The Modified Diagonalization Method for Analysing Clusters within Economies

Henryk Gurgul
Paweł Majdosz

In this paper a modification of the diagonalization method, originally put forward by Hoen (2002), is suggested which is aimed at uncovering clusters of sectors within an input-output framework. Our interest in this subject was largely motivated by the fact that the preceding method appears to be incapable of providing us with an accurate representation of the real cluster structure that exists in an economy, as a consequence of missing the position at which a given inter-sectoral flow stands in the hierarchy of the purchasing industry and the supplying industry. By making a distinction between an internal and external relationship, when it comes up at the moment of deciding whether each pair of industries is categorized as belonging to the same or different clusters, the proposed alternative, which will be referred to as the modified diagonalization method, seems to be superior to its predecessor. Such a conclusion is supported by the results of comparison of the relative performance of the rival methods (i. e. the original and modified diagonalization method) which show, among other things, that the average value of flows between industries grouped into clusters is higher in the case of the proposed method.

Key Words: internal and external interindustrial relationships, diagonalization method, clusters
JEL *Classification:* B41

Introduction

Industry clusters are nowadays an intrinsic element of the economic landscape of almost every country all over the world. Cluster-related problems have been viewed from various perspectives (spatially, inter-industrially and intra-industrially) and in varying contexts (see e. g.

Dr Henryk Gurgul is a Professor at the Department of Economics and Econometrics, University of Science and Technology, Poland.

Dr Paweł Majdosz is an Assistant Professor at the Department of Economics and Econometrics, University of Science and Technology, Poland.

Munroe and Hewings 2000). Theoretical interest in the concept of clustering is first and foremost associated with classical work on agglomeration in which the process of clustering is typically explained by the presence of externalities such as economies of scale and scope, which give economic advantages to firms agglomerating in a certain locality (see e. g. Hoover 1937; 1948; Marshall 1890; Ohlin 1933). Over the course of decades, there has been a tremendous development in the literature concerning this subject, and a comprehensive review of it is provided by Bekele and Jackson (2006). In the work of the mentioned authors the interested reader can find more details.

These theoretical approaches have been accompanied by empirical investigations of the clustering phenomenon which have found evidence supporting the presumption that an industry cluster allows enterprises to reduce costs, uncertainties and risks (see e. g. Antonelli 1999; Krugman 1991; Krugman and Venables 1996; Porter, 1998). Re-orienting economies towards a knowledge-based model has entailed a growing awareness that competitive success increasingly involves innovations and continuous quality improvements. The role of industry clusters has again appeared to be of crucial importance, this time, as a mechanism that enables innovations to spread throughout the economy (see e. g. Hauknes 1998; Martin and Sunley 2003).

Despite the importance of cluster analysis in explaining and exploring real economic structure, the effort which has been undertaken to improve the methods of cluster identification hardly gives an answer to the most urgent questions and there are still many problems unsolved in the face of which researchers may rely only on their own intuition. Overall, the practical application of cluster analysis poses many difficulties, especially since there is no uniform definition of what we should consider as a cluster. Definitions used in practice frequently depend on the particular ends to which a given study is subordinated. On the basis of the applied definition of cluster, one then decides which of the alternative methods of grouping sectors into clusters is employed. Confining our primary interest to the input-output context exclusively, it might be noticed that if great emphasis is placed in the definition upon intermediate deliveries between sectors within the economy, it is likely that the cluster identification method will be based on the matrix of intermediate deliveries. If attention is focused on the inputs of the ith product which are required for a one-unit increase in the production of the jth sector, the matrix of input coefficients may be selected as a base. If what is more important are inputs of the *i*th product required to meet the demand of (all) other sectors associated with a one-unit increase in the final demand of the *j*th sector, rather than those necessary for such an increase in the production of the *j*th sector, the inverse Leontief matrix will be probably used in place of the matrix of input coefficients. And finally, if the cluster definition stresses the extent to which the production of the *i*th sector is used by the *j*th sector in order to contribute to further production, the matrix of output coefficients may appear to the researcher to be particularly useful for dealing with the cluster identification problem.

The existence of many methods that can be used in practice for identifying inter-industrial relationships composing clusters contributes to the emergence serious difficulties with result comparability and could lead to a misinterpretation of the obtained evidence. It seems to be particularly controversy-prone when one of the rules underlying a technique employed is an economic meaning of the resulting clusters. By allowing this principle, any potential cluster that is an outcome of the formal procedure used in a first stage can then be cast away by the researcher as being deprived of precise economic meaning. The most important question that must be answered in all those cases where an economic meaning is a valid rule of the overall cluster identification procedure is to which extent the obtained results describe the actual cluster structure within the economy or, speaking more clearly, whether an exclusive goal of it is not, in fact, to support the presupposed statements and beliefs of the researcher. As long as different researchers will be tempted to base their work on the questionable principle of the economic meaning of identified clusters, any effort to improve and develop the rigorous methods for uncovering cluster structure, which could restrict the scope of arbitrariness related to the empirical studies that are undertaken in this momentous field, should be welcome.

In this paper a method designed to identify industry clusters within the economy is suggested. The main advantage of the method stems from the fact that it relies on a three-level classification of interindustrial linkages, thereby making it possible to discern between internal (or intra-cluster) relationships and external ones which connect industries belonging to different clusters, or industries within clusters with those outside. The suggested method is presented in the context of its predecessors, such as the diagonalization method (DM) proposed by Hoen (2002), and an extensive comparison using data from different countries and dates is carried out to enable the relative performances of these methods to be ascertained. The major conclusion that can be drawn from the results is that the proposed method tends to produce structures of industry clusters which seem to be more plausible and, at the same time, the level of internal relationships measured by the average value of flows among industries grouped into clusters is even higher than that of those obtained by means of the DM.

