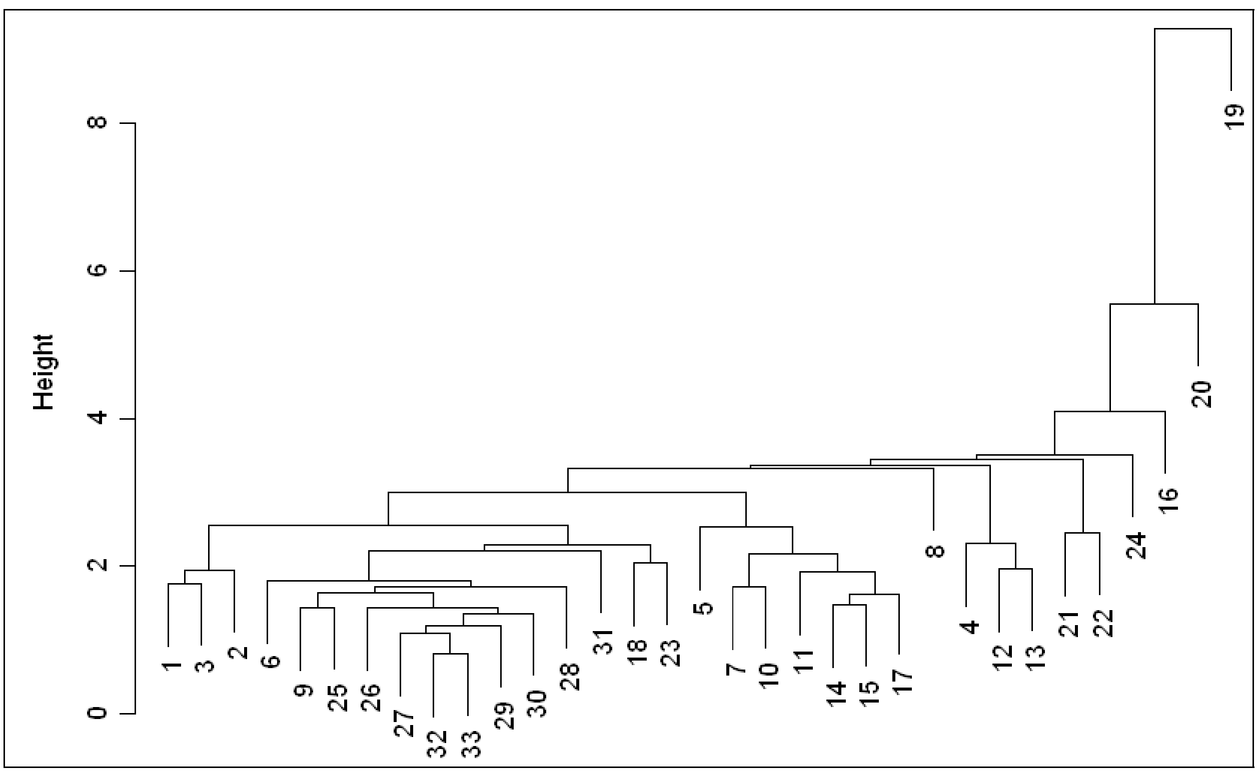


Figure 2: Dendogram of Ward's agglomerative clustering algorithm for the sales profile

