Control and Cybernetics

vol. **24** (1995) No. 4

Optimal industrial classification by threshold accepting[1]

by

John S. Chipman

Department of Economics
University of Minnesota
1035 Management and Economics
271 19th Avenue South
Minneapolis, MN 55455
U.S.A.

Peter Winker

Fakultät für Wirtschafts-
wissenschaften und Statistik
Universität Konstanz
Postfach 5560
78434 Konstanz
F.R.G.

In econometric models, the variables of economic theory are typically replaced by a much smaller set of aggregated variables, while the structure of the model remains unchanged. The manner in which the variables are partitioned into groups to be aggregated is usually based on intuition or convenience. In this paper we propose to carry this out in an optimal manner, the criterion being minimization of mean-square forecast error. This leads to an integer programming problem of high computational complexity. The optimization heuristic Threshold Accepting is implemented for the optimal partition and aggregation of a long monthly series of Swedish internal and external price indices. To correct for heteroskedasticity resulting from inflation, the sample variance matrix is assumed proportional to a diagonal matrix whose diagonal elements are the sums of squares of the external prices. This is compared with results obtained by using a scalar variance matrix and replacing Euclidean by Mahalanobis distance. The algorithm and the resulting groupings are presented.

Key words: Aggregation; integer programming; optimization heuristics; industrial classification.

1. Introduction

It is common practice in econometrics to base model specification on economic theory, yet to replace the large number of variables of the pure theory by a smaller set of aggregates of them. Such practice has often been unavoidable

1 Research was supported by a Humboldt Research Award for Senior U.S. Scientists and by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 178, "Internationalisierung der Wirtschaft" at the University of Konstanz. Thanks are due to Jane Ihrig for her preparation of the data set and to an anonymous referee for his comments.

owing either to the lack of sufficiently detailed data or to the computational difficulties of dealing with large-scale systems (or both). With the advance of the information revolution there is less and less reason for either of these considerations to apply. Yet, the practice remains, in large part probably because small models are easier to understand than large ones. But if one must aggregate, there are many alternative ways of doing so; however, until very recently the problem of aggregating or classifying variables in an optimal way has been totally intractable. In this paper we attempt to show a way to accomplish this. The method is applied to a particular problem, namely the study of the international transmission of price changes.

The basic idea of our approach is easily explained. One wishes to find a partition of industries into a certain number of groups so as to obtain the best possible prediction of the resulting indices of prices of the corresponding commodity groups within a country, given data on the corresponding indices of external (import and export) prices. The criterion for the optimal prediction is *mean-square forecast error*, denoted by ϕ, which is to be minimized. We assume a linear-homogeneous relationship between internal and external prices; for the theoretical justification of this specification we refer to Chipman and Winker (1994), where this approach was applied to German data. In that study the definition used of mean-square forecast error involved the concept of Mahalanobis distance; in the present study, which uses a comparable set of Swedish data, we introduce a modified definition of mean-square forecast error which uses Euclidean distance combined with a correction for heteroskedasticity as an alternative way of allowing for the effects of general inflation. As in the previous study, we also limit ourselves to the problem of optimally partitioning a set of medium-level categories (three- and some four-digit categories of the Swedish industrial classification system) into a specific number of groups, namely six (as corresponding to the number of groups in the official classification of the respective German price indices).

As observed above, the problem of finding a partition of a given number of industries into a smaller number of groups that minimizes mean-square forecast error falls under the heading of integer programming problems.

Simple enumeration is not feasible, since the number $P(m, m^*)$ of ways of partitioning m objects into m^* groups,

$$P(m, m^*) = \frac{1}{m^*!} \sum_{i=0}^{m^*} (-1)^i \binom{m^*}{i} (m^* - i)^m$$

(Chipman 1975, p. 151), is typically enormous. For the application to the Swedish price data analyzed in this paper this amounts to $P(25, 6) = 3.7026 \times 10^{16}$.

With regard to its computational complexity the problem is similar to problems such as the classic travelling salesman problem. In fact, it falls into the class of so-called NP-complete problems, see Winker (1992) for a proof, which

means that there is probably no exact optimization algorithm that works in a reasonable amount of computing time.

We by-pass this problem by the use of a "heuristic" combinatorial optimization algorithm, i.e., one that provides solutions sufficiently close to but not necessarily achieving the optimal value. The basic advantage of heuristic algorithms is their speed which allows for the calculation of approximative solutions even for large complex problems, when exact algorithms cannot give any solution at all in reasonable computing time. We use a refined local-search algorithm similar to the Simulated Annealing approach, (see Kirkpatrick et al., 1983 and Aarts and Korst, 1989), which is known as the Threshold Accepting algorithm. This algorithm was introduced by Dueck and Scheuer (1991) for the travelling salesman problem. Other successful implementations include integer knapsack problems (Dueck and Wirsching, 1991), or the identification of multivariate lag structures (Winker, 1995). See also Nissen and Paul (1995) for other applications.