The remainder of the paper is organized as follows. In the next section several customarily used methods of identifying industry clusters are outlined and the alternative, i. e. the modified diagonalization method (MDM) is suggested. The third section contains a brief description of the data used in comparing the relative performances of the alternative methods. The empirical results are reported in the fourth section, while the fifth section concludes the paper.

Cluster Identification Methods

To facilitate a comparison of the proposed method with those hitherto-existing, we start with an outline of the latter, thereby giving a background against which the method suggested in this paper is formed as a remedy for some of their shortcomings. As was mentioned in the previous section, one characteristic that may make one method for identifying clusters different from another is that it may use the input-output matrix of a different sort as the base for computation. Let $\boldsymbol{x}$ denote the vector of gross output and Z be the matrix of intermediate deliveries whose elements show the amounts of output of the ith sector that are sold to the jth industry to maintain its own production. Then, the related matrices, i. e. the matrix of input coefficients (A), the Leontief inverse (D), and the matrix of output coefficients (B), can be derived, in turn, as follows:

$$\mathrm{A} = \mathrm{Z}\hat{\boldsymbol{x}}^{-1} \tag{1}$$

$$D = (I - A)^{-1} = L^{-1} \tag{2}$$

$$B = \hat{\boldsymbol{x}}^{-1} Z \tag{3}$$

where a hat is used to denote the diagonal matrix with elements of the vector on its main diagonal and all other entries equal to zero.

The simplest algorithm of grouping sectors into clusters, and the most frequently used in practice, the so-called method of maximization (MM), involves finding the largest off-diagonal element of a selected matrix (suppose that this matrix is A) and joining the sectors with the largest amount of intermediate deliveries until the number of clusters identified in such a way reaches a previously fixed threshold. The only advantage of this algorithm arises from its simplicity. However, it suffers from the so-called mega cluster problem and the fact that the obtained solution is sensitive to the matrix chosen as a basis of computation. In other words, although there are relationships among the matrices (as expressed by (1)–(3)), each of them embodies different information, and it may happen that the clusters yielded by this algorithm will be completely different when we use, for example, the Z in place of A or D instead of the B matrix.

Similar to the above-described method of uncovering clusters within the economy is the method of restricted maximization (RM). Roughly speaking, whilst the former takes into account all off-diagonal elements of the respective matrix to form the clusters, the latter focuses only on those which are large enough to satisfy the imposed restrictions. Such restrictions may be expressed in different ways, but without knowledge of the distribution of the matrix elements, it is usually formulated as a multiplication factor of the matrix average. This approach appears to be more flexible for several reasons. It allows the involvement of two or even all the matrices under consideration simultaneously, by imposing a conjunction of the respective restrictions for the single matrix. The method does not require the number of clusters to be determinate at the beginning of the investigation; instead one can adjust the value of the multiplication factor to obtain the same effect. Conversely, it does not deal with the possibility of varying clusters when different matrices are used as the basis for calculation.

Hoen (2002) put forward the DM which appears to be superior to the above-described ones. The first stage of this approach involves formulating a binary matrix (R), holding ones and zeros, which is given as:

$$r_{ij} = \begin{cases} 1 \text{ if } a_{ij} > q(A, 1-\alpha) \wedge b_{ij} > q(B, 1-\alpha) \\ 0 \text{ otherwise} \end{cases}, \tag{4}$$

where $q(X, p)$ stands for quantile of the elements of X at the order p.

Then, by permuting its rows and columns, these are an attempt to transform matrix R into a block-diagonal matrix so that each of its blocks represents one cluster. The reasoning underlying this method is that the cluster should encompass all sectors which are connected to each other and, at the same time, unattached from the rest of the economy. Changing the level of significance α, we decide which intermediate deliveries, input coefficients and output coefficients are regarded as significant and which are insignificant and to be fixed at zero. The higher the level of α, the more linkages there are among the sectors within the economy, but not always more clusters are identified as Hoen himself would wish. This issue will be given more attention later on.

As Hoen (2002) documented, the DM solves the problem of changes in the composition of clusters with respect to the matrix used as a basis, and facilitates a better insight into the structure of an economy by not requiring the number of existing clusters to be specified at the beginning of the investigation. However, this approach also has several drawbacks. One is that the respective quantiles are obtained on the basis of all matrix elements, including those located on the principal diagonal. When this method is used for comparative purposes, i. e. to examine how far the composition of clusters is unchangeable over time, it seems to be reasonable to fix parameter α at the same level during the whole analysed period. But the two sets of results obtained by this means from different points are comparable only if there is no change in the proportion of elements placed on the principal diagonal to the off-diagonal elements. This is probably not a problem in the case of developed countries where the off-diagonal elements of the respective matrices dominate those located on the principal diagonal, indicating that each sector assigns only a small portion of its own current outputs in order to contribute to further production, but for transitional countries this ratio is reversed (see e. g. Ćmiel and Gurgul 2002; Gurgul and Majdosz 2005). Within a few years of market-oriented reforms the share of off-diagonal flows within the economies of the countries in transition is expected to increase and the problem will vanish when it reaches a level which is typical of developed countries all over the world. Until then however, it is recommended that all elements on the diagonal be set at zero before calculating the respective quantiles.