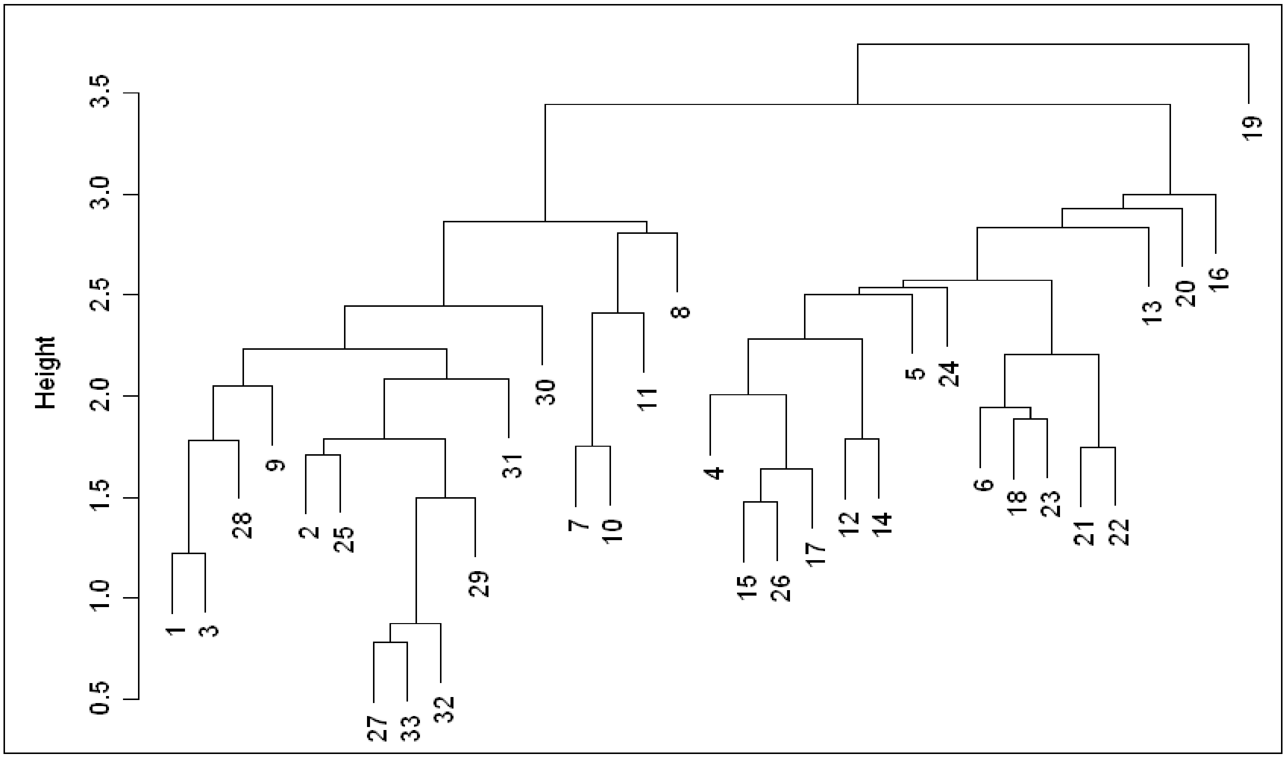
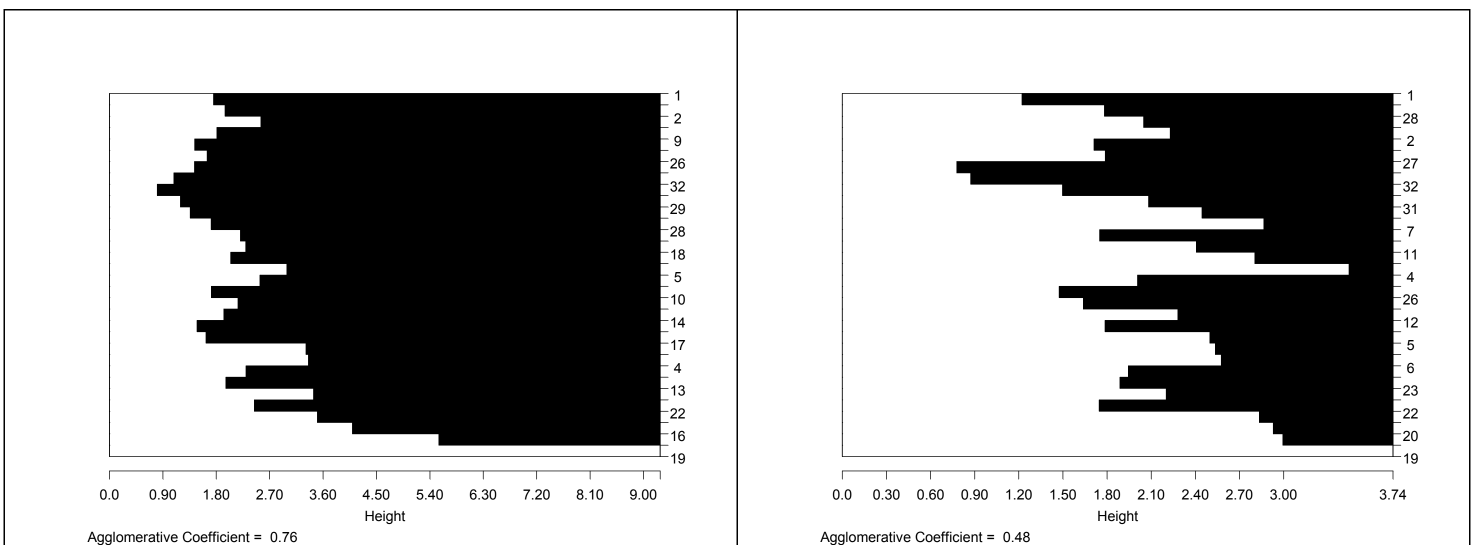