In this paper we study a problem of optimal grouping of 25 industries or commodity categories into six sectors for the purpose of analyzing the international transmission of price changes. The internal Swedish producer-price indices of 25 commodity categories are put into relation with the corresponding indices of import and export prices. Following the procedures of other statistical agencies we generated a "pseudo-official" grouping into six sectors by stage of production. Using a TA implementation we have calculated other groupings that minimize the objective function ϕ.

The rest of the paper is organized as follows. The next section provides a short summary of the theory of approximate and optimal aggregation leading to the objective function for optimization. Special attention is paid to the problem of heteroskedasticity arising for price data due to inflationary processes. In this section the application to price indices for Sweden is also introduced. Section 3. is devoted to the heuristic optimization algorithm Threshold Accepting and Section 4. to the results achieved with the method of optimal aggregation for the problem of price indices. The paper concludes with a summary.

2. Optimal Aggregation

We may formulate the problem of optimal aggregation in terms of the multivariate multiple-regression model

$$Y = XB + E \tag{1}$$

where Y is an $n \times m$ matrix of n observations on m endogenous variables, X is an $n \times k$ matrix of n observations on k exogenous variables, B is a $k \times m$ matrix of unknown regression coefficients to be estimated, and E is a random $n \times m$ matrix of error terms with zero mean and covariance

$$\mathcal{E}\{(\text{col } E)(\text{col } E)'\} = \Sigma \otimes V, \tag{2}$$

where "col E" denotes the column vector of successive columns of E, Σ is the $m \times m$ simultaneous covariance matrix and V the $n \times n$ sample covariance matrix. $\mathcal{E}$ denotes the expectation operator. We shall assume that V is positive definite.

Letting G and H respectively denote $k \times k^*$ and $m \times m^*$ (proper) *grouping matrices*, i.e., matrices with exactly one nonzero (in fact, positive) element in each row and at least one nonzero element in each column, it is customary to deal with an aggregative model

$$Y^* = X^*B^* + E^* \tag{3}$$

mimicking the true one (cf. Theil (1954)), where

$$X^* = XG \quad \text{and} \quad Y^* = YH$$

are $n \times k^*$ and $n \times m^*$ matrices of observations on k^* and m^* aggregative exogenous and endogenous variables respectively. The situation may be depicted in the commutative diagram of Figure 1 as first done by Malinvaud (1956) (The meaning of the reverse mapping $G^{\#}$ appearing in the figure will be explained later, see equation (5) below).

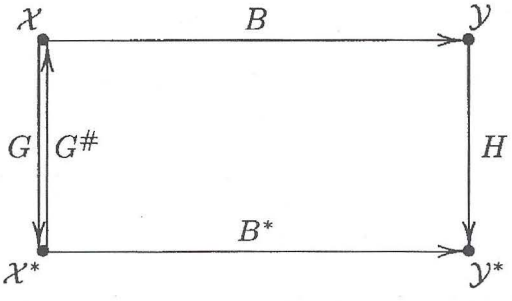

Figure 1. Commutative diagram for the aggregation problem

The case of perfect aggregation in which the diagram in Figure 1 actually commutes (i.e., $GB^* = BH$) cannot be expected to hold exactly for real data. Therefore, the approach of best approximate aggregation introduces a suitable measure of aggregation error. This is based on the discrepancy between the random variable $Y^* = YH$ to be forecast and its forecast by X^*B^* on the assumption that the model (3) is true, namely

$$Y^* - X^*B^* = (XB + E)H - XGB^* = X(BH - GB^*) + EH.$$

The *mean-square forecast error* is then defined in terms of this discrepancy as the matrix

$$\begin{aligned} F &= \mathcal{E}\{(Y^* - X^*B^*)'V^{-1}(Y^* - X^*B^*)|X\} \\ &= (BH - GB^*)'X'V^{-1}X(BH - GB^*) + nH'\Sigma H \end{aligned} \tag{4}$$

(cf. Chipman 1975, pp. 125-6). B^* is chosen in such a way as to minimize (4) for predefined modes of aggregation G and H. As shown in Chipman (1976, p. 668) this minimum is achieved by

$$B^* = G^{\#}BH \tag{5}$$

where

$$G^{\#} = (G'X'V^{-1}XG)^{-}G'X'V^{-1}X \tag{6}$$

and A^- denotes any matrix such that $AA^-A = A$. This reverse mapping (6) and the solution (5) are depicted in Figure 1; the diagram may be interpreted as commuting approximatively in the sense of minimizing the distance between GB^* and BH as defined by (4).

