Optimal Industrial Classification[1]

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
Department of Economics
University of Mannheim
68131 Mannheim
F.R.G.

A widely used method in the analysis of large-scale 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 modes of aggregation of the independent and dependent variables may in principle be chosen optimally by minimizing a measure of mean-square forecast error in predicting the dependent variables from the independent variables by using the aggregative rather than detailed variables. However, this results in an optimization problem of a high degree of complexity. Nevertheless, many efficient optimization heuristics have been developed for these kinds of complex problems. We implement the Threshold Accepting heuristic for the problem of optimal aggregation of price indices in a model of the transmission of external (import and export) prices on internal prices, using German data. The algorithm and the resulting groupings are presented. The results suggest 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.

[1]Research was supported by the Deutsche Forschungsgemeinschaft, Long Term Research Project 178, 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.

1 Introduction

A widely used method in the analysis of large-sale 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)). Moreover, the aggregative model is generally treated as if the structural characteristics of the detailed 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 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, 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 model leads to a criterion of mean-square forecast error for a given grouping

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

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. Leamer (1990) studied the mapping from a country's factor endowments to net exports and 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.

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. 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. However, owing to the tremendous complexity of finding a complete and optimal hierarchical classification system they content themselves with a sequential procedure which cannot guarantee good or even optimal groupings at any stage of the classification system. As our aim is to generate good or optimal groupings, we have to restrict our application in this paper to the more man-

[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 $P(42, 6) = 6.665 \times 10^{29}$. For Leamer's application it is still higher, namely $P(56, 9) = 7.455 \times 10^{47}$.

ageable 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 travelling salesman problem. In fact, it falls into the class of so-called NP-complete problems,[4] which means that there is probably no exact optimization algorithm that works in economic computing time.[5]

In order to close this gap, we pursue our problem by employing a local search heuristic known as Threshold Accepting (TA).[6] "Heuristic" means that for a given input of computer resources using these algorithms it cannot be guaranteed that the exact optimum will be found. However, their velocity 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.

In this paper we study a problem of optimal grouping of 42 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 42 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 42 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 computa-

[4]For discussion of NP-completeness ("NP" stands for "nondeterministic polynomial-time") see for example Garey and Johnson (1979). 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.

[5]By "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."

[6]This algorithm was introduced by Dueck and Scheuer (1990) for the travelling salesman problem. See also Dueck and Winker (1992), Winker (1995,2000), Winker and Fang (1997), and Fang et al (2000) for some successful implementations in statistics and econometrics.

tional 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; the application is not restricted to a static model, but allows for a dynamic specification of the exogenous variables (the technical details for the dynamic version are provided in Appendix A). Section 4 is devoted to the heuristic optimization algorithm TA and Section 5 to the results achieved with the method of optimal aggregation for the problem of price indices both in a static and a dynamic setting. The paper concludes with a summary of the basic findings and an outlook to future research.

2 Optimal Aggregation

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

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

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

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

where "col E" denotes the column vector of successive columns of E, Σ is the $m \times m$ simultaneous covariance matrix and V the $n \times n$ sample covariance matrix. $\mathcal{E}$ denotes the expectation operator. We shall assume that V is positive definite.[7]

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

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

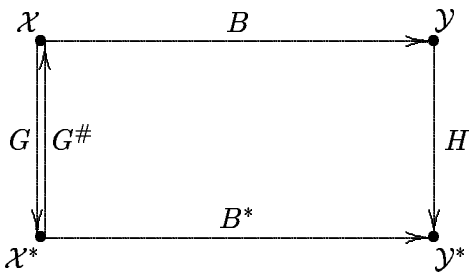
[7]The more general case rank $V \leq n$ is treated in Chipman (1975).

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).[8] 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

$$XGB^* = \mathcal{E}^*Y^* = \mathcal{E}YH = XBH, \tag{4}$$

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

$$GB^* = BH. \tag{5}$$

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).

