

Chipman, John Somerset; Winker, Peter

Working Paper

Optimal aggregation by threshold accepting: An application to the German industrial classification system

Diskussionsbeiträge - Serie II, No. 180

Provided in Cooperation with:

Department of Economics, University of Konstanz

Suggested Citation: Chipman, John Somerset; Winker, Peter (1992) : Optimal aggregation by threshold accepting: An application to the German industrial classification system, Diskussionsbeiträge - Serie II, No. 180, Universität Konstanz, Sonderforschungsbereich 178 - Internationalisierung der Wirtschaft, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/101634>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

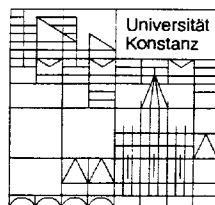
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Sonderforschungsbereich 178
„Internationalisierung der Wirtschaft“

Diskussionsbeiträge



Juristische
Fakultät

Fakultät für Wirtschafts-
wissenschaften und Statistik

John S/Chipman
Peter Winker

**Optimal Aggregation
by Threshold Accepting**

An Application to the
German Industrial Classification System

OPTIMAL AGGREGATION BY THRESHOLD ACCEPTING

An Application to the German Industrial Classification System*

John S. Chipman
University of Minnesota

Peter Winker
Universität Konstanz

Serie II - Nr. 180

Juni 1992

- * *Research was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 178, "Internationalisierung der Wirtschaft" at the University of Konstanz; National Science Foundation grant SES-8607652 at the University of Minnesota; and a Senior U.S. Scientist Award from the Humboldt Foundation. Research was conducted using the Cornell National Computer Facility, a resource of the Cornell Theory Center, which receives major funding from the National Science Foundation and the IBM Corporation, with additional support from New York State and Members of its Corporate Research Institute.*

Abstract

A widely used method in the analysis of complex econometric models is to replace the “true model” by a highly simplified aggregative one in which the variables are grouped and replaced by sums or weighted averages of the variables in each group.

The analysis of the problem of choosing an aggregative model optimally for modes of aggregation specified in advance leads to a formula for the aggregation bias. Taking this formula as objective function one would wish to choose a grouping that minimizes aggregation bias.

Unfortunately this results in an optimization problem of a high degree of complexity, which means that there is probably no exact optimization algorithm that works in economic computing time.

In the last few years however, many efficient multiple-purpose optimization heuristics have been developed for complex problems as the traveling salesman problem, optimal chip layout or optimal portfolio composition. One example of such an algorithm is the Threshold-Accepting Algorithm (TA). We implement TA for the optimal aggregation of price indices. The algorithm and the resulting groupings are presented. The results show that the use of standard or “official” modes of aggregation will in general be far from being optimal.

1 Introduction

A widely used method in the analysis of complex econometric models is to replace the “true model” by a highly simplified aggregative one in which the variables are grouped and replaced by sums or weighted averages of the variables in each group. These aggregative variables are put into relation with one another in a way that mimics the corresponding relation in the “true model”. Moreover, the aggregative model is generally treated as if the structural characteristics of the complex model carry over to it without change, enabling one to have—or to believe one has—an understanding of how the economy operates as seen through the model.

When there is no way to avoid this common practice, for example for the simple reason that the number of explanatory variables in the real data set exceeds by far the number of available observations, it should at least be carried out intelligently. Chipman (1976) gives an analysis of the problem of choosing an aggregative model optimally when the modes of aggregation are specified in advance. For an analysis of the problem of choosing the modes of aggregation optimally the reader is referred to Chipman (1975).¹

This analysis leads to a formula for aggregation bias for a given grouping of the real data. Given this objective function—which we shall denote by α in the sequel—one would wish to choose a grouping that minimizes aggregation bias.

Unfortunately this results in an optimization problem of high complexity. We believe that even the simple case we have used as an example is in the class of NP-complete problems,² which means that there is probably no exact optimization algorithm that works in economic computing time.³

In order to close this gap, we have studied the behavior of optimization heuristics for the problem of optimal grouping. In the last few years, many powerful multiple-purpose optimization heuristics have been developed for the needs in optimizing telephone nets, chip layout, job shop scheduling,

¹See Chipman (1975), p. 144.

²For discussion of NP-completeness (“NP” stands for “nondeterministic polynomial-time”) see for example Aho, Hopcroft and Ullman (1974, pp. 364ff), Garey and Johnson (1979), and Wilf (1986). In nontechnical language, an optimization problem is said to be NP-complete if the problem of finding the optimal solution with certainty is intractable. See Winker (1992) for a proof that a simplified grouping problem is NP-complete.

³By “economic computing time” we mean a “reasonable” consumption of computer resources, i.e., one that would today be feasible in terms of time and financial resources. We hesitate to specify a precise definition, but for example a time of 10^{26} years for the “correct” solution would certainly not be considered “reasonable.”

portfolio optimization, etc., as well of course as the classical travelling-salesman problem.⁴ Most of these real-life problems are also mathematically ugly and complex and do not fit into elegant mathematical models. Nevertheless, heuristic algorithms can overcome these difficulties. "Heuristic" means that these algorithms do not compute exact optima, but solutions sufficiently near to the optimal value. The basic advantage of heuristics is their velocity which makes it possible to find approximative solutions even for problems of a very high degree of complexity, when deterministic algorithms cannot give any solution at all in economic computing time.

A famous heuristic is the classical Simulated Annealing approach.⁵ In Dueck and Scheuer (1990), Dueck and Wirsching (1989) and Dueck and Winker (1990) an even more efficient form, the Threshold Accepting algorithm (TA), was introduced. TA is able to minimize almost any objective function out of almost any set of admissible solutions under almost arbitrary constraints.

In this paper we study a problem of optimal grouping of industries or commodity categories into sectors for the purpose of analyzing the international transmission of price changes. The internal German producer-price indices of 37 commodity categories are put into relation with the corresponding indices of import and export prices. The Statistisches Bundesamt, Wiesbaden, which issues these data, provides an official grouping of these 37 commodity categories into six sectors. Using a TA implementation we have calculated other groupings that minimize the objective function α in an adapted form.

Unfortunately, the objective function considered for this problem includes some matrix inversions. Thus, even with the TA algorithm we were restricted by computing time. Nevertheless, we can report some computational results which show in particular that the search for optimal aggregation turns out to give better regression results than the use of some "official" grouping. The resulting groupings are "vertical" as opposed to the official "horizontal" grouping by stages of production.

The rest of the paper is organized as follows. The next section provides an introduction to the theory of aggregation bias and optimal aggregation leading to the objective function for optimization. In Section 3 the application to price indices for the Federal Republic of Germany is introduced.