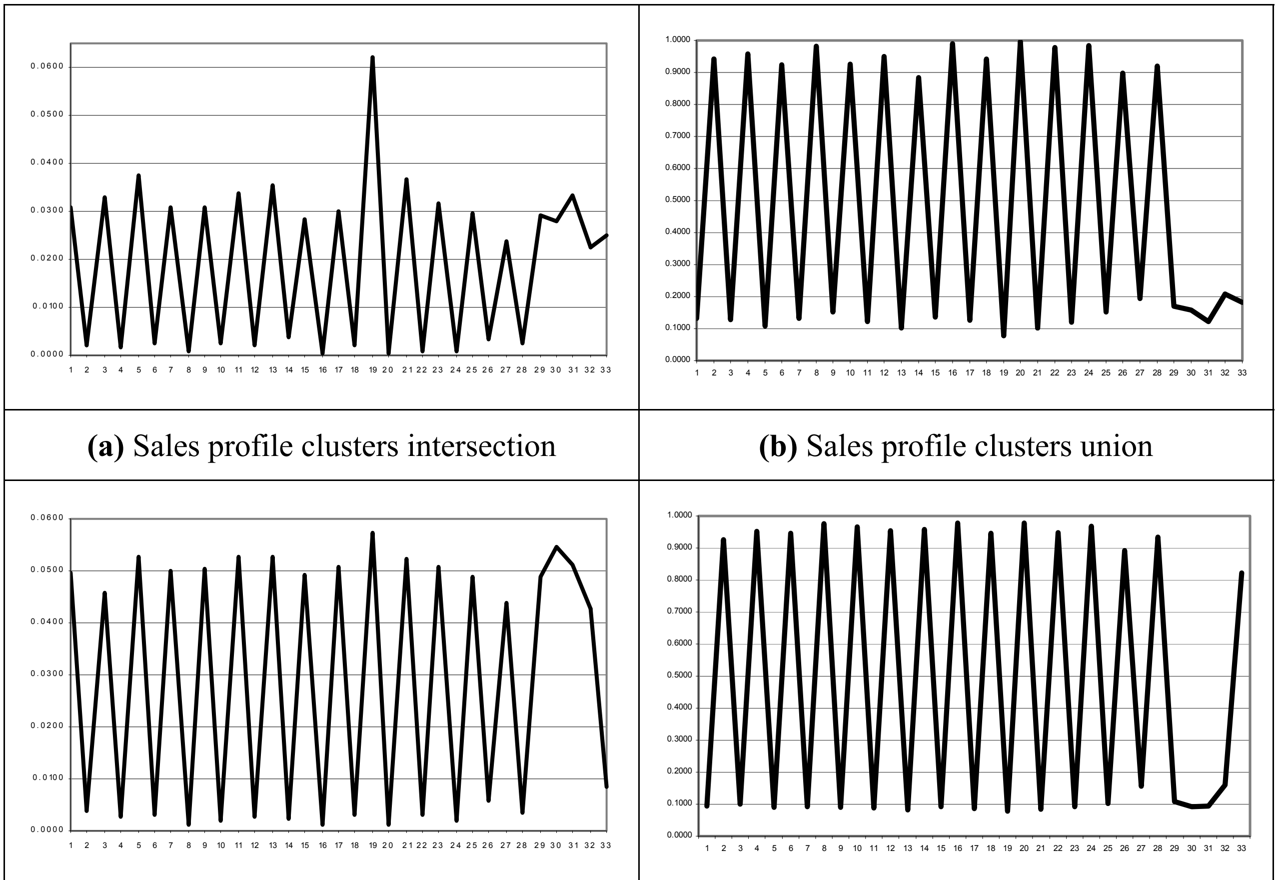


