

Chipman, John Somerset; Winker, Peter

Working Paper

Optimal industrial classification: [an application to the German industrial classification system]

Diskussionsbeiträge - Serie II, No. 236

Provided in Cooperation with:

Department of Economics, University of Konstanz

Suggested Citation: Chipman, John Somerset; Winker, Peter (1994) : Optimal industrial classification: [an application to the German industrial classification system], Diskussionsbeiträge - Serie II, No. 236, Universität Konstanz, Sonderforschungsbereich 178 - Internationalisierung der Wirtschaft, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/101775>

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:

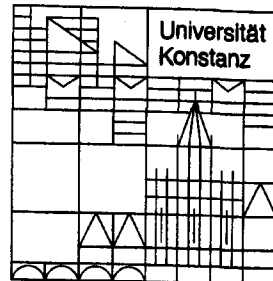
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

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 Industrial Classification

Optimal Industrial Classification

John S. Chipman

Peter Winker

Serie II - Nr. 236

Oktober 1994

Optimal Industrial Classification¹

An Application to the German Industrial Classification System

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.

A widely used method in the analysis of complex econometric models is to replace the "true model" by an 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 based on the mean-square forecast error. 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 such 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.

KEYWORDS: Aggregation; integer programming; optimization heuristics; industrial classification.

¹Research 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 Supercomputer 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.

1 Introduction

A widely used method in the analysis of complex econometric models is to replace the “true model” by an 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” (cf. Theil (1954), Malinvaud (1956), Fisher (1962, 1987)). 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. As Geweke (1985) has recently pointed out, the distortions introduced by the assumption of perfect aggregation—known as that of the “representative agent” in current macroeconomic models—may be of the same order of magnitude as the much-studied distortions introduced by ignoring expectations.

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, or alternatively, for the practical reason that only data on aggregates are available, it should at least be carried out intelligently. Two distinct problems arise: The first is that of choosing an aggregative model that *best approximates* the “true model” when the modes of aggregation are specified in advance; the second, which chiefly concerns us in this paper, is that of choosing the modes of aggregation optimally.²

In this paper we adopt an objective criterion of industrial classification along the following lines. We suppose, as Samuelson’s (1953) theory predicts, that within a country following fairly liberal trading policies, domestic price movements will closely follow movements in world prices, independently of consumer preferences. On the assumption of fixed technical coefficients, a linear-homogeneous multivariate multiple-regression model is postulated with the detailed average import and export prices as exogenous (independent) variables and the detailed average domestic prices of these same groups of commodities as endogenous (dependent) variables. The objective is to partition these industries into a smaller number of groups at a higher level of aggregation. Comparison of the aggregative endogenous variables with the conditional predictions of these variables from the aggregative

²For analyses of these two problems see Chipman (1976) and (1975) respectively.

model leads to a criterion of mean-square forecast error for a given grouping of the data. Given this objective function—which we shall denote by ϕ in the sequel—one wishes to choose a grouping that minimizes mean-square forecast error.

Of course, a classification system is designed to serve a wide variety of uses, and the criterion used in this paper refers to only one of these possible uses. However, most other uses that come to mind are closely related to this one; for example, one may wish to study the relations between quantities instead of prices. As Samuelson (1953) showed, the Jacobian of the “Stolper-Samuelson” mapping from commodity prices to factor rentals is simply the transpose of the Jacobian of the linear-homogeneous “Rybczynski” mapping from factor endowments to commodity outputs at given world prices. If one assumes identical homothetic preferences, consumer demands are linear-homogeneous functions of income—which is in turn a linear-homogeneous function of factor endowments; in this case the excesses of production over consumption, or net exports, are a linear-homogeneous function of a country’s factor endowments. It is this mapping that has recently been studied by Leamer (1990), who found that the nine one-digit SITC (Standard International Trade Classification) groupings of the 56 two-digit SITC categories formed a far-from-optimal classification. Remarking (p. 157) that the “calculation costs of a global minimization ... will ... be unacceptably high” he settled on a local optimization algorithm.³ In this paper we apply a heuristic global optimization algorithm.⁴

³ As pointed out in Leamer (1990, p. 157), the number of $m \times m^*$ proper grouping matrices for modestly large m^* is enormous. In fact, the restriction on exactly one nonzero entry per row and at least one per column leads to the following combinatorial expression for the number $P(m, m^*)$ of equivalence classes of $m \times m^*$ proper grouping matrices (considered as unordered sets of m^* column vectors each of order $m \times 1$), i.e., for the number of ways of partitioning m objects into m^* groups (cf. Chipman (1975, p. 150)):

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

For the application to the German price data analyzed in this paper this amounts to $P(37, 6) = 8.535 \times 10^{25}$. For Leamer’s application it is still higher, namely $P(56, 9) = 7.455 \times 10^{47}$.

⁴ It might be added that Leamer also makes strong assumptions that enable him to avoid having to identify and measure the “factors”; in effect, then, these are “factors” in the psychometric as well as economic sense, and each commodity grouping is associated with an implicit factor.

In this paper, by dealing with the composed mapping from external prices to factor

We limit ourselves to the problem of optimally partitioning a set of medium-level categories (two- and some three-digit categories) into a specific number of groups, namely six, corresponding to the official classification. It is obvious, however, that a complete solution of the problem of optimal industrial classification would entail derivation of a hierarchical classification system at many levels. An approach to this problem has been carried out by Cotterman and Peracchi (1992), who stress the importance of “consistency,” i.e., the requirement that categories once combined should not be broken up at a coarser level of aggregation. This could in principle be done in two ways: (1) One could take the number of groups at any level as given, and either (a) progressively reduce it and proceed sequentially as in Cotterman and Peracchi (in which case no optimization algorithm is needed, but due to the restrictions imposed by the results achieved on prior aggregation levels only a miniscule fraction of the sample space of all possible aggregation sequences is taken into account. Hence, one should not expect to obtain a very good hierarchical classification system from the statistical point of view), or (b) optimize the whole sequence, imposing the necessary restrictions on the grouping matrices (resulting in a tremendously large problem). In between (a) and (b) one could reduce the number of groups in larger steps. (2) One could as in Cotterman and Peracchi set up the objective function so that the number of groups at any level is an outcome of the problem. This is done by allowing comparison of mean-square forecast error at different levels of aggregation (i.e., different numbers of aggregated groups, m^*), and adding to the objective function a penalty for complexity, as in Akaike (1969, 1971, 1973), Hannan and Quinn (1979), and Schwarz (1978). In our formulation, however, it would be difficult to justify using the same measure to compare mean-square forecast error at different levels of aggregation; an alternative, statistical, criterion, is suggested in footnote 17 below. The main problem with deriving an optimal hierarchical classification system, however, is computational. Finding an optimal partition of 6561 three-digit categories into 81 two-digit groups is an immensely more complex problem than that of finding an optimal partition of 81 two-digit categories into 9 one-digit groups; in fact, it is simply infeasible given today’s resources.⁵ For these reasons we

rentals to internal prices, we also avoid the problem of having to find data on factor rentals, and our optimal groups could also be interpreted as representing implicit factors.