⁴See for example Kirkpatrick et al. (1983), Dueck and Scheuer (1990) and Dueck and Winker (1990).

⁵See Kirkpatrick et al. (1983) and Aarts and Korst (1989).

Section 4 is devoted to the heuristic optimization algorithm Threshold Accepting and Section 5 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

Following Chipman (1975) we may formulate the problem of optimal aggregation in terms of the multivariate multiple-regression model

$$(1) \quad Y = XB + E$$

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

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

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.⁶

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,⁷ it is customary to deal with an aggregative model

$$(3) \quad Y^* = X^*B^* + E^*$$

mimicking the true one, where

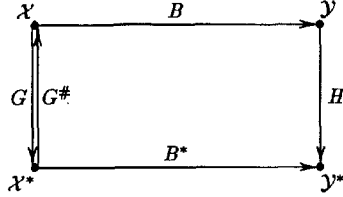
$$X^* = XG \quad \text{and} \quad Y^* = YH$$

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). We may consider three aggregation concepts in connection with this model:

⁶The more general case rank $V \leq n$ is treated in Chipman (1975).

⁷Cf. Chipman (1975), p. 135.

Figure 1: Commutative Diagram for the Aggregation Problem



1. *Perfect aggregation.* For the original detailed model (1) and the aggregative one (3) to be consistent with one another, one must have

$$(4) \quad XGB^* = \mathcal{E}^*Y^* = \mathcal{E}YH = XBH,$$

where \mathcal{E}^* denotes the expectation operator associated with the aggregative model. This can happen in two ways, as first observed by Theil (1954):

(a) *Structural similarity.* There exists a solution, B^* , to (4), for all X , hence to the equation

$$(5) \quad GB^* = BH.$$

Referring to Figure 1, this is the case in which the diagram commutes. Equation (5) is known in the literature as the “Hatanaka condition” (cf. Hatanaka, 1952). As shown in Chipman (1976, p. 720), a necessary and sufficient condition (following Penrose, 1955) for the solvability of (5) is that B should satisfy the bilinear restriction

$$(6) \quad (I - GG^-)BH = 0,$$

where G^- is any generalized inverse of G in the sense of Rao (1966), i.e., any matrix G^- satisfying $GG^-G = G$ (such a matrix always exists). This may also be written in the form $RBH = 0$ where R is an $r \times k$ matrix ($r = k - k^*$) whose rows form a basis for the row space of $I - GG^-$. If the nonzero elements of G are assumed to be ones, this restriction implies that B is partitioned into submatrices each of which has row sums equal to one another, these row sums being the respective elements of the matrix B^* . Under the restriction (6) it follows from Penrose’s theory that (5) has

$$(7) \quad B^* = G^-BH$$

as a solution.

In practice, the parameter matrices B and B^* are unknown, and one must deal with estimates. Denoting their generalized least-squares estimators by

$$\tilde{B} = X^\dagger Y, \quad \text{where} \quad X^\dagger = (X'V^{-1}X)^{-1}X'V^{-1}$$

and

$$\tilde{B}^* = X^{*\dagger}Y^*, \quad \text{where} \quad X^{*\dagger} = (X^{*'}V^{-1}X^*)^{-1}X^{*'}V^{-1}$$

respectively, it is shown in Chipman (1976, pp. 722-3) that the latter is related to the restricted generalized least-squares estimator of B subject to the bilinear restriction $RBH = 0$ —which is

$$\hat{B} = \tilde{B} - R^\dagger R \tilde{B} H H^\dagger$$

where

$$R^\dagger = (X'V^{-1}X)^{-1}R'[R(X'V^{-1}X)^{-1}R']^{-1} \quad \text{and} \quad H^\dagger = (H'\Sigma H)^{-1}H'\Sigma$$

—by a formula analogous to (7), namely

$$\tilde{B}^* = G^{-1} \hat{B} H.$$

(b) *Multicollinearity.* The domain, \mathcal{X} , of the mapping $B : \mathcal{X} \rightarrow \mathcal{Y}$ is restricted by

$$(8) \quad X = X^* \tilde{G} = X G \tilde{G},$$

where \tilde{G} is a $k^* \times k$ matrix such that $\tilde{G} G \tilde{G} = \tilde{G}$. An example of such a matrix \tilde{G} is

$$G_D^\dagger = (G' D G)^{-1} G' D,$$

where D is a diagonal matrix with positive diagonal elements. Then (8) has the interpretation given by Theil (1954, p. 32) that the “microvariables [are proportional to] the corresponding macrovariables.” For there to exist a B^* satisfying (4) for X satisfying (8) we require that there exist a solution, B^* , to

$$(9) \quad G \tilde{G} G B^* = G \tilde{G} B H.$$

This holds automatically, since one may choose $(G \tilde{G} G)^- = \tilde{G}$ and the Penrose solvability condition

$$[I - (G \tilde{G} G) \tilde{G}] G \tilde{G} B H = 0$$

is verified to hold.

It is shown in Chipman (1976, p. 726) that under (8) with $\bar{G} = G_D^+$ and R a basis for the row space of $I - GG_D^+$, a Gauss-Markoff estimator of B is given by

$$\hat{B} = (X'V^{-1}X + R'R)^{-1}X'V^{-1}Y$$

and the generalized least-squares estimator of B^* is related to it by a formula analogous to (7), namely

$$\tilde{B}^* = G_D^+ \hat{B} H.$$

(c) *Mixed cases.* There can be many cases of partially restricted structure complemented by partially restricted domain (cf. Chipman (1976, pp. 657-665, 726)).

2. *Best approximate aggregation.* Since perfect aggregation is an ideal situation that cannot be expected to be fulfilled in practice, the approach of best approximate aggregation is to define a suitable measure of aggregation bias and choose B^* in such a way as to minimize this bias. If the aggregation bias is zero, this approach reduces to the previous one.

As a measure of aggregation bias we may choose the nonnegative-definite symmetric matrix

$$(10) \quad A = (BH - GB^*)'X'V^{-1}X(BH - GB^*).$$

If either there exists a solution B^* to (5), or X satisfies (8) hence there exists a solution B^* to (9), then for such B^* , $A = 0$. Clearly there could be combinations of partial bilinear restrictions on B and partial restrictions on the domain of variation of X for which one would also have $A = 0$. Thus, best approximate aggregation includes perfect aggregation as a special case.