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

$$X = X^*\bar{G} = XG\bar{G}, \tag{6}$$

[8] The meaning of the reverse mapping $G^{\#}$ appearing in the figure will be explained later (see equation (10) below).

where $\bar{G}$ is a $k^* \times k$ matrix such that $\bar{G}G\bar{G} = \bar{G}$. Then (6) 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 (6) we require that there exist a solution, B^*, to

$$G\bar{G}GB^* = G\bar{G}BH. \tag{7}$$

This holds automatically, since one may choose $(G\bar{G}G)^- = \bar{G}$ and the Penrose (1955) solvability condition $[I - (G\bar{G}G)\bar{G}]G\bar{G}BH = 0$ 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 $Y^* = YH$ to be forecast and its forecast by X^*B^* on the assumption that the model (3) is true; this discrepancy is

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

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

$$F = \mathcal{E}\{(Y^* - X^*B^*)'V^{-1}(Y^* - X^*B^*)|X\} = A + nH'\Sigma H\,, \tag{8}$$

where

$$A = (BH - GB^*)'X'V^{-1}X(BH - GB^*). \tag{9}$$

is the matrix of "aggregation bias" (cf. Chipman, 1975, pp. 125ff).

If either there exists a solution B^* to (5), or X satisfies (6)—hence there exists a solution B^* to (7)—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

$$B^* = G^{\#} B H, \tag{10}$$

where

$$G^{\#} = (G'X'V^{-1}XG)^{-1}G'X'V^{-1}X\,, \tag{11}$$

if, as may be expected in practice, the matrix $X^* = XG$ has full rank k^*.[9] (11) corresponds to the interpretation given by Theil (1954, p. 65) as the "auxiliary least-squares regression equations" of the microvariables X on the macrovariables X^*. In Figure 1 one may read off (10) as the composition of the mapping B^* into the three mappings shown.[10]

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

$$F^* = H'B'(I - GG^{\#})'X'V^{-1}X(I - GG^{\#})BH + nH'\Sigma H, \tag{12}$$

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 (12) 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

$$\phi = \alpha + n \text{ tr } H'\Sigma H W^*. \tag{13}$$

where

$$\alpha = \text{tr } H'B'(I - GG^{\#})'X'V^{-1}X(I - GG^{\#})BHW^*. \tag{14}$$

[9] For the general case see Chipman (1976, p. 668.)

[10] Noting that $X^{*\ddagger}XX^{\ddagger} = X^{*\ddagger}$, where $X^{\ddagger} = (X'V^{-1}X)^{-}X'V^{-1}$ and $\tilde{B} = X^{\ddagger}Y$, etc., and X^{-} denotes any generalized inverse of X in the sense of Rao (1966), it follows that a formula analogous to (10) 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.

We choose the Euclidean metric $W^* = I_{m^*}$ for the application.[11] This choice seems to be the most suitable one as the price data used in our application are already measured in a natural common unit. However, as a time period of nearly 30 years is covered by the data, undue weight might be given to commodity categories whose value is subject to considerable fluctuation and, in particular, inflationary processes may overshadow the more fundamental effects we are interested in. The standard procedure to cope with heteroskedasticity introduced by price inflation is not applicable here for obvious reasons, as it consists in dividing the time series by some price index. Consequently, we choose

$$V = \operatorname{diag}\{XX'\} = \operatorname{diag}\left\{\sum_{i=1}^{k} x_{ti}^2\right\} \tag{15}$$

to correct for possible heteroskedasticity.

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.

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

[11] Alternatively, one could choose the "Mahalanobis distance" defined by the choice $W^* = (H'\Sigma H)^{-1}$; cf. Chipman (1975).

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 can be regarded as a "generalized Stolper-Samuelson mapping". 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.[12]

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 then

$$Y = XB + E = [X_1, X_2] \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} + E, \tag{16}$$

where E is a random $n \times m$ matrix with zero mean and covariance given by (2), where V is given by (15). 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

[12] 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.