Figure 3: Dendogram of Ward's agglomerative clustering algorithm for the purchases profile



(a) Sales profiles	**(b)** Purchases profiles

Figure 4: Banner plots



(a) Sales profile clusters intersection

(b) Sales profile clusters union

(c) Purchases profile cluster intersection

(d) Purchases profile clusters union

Figure 5: Intersection and Union operation on US fuzzy clusters

APPENDIX

Brief description of the dual scaling technique (Dridi and Hewings, 2002b)

Nishisato (1980, 1994) presented the dual scaling technique as a method that applies to qualitative data arranged in a contingency table. However, if we consider that the monetary values in an input-output table cells are an indicator of the frequency of exchanges between industries, then the use of the technical block of input-output tables constitutes a contingency table well suited for dual scaling technique. In Dridi and Hewings (2002b) the interindustry flows where augmented by a row for the primary factors and a column for the final demands to represent a full input-output table as a contingency table, however as the level of aggregation decreases negative values in the final demands column might appear if imports of final goods and services were too high. Having negative values in a contingency table disqualifies the use of dual scaling as a tool to derive the rows and columns weights. In order to avoid the problem we limit ourselves to the interindustry flows, this causes a loss of information by dropping the primary demands and final demand from the analysis however, the gain derived from a more detailed table would be more important in this case. With $i = 1,...,r$, and $j = 1,...,c$, the size of the contingency table will be *r* rows by *c* columns:

$$\begin{pmatrix} f_{1,1} & \cdots & f_{1,j} & \cdots & f_{1,c} \\ \vdots & & \vdots & & \vdots \\ f_{i,1} & \cdots & f_{i,j} & \cdots & f_{i,c} \\ \vdots & & \vdots & & \vdots \\ f_{r,1} & \cdots & f_{r,j} & \cdots & f_{r,c} \end{pmatrix} \tag{a.1}$$

Let $f_{i,j}$ be the monetary value of flows between industries *i* and *j*. The approach of dual scaling consists in determining a vector of columns weight and a vector of rows weight to maximize the ratio $\eta^2 = \frac{SS_b}{SS_t}$, with:

$\mathbf{F} = [f_{i,j}]_{(r+1)\times(c+1)}$; the matrix of flows in an input-output table.

$\mathbf{f}_r$; the vector of total outputs of the input-output table.

$\mathbf{f}_c$ **;** the vector of total inputs same as $\mathbf{f}_r$ for input-output.

$\mathbf{D}_r$; the diagonal matrix with row totals in the main diagonal.

$\mathbf{D}_c$; the diagonal matrix with column totals in the main diagonal

$\mathbf{y}$; a vector of weights for the supplying sectors.

$\mathbf{x}$; a vector of weights for the demanding sectors.

f_t ; the total value or intensity of the input-output table.

and $SS_b = \mathbf{x}'\mathbf{F}'\mathbf{D}_r^{-1}\mathbf{F}\mathbf{x}$ expresses the variation between the rows of $\mathbf{F}$ and $SS_t = \mathbf{x}'\mathbf{D}_c\mathbf{x}$ expresses the total variation in the full input-output table.

One way to maximize $\eta^2 = \dfrac{SS_b}{SS_t}$, is to set $SS_t = f_t$ and to maximize SS_b. The Lagrangian function of the problem will be:

$$L(\mathbf{x},\lambda) = \mathbf{x}'\mathbf{F}'\mathbf{D}_r^{-1}\mathbf{F}\mathbf{x} - \lambda(\mathbf{x}'\mathbf{D}_c\mathbf{x} - f_t) \tag{a.2}$$

with first order conditions:

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{F}'\mathbf{D}_r^{-1}\mathbf{F}\mathbf{x} - \lambda\mathbf{D}_c\mathbf{x} = 0 \tag{a.3}$$

$$\frac{\partial L}{\partial \lambda} = \mathbf{x}'\mathbf{D}_c\mathbf{x} - f_t = 0 \tag{a.4}$$

If we pre-multiply (a.3) by $\mathbf{x}'$ and rearrange, we get:

$$\lambda = \frac{\mathbf{x}'\mathbf{F}'\mathbf{D}_r^{-1}\mathbf{F}\mathbf{x}}{\mathbf{x}'\mathbf{D}_c\mathbf{x}} = \eta^2 \tag{a.5}$$

The Lagrangian multiplier is nothing but the squared correlation ratio, η^2. Equation (a.3) can be rewritten into:

$$\left(\mathbf{F}'\mathbf{D}_r^{-1}\mathbf{F} - \eta^2\mathbf{D}_c\right)\mathbf{x} = 0 \tag{a.6}$$

which if pre-multiplied by $\mathbf{D}_c^{-1}$ yields the eigenequation:

$$\left(\mathbf{D}_c^{-1}\mathbf{F}'\mathbf{D}_r^{-1}\mathbf{F} - \eta^2\mathbf{I}\right)\mathbf{x} = 0 \qquad \text{(a.7)}$$

Once the trivial[5] solution of η^2 is excluded; an eigenvector $\mathbf{x}$, associated with the highest value of η^2 is found from (a.7), $\mathbf{y}$ can be found using the following dual relationship, which justifies the use of 'dual scaling' to label this approach:

$$\mathbf{y} = \left(\frac{1}{\eta}\right)\mathbf{D}_r^{-1}\mathbf{F}\mathbf{x} \qquad \text{(a.8)}$$

At this level, we obtain what is referred to as the first solution with a percentage of total information explained of $\delta_1 = \frac{100\eta_1^2}{\sum_i \eta_i^2}$. Nishisato (1994) offers a different formulation to δ_1, but it provides the same result since every eigenvalue explains part of the association and the sum of the non-trivial eigenvalues exhausts all the association. If the first solution is judged insufficient to explain the correlation between rows and columns then a second or more solutions can be found by calculating the associated eigenvector **x**, and the vector **y**, by taking decreasing non-trivial eigenvalues[6]. In a general contingency *r*-by-*c* table, the number of possible non-trivial solutions is $s = \min(r-1, c-1)$.