The matrices A may be ranked in terms of the nonnegative definiteness of their differences. It is shown in Chipman (1976, p. 668) that A is minimized with respect to B^* when

$$(11) \quad B^* = G^\# B H,$$

where $G^\#$ is any matrix satisfying

$$X'V^{-1}XGG^\#G = X'V^{-1}XG \quad \text{and} \quad X'V^{-1}XGG^\# = (X'V^{-1}XGG^\#)'$$

If, as may be expected in practice, the matrix $G'X'V^{-1}XG$ has full rank k^* , we have

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

Writing this in the form

$$G^\# = (X^{**}V^{-1}X^*)^{-1}X^{**}V^{-1}X = X^{*+}X,$$

it has the interpretation given by Theil (1954, p. 65) as the “auxiliary least-squares regression equations” of the microvariables on the macrovariables. In Figure 1 one may read off (11) as the composition of the mapping B^* into the three mappings shown.

Noting that $X^{*\dagger}XX^\dagger = X^{*\dagger}$,⁸ it follows that once again a formula analogous to (7) holds for the estimated matrices, namely

$$\tilde{B}^* = G^* \tilde{B} H.$$

3. *Optimal aggregation.* In perfect aggregation and best approximate aggregation, the grouping mappings G and H are taken as given. In optimal aggregation, G and H are chosen optimally. For each pair $\langle G, H \rangle$ in a set \mathcal{G} one determines B^* so as to minimize the matrix (10) of aggregation bias, so as to obtain the minimizing bias matrix

$$(12) \quad A^* = H' B' (I - GG^*)' X' V^{-1} X (I - GG^*) B H.$$

The latter may then be used to determine G and H optimally. However, the problem of minimizing (12) with respect to G and H is ill posed: in general, there will not exist a minimizing A^* matrix. A scalar-valued objective function must therefore be chosen. Now, the problem of best approximate aggregation remains invariant with respect to replacement of A by $W^{*1/2} A W^{*1/2}$, where W^* is some symmetric positive-definite matrix. In general, therefore, one may choose as criterion function

$$(13) \quad \alpha = \text{tr } H' B' (I - GG^*)' X' V^{-1} X (I - GG^*) B H W^*$$

One may choose the Euclidean metric $W^* = I_m$; alternatively—and this is the option chosen in Chipman (1975)—one may use the “Mahalanobis distance” defined by the choice

$$(14) \quad W^* = (H' \Sigma H)^{-1}.$$

The main advantage of the Mahalanobis metric is that the measure of aggregation bias is independent of units of measurement in the following sense: Suppose that Σ is diagonal and that H is block diagonal with blocks equal to columns of ones. One of these blocks might correspond to summing microvariables measured in a unit of length, another to summing microvariables

⁸This follows from the fact that $X'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} = X'V^{-1}$ (cf., e.g., Chipman, 1976, p. 569).

measured in a unit of weight, etc. In the absence of any *a priori* criterion for comparing the importance of length and weight, the Mahalanobis metric weights them inversely to their variability, thus neutralizing the effect that high variability of a macrovariable might otherwise have on the aggregation bias. The Mahalanobis metric would then be unaffected by changes in the units of measurement, i.e., from grams to kilograms or from the avoirdupois to the metric system. On the other hand, in economic applications one might wish to assign subjective welfare weights to disparate aggregative variables such as employment or the price level; these could be reflected in the diagonal elements of a specified matrix W^* . Alternatively, the aggregative variables might already be measured in some natural common unit such as value in terms of a stable currency, in which case the Euclidean metric might be the most suitable; but this might give undue weight to commodity categories whose value is subject to considerable fluctuation.

It is doubtful whether use of the objective function (13) could be justified for sets of G and H matrices of different dimensions k^* and m^* , hence it will be assumed that these dimensions are given.⁹ In general, one could (in principle) follow a two-step procedure of optimizing over the set of $k \times k^*$ matrices G for each fixed $m \times m^*$ matrix H , then optimizing over the set of matrices H . In the application to be considered in the next section the problem is simplified by the fact that G is dependent upon H .

From the discussion of conditions for perfect and best approximate aggregation it is clear that the process of optimal aggregation selects grouping matrices G and H that will approximate the conditions for perfect aggregation as closely as possible. We do not know how compilers of commodity and industry classification systems in international agencies and national statistical offices select (in effect) these grouping matrices; but we suspect that the intuitive criteria used correspond more to structural similarity than to multicollinearity. This may help serve to explain the substantial departures reported in the following sections of this paper between official classification systems and those that would be adopted in accordance with the criterion

⁹In Chipman (1985) an estimation criterion was used to decide whether aggregation was desirable in the sense of providing "blown-up" aggregative least-squares estimates with lower mean-square error than the ordinary (unaggregated) least-squares estimates. In terms of this criterion one could choose the optimal degree of aggregation according to the dimensionality that provides the best blown-up estimates in terms of mean-square error. The aggregated models considered in that paper were, however, not chosen optimally; the approach could obviously be improved by combining it with the methods of the present paper. This is left for future research.

of optimal aggregation as specified in this section.

3 An Application to Price Indices

The most convenient data set available for a first implementation of Threshold Accepting for optimal aggregation 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). Since the natural way to group them is by forming weighted averages with given weights, it was most convenient to work with the price indices multiplied by their weights. Unpublished import and export price-index data of this type, called "Wertziffern," have been furnished by the Statistisches Bundesamt, Wiesbaden, for the Federal Republic of Germany.¹⁰

Then aggregation means just summation and the nonzero elements of the grouping matrices are all ones. We considered the series of 37 commodity groups to be aggregated into six groups. There exists an official method of grouping these 37 industries into six groups which makes it possible to compare our results with results based on the official grouping.¹¹

One problem with the available data set is that the price-index series come in blocks of time periods with different base years. We performed our calculations first with a data set which includes a total of 85 months from January 1976 to January 1983 having 1976 as base. Using 37 commodity categories, and thus 74 independent variables (import and export prices), leaves insufficiently many degrees of freedom for reliable estimates based on the inversions in the formula for α . Therefore we had to use generalized inverses based on singular-value decomposition in this case. Merging this series with the 1970-base series (starting in January 1970) results in a series of 157 observations. Of course, there are two natural ways to merge the series: one might calculate on the 1970- or the 1976-base. Fortunately,

¹⁰The published price-index data consist of these *Wertziffern* each divided by the weight of the respective commodity category, and then rounded to one digit after the decimal point. Because of the rounding error, accuracy is lost especially in the case of the most important (high-weight) commodity groups. In the case of the internal producer-price index, *Wertziffern* were not available, and the series used were the published price indices multiplied by their weights.

¹¹The classification system used is the *Güterverzeichnis für Produktionsstatistiken*, formerly known as the *Warenverzeichnis für die Industriestatistik*. The 37 industries are two- and three-digit categories (and because of lack of data in some mining categories, combinations of some three-digit categories) called *Güterzweige*, and the six groups of industries are called *Gütergruppen*.

groupings obtained by optimization using one of the two bases, i.e. with a low value of the objective function α for this base, turned out to have a low value of α for the other base, too. The results presented in the sequel have been achieved on the 1976-base, with the two series linked at January 1976.