$m \times m^*$ grouping matrix. We define the $k \times k^*$ grouping matrix G by

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

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 generalized least-squares estimator

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

and Σ by the pseudo-maximum-likelihood estimator S/n,[13] where

$$S = Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y$$

(from the given data set these can be computed once and for all, if necessary using generalized inverses based on singular-value decomposition[14]) the objective function to be minimized corresponding to the criterion of mean-square forecast error is, in accordance with (13) above,

$$(17) \quad \tilde{\phi} = \mathrm{tr}\{H'\tilde{B}'(I - G\tilde{G}^{\#})'X'V^{-1}X(I - G\tilde{G}^{\#})\tilde{B}H\} + \mathrm{tr}\{H'SH\},$$

and

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

in accordance with (11) above.

The most convenient data set available for a first implementation of TA 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

[13] If 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 $m^*n/(n-k)$.

[14] I.e., using the oblique generalized inverse $X^{\ddagger} = \overset{\circ}{X}{}^{\dagger}V^{-1/2}$, where $\overset{\circ}{X} = V^{-1/2}X$, in place of $(X'V^{-1}X)^{-1}X'V^{-1}$, where $\overset{\circ}{X}{}^{\dagger}$ is the Moore-Penrose generalized inverse of $\overset{\circ}{X}$.

of this type, called "Wertziffern," have been furnished by the Statistisches Bundesamt, Wiesbaden, for the Federal Republic of Germany.[15]

Then aggregation means just summation and the nonzero elements of the grouping matrices are all ones. We considered the series of $m = 42$ commodity categories to be aggregated into $m^* = 6$ groups.[16] There exists an official method of grouping these 42 industries into six groups which makes it possible to compare our results with results based on the official grouping.[17]

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 328 months from January 1968 to April 1995.

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, an optimal aggregation cannot be found by enumeration of all possible grouping matrices. In fact, 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 $I_2 \otimes H$, is NP-complete.[18] There is a nearly general consensus that no deterministic algorithm can give an exact solution to such roblems without using computer resources—i.e. computing time or storage capacity—that grow faster than

[15] The 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.

[16] Appendix C contains a table of the commodity groups and their weights in imports, exports, and internal production.

[17] The classification system used is the *Güterverzeichnis für Produktionsstatistiken*, formerly known as the *Warenverzeichnis für die Industriestatistik*. The 42 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.

[18] For a formal definition and a discussion of NP-completeness see Garey and Johnson (1979).

every polynomial in the size of the problem. 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. We implemented the local search optimization heuristic TA as introduced in Dueck and Scheuer (1990). TA is a descendant of the Simulated Annealing algorithm discussed in Kirkpatrick, Gelatt und Vecchi (1983). Successful implementations include applications to the travelling salesman problem (Dueck and Scheuer (1990)), optimal portfolio selection (Dueck and Winker (1992)), lag structure identification in VAR models (Winker (1995, 2000)), and experimental design (Winker and Fang (1997), Fang et al (2000)).

Since the focus of this paper is on the application of the theory of optimal aggregation, the optimization method used for this purpose is of interest only in as much as an optimization heuristic such as TA enables actual calculations for real data. Furthermore, the implementation of the algorithm and some discussion of its tuning for specific problem instances is documented in the literature.[19] Therefore, we may content ourselves with a short layout of the basic principles of the algorithm and its application to the optimal aggregation problem.

Given an objective function $\tilde{\phi}$ to be minimized over a set of possible solutions $\mathcal{H}$, any local search heuristic like TA proceeds as follows. It starts with a randomly selected element $H_0 \in \mathcal{H}$. In each iteration, the value of the objective function for the current candidate solution $\tilde{\phi}(H_c)$ is compared to $\tilde{\phi}(H_n)$ for an element H_n, which, in some sense to be defined later, is close to the previous element. In our implementation, the H_n are selected randomly in a neighbourhood of H_c. The local search algorithms differ mainly in the rules for deciding when to accept the new element as the current solution.