If G and H are not given, but are to be chosen optimally, the above minimization problem is ill-defined; a scalar measure of error is then required. Substituting the optimal solution (5) in (4) we obtain

$$F^* = H'B'(I - GG^{\#})'X'V^{-1}X(I - GG^{\#})BH + nH'\Sigma H. \tag{7}$$

Since the previous minimization problem is invariant with respect to replacement of F by $W^{*1/2}FW^{*1/2}$, where W^* is some $m^* \times m^*$ symmetric positive-definite matrix, we may perform this replacement and then take the trace of the resulting matrix, to obtain our criterion function

$$\phi = \text{tr } H'B'(I - GG^{\#})'X'V^{-1}X(I - GG^{\#})BHW^* + n \text{ tr } H'\Sigma HW^*. \tag{8}$$

For an alternative objective function the Reader is referred to Fisher (1962, 1969) and the discussion in Chipman (1976, pp. 665, 707-710).

In the computations to be reported below, we adopt two alternative choices for W^* and V:

(a) *Euclidean metric with correction for heteroskedasticity.*

$$W^* = I_{m^*} \quad \text{and} \quad V = \text{diag}\{XX'\} \tag{9}$$

This choice of V corrects for possible heteroskedasticity due to inflationary processes.

(b) *Mahalanobis metric with no correction for heteroskedasticity.*

$$W^* = (H'\Sigma H)^{-1} \quad \text{and} \quad V = I_n. \tag{10}$$

By weighting the variables inversely to their variability, use of the Mahalanobis distance neutralizes the effect that high variability of one variable might otherwise have on the objective function. In fact, it might partially correct for problems of heteroskedasticity. This was the procedure followed in Chipman and Winker (1994).

We calculate optimized groupings for both objective functions which are presented in Section 4.

Since B and Σ are unknown, we replace them by estimates. In the case (9) of Euclidean metric and correction for heteroskedasticity, we choose for B the generalized least-squares estimator

$$\tilde{B} = (X'V^{-1}X)^{-1}X'V^{-1}Y \tag{11}$$

and for Σ the pseudo-maximum-likelihood estimator

$$\hat{\Sigma} = n^{-1}(Y - X\tilde{B})'V^{-1}(Y - X\tilde{B}) = S/n \tag{12}$$

where

$$S = Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y. \tag{13}$$

Our estimate of ϕ then becomes

$$\hat{\phi} = \text{tr } H'\tilde{B}'(I - G\tilde{G}^{\#})'X'V^{-1}X(I - G\tilde{G}^{\#})\tilde{B}H + \text{tr}\{H'SH\}. \tag{14}$$

In the case (10) of Mahalanobis metric and no correction for heteroskedasticity, our estimates for B and Σ are just as in (11) and (12) but with V set equal to I_n, and our estimate of ϕ then becomes

$$\hat{\phi} = n \text{ tr } H'\tilde{B}'(I - G\tilde{G}^{\#})'X'X(I - G\tilde{G}^{\#})\tilde{B}H(H'SH)^{-1} + nm^*. \tag{15}$$

Dividing through by n we have

$$\tilde{\phi} = \tilde{\alpha} + m^* \tag{16}$$

where

$$\tilde{\alpha} = \text{tr } X(I - G\tilde{G}^{\#})\tilde{B}H(H'SH)^{-1}H'\tilde{B}'(I - G\tilde{G}^{\#})'X'. \tag{17}$$

Since m^* is constant in our applications (namely $m^* = 6$), we shall use $\tilde{\alpha}$—which we refer to as the "aggregation bias"—as our criterion function in the Mahalanobis case.

In our particular application the problem is simplified because of its special structure. The $n \times k$ matrix X has the special form $X = [X_1, X_2]$ where X_1 and X_2 are $n \times m$ matrices of n consecutive monthly observations on import and export price indices of m commodity categories, respectively, and Y denotes the $n \times m$ matrix of internal producer prices for the same commodity categories. The regression model (1) then becomes

$$Y = XB + E = [X_1, X_2]\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} + E. \tag{18}$$

Defining H to be an $m \times m^*$ grouping matrix, we define the $k \times k^*$ grouping matrix G by

$$G = \begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix} = H \otimes I_2$$

where $k = 2m$ and $k^* = 2m^*$. The object is then to choose the optimal H out of the class of $m \times m^*$ proper grouping matrices.