⁵In practice, the number of commodity groups represented at the d -digit level is considerably less than 9^d . However, the 1982 edition of the German commodity classification system (Statistisches Bundesamt (1981, p. VIII)) lists 6293 commodity groups at the 6-digit level, 1104 at the 4-digit level, 268 at the 3-digit level, and 41 at the 2-digit level

content ourselves in this paper with the more manageable problem of classifying a given number of industries into a specific, relatively small, number of groups.

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. With regard to its computational complexity the problem is similar to problems such as the classic traveling salesman problem. In fact, it falls into the class of so-called NP-complete problems,⁶ which means that there is probably no exact optimization algorithm that works in economic computing time.⁷

In order to close this gap, we pursue our problem by employing a heuristic integer optimization algorithm known as Threshold Accepting.⁸ This is a refined local-search algorithm similar to the Simulated Annealing approach.⁹ Successful implementations of Threshold Accepting include the classic traveling salesman problem, knapsack problems, portfolio optimization or the identification of multivariate lag structures.¹⁰ Most of these real-life problems are also mathematically ugly and complex and do not fit into elegant mathematical models. Nevertheless, heuristic algorithms like the Threshold Accepting algorithm 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

(where chemicals—category 40—are reckoned as a single 2-digit category). Optimally partitioning the 268 3-digit categories into 41 2-digit groups is already a huge problem, since $P(268, 41) = 4.761 \times 10^{382}$. Even optimally partitioning the 1104 4-digit categories into 268 3-digit groups is probably not feasible at the present time with our heuristic methods.

⁶For 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 the unrestricted problem of optimal aggregation is NP-complete.

⁷By "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^{28} years for the "correct" solution would certainly not be considered "reasonable."

⁸This algorithm was introduced by Dueck and Scheuer (1991) for the traveling salesman problem.

⁹See Kirkpatrick et al. (1983) and Aarts and Korst (1989).

¹⁰Cf. Dueck and Scheuer (1991), Dueck and Winker (1992), Winker (1994a,b).

give any solution at all in economic computing time.

In this paper we study a problem of optimal grouping of 37 industries or commodity categories into six 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 ϕ .

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 the official grouping. The resulting groupings tend to be "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 approximate 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. 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

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.¹¹

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

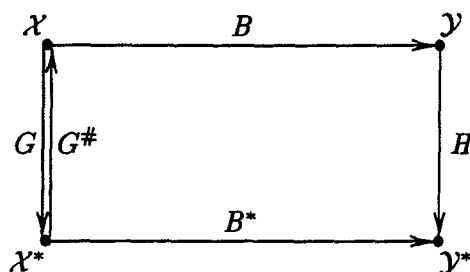
$$(3) \quad Y^* = X^* B^* + E^*$$

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).¹² We may consider three aggregation concepts in connection with this model:

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

$$(4) \quad XGB^* = \mathcal{E}^* Y^* = \mathcal{E} Y H = X B H,$$

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

¹¹The more general case $\text{rank } V \leq n$ is treated in Chipman (1975).

¹²The meaning of the reverse mapping $G^\#$ appearing in the figure will be explained later (see equation (12) below).

(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^* .¹³ This equality of row sums within each cluster ensures that the effect on the aggregate of the dependent variables in each cluster of a change in one of the independent variables in that cluster be the same as the effect of a change in any of the other independent variables in that cluster, hence we need only know the change in the sum of the independent variables in the cluster to determine the effect on the sum of the dependent variables in the cluster. Under the restriction (6) it follows from Penrose’s theory that (5) has

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

as a solution.

¹³These bilinear restrictions $RBH = 0$ on the multivariate model (1), which may be written in the form $(H' \otimes R)\text{col } B = 0$ as applied to the corresponding univariate model $\text{col } Y = (I \otimes X)\text{col } B + \text{col } E$ (cf. Chipman (1976, p. 714)), are of course capable of being statistically tested if sufficient detailed data are available. For such tests in the case of a related model in which the Kronecker product $I \otimes X$ is replaced by a block diagonal matrix whose i th diagonal element is X_i , see Zellner (1962), Pesaran, Pierse, and Kumar (1989), Pesaran and Pierse (1989), and Thompson and Lyon (1992). (For a discussion of the relation between these two models see Chipman (1975, pp. 155–7).) Alternatively, and more in the spirit of Grunfeld and Griliches (1960), one may wish to test the hypothesis that the imposition of the assumption of perfect aggregation, even if it is untrue, lowers the mean-square error of the estimates of the detailed structural coefficients; for this approach see Chipman (1985).

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

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

where \bar{G} is a $k^* \times k$ matrix such that $\bar{G} G \bar{G} = \bar{G}$.¹⁴ 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

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

This holds automatically, since one may choose $(G \bar{G} G)^- = \bar{G}$ and the Penrose solvability condition $[I - (G \bar{G} G) \bar{G}] G \bar{G} B H = 0$ is verified to hold.

(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 error and choose B^* in such a way as to minimize this error. If the aggregation error achieves its minimum possible value, this approach reduces to the previous one.

To arrive at a criterion of forecast error, we may consider the discrepancy between the random variable $Y^* = Y H$ to be forecast and its forecast by $X^* B^*$ on the assumption that the model (3) is true; this discrepancy is

$$Y^* - X^* B^* = (X B + E) H - X G B^* = X (B H - G B^*) + E H.$$

Alternatively, one might prefer to limit oneself to the discrepancy between the *expected value* $\mathcal{E}\{Y^*|X\} = X B H$ of Y^* and its forecast by $X^* B^*$, which is just $X (B H - G B^*)$. In terms of the latter criterion we may define the

¹⁴ An example of such a matrix \bar{G} is

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

where D is a diagonal matrix with positive diagonal elements.

matrix of *aggregation bias*¹⁵ as

$$(10) \quad \begin{aligned} A &= (\mathcal{E}\{Y^*|X\} - X^*B^*)'V^{-1}(\mathcal{E}\{Y^*|X\} - X^*B^*) \\ &= (BH - GB^*)'X'V^{-1}X(BH - GB^*). \end{aligned}$$

In terms of the former notion of discrepancy we define the *mean-square forecast error* as the matrix

$$(11) \quad F = \mathcal{E}\{(Y^* - X^*B^*)'V^{-1}(Y^* - X^*B^*)|X\} = A + nH'\Sigma H$$

(cf. Chipman (1975, pp. 125–6)).

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 F and A may be ranked in terms of the nonnegative definiteness of their differences. For fixed G and H , minimization of F is clearly equivalent to minimization of A . It is shown in Chipman (1976, p. 668) that A is minimized with respect to B^* for fixed G and H when

$$(12) \quad B^* = G^\# BH,$$

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

$$(13) \quad G^\# = (G'X'V^{-1}XG)^{-1}G'X'V^{-1}X.$$

Writing this in the form

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

it has the interpretation given by Theil (1954, p. 65) as the “auxiliary least-squares regression equations” of the microvariables on the macrovariables.