We mentioned earlier that the maximum number of solutions we can find is $\min(r-1, c-1)$, and since the number of rows and columns in an input-output table are equal, then the number of solutions is simply $r-1$ or $c-1$, the size of the technical block matrix in an input-output system.

The application of the above technique to extract the *s* solutions of the rows and columns of the input-output table or the contingency table in general provides two matrices, the first of which holds the s weights for the columns and its dimensions are $c \times s$, while the second matrix holds the s weights or solutions for the rows and its dimension is $r \times s$. With $k = 1,...,s$, the horizontal concatenation of the column and the row solutions produce respectively the following matrices; matrix **X** of weights for the columns, and the matrix **Y** of weights for the rows having the following configurations:

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,k} & \cdots & x_{1,s} \\ \vdots & & \vdots & & \vdots \\ x_{j,1} & \cdots & x_{j,k} & \cdots & x_{j,s} \\ \vdots & & \vdots & & \vdots \\ x_{c,1} & \cdots & x_{c,k} & \cdots & x_{c,s} \end{pmatrix} \qquad \text{(a.9)}$$

$$\mathbf{Y} = \begin{pmatrix} y_{1,1} & \cdots & y_{1,k} & \cdots & y_{1,s} \\ \vdots & & \vdots & & \vdots \\ y_{i,1} & \cdots & y_{i,k} & \cdots & y_{i,s} \\ \vdots & & \vdots & & \vdots \\ y_{r,1} & \cdots & y_{r,k} & \cdots & y_{r,s} \end{pmatrix} \qquad \text{(a.10)}$$

The matrices **X** and **Y** can be used to compute the inter-rows and inter-columns Euclidian distances in the space of dimension s.

$$d_{i,i'} = \sqrt{\sum_{k=1}^{s} \left(y_{i,k} - y_{i',k} \right)^2} \qquad ; \forall i = 1,...,r \; ; \forall i' = 1,...,r \qquad \text{(a.11)}$$

$$d_{j,j'} = \sqrt{\sum_{k=1}^{s} \left(x_{j,k} - x_{j',k} \right)^2} \qquad ; \forall j = 1,...,c \; ; \forall j' = 1,...,c \qquad \text{(a.12)}$$

[5] Trivial because it assigns the same weight to all elements of **x**, which prevents any further analysis.
[6] For our purpose we extract all possible solutions.

Tables

Table A.1: Input-output aggregation for the US economy

Ind.	Description
1	Agriculture, forestry & fishing
2	Mining & quarrying
3	Food, beverages & tobacco
4	Textiles, apparel & leather
5	Wood products & furniture
6	Paper, paper products & printing
7	Industrial chemicals
8	Drugs & medicines
9	Petroleum & coal products
10	Rubber & plastic products
11	Non-metallic mineral products
12	Iron & steel
13	Non-ferrous metals
14	Metal products
15	Non-electrical machinery
16	Office & computing machinery
17	Electrical apparatus, nec
18	Radio, TV & communication equipment
19	Shipbuilding & repairing
20	Other transport
21	Motor vehicles
22	Aircraft
23	Professional goods
24	Other manufacturing
25	Electricity, gas & water
26	Construction
27	Wholesale & retail trade
28	Restaurants & hotels
29	Transport & storage
30	Communication
31	Finance & insurance
32	Real estate & business services
33	Community, social & personal services

Table A.2: Fifteen crisp clusters for row and column profiles

Ind.	Row Cluster	Col. Cluster
1	4	3
2	4	2
3	4	3
4	6	4
5	7	7
6	5	1
7	1	6
8	8	8
9	5	3
10	1	6
11	1	9
12	3	5
13	3	10
14	1	5
15	1	4
16	9	11
17	1	4
18	2	1
19	10	12
20	11	13
21	12	1
22	13	1
23	2	1
24	14	14
25	5	2
26	5	4
27	5	2
28	5	3
29	5	2
30	5	15
31	15	2
32	5	2
33	5	2