As mentioned above, using the DM does not require the number of clusters to be determined by the researcher. Instead, the level of significance α can be adjusted for the same result, i. e. to obtain the predetermined number of clusters identified within the economy. However, there is no unambiguous relationship between the level of significance α and the number of uncovered clusters. With a higher value of parameter α, more entries in the matrix of interest emerge as important linkages and nothing further. Under some particular circumstances, this may lead to the inclusion within clusters of sectors previously excluded at the lower level of significance α, or even to identifying new clusters. Nonetheless, it should be stressed that the relationship: the higher parameter α, the more clusters uncovered is not automatically true. More strictly speaking, for some range of values of parameter α such a relationship holds true, but for another it does not. To realise this, suppose that parameter α is equal to zero, implying that matrix R consists of zeros, exclusively. Without significant entries in the respective matrix, no cluster will be identified. In contrast, when parameter α equals 100%, matrix R is formed with ones. Having only sectors with significant linkages with each other, all the sectors are now included within the same identified single cluster. It becomes immediately obvious that for α to belong to the range from 0 to k the number of identified clusters is a non-decreasing function of α, but if α belongs to the range from k to 1, the number of identified clusters is a non-increasing function of α. For the sake of simplicity, we abstract from a situation where the number of clusters is serially a non-decreasing and non-increasing function of α for $\alpha \in \langle 0, k_1) \cup \langle k_1, k_2) \cup \ldots \cup \langle k_n, 1\rangle$.

Therefore, k is a threshold, in excess of which the number of uncovered clusters within the economy can only diminish. This is accompanied by a trend towards the joining of clusters, thus obscuring the real cluster structure we want to explain and explore. Theoretically, with a simplified example like this, it is possible to determine a threshold k, not analytically of course but only practically by trial-and-error, and by selecting a level of significance α which is below or even equal to this threshold to avoid blurring the economy's structure. In practice, however, when it is likely that the number of clusters is serially a non-decreasing and non-increasing function of α, we would have to estimate the n threshold of ki (where n is the number of changes in direction). Even if the values of the threshold for any i are known, a still unanswered question is which value out of the n thresholds should be used to maximize the transparency of the identified relationships among the sectors operating within the real-world economy. In the light of the above-mentioned difficulties, it is obvious that an effort should be undertaken to develop another approach which, being aimed at identifying clusters of sectors within the economy, distinguishes between linkages of sectors belonging to the same cluster, relationships between two sectors arising from different clusters, and the linkages of sectors within clusters with the rest of the sectors classified as outside the clusters.

Defining the problem, it should be stated that the main shortcoming of the DM stems from its inability to classify sectors which may potentially be assigned to more than one cluster. Without a rule for grading linkage strength, in the case when a given sector has relationships with two different clusters, the DM blurs the economy's structure by joining these clusters into one. Hence, the method proposed by us, the MDM, should be first and foremost provided with an operational principle which enables the alternative allocations of a given sector to be ranked in respect to the strength of inter-industry linkages associated with each of them.

Seeking a solution to this problem, it becomes apparent immediately that the application by the DM simple categorization of inter-industrial relationships, according to their magnitudes, as significant or not is no longer valid, and that the significant relations among industries have to be further broken down if the proposed method is to give any advantages to the researcher who is irritated by the necessity to choose the threshold k in such a way so as to maximize the number of clusters uncovered without any, even the slightest, guide as to how it should be reached in practice. We would gain little, if anything, by setting leaning a division of the significant – in terms of magnitude – inter-industrial relationships on the absolute value principle since this amounts to introducing two unattached thresholds, instead of the one acting under the original diagonalization method, which would be hardly able to avoid the tendency to amalgamate different clusters into one with α increasing. What we propose in this paper is to look at a given connection from the perspective of co-relative industries. Each inter-sectoral relationship expressing the actual flow of goods and services which takes place between two involved industries can be considered, at least, from two different viewpoints, namely, the side of industry that sells its output and the side of industry that purchases it. Whereas in the case of the former, a question we have to answer concerns an issue as to what is the position of the flows in the hierarchy of the selling industry. In the case of the latter, we will be rather concerned with the degree to which that same flow is crucial for the purchasing industry.

Suppose that the element located in the ith row and the jth column of the matrix Z, i. e. z_{ij}, is deemed as significant, no matter what the term 'significant' exactly means here. Then the ith sector is a supplier (seller) whilst the jth sector represents the demand side (buyer). Note that under the DM, such sectors connected by the element z_{ij} would be automatically considered as composing the same cluster. But now, such a conclusion will be valid if, and only if, in the ith row there is no element that would be greater than the element placed on the jth column and, at the same time, in the jth column there is no element exceeding those located in the ith row. This means that here we focus our attention on the relative position of the flow at hand, on the list of deliveries of the selling industry as well as on the list of purchases of the buying industry. Only if the considered flow is ranked first in terms of both the lists, will the situation be equivalent to those presupposed by the DM, and the corresponding relationship will be referred to as internal or primary (from a given cluster's point of view). On the other hand, if at least one requisite mentioned above is violated, i. e. if there is a larger flow in the ith row or in the jth column than the element z_{ij}, then such a relationship, its significance remaining, will be termed as external or secondary.