From the point of view of economic theory, it would make sense to posit either a linear-homogeneous relationship between the internal and external prices, or a linear-nonhomogeneous (affine) relationship between their logarithms. The former would correspond to the case of a Leontief fixed-coefficients technology and the latter to a Cobb-Douglas technology. However, in the latter case the aggregate price indices would be geometric rather than arithmetic means; while this makes sense from the point of view of economic theory, statistical agencies by tradition publish price indices only as arithmetic means. Therefore we have used the linear-homogeneous model with no constant term as our regression model.¹²

The problem may be stated as follows. Let X_1 and X_2 denote $n \times m$ matrices of n consecutive monthly observations on import and export price indices ("Wertziffern") of m commodity categories, respectively, and let Y denote the $n \times m$ matrix of internal producer prices for the same commodity categories.¹³ Let $X = [X_1, X_2]$ denote the $n \times k$ matrix of observations on the $k = 2m$ independent variables. The regression model is then

$$(15) \quad Y = XB + E = [X_1, X_2] \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} + E,$$

where E is a random $n \times m$ matrix with zero mean and covariance

$$(16) \quad \mathcal{E}\{(\text{col}E)(\text{col}E)'\} = \Sigma \otimes I_n.$$

The natural aggregation process is quite simple. We define H to be an $m \times m^*$ grouping matrix¹⁴ where, say, $m = 37$ and $m^* = 6$ in this application. We define the $k \times k^*$ grouping matrix G by

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

¹²Such a model was employed in Chipman (1983) to which the reader may be referred for more details concerning the classification system.

¹³A few commodity categories, such as that of electricity, gas, central heating, and water, as well as watercraft and aircraft, are not represented in the import- and export-price-index series, and have therefore been omitted from the producer-price-index series.

¹⁴A *grouping matrix* is defined as a rectangular matrix containing at most one nonzero entry in each row, this entry being (in the present application) unity. A proper grouping matrix is a grouping matrix with exactly one unit element in each row and at least one unit element in each column. Cf. Chipman (1975), p. 135.

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

Defining $\tilde{B} = (X'X)^{-1}X'Y$ and

$$S = (Y - X\tilde{B})'(Y - X\tilde{B}) = Y'Y - Y'X(X'X)^{-1}X'Y$$

(from the given data set these can be computed once and for all, if necessary using generalized inverses based on singular-value decomposition¹⁵), the objective function to be minimized is

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

—in accordance with (4.40) on p. 144 of Chipman (1975), where S may be replaced by $\tilde{\Sigma} = S/(n - k)$ —where

$$\tilde{G}^\# = (G'X'XG)^{-1}G'X'X = (X''X^*)^{-1}X''X,$$

in accordance with (4.1) on p. 134 of Chipman (1975).

The above definition (17) of the objective function uses the Mahalanobis metric (14). As mentioned in the previous section, if the variables are measured in a natural common unit such as value in terms of a stable currency, the Euclidean metric would be the most reasonable one to use. In the present application, while the variables are measured in the same units (D-marks), the value of the currency erodes through time with inflation. This would best be handled by allowing for heteroskedasticity as in the general formula (2) instead of assuming homoskedasticity as in (16). We plan to do this in future work. With the specification (16), a distortion is introduced which is partially compensated for by using the Mahalanobis metric.

4 Optimization

It is shown in Winker (1992) that the problem of choosing a proper grouping matrix H minimizing the objective function

$$(18) \quad \text{tr}\{\tilde{B}H(H'SH)^{-1}H'\tilde{B}'\}$$

belongs to the class of NP-complete problems. Furthermore, it is argued that most likely the problem of optimal aggregation belongs to this class as

¹⁵I.e., using the generalized inverse X^\dagger in place of $(X'X)^{-1}X'$.

well. This means that the problem of finding the optimal solution with certainty is intractable. There is a nearly general consensus that no deterministic algorithm can give a solution to this problem without using computer resources, i.e. computing time or storage capacity, growing faster than every polynomial in the size of the problem.¹⁶

A way out of this dead end for practical applications of the theory of optimal aggregation is the use of optimization heuristics. These algorithms do not give the solution to an optimization problem with certainty but in general perform well in giving a good approximation to this solution. We use for the solution of the problem of optimal aggregation the multiple purpose algorithm Threshold Accepting (TA) as introduced in Dueck and Scheuer (1990) and Dueck (1989). TA is a descendant of the Simulated Annealing algorithm discussed in Kirkpatrick, Gelatt und Vecchi (1983). In many applications it turned out to be even superior to Simulated Annealing, i.e. it gave better results with less computation time. Successful implementations exist for the traveling salesman problem (Dueck and Scheuer (1990)), multiple constraint knapsack problems (Dueck and Wirsching (1991)), optimal portfolio selection (Dueck and Winker (1990)) and many other problems of high complexity.

TA like many other optimization heuristics is a local search algorithm. It acts on a given set of feasible solutions. The algorithm starts with an arbitrary element of this set. In each iteration an element is chosen randomly or in a deterministic way out of a predefined neighborhood of the current solution. Then, the value of the objective function is calculated for the new element. If it is *not much worse* then the old value the new element is accepted as current solution. Therefore, the performance of the algorithm depends crucially on the choice of the local neighborhoods, the levels which determine when to accept a new element as the current solution (threshold sequence) and the number of iterations of the algorithm. The basic structure of the algorithm is presented in figure 2.

In the case of optimal aggregation, the set of feasible solutions consists of all proper grouping matrices.¹⁷ Thus, a feasible solution is a matrix in $\{0,1\}^{m \times m^*}$ of full rank, and the set of feasible solutions is the union of subspaces of $\{0,1\}^{m \times m^*}$ of codimension 1 given by the full rank condition.

¹⁶This assumption is known as Cook's hypothesis.

¹⁷Of course, it is no problem to introduce additional constraints on this set. For example, one might wish to require that a certain set of commodity categories always be kept within the same group. Naturally, this would be tantamount to treating them as a single aggregated category.

Figure 2: TA Algorithm for minimization

```

choose an initial feasible solution
choose an initial THRESHOLD  $T > 0$ 
Opt: choose a new element in a neighborhood of
      the current solution
      compute  $\Delta E := \text{objective function}(\text{new element}) -$ 
                   $\text{objective function}(\text{current solution})$ 
      IF  $\Delta E < T$ 
        THEN current solution := new element
      IF too many iterations
        THEN change THRESHOLD  $T$ 
      IF the threshold is too low to promise further improvements
        THEN stop
GOTO Opt

```

Hence, any randomly generated proper grouping matrix¹⁸ can serve as an initial solution to the algorithm.