The simplest local search algorithm accepts H_n if and only if it leads to a decrease in the objective function. Consequently, it convergences to a local minimum. However, for problems with many local minima such as 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 refined versions of local search algorithms accept

[19] See the references in Winker and Fang (1997) and Winker (2000).

a temporary worsening in order to escape such local minima. Hence, these algorithms may exhibit a "hill-climbing" behavior[20]

In the case of TA this idea is implemented using the so–called threshold sequence. This sequence of non negative numbers decreases to zero during the iterations of the algorithm. At each iteration the current threshold determines up to what amount a worsening of the objective function $\tilde{\phi}$ is accepted when moving from H_c to H_n. Given a treshold T, H_n is accepted if and only if $\tilde{\phi}(H_n) - \tilde{\phi}(H_c) < T$, i.e. always if H_n is "better" than H_c and up to the threshold T if it is worse. This way the algorithm escapes local minima, but may happen to jump back in the next iteration. However, this is unlikely to happen if the neighbourhoods are large enough.[21] Then, TA converges eventually to a solution arbitrarily close to the global optimum with probability larger than $1 - \delta$ for any given positive δ (Althöfer and Koschnick (1991)). Unfortunately, the number of iterations has to grow very fast to obtain this result.

Given this general outline of the algorithm, the implementation to the problem of optimal aggregation requires to impose a neighbourhood structure on $\mathcal{H}$, which is the set of proper grouping matrices in our application, and to select a threshold sequence.

The choice of neighbourhoods or a metric on $\mathcal{H}$ is crucial for the performance of TA. If they are chosen too small, the risk of getting stuck in bad local minima increases, if they are too large, TA will perform like a random search procedure, i.e. loose efficiency. For the application to proper grouping matrices, the Hamming distance (Hamming (1950)) seems a natural and appropriate concept. 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:[22]

$$d_H(H, \tilde{H}) = \sum_{i=1}^{m} \sum_{j=1}^{m^*} \mid h_{ij} - \tilde{h}_{ij} \mid \tag{18}$$

For a given proper grouping matrix an element 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

[20] Cf. figure 2 in section 5.

[21] In addition, TA could be combined with ideas from tabu search to avoid such cycling.

[22] E. 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.

generated by simultaneously moving two commodities to different groups. The choice of spheres of radius 4 with regard to d_H proved to be a good choice for the application to proper grouping matrices.[23]

The final ingredient to our TA implementation consists of the choice of a threshold sequence. We use a relative definition of the threshold values, i.e. the threshold describes up to what percentage a worsening of $\tilde{\phi}$ is admitted when moving from H_c to H_n in one iteration step. We generate the threshold sequence from an empirical distribution of local relative deviations. This empirical distribution is generated by randomly generating a large number of proper grouping matrices H_i and neighbouring matrices $\tilde{H}_i$. The relative deviations are calculated as $(\tilde{\phi}(\tilde{H}_i) - \tilde{\phi}(H_i))/\tilde{\phi}(H_i)$. From the resulting absolute values of relative deviations sorted in decreasing order the values of some lower quantile (between 25 and 50 per cent) are chosen as the threshold sequence.[24]

Finally, the number of iteration has a positive influence on the quality of the obtained results though with a decreasing rate. The result of the algorithm is the minimum of all past evaluations and the associated H matrix. If the number of iterations is large engough and the algorithm is well tuned, in general, this best result corresponds to the final candidate solution.

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 has been coded in FORTRAN using some ESSL-subroutines and was run on IBM RS 6000/3AT workstations at the University of Konstanz. The optimized grouping matrix presented in the sequel has been achieved by 250,000 iterations in about 20,000 CPU-seconds for one of 10 trials.