Before going on to outline the procedure of the suggested approach, which is based on a distinction made between internal and external (or primary and secondary) inter-industrial relationships, it is necessary to point out one, rather practical, issue. Consider a situation in which z_{ij} and z_{ji} are both significant and z_{ij} is larger than z_{ji}. It might well happen that z_{ij} would be categorized as a secondary relation and z_{ji} as a primary one since there is nothing that would guarantee that both the relationships will be simultaneously considered as internal or external and this may occur only by chance. In order to prevent such an ambiguous finding, it is necessary to introduce the additional conventional principle that with z_{ij} and z_{ji} being significant, only that relationship is subject to further consideration which is larger, i. e. z_{ij} in our simple example.

The proposed approach draws upon several elements that have been utilized under the original diagonalization method. The first step consists of creating the restriction matrix (Q), but unlike its predecessor which yields a binary matrix, a three-value-coding is used here to allow for discerning between external and internal linkages. Assuming the matrix of deliveries (Z) is to be chosen as a basis for calculation, which however does not impair its generality, we can express the restriction matrix as:

$$q_{kl} = \begin{cases} 2 \text{ if } a_{kl} > q(A, 1-\alpha) \wedge b_{kl} > q(B, 1-\alpha) \\ \qquad \wedge z_{kl} > z_{kj} > \wedge z_{kl} > z_{il}, \forall i \neq k, j \neq l \\ 1 \text{ if } a_{kl} > q(A, 1-\alpha) \wedge b_{kl} > q(B, 1-\alpha) \\ 0 \text{ otherwise} \end{cases} . \tag{5}$$

Note that all the input-output matrices involved in (5), i. e. A, B, and also Z in our example, should be at first prepared in such a way that the elements located on their main diagonal are all set to zero. This is a prerequisite for any further calculation, because without neutralizing the effect of the main-diagonal elements of the respective matrices, the method might produce erroneous results. A reason for this stems not only from the fact that an inclusion of the main-diagonal elements in the calculation of the quantiles in (5) may lead to overestimation of their values, but large entries placed on the main diagonal of the matrix used as a base, are able to change what we will regard as an internal relation or external relation.

The second step is almost identical to the corresponding step under the DM with one exception. Starting with the restriction matrix and using Hoen's algorithm (see Hoen 2002) we then try to transform it into a block-diagonal matrix with respect to the internal inter-industrial relationships only (the entries with a '2' digit) and allow the elements representing the external relations (with a '1' digit) to change their positions freely with row-wise and columnwise permutations that are necessary to complete Hoen's algorithm. In other words, while transforming the restriction matrix into a block-diagonal one, we treat all the entries which, according to (5), are equal to one as if they are set to zero, but without losing information concerning their positions once the algorithm is terminated.

The transformed restricted matrix can be interpreted in terms of external and internal relationships as follows. Each block, as discovered by Hoen's algorithm, represents a single cluster of industries among which only internal relationships occur. The elements of the restriction matrix pertaining to external relationships indicate, therefore, either an inter-cluster linkage or connection of a given cluster with the rest of the economy that are composed of the industries not being assigned to any cluster.

What do we gain from using the suggested method? For instance, if a certain sector is significantly connected with two other sectors belonging to different clusters, it will be assigned to the cluster whose linkage is stronger. Information about the existence of a significant, although relatively weaker linkage with the sector belonging to other cluster is not lost, however, because such a linkage is automatically classified as an external one. In this way the proposed method prevents an unreasonable joining of clusters without omitting significant relationships among the clusters (or between the clusters and the rest of the economy) and, therefore, offers a better insight into the real structure of the economy. These benefits would, however, appear illusory, if the method does worse in terms of other properties such as, for example, its soundness in the selection of a basic matrix for calculation, or the ratio of average flows among the sectors included within clusters to the analogous value for sectors outside the clusters. Therefore, in the following sections we empirically test the properties of the suggested method and compare the obtained results with those produced by its predecessor, i. e. the DM.

Data Description

When illustrating the MDM's capacity for uncovering the cluster structure as compared with that of the DM, a sufficiently high level of disaggregation of the input-output tables used as the basis for computation is of prime importance. Although the initial number of distinct sectors within an economy differs significantly across empirical investigations concerning the problem of identifying clusters of industries, it can be found that such tables almost always distinguish no less than a hundred sectors (see e. g. Hauknes 1998). In order to find out the relative performance of the method discussed above, we therefore used several national input-output tables for different dates which deal with at least one hundred different sectors and which, of course, were available to us. To be more precise, the tables from three countries were engaged in the sample, namely the US tables for various dates, the Danish tables for various dates, and the UK tables for 1995. In order to keep the presentation of results short, and also because the main conclusions about the promising nature of the MDM remain basically the same irrespective of which country and which date are selected, we decided not to present all outcomes but only those being obtained by using the UK input-output tables for 1995. Other results are available from the authors on request.