A much more important aspect for this local search optimization heuristic is the choice of the definition of neighborhoods. When will we regard two proper grouping matrices as “close” to each other? There are two trivial concepts of neighborhoods: the only neighbor of an element is the element itself, or every element of the set is a neighbor to all elements. The former concept would chain the algorithm to its initial solution not allowing any improvement, whereas for the latter the probability of finding an acceptable new element in this huge neighborhood is rather small requiring a tremendous number of iterations in order to achieve good approximations to an optimal solution. Figure 3 might give an idea of the structure of this trivial neighborhood. This figure has been obtained by randomly generating 500 proper grouping matrices H_i and plotting the corresponding values of the objective function $\alpha(H_i)$.¹⁹

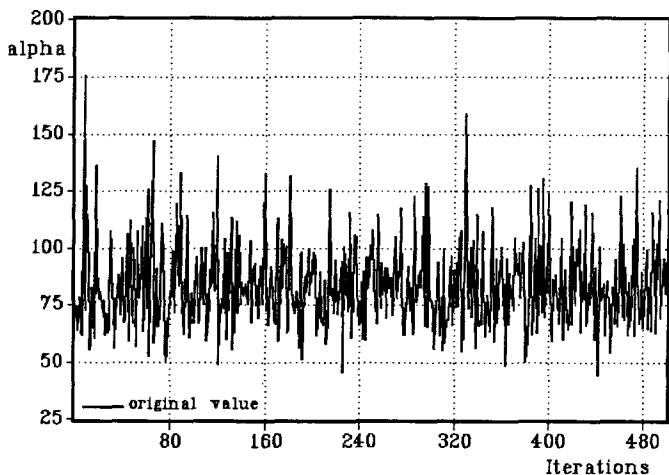
A well-known concept in the theory of $\{0, 1\}$ vector spaces is the Hamming distance introduced in Hamming (1950).²⁰ It seems natural and ap-

¹⁸In our application a proper grouping matrix is a matrix with exactly one unit element per row and at least one unit element in each column.

¹⁹See equation 13 for the definition of α .

²⁰For a definition in the context of information theory, the reader is referred to Yaglom

Figure 3: Randomly chosen grouping matrices



appropriate to use this metric for our purposes. The Hamming distance d_H between two grouping matrices $H = (h_{ij})$ and $\tilde{H} = (\tilde{h}_{ij})$ is given by the number of differing entries:

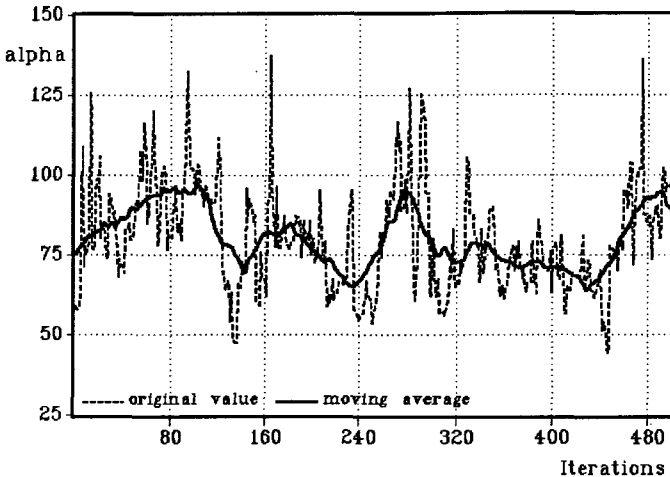
$$(19) \quad d_H(H, \tilde{H}) = \sum_{i=1}^m \sum_{j=1}^{m^*} |h_{ij} - \tilde{h}_{ij}|$$

As our set of feasible solutions is the set of proper grouping matrices, a Hamming distance of 2 between two elements of this sets means that exactly one commodity has moved from one group to another and a Hamming-distance of 4 that two commodities have moved to different groups. In our application the use of neighborhoods defined as spheres of radius 4 with regard to d_H proved to be a good choice. A heuristic argument for this choice is given by Figure 4. This figure is based on a simulation comparable to the one used for Figure 3. Starting with a randomly chosen initial grouping matrix, in each iteration an element of the sphere of radius 4 with regard to

and Yaglom (1983), p. 338.

d_H in the set of proper grouping matrices has been randomly selected and became the new current solution. As in Figure 3, for each grouping matrix the corresponding value of the objective function is plotted resulting in the dashed line. The solid line has been obtained as a moving average of order 20 of the actual solutions. While such a figure can only give a very coarse approximation of the local behavior of the objective function α on the set of proper grouping matrices, it exhibits the so-called “dune structure”²¹ of the optimization problem for this neighborhood definition. The applications of TA algorithms implemented so far show that TA implementations work well on problems which exhibit this dune structure for a given definition of local neighborhoods.

Figure 4: Dune structure



²¹This term was introduced by Dueck to describe the local behavior of an objective function as evidenced in Figure 4, i.e. when the values of the objective function for elements in a neighborhood show a common order of magnitude. He assumes that all problems exhibiting such a “dune structure” for properly defined neighborhoods are tractable by TA. A detailed analysis of the convergence properties of TA is given by Althöfer und Koschnik (1989).

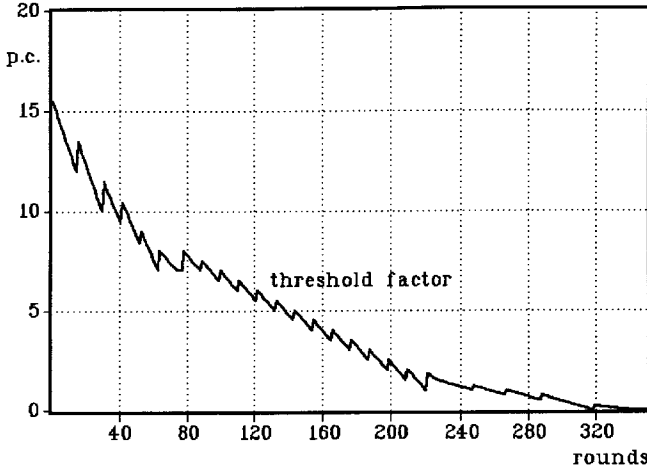
The difference between the Threshold Accepting algorithm and a trivial local search algorithm is that TA accepts a new element in a neighborhood of the current solution if the corresponding value of the objective function is *not much worse* than the old one. Obviously, if the algorithm accepts only “better” elements it will end in a local minimum nearly with certainty. For problems of high complexity such as the optimal aggregation problem this local minimum in general is far away from the global minimum. Therefore, the TA algorithm allows a deterioration of the objective function during the optimization procedure in order to escape such local minima. This ability to escape local minima can be seen in Figures 7 and 8 and might be described as “hill climbing”. During the optimization procedure, the threshold factor changes and decreases to zero at the very end. It describes 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 10 per cent means that every 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 1.1 times the value of the old current solution. Figure 5 shows the threshold sequence used to obtain the result presented in the next section.