Table A.3: Fuzzy clusters for the sales profile

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15
Ind. 1	0.1321	0.0774	0.0665	0.0744	0.0482	0.0750	0.0641	0.0843	0.0396	0.0633	0.0307	0.0489	0.0466	0.0770	0.0719
Ind. 2	0.0077	0.9422	0.0040	0.0048	0.0032	0.0046	0.0040	0.0050	0.0025	0.0040	0.0020	0.0031	0.0030	0.0051	0.0047
Ind. 3	0.1276	0.0752	0.0664	0.0729	0.0515	0.0742	0.0672	0.0810	0.0402	0.0639	0.0331	0.0491	0.0486	0.0777	0.0715
Ind. 4	0.0048	0.0031	0.9589	0.0030	0.0022	0.0032	0.0034	0.0037	0.0020	0.0031	0.0015	0.0024	0.0022	0.0033	0.0031
Ind. 5	0.1077	0.0666	0.0636	0.0735	0.0538	0.0753	0.0680	0.0936	0.0457	0.0684	0.0373	0.0558	0.0513	0.0708	0.0686
Ind. 6	0.0108	0.0059	0.0047	0.9231	0.0041	0.0064	0.0050	0.0069	0.0032	0.0060	0.0025	0.0042	0.0039	0.0067	0.0066
Ind. 7	0.1314	0.0703	0.0559	0.0795	0.0510	0.0825	0.0627	0.0864	0.0407	0.0657	0.0308	0.0518	0.0468	0.0727	0.0717
Ind. 8	0.0019	0.0014	0.0012	0.0015	0.9813	0.0014	0.0013	0.0015	0.0010	0.0014	0.0007	0.0012	0.0011	0.0017	0.0015
Ind. 9	0.1523	0.0777	0.0580	0.0736	0.0464	0.0689	0.0590	0.0801	0.0367	0.0641	0.0308	0.0465	0.0443	0.0857	0.0757
Ind. 10	0.0102	0.0054	0.0050	0.0063	0.0039	0.9252	0.0055	0.0070	0.0033	0.0057	0.0025	0.0043	0.0035	0.0061	0.0060
Ind. 11	0.1225	0.0681	0.0630	0.0736	0.0481	0.0779	0.0629	0.0991	0.0428	0.0669	0.0336	0.0529	0.0486	0.0693	0.0709
Ind. 12	0.0059	0.0034	0.0039	0.0035	0.0027	0.0039	0.9508	0.0046	0.0023	0.0039	0.0020	0.0029	0.0026	0.0039	0.0037
Ind. 13	0.1023	0.0660	0.0726	0.0661	0.0513	0.0700	0.0894	0.0796	0.0466	0.0704	0.0356	0.0540	0.0525	0.0739	0.0698
Ind. 14	0.0177	0.0083	0.0080	0.0093	0.0057	0.0098	0.0090	0.8849	0.0048	0.0089	0.0036	0.0062	0.0057	0.0093	0.0089
Ind. 15	0.1352	0.0647	0.0663	0.0706	0.0446	0.0732	0.0692	0.0954	0.0404	0.0701	0.0285	0.0515	0.0459	0.0742	0.0703
Ind. 16	0.0010	0.0007	0.0008	0.0008	0.0007	0.0008	0.0008	0.0008	0.9892	0.0009	0.0005	0.0008	0.0007	0.0008	0.0008
Ind. 17	0.1258	0.0644	0.0667	0.0700	0.0468	0.0785	0.0673	0.0900	0.0429	0.0711	0.0299	0.0541	0.0469	0.0750	0.0707