As one aim of this paper is to assess the robustness of the outcomes of an earlier analysis for German price data in Chipman and Winker (1994), we use a similar set of price data for Sweden. Unpublished import, export, and domestic producer price-index data have been furnished by the Statistiska centralbyrån, Stockholm. The data set consists of monthly observations on import and export price indices (which are formed as weighted averages of prices with fixed weights) and internal producer-price indices (formed the same way), covering the period 1971-1992. Since the natural way to group them is by forming weighted averages with the given weights, it has been most convenient to work with the price indices multiplied by their weights. Then aggregation means just summation and the nonzero elements of the grouping matrices are all ones. We considered the series of $m = 25$ commodity categories to be aggregated into $m^* = 6$ groups. There exists no official method of grouping these 25 industries into a smaller number of groups for Sweden. Hence, in order to compare our results with those of our German study in which 37 commodity categories were partitioned into six groups according to the official classification system, we generated a "pseudo-official" grouping similar to the German one which essentially is a grouping by stage of production. We present this horizontal grouping together with the optimized groupings in Section 4..

3. Optimization

The use of simple enumeration algorithms is completely infeasible for the problem of optimal aggregation owing to the enormous number of proper grouping matrices. Furthermore, we are not aware of any standard software for integer programming problems of this magnitude with a non-linear objective function. As the problem is NP-complete, there exists no feasible deterministic algorithm giving the exact optimal solution with certainty using only a reasonable amount of computer resources.

Consequently, as in Chipman and Winker (1994) we use the refined local search heuristic Threshold Accepting to achieve good approximations to the global optimum, which we call "optimized groupings".

The pivotal step of TA like any other local search algorithm consists in comparing a given grouping H in the set of feasible groupings $\mathcal{H}$ with other groupings in a neighborhood $H \in \mathcal{U}(H) \subset \mathcal{H}$. The comparison is based on the objective function $\tilde{\phi}$.

A trivial local search algorithm accepts a new element in the neighbourhood if and only if it leads to a reduction of the objective function. Consequently, it shows a strict "down-hill" behavior which results in convergence to some local minimum, which in general is a rather poor local minimum. The mean performance of this algorithm is not satisfactory. The central idea of the Threshold Accepting approach is to accept a temporary worsening in order to escape such

local minima. Hence, these algorithms show a "hill-climbing" behavior (see Figure 3). For problems with known optimal solution this approach proved to give very good approximations to the global optimum.

A problem arising in all implementations of local search algorithms on integer sets is the notion of "neighborhoods" as it is not given in a standard manner as e.g. for Euclidean spaces (ε-spheres in $\mathbb{R}^n$). They should be defined in a way to secure that elements of a neighborhood $\mathcal{U}(H)$ are "close" to H. As the set of all grouping matrices defines a $\{0,1\}$ vector space, the set of proper grouping matrices being a subset, the Hamming distance, Hamming (1950), seems to be a natural and appropriate choice. For two grouping matrices $H = (h_{ij})$ and $\tilde{H} = (\tilde{h}_{ij})$ the Hamming distance d_H is defined by

$$d_H(H, \tilde{H}) = \sum_{i=1}^{m} \sum_{j=1}^{m^*} \mid h_{ij} - \tilde{h}_{ij} \mid . \tag{19}$$

Following the analysis and results of some simulations in Chipman and Winker (1994) we use a Hamming distance of 4 to define neighborhoods for our application on Swedish price data. It should be stressed that the performance of local search algorithms depends crucially on the choice of the local structure.

As the Threshold Accepting algorithm accepts a temporary worsening of the current solution during the iteration process, it has to impose some criterion on when to accept a randomly chosen element in the neighbourhood of the current solution. This task is fulfilled by the threshold values. During the optimization procedure, the threshold values decrease to zero. They describe up to what amount a worsening of the objective function will be accepted when moving from the current solution to a new element in the neighborhood. For example, a threshold factor of 2 per cent means that a new element in the neighborhood of a current solution will be accepted as the new current solution, if the corresponding value of the objective function is not higher than 1.02 times the value of the old current solution.

In the paper of Dueck and Scheuer (1990) introducing the Threshold Accepting algorithm for a large scale travelling salesman problem the threshold sequence is exogenously given. The paper of Nissen and Paul (1995) represents a first step in letting the choice of these parameters be based on the data. The approach followed here may be regarded as a second step, since the threshold sequence used to obtain the results presented in the next section was generated endogenously. It was created from an empirical distribution of local relative deviations. To that end a large number (this number depending on the total number of iterations) of proper grouping matrices H was chosen randomly. Then, an element of the neighborhood of H was chosen and the relative deviation of the objective function evaluated for the two grouping matrices calculated. The resulting absolute values of relative deviations were sorted in decreasing order and the lower 50 per cent were chosen as the threshold sequence.