The saw-tooth like sequence proved to be superior to just a linear threshold sequence in our application. In analogy to the Simulated Annealing algorithm the threshold sequence might be thought of as a “cooling schedule”. In the beginning, the system is assumed to be “hot”, there is no structure or preferred element in it. As the threshold is reduced during the optimization procedure the system gets “colder” and tends to more stable structures. Now, there are preferred elements as can be seen in Figure 8 on the straight segments of the plot when no new elements were accepted. The system stayed for a while in a local minimum. In the end, the threshold is reduced to zero and the system stops at the global minimum or a local minimum with a value of the objective function close to the global minimum. Indeed, applications of the TA algorithm to large travelling-salesman problems with a known global minimum have shown that it ends with a value in the range of 1 per cent from the global optimum.²²

We now give the specific TA implementation for optimal aggregation. It enables us to compute grouping matrices which result in a value of the objective function α close to the optimum. This may be done as well under almost any constraint on the grouping matrix. For example, the *a priori*

²²Cf. Dueck and Scheuer (1990).

Figure 5: Threshold sequence



knowledge of the model under consideration might lead to some restrictions on the grouping. Then TA calculates a close-to-optimum grouping matrix for the optimal aggregation problem with constraints.

The TA algorithm being described in the sequel is a local search or exchange algorithm. It starts with an arbitrary proper grouping matrix as current solution. In each iteration an element of the neighborhood of the current solution is randomly chosen. The neighborhoods are given by the spheres for the Hamming metric as described above. A more vivid idea of this specification is that one tries to remove two commodity categories out of their groups and into two randomly chosen new groups. Then, the value of the objective function is calculated for this element. The objective function used for the optimization of aggregation is the function $\tilde{\alpha}$ from equation (17). If the new value of the objective function is less than the one for the current solution times the threshold factor the new element becomes the current solution. If it is not, another element of the neighborhood of the current solution will be chosen. In each round of the algorithm this procedure is repeated *steps* times. The threshold factor evolves from one

Figure 6: TA Algorithm for Optimal Aggregation

```

choose an initial proper grouping matrix  $H$ 
choose an initial THRESHOLD  $T > 0$ 
FOR  $i := 1$  TO rounds DO
BEGIN
  FOR  $j := 1$  TO steps DO
  BEGIN
    UNTIL  $\text{rank}(\tilde{H}) = m^*$  DO
    BEGIN
      choose  $\tilde{H}$  with  $d_H(H, \tilde{H}) \leq 4$ 
    END
    compute  $\Delta E := \alpha(\tilde{H}) - \alpha(H)$ 
    IF  $\Delta E < T$  THEN  $H := \tilde{H}$ 
  ENDDO
  change THRESHOLD
ENDDO

```

round to the next. The algorithm that was finally implemented is presented in Figure 6.

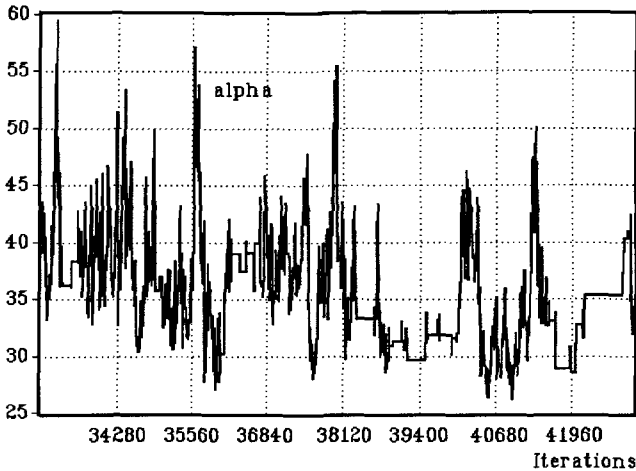
The parameters *rounds* and *steps* define parameters for the running time of the optimization, i.e. more “rounds” and more “steps” lead to a higher quality of the results. Once the initial parameters are fixed and a feasible H has been chosen the algorithm runs *rounds* rounds. In any round it chooses—*steps* times—two rows and two columns, and tries to bring the corresponding commodity categories into the so-defined new groups. The new grouping matrix is accepted if it is not *much* worse than the old in terms of the given objective function. The threshold is lowered in proportion to the improvement in the construction of the grouping and evolves as i increases.

5 Computational Results

Figures 3 and 4 in the last section tried to give some insight into the local features of our application to price indices. Due to the fact that the space of proper grouping matrices as a subset of $\{0, 1\}^{m \times m^*}$ can hardly be represented graphically this has been done by rather heuristic simulations. It is

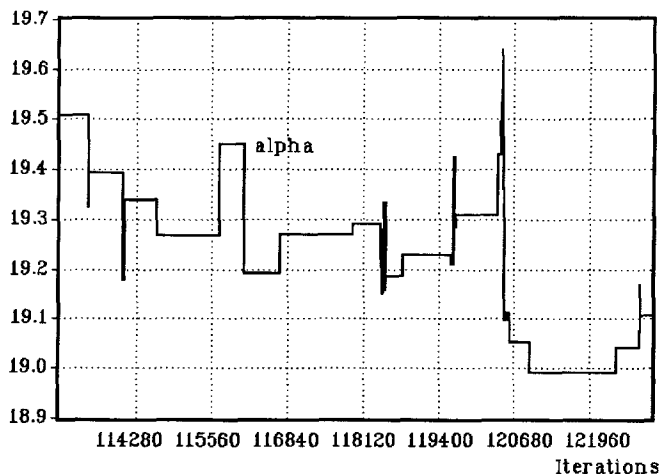
a task for further research to study more carefully these features as well as the behavior of the TA algorithm under these conditions.

Figure 7: The way to an optimal solution 1



In this section we shall restrict ourselves to the presentation of computational results achieved with a TA implementation on optimal aggregation. The TA algorithm introduced in Figure 6 has been coded in FORTRAN and was run on the IBM 3090 vector facility at the Cornell National Computer Facility. The optimized grouping matrix presented in the sequel has been achieved by nearly 200,000 iterations in 9603.78 seconds with the threshold sequence of Figure 5. Figures 7 and 8 show details of the resulting sequence of values for the objective function α for the current solution. In the beginning of the optimization the algorithm accepts a new current solution nearly in every iteration whereas as the optimization proceeds further the current solutions become more stable. Nevertheless, in both plots 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. It is a question of experience in working with TA algorithms to choose the parameters for the algorithm in a way

Figure 8: The way to an optimal solution 2



that results in a series of decreasing local optima which is assumed to lead to an optimum near to the global one if not to the global optimum itself.²³

Runs with a lower number of iterations resulted in slightly higher values for the objective function and different grouping matrices.²⁴ Nevertheless, all these “good” grouping matrices shared some patterns and the tendency to “vertical grouping” which will be discussed in the following.