The above-mentioned tables, derived directly from the National Statistics, are evaluated at current prices from the seller's point of view (basic prices) and are fully consistent with the European System of Accounts 1995 (ESA 95). The original statistics provide coverage of the economy as a whole combined with 138 industries/products using the Standard Industrial Classification 1992 (SIC 92). The last fifteen entries of the respective tables, however, arose from dividing some industries/products associated with Government and non-profit institutions serving households into market and non-market components. Doing so, the tables give

TABLE 1 Industry classification

1	Agriculture	32	Pulp, paper and paperboard
2	Forestry	33	Paper and paperboard products
3	Fishing	34	Printing and publishing
4	Coal extraction	35	Coke ovens, refined petroleum and nuclear fuel
5	Oil and gas extraction	36	Industrial gases and dyes
6	Metal ores extraction	37	Inorganic chemicals
7	Other mining and quarrying	38	Organic chemicals
8	Meat processing	39	Fertilisers
9	Fish and fruit processing	40	Plastics and synthetic resins etc
10	Oils and fats	41	Pesticides
11	Dairy products	42	Paints, varnishes, printing ink etc
12	Grain milling and starch	43	Pharmaceuticals
13	Animal feed	44	Soap and toilet preparations
14	Bread, biscuits, etc	45	Other chemical products
15	Sugar	46	Man-made fibres
16	Confectionery	47	Rubber products
17	Other food products	48	Plastic products
18	Alcoholic beverages	49	Glass and glass products
19	Soft drinks and mineral waters	50	Ceramic goods
20	Tobacco products	51	Structural clay products
21	Textile fibres	52	Cement, lime and plaster
22	Textile weaving	53	Articles of concrete, stone etc
23	Textile finishing	54	Iron and steel
24	Made-up textiles	55	Non-ferrous metals
25	Carpets and rugs	56	Metal castings
26	Other textiles	57	Structural metal products
27	Knitted goods	58	Metal boilers and radiators
28	Wearing apparel and fur products	59	Metal forging, pressing, etc
29	Leather goods	60	Cutlery, tools etc
30	Footwear	61	Other metal products
31	Wood and wood products	62	Mechanical power equipment

Continued on the next page

a better insight into the inter-industrial relationships taking account of the differences in proportions of inputs when an industry's products are sold on market principles as opposed to the case where market mechanisms do not apply. Taking into account the purpose of our investigation, however, we decide not to distinguish market and non-market components and to aggregate the tables into 123 industries/products (see table 1 for the list of sectors), avoiding in this way the zero-row or zero-column problems which would have to be corrected if we did not reduce the number of sectors to the above-mentioned 123.

The original statistics enable the analysis to be carried out either on commodity-by-industry or commodity-by-commodity basis. With

TABLE 1 *Continued*

63	General purpose machinery	93	Railway transport
64	Agricultural machinery	94	Other land transport
65	Machine tools	95	Water transport
66	Special purpose machinery	96	Air transport
67	Weapons and ammunition	97	Ancillary transport services
68	Domestic appliances nec	98	Postal and courier services
69	Office machinery and computers	99	Telecommunications
70	Electric motors and generators etc.	100	Banking and finance
71	Insulated wire and cable	101	Insurance and pension funds
72	Electrical equipment nec	102	Auxiliary financial services
73	Electronic components	103	Owning and dealing in real estate
74	Transmitters for TV, radio and phone	104	Letting of dwellings
75	Receivers for TV and radio	105	Estate agent activities
76	Medical and precision instruments	106	Renting of machinery etc
77	Motor vehicles	107	Computer services
78	Shipbuilding and repair	108	Research and development
79	Other transport equipment	109	Legal activities
80	Aircraft and spacecraft	110	Accountancy services
81	Furniture	111	Market research, management consultancy
82	Jewellery and related products	112	Architectural activities and technical consultancy
83	Sports goods and toys	113	Advertising
84	Miscellaneous manufacturing nec and recycling	114	Other business services
85	Electricity production and distribution	115	Public administration and defence
86	Gas distribution	116	Education
87	Water supply	117	Health and veterinary services
88	Construction	118	Social work activities
89	Motor vehicle distribution and repair, automotive fuel retail	119	Sewage and sanitary services
		120	Membership organisations nec
90	Wholesale distribution	121	Recreational services
91	Retail distribution	122	Other service activities
92	Hotels, catering, pubs etc.	123	Private households with employed persons

commodity-specific technologies being a more persuasive assumption in an input-output framework, the latter form of tables are used in the entire study. It should, however, be stressed that the findings do not differ significantly with respect to both techniques of deriving input-output tables in their quadratic form.

Empirical Results

Neither of the above-outlined methods of identifying clusters necessitates the number of distinct clusters to be specified in advance. Instead, changing the level of significance α gives the same result, that is, the desired number of identified clusters is achieved. But comparing the relative performances of the methods at a fixed level of significance may be rather confusing due to the fact that each of these methods is likely to have a different range of α over which reasonable results are produced. Another, and perhaps more revealing, way of illustrating their implementation is by imposing a fixed number of identified clusters without a concern for the level of significance at which it is achieved in the case of each of the methods. Following the latter, we selected the number of clusters to be identified by means of each of the methods under consideration (16 clusters were chosen), and then the level of significance was adjusted until the specified number of clusters was reached. While the DM gives the desired number of clusters at a level of significance amounting to 0.5%, the corresponding value for the MDM is equal to 1.32%. Perhaps, it should be here remembered that α (or level of significance at which clusters are evaluated) is simply a single complement of quantile order that is calculated on the basis of the matrix elements (see the equations (4) and (5)). Note also that the input-output matrices involved in computation were first adjusted for main-diagonal element effects by imposing a zero diagonal principle (all the corresponding entries were set to zero), so that what constitutes a basis for calculation of the quantiles are only the off-diagonal elements of the respective matrices.