However, both our experience and a small simulation study suggest that the choice of the threshold parameters is not too crucial for the mean performance of the algorithm as long as it falls in a reasonable range. For example, simple linear threshold sequences might give results of similar quality than the ones presented in the next section.

A last important parameter for the overall performance of TA is the total number of iterations, i.e. local exchange trials as described above. For each of I different threshold values J exchange trials are performed. A flow chart of the implementation of the Threshold Accepting algorithm for the problem of optimal aggregation can be found in Chipman and Winker (1994).

This implementation of TA guarantees that only a finite number of iterations is performed. For a reasonable choice of tuning parameters the algorithm stops at a local minimum with respect to the chosen neighborhood definition. Asymptotically, this local minimum will be the global one as proved by Althöfer and Koschnick (1991). Unfortunately, their proof is not constructive, but it allows for the conclusion that for every $\varepsilon > 0$ and every problem size, i.e. the dimension of the grouping matrices H, there exists a threshold sequence such that the probability of ending up in a global minimum is greater or equal to $1 - \varepsilon$. Of course, the necessary number of iterations will increase as ε goes to zero.

4. Optimized Groupings

The implementation of Threshold Accepting described in the previous section was used to obtain optimized groupings for the Swedish industrial classification system. The TA algorithm has been coded in FORTRAN77 using some ESSL-subroutines for matrix operations. The program was run on different IBM RS6000 workstations.

We recall that we considered a linear-homogeneous regression model for price indices multiplied by their weights. The grouping problem consists in the aggregation of time series for 25 commodity categories into only six groups per series (internal producer price, import price, export price).

The monthly data cover the period 1971-1992. The data for the years 1971-1979 are calculated on a 1968 base; subsequent data are calculated using weights from December of the previous year. These have all been linked to the 1968 series and expressed as 1968=100, multiplied by the 1968 weights (in millions of Swedish crowns).

Before turning to the optimized groupings, the "pseudo-official" grouping we used as a benchmark should be introduced. It is given by the following list, where the code numbers refer to the Swedish industrial classification *Svensk standard för näringsgrensindelning (SNI)*, which is a refinement of the United Nations Standard Industrial Classification of All Economic Activities (ISIC):

Agricultural, hunting, forestry, and fishery products

100 Agricultural, hunting, forestry, and fishery products

Mining and quarrying products	
200	Mining and quarrying products
Basic materials	
369	Other non-metallic mineral products
371	Iron and steel
372	Non-ferrous metals
351	Industrial chemicals
352	Other chemical products
353+354	Petroleum products, lubricating oils, asphalt & coal products
331	Sawn timber, plywood and other worked wood
355	Rubber products
340	Pulp, paper, paper products and printed matter
Capital goods	
3841	Ships and boats
384\3841	Transport equipment other than ships and boats
383	Electrical products
385	Instruments, photographic and optical goods
381	Fabricated metal products
3825	Office, computing and accounting machinery
382\3825	Machinery excluding office, computing and accounting machinery
Consumer goods	
361+362	Fine ceramics, glass and glassware
330	Wood products
356	Plastic products
323+324	Leather, leatherware and footwear
321	Textiles
322	Apparel
Food, beverages and tobacco	
310	Food, beverages and tobacco

For comparability with our previous study (cf. Chipman and Winker (1994)) we have grouped the 25 commodity categories in the above list according to six "stages of production" following the procedure of the German classification ststem. We shall refer to this as the "horizontal" grouping. As in the German case this grouping turns out to be far from optimal. For the Euclidean metric its yields the value $\hat{\phi} = 701.95$ for the mean-square forecast error and for the Mahalanobis metric the value $\tilde{\alpha} = \tilde{\phi} - m^* = 18.22$ for the aggregation bias. Both values are about three times as large as the results for optimized groupings.

Since in this paper we are especially concerned with the problem of heteroskedastic error terms let us start with the results for the Euclidean metric using the estimator (9) of V to correct for heteroskedasticity.

Figures 2 and 3 may give an idea about some properties of the optimization procedure for a run of the algorithm with only 20,000 iterations leading to the best grouping for the Euclidean distance presented below. In Figure 2 we try to illustrate how the local structure near the current solution changes as the algorithm approaches its final local minimum.