¹⁵The term “bias” here must not be confused with this term as used in the theory of statistical estimation. There can be “aggregation bias” even if B and B^* are perfectly known. We retain the terminology, however, since it has been traditional since Theil (1954).

In Figure 1 one may read off (12) as the composition of the mapping B^* into the three mappings shown.¹⁶

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 (11) of forecast error, resulting in the minimizing matrix

$$(14) \quad F^* = H'B'(I - GG^\#)'X'V^{-1}X(I - GG^\#)BH + nH'\Sigma H,$$

where the first term on the right is the minimizing bias matrix, A^* . This may then be used to determine G and H optimally. However, the problem of minimizing (14) with respect to G and H is ill posed: in general, there will not exist a minimizing F^* matrix. A scalar-valued objective function must therefore be chosen. Now, the problem of best approximate aggregation remains invariant with respect to replacement of F by $W^{*1/2}FW^{*1/2}$, where W^* is some symmetric positive-definite matrix. In general, therefore, one may choose as criterion function

$$(15) \quad \phi = \alpha + n \operatorname{tr} H'\Sigma HW^*.$$

where

$$(16) \quad \alpha = \operatorname{tr} H'B'(I - GG^\#)'X'V^{-1}X(I - GG^\#)BHW^*.$$

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

$$(17) \quad W^* = (H'\Sigma H)^{-1}.$$

With the latter choice, (15) reduces to $\phi = \alpha + nm^*$, and minimization of (15) is equivalent to minimization of (16) for fixed n and m^* . 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 measured

¹⁶Noting that $X^{*\dagger}XX^\dagger = X^{*\dagger}$, where $X^\dagger = (X'V^{-1}X)^{-1}X'V^{-1}$ and $\tilde{B} = X^\dagger Y$, etc., it follows that a formula analogous to (12) holds for the estimated matrices, namely $\tilde{B}^* = G^*\tilde{B}H$; thus, generalized least-squares estimation of the aggregative model provides best linear unbiased estimation of the best approximate aggregation.

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 mean-square forecast error. 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 (15) 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.¹⁷

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.¹⁸ We do not know how compilers of commodity

¹⁷In 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, e.g., assigning the coefficient of an aggregated independent variable to all its component disaggregated independent variables by a process of optimal disaggregation. 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. However, the computational problems would be enormous and this is left for future research.

¹⁸For example, if a subset of columns of X are highly collinear, it will tend to aggregate the corresponding variables together; alternatively, if X is well conditioned, it will tend to group variables together so that the corresponding submatrices of B have row sums which are as equal to each other as possible. These conditions are closely related to the

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 of optimal aggregation as specified in this section.

3 An Application to Price Indices

For our application we examine the structural relationship between commodity prices in a country's home markets and the corresponding world prices as represented by the prices (expressed in the country's own currency) of its imports and exports. By Samuelson's (1953) and Shephard's (1953) theories, the rentals of the factors of production employed in the country's export and import-competing industries are determined from the external prices by inverting the system of minimum-unit-cost functions dual to the production functions, while the prices of non-traded commodities are determined from the factor rentals directly via the corresponding minimum-unit-cost functions. This composed mapping from external to internal prices has been called the "generalized Stolper-Samuelson mapping" in Chipman (1978). If the production technology is of the Leontief fixed-coefficients type, then the minimum-unit cost functions are linear-homogeneous. This assumption has been chosen here in order to permit the application of the theory of linear aggregation, since published price indices have, since the time of Irving Fisher, been presented by statistical agencies as weighted arithmetic means.¹⁹

Our aggregation problem may therefore be formulated as follows. X_1 and X_2 denote $n \times m$ 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. Let $X = [X_1, X_2]$ denote the $n \times k$ matrix of observations on the $k = 2m$ independent variables. The regression model is

intuitive ideas about "similarity" of commodities and of processes of production.

¹⁹ An economically more reasonable assumption might have been that of a loglinear (Cobb-Douglas) technology, in which case the minimum-unit-cost functions are of the same loglinear type. This would be practicable if the price indices issued by statistical agencies were geometric means, as used to be the case in the time of Stanley Jevons. Then all the ensuing relations would be loglinear.

then

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

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

$$(19) \quad \mathcal{E}\{(\text{col}E)(\text{col}E)'\} = \Sigma \otimes I_n.$$

From the assumptions postulated, the matrix B depends entirely on the production coefficients in the country's industries.

The natural aggregation process is quite simple. We define 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 $k^* = 2m^*$. Now the object is to choose the optimal H out of the class of $m \times m^*$ proper grouping matrices.

Replacing B by its least-squares estimator $\tilde{B} = (X'X)^{-1}X'Y$ and Σ by the usual maximum-likelihood estimator S/n ,²⁰ where

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

(from the given data set these can be computed once and for all, if necessary using generalized inverses based on singular-value decomposition²¹), and setting $V = I$, the objective function to be minimized corresponding to the criterion of mean-square forecast error is, in accordance with (15) above, $\tilde{\phi} = \tilde{\alpha} + m^*$, where

$$(20) \quad \tilde{\alpha} = \text{tr}\{X(I - G\tilde{G}^\#)\tilde{B}H(H'SH)^{-1}H'\tilde{B}'(I - G\tilde{G}^\#)'X'\},$$

and

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

in accordance with (13) above. Since we take m^* as fixed, and use the Mahalanobis distance, the criteria of minimizing the mean-square forecast error $\tilde{\phi}$ and minimizing the aggregation bias $\tilde{\alpha}$ coincide. Henceforth, therefore, we use the latter as our optimization criterion.

²⁰If the usual best quadratic unbiased estimator $S/(n - k)$ is used instead, then in the formula for ϕ below the term m^* would be replaced by $m^*n/(n - k)$.

²¹I.e., using the generalized inverse X^\dagger in place of $(X'X)^{-1}X'$.

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 the 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.²²

Then aggregation means just summation and the nonzero elements of the grouping matrices are all ones. We considered the series of $m = 37$ commodity categories to be aggregated into $m^* = 6$ 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.²³

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 with a data set which includes a total of 85 months from January 1976 to January 1983 having 1976 as base, merged with a 1970-base series (starting in January 1970), resulting 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, 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.

The above definition (20) of the objective function uses the Mahalanobis

²²The 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.

²³The classification system used is the *Güterverzeichnis für Produktionsstatistiken*, formerly known as the *Warenverzeichnis für die Industriestatistik*. The 37 industries are two- and some 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*. A 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.

metric (17). 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 (19). We plan to do this in future work. With the specification (19), a distortion is introduced which is partially compensated for by using the Mahalanobis metric.

4 Optimization

As pointed out in footnote 3 above, the number of $m \times m^*$ proper grouping matrices for modestly large m^* is enormous. Hence, a simple enumeration algorithm is completely infeasible.