Ind. 18	0.0072	0.0039	0.0040	0.0048	0.0031	0.0046	0.0044	0.0051	0.0028	0.9421	0.0020	0.0037	0.0030	0.0047	0.0046
Ind. 19	0.0773	0.0670	0.0665	0.0676	0.0648	0.0672	0.0671	0.0674	0.0620	0.0675	0.0629	0.0644	0.0651	0.0684	0.0646
Ind. 20	0.0004	0.0003	0.0003	0.0003	0.0003	0.0003	0.0004	0.0003	0.0003	0.0003	0.9954	0.0003	0.0003	0.0003	0.0003
Ind. 21	0.1028	0.0648	0.0697	0.0701	0.0526	0.0764	0.0686	0.0764	0.0501	0.0758	0.0365	0.0677	0.0504	0.0693	0.0687
Ind. 22	0.0024	0.0016	0.0016	0.0017	0.0013	0.0018	0.0016	0.0019	0.0013	0.0019	0.0009	0.9773	0.0013	0.0017	0.0017
Ind. 23	0.1210	0.0647	0.0605	0.0765	0.0507	0.0767	0.0646	0.0775	0.0436	0.0725	0.0315	0.0566	0.0502	0.0775	0.0759
Ind. 24	0.0017	0.0012	0.0012	0.0013	0.0010	0.0012	0.0012	0.0014	0.0009	0.0012	0.0007	0.0010	0.9833	0.0013	0.0013
Ind. 25	0.1527	0.0714	0.0583	0.0744	0.0460	0.0718	0.0625	0.0784	0.0369	0.0652	0.0295	0.0464	0.0438	0.0874	0.0753
Ind. 26	0.0155	0.0077	0.0065	0.0082	0.0061	0.0077	0.0068	0.0085	0.0042	0.0074	0.0033	0.0053	0.0050	0.8989	0.0091
Ind. 27	0.1948	0.0634	0.0551	0.0745	0.0396	0.0707	0.0596	0.0960	0.0316	0.0633	0.0238	0.0419	0.0383	0.0776	0.0699
Ind. 28	0.0112	0.0058	0.0051	0.0068	0.0042	0.0064	0.0054	0.0067	0.0035	0.0060	0.0026	0.0043	0.0040	0.0075	0.9204
Ind. 29	0.1696	0.0667	0.0552	0.0728	0.0438	0.0721	0.0583	0.0819	0.0343	0.0646	0.0293	0.0441	0.0413	0.0873	0.0787
Ind. 30	0.1572	0.0667	0.0558	0.0788	0.0467	0.0708	0.0588	0.0754	0.0373	0.0669	0.0281	0.0472	0.0438	0.0889	0.0776
Ind. 31	0.1229	0.0713	0.0590	0.0769	0.0504	0.0705	0.0616	0.0770	0.0456	0.0694	0.0334	0.0551	0.0497	0.0789	0.0784
Ind. 32	0.2078	0.0616	0.0497	0.0729	0.0405	0.0652	0.0530	0.0774	0.0310	0.0610	0.0226	0.0402	0.0375	0.0964	0.0832
Ind. 33	0.1827	0.0637	0.0531	0.0733	0.0426	0.0676	0.0575	0.0768	0.0335	0.0637	0.0251	0.0434	0.0419	0.0925	0.0825
Card	*2.7241*	*2.2825*	*2.1671*	*2.3676*	*1.9391*	*2.3618*	*2.2210*	*2.5121*	*1.8029*	*2.2665*	*1.6331*	*1.9888*	*1.9131*	*2.4516*	*2.3687*