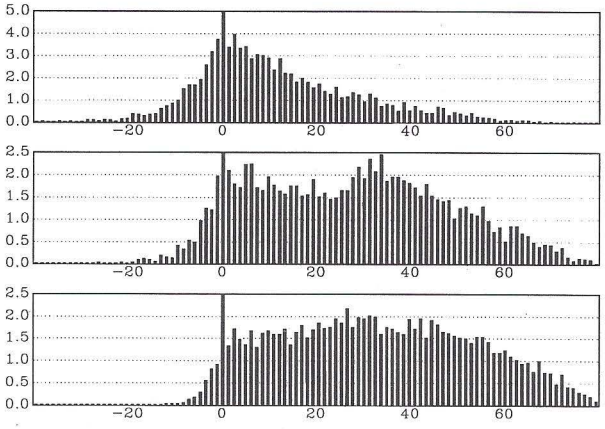

Figure 2. Local structure of aggregation problem

All plots show the empirical frequencies of the relative deviations of the objective function between the current solution of the algorithm and the tested elements in the neighborhood. The uppermost bar chart describes the distribution for the first 5,000 exchange trials, i.e. at the very beginning of the optimization. The distribution still has a significant weight for negative deviations, i.e. for possible improvements, whereas the lower charts for the exchange trials 5,001 to 10,000 and 10,001 to 15,000, respectively, demonstrate that this distribution shifts to the right as the algorithm proceeds. This means that it becomes more likely that the current solution is a local minimum with regard to the given neighborhood structure. For the last few hundred iterations the final local minimum will be achieved and the distribution of relative deviations has only positive weights for values greater or equal to zero.

Figure 3 gives some other interesting insights in the resulting sequence of values for the objective function $\tilde{\phi}$ during the optimization process. In the beginning of the optimization the algorithm accepts a new current solution nearly in every iteration resulting in the very volatile behavior in the upper part of the figure. As the optimization proceeds further the current solutions

become more stable. In particular in the lower part of the plot showing the values of $\tilde{\phi}$ for the current solutions for the iteration steps 10,001 to 15,000 the typical "hill-climbing" behavior of TA can be detected, i.e. in order to achieve a better current solution it proves to be necessary to admit a worsening of the solution first to escape local minima.

Let us indicate that one run with 20,000 iterations used about 1,600 CPU-seconds on the fastest of the workstations we used which was an IBM RS 6000/360. It should be noted that one attempt to achieve an optimized grouping consists in general of several (10) trials with different initializations of the random number generator.

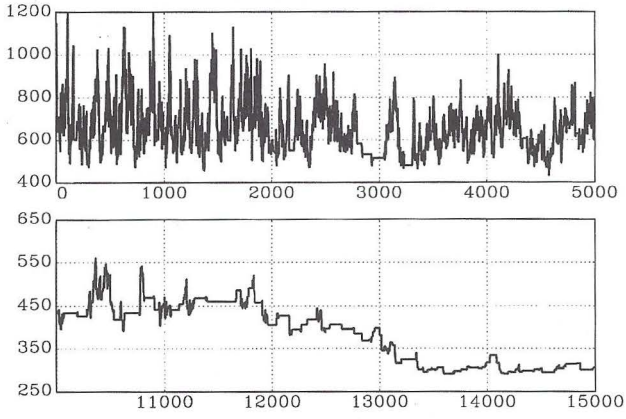

Figure 3. The Way to an Optimal Solution

The optimized grouping we can present for the Euclidean metric seems to be a very strong attractor in the underlying set of proper grouping matrices, as it was attained again and again for different random start groupings and different numbers of iterations ranging from 20,000 to 100,000. The value of the objective function $\tilde{\phi}$ for this grouping is only 279.02 as compared to the 701.95 of the pseudo-official grouping. Before discussing it in some detail we present

the optimized grouping:

Group 1	
100	Agricultural, hunting, forestry and fishery products
200	Mining and quarrying products
353-354	Petroleum products, lubricating oils, asphalt & coal products
3841	Ships and boats
384\3841	Transport equipment other than ships and boats
340	Pulp, paper, paper products and printed matter
Group 2	
371	Iron and steel
385	Instruments, photographic and optical goods
381	Fabricated metal products
382\3825	Machinery excluding office, computing and accounting machinery
356	Plastic products
323+324	Leather, leatherware and footwear
310	Food, beverages and tobacco
Group 3	
352	Other chemical products
3825	Office, computing and accounting machinery
330	Wood products
Group 4	
372	Non-ferrous metals
351	Industrial chemicals
355	Rubber products
Group 5	
369	Other non-metallic mineral products
331	Sawn timber, plywood and other worked wood
383	Electrical products
Group 6	
361+362	Fine ceramics, glass and glassware
321	Textiles
322	Apparel

Group 1 accounted for 31% of Sweden's imports and 40% of its exports in 1968. Crude petroleum accounted for 61% of category 200 imports in 1968 (and much more after the two oil shocks of 1973 and 1979); it is combined in Group 1 with petroleum products which accounted for 27% of Group-1 imports (compared with 10% for crude petroleum). The same clustering appeared for the German data in Chipman and Winker (1994). Category 100, which has a substantial forestry component (28% of its exports), is combined in this group with paper products which comprised 48% of Group-1 exports (compared with 2% from forestry). The two transport categories are also combined, which accounted for 30% of Group 1's imports and 35% of its exports.