In Winker (1992) it was proved that the problem of optimal aggregation in its most general form, i.e. without restricting G to be equal to $H \otimes I_2$, is NP-complete.²⁴ There is a nearly general consensus that no deterministic algorithm can give an exact solution to this problem with certainty without using computer resources—i.e. computing time or storage capacity—that grow faster than every polynomial in the size of the problem.²⁵ Consequently, not only the trivial enumeration algorithm is infeasible for this problem, but there exists no feasible algorithm giving the exact optimal solution with certainty.

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 global optimum to a discrete optimization problem with certainty, but in general perform well in giving a good approximation to this optimum. For the solution of the problem of optimal aggregation we implemented the multiple-purpose optimization heuristic Threshold Accepting (TA) as introduced in Dueck and Scheuer (1990). 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.

²⁴For an exact definition and a discussion of NP-completeness see for example Aho, Hopcroft and Ullman (1974, pp. 364ff), Garey and Johnson (1979), Wilf (1986) and the other references from Winker (1992).

²⁵This assumption is known as Cook's hypothesis.

Successful implementations exist for the traveling salesman problem (Dueck and Scheuer (1990), Winker (1994b)), multiple constraint integer knapsack problems (Dueck and Wirsching (1991)), optimal portfolio selection (Dueck and Winker (1992)), lag structure identification in VAR models (Winker (1994a)) and many other problems of high computational complexity.

Like the more widely used Simulated Annealing algorithm and many other optimization heuristics, TA belongs to the class of refined local search algorithms for discrete optimization problems. The central idea of a local search algorithm is to compare a given element H in the set of possible solutions \mathcal{H} with other elements in a neighborhood $H \in \mathcal{U}(H) \subset \mathcal{H}$ with regard to an objective function. The local search algorithms differ in the rules for deciding when to accept a new element as the current solution.

For the standard local search algorithm, a new element is accepted if and only if it leads to a decrease in the objective function. Consequently, it shows a strict “down-hill” behavior which results in a convergence to some local minimum. For problems with many local minima such as the traveling salesman problem or the problem of optimal aggregation, the probability of being stuck in a “bad” local minimum is high and the mean performance of the algorithm is not satisfactory. The idea of more refined versions of local search algorithms such as the Simulated Annealing or the Threshold Accepting approach is to accept a temporary worsening in order to escape such local minima. Hence, these algorithms show a “hill-climbing” behavior²⁶ leading to good approximations to a global optimum.

For the problem of optimal aggregation the set of possible solutions is given by the proper grouping matrices, i.e. the matrices in $\{0, 1\}^{m \times m^*}$ of full rank.²⁷ Consequently, any randomly generated proper grouping matrix can serve as an initial solution to the algorithm. The objective function is given by the measure for the aggregation bias $\tilde{\alpha}$ as introduced in the previous section.

Finally, as the meaning of “neighborhoods” is not given in a standard manner as e.g. for euclidian spaces (ε -spheres in \mathbb{R}^n), they must be defined explicitly. They should be defined such that elements of a neighborhood $\mathcal{U}(H)$ are “close” to H . When will we regard two proper grouping matrices as “close” to each other? There are two trivial concepts of neighborhoods: the

²⁶ Cf. figure 5 in section 5.

²⁷ Of course, it is no problem to introduce additional restrictions based on prior information. Then, the set \mathcal{H} becomes smaller. Only for the case when the prior restrictions reduce \mathcal{H} to a very small set, the algorithm has to be adapted slightly as done in Dueck and Winker (1992).

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. The upper half of Figure 2 may give an idea of the structure of this trivial neighborhood. This figure has been obtained by randomly generating proper grouping matrices H and \tilde{H} in the trivial global neighborhood of H . For 5,000 random drawings the relative deviations in the values of the objective function $\tilde{\alpha}$ were calculated and the plot gives a histogram of their empirical distribution. The large dispersion of the relative deviations indicates that the probability of finding an acceptable new grouping in this neighborhood is rather small, since no really local structure is imposed.

A well-known concept in the theory of $\{0,1\}$ vector spaces is the Hamming distance introduced in Hamming (1950).²⁸ It seems natural and 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:²⁹

$$(21) \quad d_H(H, \tilde{H}) = \sum_{i=1}^m \sum_{j=1}^{m^*} |h_{ij} - \tilde{h}_{ij}|$$

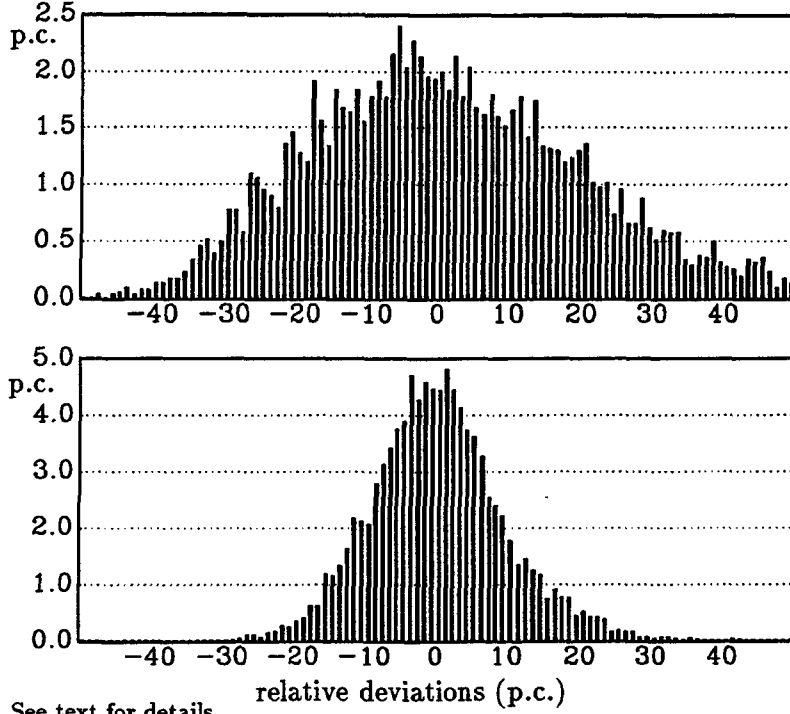
As our set of feasible solutions is the set of proper grouping matrices, two elements with a Hamming distance of 2 can be obtained for example by moving one commodity from one group to another. Likewise, two elements with a Hamming distance of 4 might be generated by simultaneously moving two commodities to different groups.³⁰

²⁸For a definition in the context of information theory, the reader is referred to Yaglom and Yaglom (1983), p. 338.

²⁹E. Ronchetti proposed to us a modification of this standard concept of Hamming distance by weighting the differing entries with their base-year weights. This modification did not lead to a very different local behavior. Hence, the application is based on the standard concept.

³⁰Note, however, that if two columns of an $m \times m^*$ grouping matrix are interchanged, the Hamming distance between the original and altered grouping matrix becomes $2m$, even though nothing has changed but the position or "name" of the group. This anomaly could be avoided by applying the concept of Hamming distance to equivalence classes of grouping matrices (regarded as unordered columns). However, as we use only small Hamming distances the differences in the topology are not important and the generation is faster using the standard notion.

Figure 2: Histograms of Relative Deviations^{a)}



a) See text for details.