Figure 1 depicts the cluster structure of the UK economy for 1995 based on the DM. With the help of this it becomes immediately obvious why as many as 16 clusters need to be identified to compare the alternative approaches. Of the 16 uncovered clusters only 7 are composed of three or more industries. Surprisingly, as will be shown later on, the same proportion of so-called mini-clusters, where only one inter-industry relationship constitutes a cluster, applies to the results obtained by means of the MDM. Consideration of such mini-clusters is not of interest from the practitioner's point of view, though they have some informational content. Searching for significant levels at which the problem of mini-clusters does not exist, although theoretically possible, would lead to a worse transparency of results and compromise the overall comparison of alternative methods of identifying clusters, which is the main aim of this study. Preferring to preserve the transparency of our results as far as possible, we decide not to adjust for mini-clusters, but hardly any attention will be given to such clusters due to their minimal economic importance for a practitioner. Despite this, the full results are presented, including mini-clusters, to facilitate for interested readers an insight into the real economic landscape of the economy under study. One important consideration which can be drawn from the relative high ratio of mini- to all

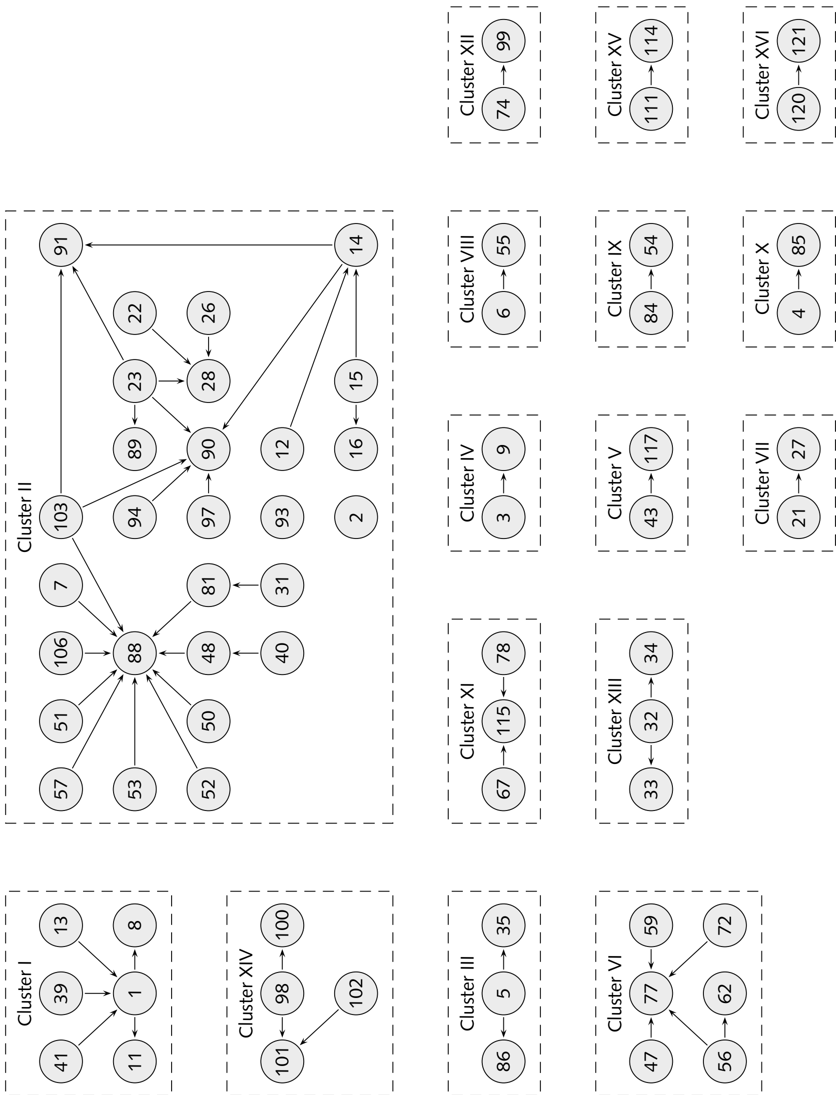

FIGURE 1 Clusters in the UK economy for 1995 based on the Diagonalization Method

identified clusters is that this problem remains unsolved no matter which of the methods is used.

The largest cluster (II) consists of as many as 28 industries which are rather diverse in terms of their activities, ranging from Forestry through Textiles and Distribution (both wholesale and Retail) to Con-