Group 2 accounted for 30% of imports and 32% of exports in 1968. Categories 371, 381, and 382\3825, consisting of iron and steel and their products,

accounted for 60% of Group 2's imports and 87% of its exports. Again, we found the same grouping of iron and steel together with instruments and plastic products for the German data. While some of the combinations in Groups 3 to 5 are hard to explain, textiles and apparel are together in Group 6.

The following list presents the best grouping we could achieve for the Mahalanobis distance. However, it seems to be a weaker attractor than the grouping for the Euclidian metric as it has been attained only once for one out of ten trials. As one of the basic novelties of this paper is the use of the heteroskedasticity-corrected Euclidian metric, we concentrated our computing effort on this objective function. Consequently, the total number of attempts for the Mahalanobis metric was smaller. Nevertheless, the value of the Mahalanobis objective function $\tilde{\alpha}$ for this grouping amounts to 6.41 as compared to the 18.22 of the pseudo-official grouping.

Group 1

100	Agricultural, hunting, forestry and fishery products
200	Mining and quarrying products

Group 2

369	Other non-metallic mineral products
384\3841	Transport equipment other than ships and boats
383	Electrical products
361+362	Fine ceramics, glass and glassware
340	Pulp, paper, paper products and printed matter

Group 3

352	Other chemical products
355	Rubber products
385	Instruments, photographic and optical goods
382\3825	Machinery excluding office, computing and accounting machinery
310	Food, beverages and tobacco

Group 4

351	Industrial chemicals
353+354	Petroleum, lubricating oils, asphalt & coal products
331	Sawn timber, plywood and other worked wood
356	Plastic products
323+324	Leather, leatherware and footwear

Group 5

3825	Office, computing and accounting machinery
321	Textiles

Group 6

371	Iron and steel
372	Non ferrous metals
3841	Ships and boats
381	Fabricated metal products
330	Wood products
322	Apparel

In contrast to our previous German results, the two primary sectors (100 and 200) are combined together in Group 1. While the combination of petroleum products, industrial chemicals and plastic products in Group 4 and of metals and their products in Group 6 is convincing, on the whole the combinations are harder to explain than those obtained with the Euclidean metric with correction for heteroskedasticity. But for the reasons explained above, we feel less confident about the optimality of this grouping compared with the previous one.

As the data might be revised or the data sample might change, e.g. owing to new incoming observations, we are interested not only in the best groupings we could achieve with the methodology presented above, but as well in the robustness of these results. Will the groupings or even the minimal values of the objective function change dramatically for a small change in the data base?

A first result on robustness is quite easy to obtain and seems to be a very strong one, but unfortunately is less useful in practice. It draws on the discrete character of the set of proper grouping matrices and the continuity of our objective function. It can be concluded (Chipman and Winker, 1994, pp. 28ff). that a small enough change in the data will result in no change of the optimal grouping. However, we neither know how small is "small enough" nor do we know the global optimum with certainty.

A more practical approach uses a somewhat different understanding of the meaning of robustness. Here, we are interested in knowing whether a slight change in the data or in the parameters of the algorithm will lead to completely different outcomes with regard to the values of the objective function $\tilde{\phi}$ and to the main features of the resulting groupings.

To begin with the optimization parameters, we tried a huge bundle of different threshold sequences, used different numbers of iterations from 10,000 to 200,000 and many different initial values for the random number generator. The general impression is a negative correlation between the number of iterations and the achieved values for $\tilde{\phi}$, a rather weak influence of different forms for the threshold sequence – as long as the thresholds are not too small – and optimal values for $\tilde{\phi}$ nearly always in the same order of magnitude. The run with 10 trials leading to the optimal grouping for the Euclidean metric presented above gave a mean value of $\tilde{\phi}$ of 306.24 with standard deviation of 22.71.

Furthermore, all these "good" grouping matrices shared some patterns and the same tendency to "vertical grouping" as the best grouping presented above, except for Group 1 in the Mahalanobis case.

Finally, we can compare the groupings obtained by optimization with regard to different objective functions: Euclidean and Mahalanobis metric, respectively. The idea is also to calculate the value of the other objective function. Table 1 shows the resulting values for $\tilde{\phi}$ and $\tilde{\alpha}$.