The lower half of Figure 2 shows the empirical distribution of relative deviations obtained as above with the sole difference that now the matrices H were still randomly generated whereas \tilde{H} was chosen randomly in the sphere of radius 4 around H with regard to the Hamming distance. Comparing the two parts of the plot it is worth noting that a shrinking of the neighborhoods leads to a concentration of the empirical distribution of relative deviations around 0 per cent, i.e. to a more locally oriented behavior of the algorithm, but at the same time reduces the number of feasible moves in each iteration. Consequently, the risk of being stuck in a local minimum increases—at least for the standard local-search algorithm—with shrinking neighborhoods.

In our application the use of neighborhoods defined as spheres of radius 4 with regard to d_H proved to be a good choice, although the quality of the results did not decrease dramatically when choosing spheres of radii 2

or 6. Nevertheless, the performance of all applications of refined local search algorithms we know depends strongly on the definition of the local structure.

A further important set of parameters for the Threshold Accepting algorithm are the threshold values, i.e. the levels determining when to accept a new element as the current solution. 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 4 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 1.04 times the value of the old current solution.

Figure 3 shows the threshold sequence used to obtain the result presented in the next section. It was created from an empirical distribution of local relative deviations as in the lower part of Figure 2. Then, from the resulting absolute values of relative deviations sorted in decreasing order the lower 50 per cent were chosen as the threshold sequence. This presents a method for generating the threshold sequence automatically based on the characteristics of the data. The results obtained with this method were at least as good as results obtained with manually tuned sequences.

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 5 in the next section 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 traveling salesman problems with a known global minimum have shown that it ends with a value in the range of one percent from the global optimum.³¹

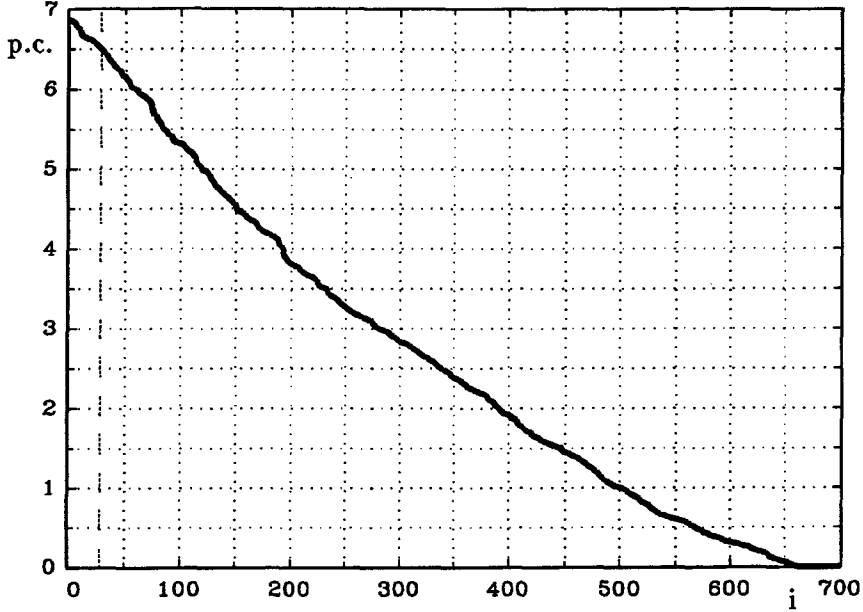
Finally, the number of iterations—given by $I \times J$ in the algorithm³²—has a positive influence on the quality of the obtained results though with a decreasing rate.³³ The flow chart of the implementation of the Threshold

³¹Cf. Dueck and Scheuer (1990), Winker (1994b).

³²For each of I different threshold values J exchange trials are performed. See figure 4.

³³Cf. the simulation performed for a large scale traveling salesman problem in Winker

Figure 3: Threshold sequence



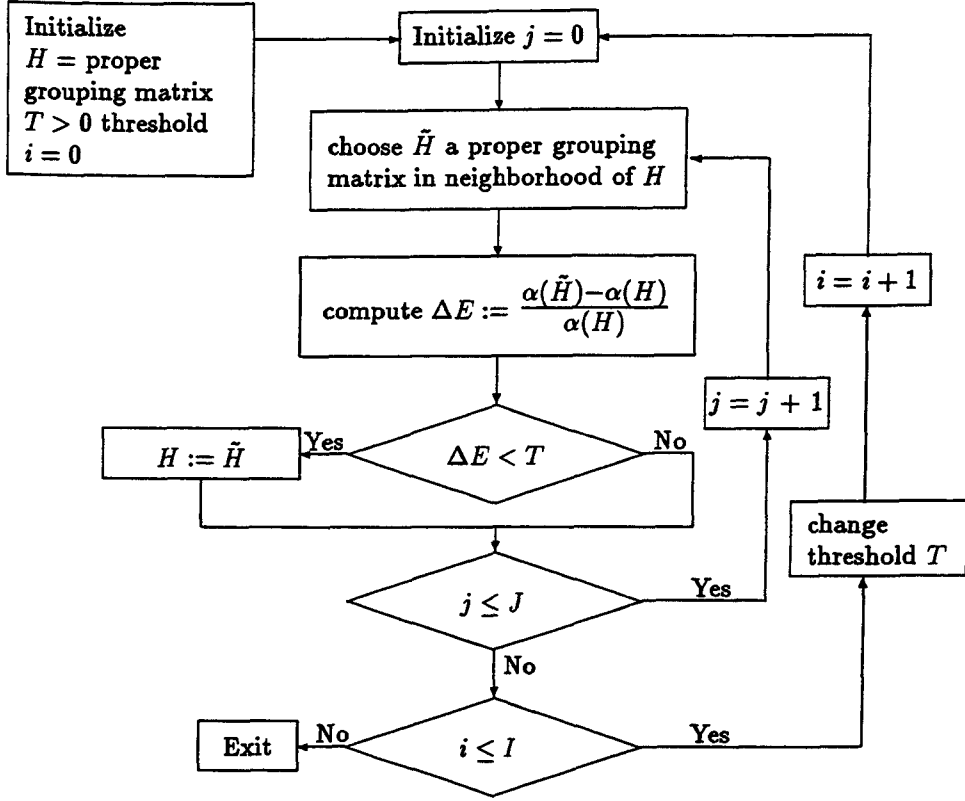
Accepting algorithm for the problem of optimal aggregation is given by figure 4.

Given this implementation, it can be guaranteed that the algorithm stops after a finite number of iterations. In general, the algorithm stops at a local minimum with respect to the chosen neighborhood definition. In the best case, this local minimum is also a global minimum. Althöfer and Koschnick (1991) proved some convergence results for TA based on similar results for the Simulated Annealing algorithm.³⁴ Although their proofs are not constructive they allow 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 (T_0, \dots, T_I) 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.

(1994b).

³⁴Cf. Aarts and Korst (1989), pp. 33ff.