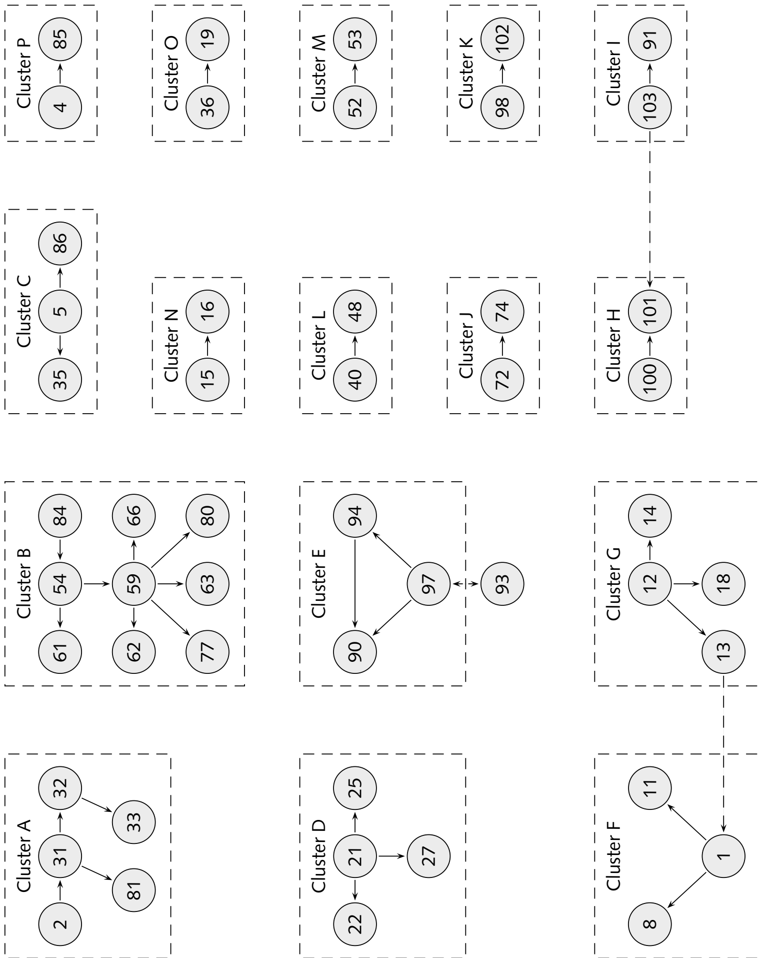

FIGURE 2 Clusters in the UK economy for 1995 based on the Modified Diagonalization Method

struction and Transportation. The high level of activity diversification in this mega-cluster poses some difficulties when one tries to give it a name. This cluster could be, for example, referred to as the Wood-textile-construction cluster, because all of these kinds of activities are strongly represented within its structure. However, almost everyone will agree with us that this name is rather ungainly. No matter whatever name this cluster is given, it is more important that it, in fact, obscures the actual relationships among industries constituting the cluster structure since some sub-clusters are likely to be sensibly singled out. One such sub-cluster might include, for example, such industries as Forestry (2), Wood and wood products (31) and Furniture (91). Another one might consist of Textile weaving (22), Textile finishing (23) and Other textiles (26). Furthermore, Sugar (15) and Confectionery (16), Cement, lime and plaster (52) and Articles of concrete (53) as well as Plastics and Synthetic resins (40) and Plastic products (48) are other exemplifications of pairs of industries which should be probably considered as sub-clusters of the mega Wood-textile-constriction cluster.

Other clusters, excluding the mini-clusters mentioned above, suggested by the DM seem to be better defined. The Agro cluster (I) includes such industries as Agriculture (1), Meat processing (8), Dairy products (11), and Animal feed (13), Fertilisers (39) and Pesticides (41) as well. Included in the Energy cluster (III) are only three industries, namely Oil and gas extraction (5), Coke ovens and refined petroleum (35), and Gas distribution (86). As large as the Agro cluster is the Metal and Machinery cluster (IV) with six industries, which is followed by the Connection and Financial cluster (XIV) with four industries: Banking and finance (100), Insurance (101), Auxiliary financial services (102) as well as Postal and courier services (98). Two other clusters identified by the DM are the Paper cluster (XIII) and the Weapons and Shipbuilding cluster (XI) each of which consists of three industries.

Figure 2 shows the clusters obtained by means of the MDM. As mentioned above, the number of clusters identified is the same as with the DM, but one can see that there exist substantial differences in terms of the composition of each of them when comparing the two alternative approaches presented here. Note also that unlike the DM, the MDM provides information about the inter-industry relationships of two kinds. In order to distinguish them, a full line is used to denote inter-cluster relationships among industries, whereas a dotted line means external relationships where out of two intertwined industries one is outside clusters. There are only three such external relationships in the figure.

One can see that now the largest is the Metal and Machinery cluster (B) with nine intertwined industries. It is useful to state right at the beginning that the results based on the MDM, with the exception of the Energy cluster (C), in which the component industries are exactly the same as those obtained by means of the DM, and despite the same labels (names) in some cases as previously used, the component industries of the respective clusters are completely different under the MDM. The current Metal and Machinery cluster is composed of such industries among others as Metal forging and pressing (59), Mechanical power equipment (62) and Motor vehicles (77), which were previously classified as belonging to cluster IV as well as Iron and steel (54) and Miscellaneous manufacturing and recycling (84) grouped the first time into the two-element cluster IX. It also deserves to be emphasized that some industries in this cluster, mainly Other metal products (61), General purpose machinery (63), Special purpose machinery (66), and Aircraft and spacecraft (80), were set aside when using the DM.

The subsequent Wood and paper cluster (A) is ranked second with respect to its size. Out of five industries included within it, three (Forestry (2), Wood and wood products (31), and Furniture (81)) were originally in cluster II, whereas the following two (Pulp, paper and paperboard (32) and Paper and paperboard products (33)) were previously grouped into cluster XIII. The Textiles cluster (D), on the other hand, is formed partially from cluster VII (Textile fibres (21) and Knitted goods (27)) and partially from the mega cluster in which Textile weaving (22) was previously classified.

Interestingly, the Transportation and telecommunication cluster (E) is entirely formed by breaking down the mega cluster identified by the DM. One can, however, see that Railway transport (93) and Ancillary transport services (97) are here connected via an external bi-directional relationship, and not what is suggested by the DM. Also, in the case of mini-cluster I we find that it emerged from the same mega cluster, but that now there exists an external relation between this cluster and cluster H entirely formed by breaking down the Connection and Financial cluster (XIV).