Of course, the values in fields (1,1) and (2,2), respectively, must be the smallest in each column. Furthermore, a grouping optimized with regard to the Mahalanobis distance also tends to give low values for the Euclidean distance, whereas the difference in $\tilde{\alpha}$ is less clearcut between the grouping optimized for

Table 1. Comparison of Different Groupings

		$\tilde{\phi}$	$\tilde{\alpha}$
grouping	$\tilde{\phi}$	279.02	14.36
optimized for	$\tilde{\alpha}$	342.08	6.41
pseudo-official grouping		701.95	18.22

the Euclidean distance and the pseudo-official grouping. The same tendency can be found for other good groupings with regard to the two objective functions.

5. Conclusion

In this paper we have studied the problem of optimal aggregation of commodity categories for a Swedish data set. The aggregation criterion is based on the quality of the model in forecasting internal from external price indices.

The results obtained in this and an earlier paper using German data allow for the conclusion that the mode of classification matters. It can have a strong effect on the outcomes of econometric modelling. Furthermore, standard modes of industrial classification used by official agencies may be far from optimal for estimation and forecasting purposes. The use of a Threshold Accepting implementation has led to a considerable reduction in the value of the objective function as compared to some official groupings.

However, the economic interpretation of the "improved groupings" achieved by the use of optimization heuristics is not yet completely obvious. Although the results for the German and the Swedish data both exhibit many "vertical clusters"—groupings which take account of input-output relationships—some "horizontal clusters" persist and some clusters cannot be easily explained by intuitive reasoning.

The present study might be regarded as a second step in the exploration of the aggregation of price indices. The model is still extremely simple and does not yet allow for dynamic effects which might be captured by using distributed lags of the exogenous variables. However, the assumption of homoskedastic residuals which was implicit in Chipman and Winker (1994) has been relaxed by positing heteroskedastic residuals, with variances proportionate to the sums of squares of the external prices, and using Euclidean instead of Mahalanobis distance. The comparison of results achieved with the two objective functions show important differences of the resulting groupings. We plan to apply the methods presented in this paper to other data sets, in particular to Dutch price-index data, in

order to obtain further insights in the economic meaning and robustness of the resulting clusters.

References

AARTS, E., AND J. KORST (1989) *Simulated Annealing and Boltzmann Machines.* Chichester: John Wiley & Sons.

ALTHÖFER, I., AND K.-U. KOSCHNICK (1991) On the Convergence of 'Threshold Accepting'. *Applied Mathematics and Optimization*, 24, 183-195.

CHIPMAN, J. S. (1975) Optimal Aggregation in Large-Scale Econometric Models. *Sankhyā*, Series C, 37, 121-159.

CHIPMAN, J. S. (1976) Estimation and Aggregation in Econometrics: An Application of the Theory of Generalized Inverses, in *Generalized Inverses and Applications,* ed. by M. Z. Nashed. New York: Academic Press, 553-773.

CHIPMAN, J. S., AND P. WINKER (1994) Optimal Industrial Classification, Diskussionsbeiträge No. 236, Universität Konstanz.

DUECK, G., AND T. SCHEUER (1991) Threshold Accepting: A General Purpose Algorithm Appearing Superior to Simulated Annealing. *Journal of Computational Physics*, 90, 161-175.

DUECK, G., AND J. WIRSCHING (1991) Threshold Accepting Algorithms for 0-1 Knapsack Problems, in *Proceedings of the Fourth European Conference on Mathematics in Industry,* ed. by H. Wacker and W. Zulehner. Stuttgart: B. G. Teubner and Kluwer Academic Publishers, 255-262.

FISHER, W. D. (1962) Optimal Aggregation in Multi-Equation Prediction Models. *Econometrica*, 30, 774-769.

FISHER, W. D. (1969) *Clustering and Aggregation in Economics.* Baltimore: The Johns Hopkins Press.

HAMMING, R. W. (1950) Error Detecting and Error Correcting Codes. *Bell System Technical Journal*, 29, 147-160.

KIRKPATRICK, S., C. GELATT AND M. VECCHI (1983) Optimization by Simulated Annealing. *Science*, 220, 671-680.

MALINVAUD, E. (1956) L'agrégation dans les modèles économiques. *Cahiers du Séminaire d'Économétrie*, 4, 69-146.

NISSEN, V., AND H. PAUL (1995) A Modification of Threshold Accepting and its Application to the Quadratic Assignment Problem. *OR-Spektrum*, **17**, 205-210.

THEIL, H. (1954) *Linear Aggregation of Economic Relations.* Amsterdam: North-Holland.

WINKER, P. (1992) Some Notes on the Computational Complexity of Optimal Aggregation. Diskussionsbeiträge No. 184, Universität Konstanz.

WINKER, P. (1995) Identification of Multivariate AR-Models by Threshold Accepting. Computational Statistics & Data Analysis, **20**, 295-307.