Figure 4: Threshold Accepting Algorithm for Optimal Aggregation



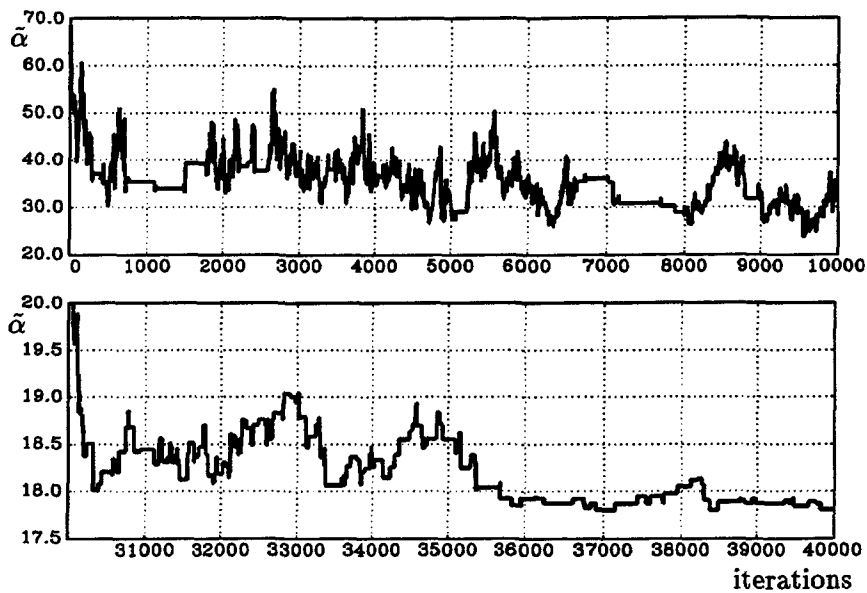
5 Computational Results

In this section some computational results achieved with the TA implementation for optimal aggregation are presented and some remarks on the robustness of these results are made. To begin with the computational results, the TA algorithm as given by the flow chart in figure 4 has been coded in FORTRAN using some ESSL-subroutines and was run first on the IBM 3090 vector facility at the Cornell National Supercomputer Facility and later on an IBM RS 6000/340 workstation at the University of Konstanz. The optimized grouping matrix presented in the sequel has been achieved by 50,000 iterations in about 5.000 CPU-seconds for one of 10 trials with the

threshold sequence presented in figure 3 above.

Figure 5 shows some details of the resulting sequence of values for the objective function $\tilde{\alpha}$ for the current solutions during the optimization process. 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. Furthermore, in both parts of the plot 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.

Figure 5: The Way to an Optimal Solution^{a)}



a) See text for details.

Although the threshold sequence used for the optimization was automatically generated by the data, it is still a question of experience in working with optimization heuristics to choose the parameters for the algorithm in a way that results in a sequence of decreasing local optima which may lead to a final local optimum near to the global one if not to the global optimum

itself.³⁵

We recall that we considered a *linear-homogeneous regression model* for price indices given by the “Wertziffern”. The grouping problem consists in the aggregation of time series for 37 commodity categories into only six groups per series (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.

Agricultural, forestry, and fishery products	
0	Agricultural, forestry, and fishery products
Mining products	
211	Coal
212	Crude oil and natural gas
—	Other mining products
Basic materials	
22	Petroleum products
25	Quarrying products
27	Iron and steel
28	Nonferrous metals
291	Iron, steel and malleable cast iron products
295	Nonferrous metal foundry products
301	Products of drawing and cold-rolling mills
40	Chemical products including nuclear fuel
53	Sawn timber, plywood, and other worked wood
55	Wood pulp, cellulose, paper and paperboard
59	Rubber products
Capital goods	
302	Steelworking products
31	Structural-steel products and rolling stock
32	Machinery (including farm tractors)
33	Road vehicles (excluding farm tractors)
36	Electrical products
37	Precision and optical goods, clocks and watches
38	Ironware, sheet-metal ware, and hardware
50	Office machinery and data-processing equipment
Consumer goods	
39	Musical instruments, toys, sporting goods, jewelry, film, etc.
51	Fine ceramics

³⁵For a discussion of the choice of threshold or cooling sequences see Aarts and Korst (1989) and Althöfer and Koschnik (1991). See also the results of the simulation in Winker (1994b).

52	Glass and glassware
54	Wood products
56	Paper and paperboard products
57	Printed and duplicated matter
58	Plastic products
61	Leather
621	Leatherware (including travelware)
625	Footware
63	Textiles
64	Apparel
Food, beverages and tobacco	
68	Food and beverages
69	Tobacco products

This grouping can be classified as a grouping by stage of production, or a "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 $\tilde{\alpha}$. For $\tilde{\alpha}$ 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 $\tilde{\alpha}$ for the optimized³⁶ grouping is four times lower than for the official grouping.

In contrast to the official grouping, most of the groupings obtained with the optimization procedure can be described as "vertical groupings". For a group of commodities they tend to contain the products of the preceeding stages of production rather than all the commodities at the same stage of production. Of course, for real data one should not expect to find a completely obvious grouping. Especially, it seems to be difficult to find smaller categories such as fine ceramics or leatherware in the same group for different optimization runs. Unfortunately, the best grouping we found with regard to the given criterion $\tilde{\alpha}$ is less clear cut than some of the other very good groupings not presented. Before discussing it in some detail we present the optimized grouping:

Group 1	
211	Coal
28	Nonferrous metals
31	Structural-steel products and rolling stock

³⁶Note that "optimized" in this context means with regard to the objective function $\tilde{\alpha}$ for the 1976 base.

37 Precision and optical goods, clocks and watches

56 Paper and paperboard products

Group 2

0 Agricultural, forestry, and fishery products

291 Iron, steel and malleable cast iron products

302 Steelworking products

39 Musical instruments, toys, sporting goods, jewelry, film, etc.

51 Fine ceramics

58 Plastic products

59 Rubber products

Group 3

212 Crude oil and natural gas

22 Petroleum products

27 Iron and steel

301 Products of drawing and cold-rolling mills

36 Electrical products

38 Ironware, sheet-metal ware, and hardware

40 Chemical products including nuclear fuel

53 Sawn timber, plywood, and other worked wood

63 Textiles

68 Food and beverages

Group 4

25 Quarrying products

52 Glass and glassware

54 Wood products

55 Wood pulp, cellulose, paper and paperboard

57 Printed and duplicated matter

64 Apparel

69 Tobacco products

Group 5

— Other mining products

295 Nonferrous metal foundry products

32 Machinery (including farm tractors)

50 Office machinery and data-processing equipment

621 Leatherware (including travelware)

Group 6

33 Road vehicles (excluding farm tractors)

61 Leather

625 Footware

In our optimized grouping coal is not in the same group as crude oil and natural gas. In our model of import, export and internal producer prices this can be explained by 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 and chemical products on the other side in distinct groups. The inclusion of the other commodities in the first group appears to be less convincing. It might be stated that in some groupings with a value of the aggregation bias $\bar{\alpha}$ only slightly higher than for the one presented, coal formed a grouping by itself.