Table A.4: Fuzzy clusters for the purchases profile

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
Ind. 1	0.0940	0.0774	0.0706	0.0871	0.0686	0.0530	0.0606	0.0642	0.0623	0.0497	0.0649	0.0498	0.0644	0.0586	0.0748
Ind. 2	0.0081	0.9262	0.0053	0.0070	0.0054	0.0041	0.0046	0.0053	0.0050	0.0040	0.0053	0.0040	0.0053	0.0045	0.0060
Ind. 3	0.1001	0.0762	0.0701	0.0894	0.0695	0.0505	0.0571	0.0644	0.0601	0.0458	0.0637	0.0481	0.0618	0.0552	0.0879
Ind. 4	0.0045	0.0039	0.9511	0.0044	0.0033	0.0027	0.0029	0.0037	0.0037	0.0028	0.0036	0.0029	0.0035	0.0031	0.0039
Ind. 5	0.0821	0.0682	0.0664	0.0908	0.0687	0.0567	0.0629	0.0670	0.0672	0.0529	0.0674	0.0584	0.0661	0.0602	0.0652
Ind. 6	0.0056	0.0043	0.0036	0.0050	0.9458	0.0033	0.0037	0.0036	0.0034	0.0029	0.0043	0.0029	0.0040	0.0035	0.0042
Ind. 7	0.0922	0.0731	0.0590	0.0875	0.0736	0.0568	0.0769	0.0629	0.0575	0.0502	0.0653	0.0504	0.0658	0.0593	0.0696
Ind. 8	0.0022	0.0018	0.0016	0.0020	0.0018	0.9760	0.0017	0.0017	0.0017	0.0014	0.0017	0.0013	0.0017	0.0015	0.0019
Ind. 9	0.0892	0.0825	0.0681	0.0816	0.0710	0.0560	0.0600	0.0662	0.0615	0.0503	0.0643	0.0504	0.0637	0.0589	0.0765
Ind. 10	0.0030	0.0026	0.0023	0.0030	0.0026	0.0022	0.9663	0.0023	0.0022	0.0020	0.0024	0.0020	0.0025	0.0022	0.0025
Ind. 11	0.0868	0.0746	0.0651	0.0881	0.0713	0.0545	0.0652	0.0607	0.0609	0.0528	0.0667	0.0569	0.0663	0.0588	0.0714
Ind. 12	0.0043	0.0037	0.0035	0.0040	0.0031	0.0026	0.0028	0.9531	0.0040	0.0026	0.0034	0.0032	0.0033	0.0030	0.0036
Ind. 13	0.0777	0.0727	0.0729	0.0741	0.0627	0.0545	0.0578	0.0750	0.0818	0.0547	0.0677	0.0525	0.0646	0.0643	0.0669
Ind. 14	0.0037	0.0032	0.0033	0.0035	0.0028	0.0024	0.0026	0.0037	0.9573	0.0024	0.0032	0.0029	0.0030	0.0028	0.0032
Ind. 15	0.0886	0.0744	0.0720	0.0922	0.0638	0.0491	0.0557	0.0677	0.0687	0.0510	0.0700	0.0492	0.0681	0.0578	0.0717
Ind. 16	0.0018	0.0016	0.0015	0.0017	0.0014	0.0013	0.0014	0.0015	0.0015	0.9787	0.0017	0.0013	0.0018	0.0014	0.0015
Ind. 17	0.0861	0.0711	0.0696	0.0861	0.0650	0.0516	0.0613	0.0696	0.0677	0.0529	0.0724	0.0507	0.0687	0.0586	0.0687
Ind. 18	0.0052	0.0041	0.0038	0.0049	0.0042	0.0030	0.0032	0.0038	0.0039	0.0033	0.9456	0.0032	0.0044	0.0034	0.0040
Ind. 19	0.0754	0.0749	0.0687	0.0782	0.0673	0.0600	0.0611	0.0592	0.0687	0.0593	0.0669	0.0573	0.0685	0.0647	0.0699
Ind. 20	0.0017	0.0016	0.0016	0.0019	0.0015	0.0012	0.0014	0.0018	0.0018	0.0013	0.0017	0.9779	0.0016	0.0014	0.0017
Ind. 21	0.0836	0.0688	0.0728	0.0792	0.0656	0.0521	0.0600	0.0642	0.0619	0.0581	0.0735	0.0554	0.0818	0.0570	0.0661
Ind. 22	0.0050	0.0039	0.0036	0.0046	0.0037	0.0029	0.0033	0.0035	0.0035	0.0034	0.0043	0.0029	0.9485	0.0032	0.0037
Ind. 23	0.0913	0.0727	0.0645	0.0840	0.0700	0.0534	0.0619	0.0648	0.0621	0.0529	0.0720	0.0508	0.0704	0.0599	0.0693
Ind. 24	0.0029	0.0024	0.0023	0.0027	0.0024	0.0018	0.0020	0.0023	0.0023	0.0018	0.0023	0.0019	0.0023	0.9681	0.0024
Ind. 25	0.1030	0.0816	0.0656	0.0808	0.0685	0.0530	0.0569	0.0664	0.0598	0.0489	0.0650	0.0532	0.0634	0.0560	0.0778
Ind. 26	0.0120	0.0092	0.0078	0.8926	0.0083	0.0058	0.0068	0.0075	0.0071	0.0056	0.0082	0.0060	0.0080	0.0067	0.0084
Ind. 27	0.1560	0.0805	0.0585	0.0860	0.0678	0.0473	0.0508	0.0579	0.0554	0.0443	0.0645	0.0440	0.0626	0.0517	0.0726
Ind. 28	0.0072	0.0055	0.0049	0.0059	0.0048	0.0039	0.0040	0.0048	0.0046	0.0035	0.0047	0.0038	0.0046	0.0041	0.9336
Ind. 29	0.1078	0.0749	0.0633	0.0844	0.0666	0.0533	0.0562	0.0632	0.0594	0.0490	0.0680	0.0581	0.0645	0.0562	0.0751
Ind. 30	0.0914	0.0765	0.0644	0.0773	0.0690	0.0582	0.0591	0.0657	0.0618	0.0548	0.0693	0.0546	0.0675	0.0588	0.0715
Ind. 31	0.0937	0.0760	0.0620	0.0814	0.0716	0.0555	0.0596	0.0624	0.0601	0.0559	0.0688	0.0511	0.0701	0.0589	0.0729
Ind. 32	0.1594	0.0758	0.0584	0.0828	0.0672	0.0532	0.0503	0.0582	0.0556	0.0453	0.0640	0.0428	0.0629	0.0521	0.0719
Ind. 33	0.8222	0.0163	0.0122	0.0185	0.0143	0.0100	0.0107	0.0125	0.0115	0.0091	0.0137	0.0086	0.0134	0.0109	0.0159
Card	*2.6480*	*2.3423*	*2.2003*	*2.4724*	*2.2334*	*1.9921*	*2.0909*	*2.1707*	*2.1460*	*1.9533*	*2.2204*	*1.9584*	*2.2089*	*2.0669*	*2.2961*