We recall that we considered a linear-homogeneous regression model for price indices given by “Wertziffern”. The grouping problem comprises the aggregation of time series for 37 commodity groups into only six series per variable (internal producer price, import price, export price).

The official grouping as given by the publications of the Statistisches Bundesamt is presented in the following table.

²³ For a discussion of the choice of threshold or cooling sequences see Aarts and Korst (1989) and Althöfer and Koschnik (1991).

²⁴ The value of α rised about 3 per cent for a run with about 90,000 iterations and about 7 per cent for a run with only 10,000 iterations.

Agricultural, forestry, and fishery products

- Agricultural, forestry, and fishery products

Mining products

- Coal
- Crude oil and natural gas
- Other mining products

Basic materials

- Petroleum products
- Quarrying products
- Iron and steel
- Nonferrous metals
- Iron, steel and malleable cast iron products
- Nonferrous metal foundry products
- *Products of drawing and cold-rolling mills*
- Chemical products including nuclear fuel
- Sawn timber, plywood, and other worked wood
- Wood pulp, cellulose, paper and paperboard
- Rubber products

Capital goods

- Steelworking products
- Structural-steel products and rolling stock
- Machinery (including farm tractors)
- Road vehicles (excluding farm tractors)
- Electrical products
- Precision and optical goods, clocks and watches
- Ironware, sheet-metal ware, and hardware
- Office machinery and data-processing equipment

Consumer goods

- Musical instruments, toys, sporting goods, jewelry, film, etc.
- Fine ceramics
- Glass and glassware
- Wood products
- Paper and paperboard products
- Printed and duplicated matter
- Plastic products
- Leather
- Leatherware (including travelware)
- Footware
- Textiles
- Apparel

Food, beverages and tobacco

- Food and beverages
- Tobacco products

This grouping can be classified as a grouping by stage of production or "horizontal" grouping. As far as the regression problem, as described above, is concerned, this grouping is far from being optimal. The distance might be expressed in terms of the objective function α . For α this official grouping results in a value about four times the best value achieved by TA optimization. The superiority of the results achieved by optimal aggregation remains evident even when changing the underlying objective function. For example, when using the 1970 base instead of the 1976 base the value of α for the optimized²⁵ grouping is four times lower than for the official grouping.²⁶

In contrast to the official grouping, the optimized grouping can be described as a "vertical grouping". For a group of commodities it contains the products of the preceding stages of production rather than all the commodities at the same stage of production. Before discussing the resulting groups in detail we give the optimized grouping:

Coal

- Coal

Basic agricultural and forestry materials; precision goods

- Agricultural, forestry, and fishery products
- Sawn timber, plywood, and other worked wood
- Paper and paperboard products
- Rubber products
- Leatherware (including travelware)
- Precision and optical goods, clocks and watches
- Musical instruments, toys, sporting goods, jewelry, film, etc.

Chemical, metal, and basic consumer goods

- Crude oil and natural gas
- Petroleum products
- Chemical products including nuclear fuel
- Plastic products
- Iron and steel
- Nonferrous metals

²⁵ Remark that "optimized" in this context means with regard to the objective function α for the 1976 base.

²⁶ About the same holds for the Euclidian distance. However, there might arise some problems due to heteroskedasticity in the data. Hence, we will study this objective function and the resulting groupings in a subsequent paper.

- Steelworking products
 - Structural-steel products and rolling stock
 - Products of drawing and cold-rolling mills
 - Ironware, sheet-metal ware, and hardware
 - Electrical products
 - Textiles
 - Footware
 - Food and beverages
- Construction, housing, and other consumer goods
- Quarrying products
 - Iron, steel and malleable cast iron products
 - Wood products
 - Wood pulp, cellulose, paper and paperboard
 - Printed and duplicated matter
 - Glass and glassware
 - Apparel
 - Tobacco products

Machinery

- Other mining products
- Nonferrous metal foundry products
- Machinery (including farm tractors)
- Office machinery and data-processing equipment

Road vehicles and other products

- Road vehicles (excluding farm tractors)
- Leather
- Fine ceramics

The first group of our optimized grouping contains only one commodity, coal. This singular role of coal in our model of import, export and internal producer prices might be due to the fact that at least the internal producer prices for coal in the Federal Republic of Germany are determined more by political decisions than by market forces. Furthermore, the separate development in prices for coal and oil after the two OPEC oil price shocks makes it even more convincing to group coal on the one side and oil together with petroleum, chemical and plastic products on the other side in distinct groups.

Our second category includes materials and products from agriculture and forestry together with precision goods. It is surprising that this combination has a long tradition in the southwestern parts of Germany where the labor force shifted from agricultural and forestry work during summertime to small manufactories of clocks, musical instruments and optical goods for

the rest of the year. In 1983, the end of our data sample, 42 percent of the total turnover in precision goods, optical goods and clocks and 40 percent of the total turnover in musical instruments, toys etc. of the Federal Republic of Germany took place in the state of Baden-Württemberg, whereas its share in the total turnover of the manufacturing sector amounted to only 17 percent.²⁷ This result becomes even more striking when considering only the district "Schwarzwald-Baar-Kreis" in the Black Forest. This rather rural region accounted for 7 percent of all people working in the sector of precision goods, optical goods and clocks in 1982, whereas its share in the total labor force in mining and manufacturing amounted to only 0.5 percent.²⁸ In other terms, more than 30 percent of the labor force in mining and manufacturing in 1982 was in this sector for the "Schwarzwald-Baar-Kreis". The corresponding share for the Federal Republic came up to only 2 percent.

The third and by the number of included commodities largest groups contains chemical and plastic products with the initial products crude oil, natural gas and petroleum products. Moreover, it groups the sector of steel and metal products up to electrical products and finally, basic consumer goods like food and beverages, footwear and textiles. Of course, one might be tempted from an economic point of view to bring the last subgroup of basic consumer goods perhaps into the second or fourth group. However, not only do the statistical results advise the presented grouping, but to be fair, one should be aware of the fact that the grouping of 37 industries in only 6 groups cannot give a completely obvious solution. Nevertheless, the method of Optimal Aggregation by Threshold Accepting gives a grouping superior to the official one from the econometric point of view and at least as convincing in its economic interpretation.