We find similar difficulties in explaining the second cluster of our best grouping. However, the grouping of products from agriculture and forestry together with musical instruments and toys has a long tradition in the southwestern parts of Germany where the labor force shifted from agricultural and forestry work during summertime to small manufacturies of clocks, musical instruments and optical goods for the rest of the year. Consequently, we would have liked to see precision and optical goods, clocks and watches in the same group as is the case for a grouping with a slightly higher value of $\bar{\alpha}$.

The third group is the largest not only by the number of included commodities but also by its importance in German trade: in 1976 (the base year) it comprised 52.5 percent of German imports and 41.5 percent of German exports; in 1983 (the last year of the sample) the corresponding percentages were 45.0 and 41.2 (cf. Statistisches Bundesamt 1976, 1983). It combines chemicals and petroleum products with crude oil and natural gas, whereas in the official grouping the latter are in a separate mining category; these three commodity categories (212, 22, and 40) alone accounted for 25.3 and 34.1 percent of German imports and exports respectively in 1976. Group 3 also combines together another subgroup consisting of iron and steel, metal products and electrical goods, whereas in the official grouping iron and steel (27) and products of drawing and cold-rolling (301) mills are classified as "basic materials" while metal and electrical products (categories 38 and 36) are classified as "capital goods". Group 3 also includes textiles and food-stuffs and beverages. An interpretation is that 6 is too small a number of groups into which to partition the 37 commodity categories, and that as a result certain unrelated subgroups will necessarily be combined together into a single group.

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 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; it comprised 8.8 percent of total imports and 20.7 percent of total exports in 1976. Except for leatherware (which accounted for only 3.2 percent of the imports and 0.6 percent of the exports of this group in 1976), it comprises only machinery and office machinery and data-processing equipment with the primary products of nonferrous metals and other mining products. The last industry, other mining products, includes not only nonferrous metal ores 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 both for the car manufactures and the third industry in this group, footwear.

The aggregation bias $\tilde{\alpha}$ for the above optimized grouping amounts to 17.7517 compared with 72.9468 for the official grouping. While there is no proof that this result is optimal or at least nearly optimal, many attempts allow for the conclusion that it is a good solution. Furthermore, 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.

An interesting question concerns the robustness of the achieved groupings with regard to different data samples or base years for the weights, and to different random starting matrices and parameter sets for the TA implementation.

A first remark on the robustness of the real optimal grouping with regard to changes or errors in the data can be made from a theoretical standpoint. The optimal grouping H^* , i.e. the proper grouping matrix minimizing $\tilde{\alpha}$, is an element of the discrete space \mathcal{H} ,³⁷ and $\tilde{\alpha}$ can be thought of as a function

$$(22) \quad \mathcal{H} \times \mathcal{X} \longrightarrow \mathbb{R},$$

where \mathcal{X} denotes the space of data sets (X, Y) . $\tilde{\alpha}$ is uniformly continuous

³⁷As the exchange of two columns of a grouping matrix does not change the resulting grouping itself, we might assume \mathcal{H} to be given by the set of equivalence classes of proper grouping matrices with regard to this exchange operator.

in the second argument, as \mathcal{H} is finite. Consequently, in the general case

$$(23) \quad H \neq \tilde{H} \implies \tilde{\alpha}(H) \neq \tilde{\alpha}(\tilde{H})$$

and therefore

$$(24) \quad \delta \equiv \min_{H, \tilde{H} \in \mathcal{H}} |\tilde{\alpha}(H) - \tilde{\alpha}(\tilde{H})| > 0.$$

As $\tilde{\alpha}$ is continuous in (X, Y) there exists a small positive ε such that a perturbation of the data by less than ε in the euclidian norm will lead to a deviation in the values of $\tilde{\alpha}$ for any $H \in \mathcal{H}$ by less than $\delta/2$. Consequently, the resulting optimal grouping will remain the same for small perturbations or errors in the data.

Unfortunately, two aspects reduce the meaningfulness of this rather strong result on robustness. Firstly, a heuristic optimization algorithm does not behave in a completely deterministic manner and does not give the global optimum with certainty. Thus, a small change in the data might lead to a different outcome. Secondly, even if the algorithm behaves deterministically and always gives the global optimum, the order of magnitude of the admissible perturbations in the data, i.e. ε , remains unknown.

Therefore, we give some further empirical evidence on the robustness of the results achieved by the method of optimal aggregation with regard to 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 $\tilde{\alpha}$ 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 $\tilde{\alpha}$, a rather weak influence of different forms for the threshold sequence—as long as the thresholds are not too small—and optimal values for $\tilde{\alpha}$ nearly always in the order of magnitude of 20. The run with 10 trials leading to the optimal grouping presented above gave a mean value of $\tilde{\alpha}$ of 20.7471 with standard deviation of 2.2174.

Furthermore, all these “good” grouping matrices shared some patterns and the same tendency to “vertical grouping” as the best grouping presented above.

The price indices were supplied with the weights for two base years, namely 1970 and 1976. Hence, they could be linked using the weights of

one of these two years. The results presented above were obtained using data linked to the 1976 base. We also calculated $\tilde{\alpha}$ for this grouping using the data linked to the 1970 base. Furthermore, using the same optimization parameters as for the just-presented optimized grouping we obtained an optimized grouping for these data. Table 1 shows the resulting values for $\tilde{\alpha}$.

Table 1: $\tilde{\alpha}$ for different base years

		$\tilde{\alpha}$ calculated for	
		1970	1976
grouping	1970	17.4225	19.6763
optimized for	1976	18.5126	17.7517
official grouping		72.1548	72.9459

Of course, the values in fields (1,1) and (2,2), respectively, must be the smallest in each column. Nevertheless, it is worth noting that the differences are quite small compared to the results for the official grouping.

6 Conclusion

In this paper we have studied a particular aggregation problem, namely that of aggregating commodity categories into groups for the purpose of assessing and forecasting the impact of changes in external prices on the prices in a country's internal markets, the country chosen being Germany. However, we feel that it is appropriate to draw some general conclusions from the results obtained:

1. *Aggregation matters.* The process of aggregation, and the mode of aggregation chosen, can have a substantial impact on the results obtained in econometric research.

2. *Optimal aggregation is not trivial.* The problem of choosing an optimal mode of aggregation is far from trivial. Indeed, previous to the development of optimization heuristics it was intractible, and even with these methods it is still infeasible for very large sets and partitions.

3. *Standard methods of aggregation are far from optimal.* The modes of aggregation implied by official classification systems and the groupings provided by statistical agencies in presenting their data may be far from the optimal classification system needed for purposes of econometric estimation

and prediction.

4. *Optimization heuristics offer a way to better groupings.* The reduction in the value of the criterion function—mean-square forecast error, or aggregation bias, in our formulation—by this method can be very considerable.

5. *The economic meaning of “better groupings” is not yet completely obvious.* While we have detected a tendency for “vertical” groupings—groupings which take account of input-output relationships between industries—to outperform “horizontal” groupings—which group commodities by stage of production, certain commodity combinations which give better predictive results cannot be easily explained by intuitive reasoning.