The last two clusters that will be given attention are clusters F and G. The former emerged as a part of the Agro cluster (I), whereas the latter consists of Grain milling and starch (12), and Bread and biscuits (14) being previously grouped into the mega cluster as well as Animal feed (13) which previously belonged to the Agro cluster and is now revealed to be externally interrelated with Agriculture (1) from cluster F.

Another way of dealing with externally interrelated clusters is by treating all the individual clusters among which there exist external relationships as a single cluster with sub-clusters. Following this approach results in joining the pairs of clusters F and G as well as H and I, previously treated separately.

A desirable feature of any method for identifying clusters of indus-

TABLE 2 Comparison of the strength of identified relationships under the two alternative methods

	Diagonalization Method			Modified Diag. Method		
	(1)	(2)	(3)	(1)	(2)	(3)
Q1.	0.009	0.000	0.000	0.013	0.000	0.000
Mean	20.336	8.920	5.813	25.339	10.111	7.772
Median	1.721	0.790	0.547	2.141	0.925	0.700
Q3.	102.227	21.292	18.453	102.831	27.020	50.767
Std. deviation	796.375	93.281	96.200	486.315	117.746	744.978

Note that all figures are in £ million.

tries is that it should produce the same results independently of which of the alternative matrices is used as a basis for calculations. As mentioned above, the DM has this feature which becomes immediately obvious when taking the manner of forming the R matrix (see (4)) into account. But, it turned out that the MDM also results in always forming the same clusters for intermediate deliveries, input and output coefficient matrices. Only using the Leontief inverse matrix produces different clusters, and sometimes the differences were rather essential. Nevertheless, this disadvantage still appears to be outweighed by its benefits in uncovering clusters of better transparency when compared with the DM. A more important question, however, is whether both methods perform comparably in terms of the strength of identified relationships among industries within and outside the clusters. To answer this question some helpful descriptive statistics were computed for the intermediate deliveries matrix. The results are reported in table 2.

Due to the more careful way of assessing the significance of inter-industry relationships applied by the MDM, in that some of them are classified as external ones, it should be expected that the inter-cluster relationships will be stronger than those suggested by the DM. As one can see, this anticipation finds full confirmation from the figures in table 2. Mean inter-cluster flow under the DM equals over £20 million, whereas using the latter method it is a further £5 million greater. In addition to this, it turned out that the dispersion around average value measured by standard deviation for the MDM is approximately 60% of that obtained under the former method. On the other hand, for the same reasons we find that the average values of flows and their standard deviations between industries within clusters and outside them, as well as among industries beyond clusters, are both greater in the case of the MDM as compared to the DM.

Conclusions

This paper aims at improving methods designed to identify industry clusters by explicitly distinguishing between an internal and external relationship depending on whether two intertwined industries are grouped into the same or different clusters, or whether one of intertwined industries is classified outside the clusters. The most interesting finding of this study is that the MDM appears to produce a resultant cluster structure which is superior in some respects to that of the alternative method. In particular, the cluster structure under the MDM seems to be more transparent and more easily interpretable. Furthermore, as our experiment has shown, using this method does not necessarily entail worse performance in terms of the strength of identified relationships among industries within and outside the clusters. On the contrary, they even appear better.

We are, however, aware that the proposed method still leaves unsolved many other problems that one can encounter in investigating the industry clusters in a real-world economy. These include, for example, the so-called mini-cluster problem, and some inconvenience rooted in the fact that the choice of a suitable level of significance under the MDM may still be regarded as rather arbitrary. Perhaps further theoretical and empirical efforts in this field will help to overcome the common drawbacks of methods of identifying industry clusters, and contribute to reducing the extent of arbitrary decisions in these kind of analyses.

References

Ćmiel, A., and H. Gurgul. 2002. Application of maximum entropy principle in key sector analysis. *Systems Analysis Modelling Simulation* 42:1361–76.

Antonelli, C. 1999. *The microdynamics of technological change.* London: Routledge.

Bekele, G. W., and R. Jackson. 2006. Theoretical perspectives on industry clusters. Research Paper 2006-5, West Virginia University.

Gurgul, H., and P. Majdosz. 2005. Key sector analysis: A case of the transited Polish economy. *Managing Global Transitions* 3 (1): 95–111.

Hauknes, J. 1998. Norwegian input-output clusters and innovation patterns, STEP report R-15.

Hoen, A. 2002. Identifying linkages with a cluster-based methodology. *Economic Systems Research* 14 (2): 131–45.

Hoover, E. M. 1937. *Location theory and the shoe and leather industries.* Cambridge, MA: Harvard University Press.

Hoover, E. M. 1948. *The location of economic activity.* New York: McGraw Hill.

Krugman, P. 1991. *Geography and trade.* Cambridge, MA: MIT Press.

Krugman, P., and A. J. Venables. 1996. Integration, specialization, adjustment. *European Economic Review* 40:959–68.

Marshall, A. 1890. *Principles of economics.* London: Macmillan.

Martin, R., and P. Sunley. 2003. Deconstructing clusters: chaotic concept or policy panacea? *Journal of Economic Geography* 3 (1): 5–35.

Munroe, D. K., and G. J. D. Hewings. 2000. The role of intraindustry trade in interregional trade in the midwest of the US. Discussion Paper 99-T-7, University of Illinois.

Ohlin, B. 1933. *Interregional and international trade.* Cambridge, MA: Harvard University Press.

Porter, M. E. 1998. *On competition.* Boston: Harvard Business Review Press.