The next group covers most of the commodities one would think of in the context of construction and housing from the basic materials of the quarrying industry and ingredients for reinforced concrete, to wood products like formwork, up to wood pulp and cellulose. Moreover, it includes other consumer goods like books, newspapers, glass and tobacco which are quite different from the basic consumer goods in the previous group.

Group 5 is perhaps the most homogeneous group besides the singular

²⁷Source: Statistisches Bundesamt Wiesbaden 1983, Fachserie 4 "Produzierendes Gewerbe", Reihe 4.1.4 "Beschäftigung und Umsatz der Betriebe in Bergbau und im Verarbeitenden Gewerbe nach Bundesländern".

²⁸Source: Statistisches Bundesamt Wiesbaden 1982, Fachserie 4 "Produzierendes Gewerbe", Reihe 4.1.3 "Regionale Verteilung der Betriebe im Bergbau und im Verarbeitenden Gewerbe und deren Beschäftigte".

case of coal. It comprises only machinery and office machinery with the primary products of nonferrous metals and other mining products. The last industry, other mining products includes not only nonferrous metals but silicon, too. Thus, we find all the basic materials for the growing sector of electronic data equipment in this group.

The last group includes the important automobile industry together with the leather industry which produces some primary products for the car manufactures, and fine ceramics which could be seen as a residual. We tried to move fine ceramics to group 4, "Construction, housing and other consumer goods". This resulted in a slight increase of the objective function. Therefore, we conclude that there is no close relation between the prices of the automobile sector and fine ceramics.

The value of α for the above optimized grouping amounts to 18.3088 in comparison to 72.9468 for the official grouping. While there is no proof that this result is optimal or at least nearly optimal, several attempts allow for a strong feeling that it is a good solution. In a randomly generated example with 10 variables to be aggregated into 3 groups, which means about 60,000 possible grouping matrices the TA implementation gave the real optimum after less than a thousand iterations. This is no proof of the optimality in our real data example, where the correct optimum cannot be calculated straightforwardly because there are about $6 \cdot 10^{28}$ feasible grouping matrices. However, it is much better than the official grouping in the value of α and for other objective functions. Furthermore, Figure 4 gives an impression of the order of magnitude of α which one might expect for randomly chosen grouping matrices.

6 Conclusion

It was not the purpose of this paper to judge the method of aggregation in large-scale econometric models. But considering the computational results for the example presented of price indices one might state that results of aggregative models might depend strongly on the chosen mode of aggregation. Hence, in using the common practice of aggregation, one should be aware of the possible consequences of some official modes of grouping. For linear regression models the methods presented in this paper open a way to better groupings of given variables.

The results for the application to price indices allow the conclusion that for the aggregation of these data a vertical grouping is more appropriate

than the official horizontal one.

References

- [1] E. Aarts and J. Korst. *Simulated Annealing and Boltzmann Machines*. John Wiley & Sons, Chisester 1989
- [2] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms*. Reading, Mass.: Addison-Wesley Publishing Company, 1974.
- [3] I. Althöfer and K.-U. Koschnick. "On the Convergence of 'Threshold Accepting'." *Applied Mathematics and Optimization*, 24 (September 1991), 183-195.
- [4] J. S. Chipman. "Optimal Aggregation in Large-Scale Econometric Models." *Sankhyā*, Series C, 37 (1975), 121-159.
- [5] J. S. Chipman. "Estimation and Aggregation in Econometrics: An Application of the Theory of Generalized Inverses." In: M. Z. Nashed, ed., *Generalized Inverses and Applications*. New York: Academic Press, 1976, pp. 553-773.
- [6] J. S. Chipman. "Dynamic Adjustment of Internal Prices to External Price Changes, Federal Republic of Germany, 1958-1979. An Application of Rank-Reduced Distributed-Lag Estimation by Spline Functions." In: W. Eichhorn, R. Henn, K. Neumann, and R. W. Shephard, eds., *Quantitative Studies on Production and Prices*. Würzburg: Physica-Verlag, 1983, pp. 195-230.
- [7] J. S. Chipman. "Testing for Reduction of Mean Square Error by Aggregation in Dynamic Econometric Models." In: P. R. Krishnaiah, ed., *Multivariate Analysis - VI*. Amsterdam: North-Holland, 1985, pp. 97-119.
- [8] G. Dueck. "New Optimization Heuristics, the Great Deluge Algorithm and the Record-to-Record Travel." Technical Report 89.06.011, IBM Heidelberg Scientific Center 1989.
- [9] G. Dueck and T. Scheuer. "Threshold Accepting: A General Purpose Algorithm Appearing Superior to Simulated Annealing." *Journal of Computational Physics*, 90 (1990), 161-175.

- [10] G. Dueck and P. Winker. "New Concepts and Algorithms for Portfolio Choice." *Applied Stochastic Models and Data Analysis*, (1992), forthcoming.
- [11] G. Dueck and J. Wirsching. "Threshold Accepting Algorithms for 0-1 Knapsack Problems." In: H. Wacker and W. Zulehner (eds.). "Proceedings of the Fourth European Conference on Mathematics in Industry." B. G. Teubner, Stuttgart, and Kluwer Academic Publishers, 1991, 255-262.
- [12] W. D. Fisher. "Optimal Aggregation in Multi-Equation Prediction Models." *Econometrica*, 30 (1962), 774-769.
- [13] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: W. H. Freeman, 1979.
- [14] R. W. Hamming. "Error Detecting and Error Correcting Codes." *Bell System Technical Journal*, 29 (1950), 147-160.
- [15] M. Hatanaka. "Note on Consolidation within a Leontief System." *Econometrica*, 20 (1952), 301-303.
- [16] S. Kirkpatrick, C. Gelatt and M. Vecchi. "Optimization by Simulated Annealing." *Science*, 220 (1983), 671-680.
- [17] E. Malinvaud. "L'agrégation dans les modèles économiques." *Cahiers du Séminaire d'Économétrie*, 4 (1956), 69-146.
- [18] R. Penrose. "A Generalized Inverse for Matrices." *Proceedings of the Cambridge Philosophical Society*, 51 (1955), 406-413.
- [19] C. R. Rao. "Generalized Inverse for Matrices and its Applications to Statistics." In: F. N. David, ed., *Festschrift for J. Neyman: Research Papers in Statistics*. New York: Wiley, 1966, pp. 263-279.
- [20] H. Theil. *Linear Aggregation of Economic Relations*. Amsterdam: North-Holland, 1954.
- [21] H. S. Wilf. *Algorithms and Complexity*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1986.

- [22] P. Winker. "Some Notes on the Computational Complexity of Optimal Aggregation." Unpublished Working Paper, Universität Konstanz, 1992.
- [23] A. M. Yaglom and I. M. Yaglom. *Probability and Information*. Dordrecht/Boston/Lancaster: D. Reidel Publishing Co., 1983.