We regard the present study as an initial exploration. We have kept the model itself extremely simple, but certain obvious refinements could be introduced which might improve the results. In particular, positing heteroskedastic residuals, with variances proportionate to the sums of squares of the external prices, and using Euclidean instead of Mahalanobis distance, may prove a better way to take account of the effects of inflation. This is planned in future work. Secondly, the assumption of simultaneous monthly causation from external to internal prices could be replaced—as was done in Chipman (1983, 1985)—by a specification of distributed lags up to twelve months. Finally, we plan to apply our methods to different data sets, in particular to Swedish and Dutch price-index data.

References

- AARTS, E., AND J. KORST (1989): *Simulated Annealing and Boltzmann Machines*. Chichester: John Wiley & Sons.
- AHO, A. V., J. E. HOPCROFT, AND J. D. ULLMAN (1974): *The Design and Analysis of Computer Algorithms*. Reading, Mass.: Addison-Wesley Publishing Company.
- AKAIKE, H. (1969): “Fitting Autoregressive Models for Prediction,” *Annals of the Institute of Statistical Mathematics*, 21, 243–247.
- AKAIKE, H. (1971): “Autoregressive Model Fitting for Control,” *Annals of the Institute of Statistical Mathematics*, 23, 163–180.
- AKAIKE, H. (1973): “Information Theory and an Extension of the Maximum Likelihood Principle,” in *Second International Symposium on*

- Information Theory*, ed. by B. N. Petrov and F. Csaki, Budapest: Akademiai Kiado, 267–281.
- 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. (1978): “Towards the Construction of an Optimal Aggregative Model of International Trade: West Germany, 1963–1975,” *Annals of Economic and Social Measurement*, 6, 535–554.
- CHIPMAN, J. S. (1983): “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 *Quantitative Studies on Production and Prices*, ed. by W. Eichhorn, R. Henn, K. Neumann, and R. W. Shephard. Würzburg: Physica-Verlag, 195–230.
- CHIPMAN, J. S. (1985): “Testing for Reduction of Mean-Square Error by Aggregation in Dynamic Econometric Models,” in *Multivariate Analysis—VI*, ed. by P. R. Krishnaiah. Amsterdam: North-Holland, 97–119.
- COTTERMAN, R., AND F. PERACCHI (1992): “Classification and Aggregation: An Application to Industrial Classification in CPS Data.” *Journal of Applied Econometrics*, 7, 31–51.
- DUECK, G. (1989): “New Optimization Heuristics, the Great Deluge Algorithm and the Record-to-Record Travel.” Technical Report 89.06.011, IBM Heidelberg Scientific Center.
- 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 P. WINKER (1992): "New Concepts and Algorithms for Portfolio Choice." *Applied Stochastic Models and Data Analysis*, 8, 159–178.
- 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. (1987): "Aggregation of Economic Relations," in *The New Palgrave: A Dictionary of Economics*, ed. by J. Eatwell, M. Milgate, and P. Newman. London: The Macmillan Press, and New York: The Stockton Press, Vol. 1, 52–53.
- GAREY, M. R., AND D. S. JOHNSON (1979): *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: W. H. Freeman.
- GEWEKE, J. (1985): "Macroeconometric Modeling and the Theory of the Representative Agent." *American Economic Association Papers and Proceedings*, 75 (May), 206–210.
- GRUNFELD, Y., AND Z. GRILICHES (1960): "Is Aggregation Necessarily Bad?" *Review of Economics and Statistics*, 42, 1–13.
- HAMMING, R. W. (1950): "Error Detecting and Error Correcting Codes." *Bell System Technical Journal*, 29, 147–160.
- HANNAN, E. J., AND B. G. QUINN (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society [B]*, 41, 190–195.
- HATANAKA, M. (1952): "Note on Consolidation within a Leontief System." *Econometrica*, 20, 301–303.
- KIRKPATRICK, S., C. GELATT AND M. VECCHI (1983): "Optimization by Simulated Annealing." *Science*, 220, 671–680.

- LEAMER, E. E. (1990): "Optimal Aggregation of Linear Net Export Systems," in *Disaggregation in Econometric Modelling*, ed. by T. Barker and M. H. Pesaran. London: Routledge, 150–170.
- MALINVAUD, E. (1956): "L'agrégation dans les modèles économiques." *Cahiers du Séminaire d'Économétrie*, 4, 69–146.
- PENROSE, R. (1955): "A Generalized Inverse for Matrices." *Proceedings of the Cambridge Philosophical Society*, 51, 406–413.
- PESARAN, M. H., AND R. G. PIERSE (1989): "A Proof of the Asymptotic Validity of a Test for Perfect Aggregation," *Economics Letters*, 30 (1989), 41–47.
- PESARAN, M. H., R. G. PIERSE, AND M. S. KUMAR (1989): "Econometric Analysis of Aggregation in the Context of Linear Prediction Models." *Econometrica*, 57 (July), 861–888.
- RAO, C. R. (1966): "Generalized Inverse for Matrices and its Applications to Statistics," in *Festschrift for J. Neyman: Research Papers in Statistics*, ed. by F. N. David. New York: Wiley, 263–279.
- SAMUELSON, P. A. (1953): "Prices of Factors and Goods in General Equilibrium." *Review of Economic Studies*, 21, 1–20.
- SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.
- SHEPARD, R. W. (1953): *Cost and Production Functions*. Princeton: Princeton University Press. Reprinted, Berlin-Heidelberg-New York: Springer-Verlag, 1981.
- STATISTISCHES BUNDESAMT, WIESBADEN (1976, 1983): *Außenhandel nach Ländern und Warengruppen der Industriestatistik (Spezialhandel)*. Stuttgart und Mainz: Verlag W. Kohlhammer.
- (1981): *Systematisches Güterverzeichnis für Produktionsstatistiken*, Ausgabe 1982. Stuttgart und Mainz: Verlag W. Kohlhammer GmbH.
- THEIL, H. (1954): *Linear Aggregation of Economic Relations*. Amsterdam: North-Holland.
- THOMPSON, G. D., AND C. C. LYON (1992): "A Generalized Test for Perfect Aggregation." *Economics Letters*, 40, 389–396.

- WILF, H. S. (1986): *Algorithms and Complexity*. Englewood Cliffs, N.J.: Prentice-Hall.
- WINKER, P. (1992): "Some Notes on the Computational Complexity of Optimal Aggregation." Diskussionsbeiträge No. 184, Universität Konstanz.
- WINKER, P. (1994a): "Identification of Multivariate AR-Models by Threshold Accepting." Computational Statistics & Data Analysis, forthcoming.
- WINKER, P. (1994b): "The Tuning of the Threshold Accepting Heuristic for Travelling Salesman Problems." Paper prepared for the OR 94, Berlin.
- YAGLOM, A. M., AND I. M. YAGLOM (1983): *Probability and Information*. Dordrecht/Boston/Lancaster: D. Reidel Publishing Co.
- ZELLNER, A. (1962): "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association*, 57, 348-368.