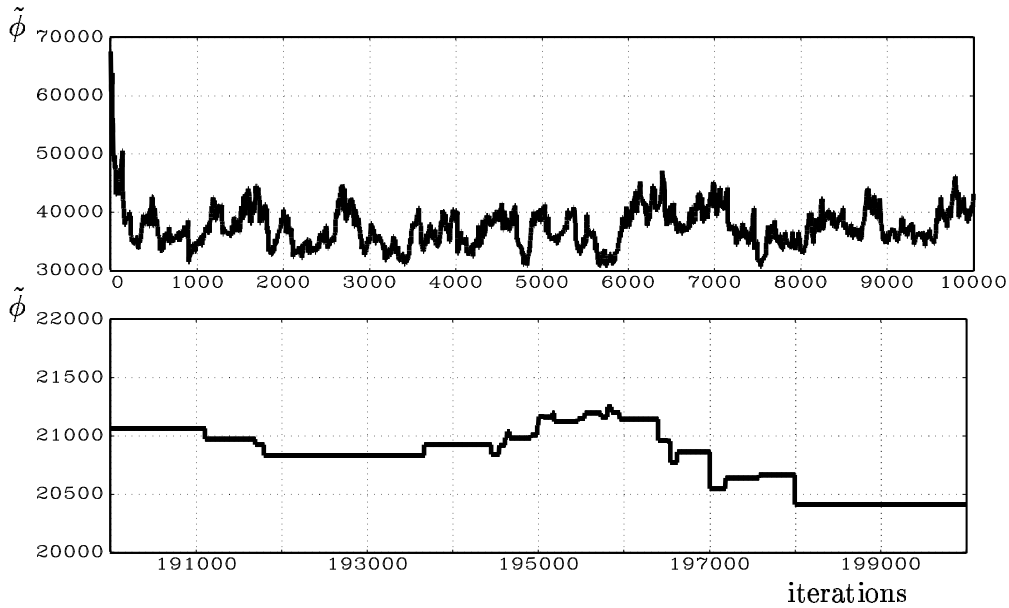
Figure 2 shows some details of the resulting sequence of values for the objective function $\tilde{\phi}$ for the current solutions during the optimization process. In the beginning of the optimization the algorithm accepts a new current solution nearly in every iteration whereas as the optimization proceeds fur-

[23] For a more detailed analysis of neighbourhood structures for TA see Fang ang Winker (1997).

[24] For the details of this data driven threshold sequence generation see Winker and Fang (1997).

ther 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 2: Aggregation bias $\tilde{\phi}(H_c)$ for first 10 000 iterations (top) and iterations 190.000 to 250.000 (bottom)



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 42 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 Table 1. This grouping can be classified as a grouping by stage of production, or a "horizontal" grouping. As far as the regression problem, as desribed above, is concerned, this grouping is far from being optimal. The distance might be expressed in terms of the objective function $\tilde{\phi}$. For $\tilde{\phi}$ this official grouping results in a value about three times the best value achieved by TA optimization.

In contrast to the official grouping, most of the groupings obtained with

Table 1: The Official Grouping

Agricultural, forestry, & fish. prod.
– Agricultural, forestry, & fish. prod.
Mining products
– Coal
– Crude oil and natural gas
Basic materials
– Petroleum products
– Quarrying products
– Iron and steel
– Nonferrous metals
– Iron, steel, and malleable cast iron
– Prod. of drawing & cold-rolling mills
– Chemical 41 (Inorganic)
– Chemical 42 (Organic)
– Chemical 43 (Fertilizer)
– Chemical 44 (Plastic)
– Chemical 45 (Chemical fibres)
– Chemical 46 (Colouring matter etc.)
– Chemical 47 (Pharmaceuticals)
– Chemical 49 (Other)
– Sawn timber, plywood etc.
– Wood pulp, cellulose, paper etc.
– Rubber products

Capital Goods
– Steelworking Products
– Structural-steel prod. & rolling stock
– Machinery (including farm tractors)
– Road vehicles (excluding farm tractors)
– Electrical products
– Precision and optical goods, clocks etc.
– Ironware, sheet-metal ware, & hardware
– Office machinery & data-processing eq.
Consumer goods
– Musical instruments, toys, 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

the optimization procedure contain at least some "vertical groupings". For a group of commodities they tend to contain the products of the preceding 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. Before discussing it in some detail we present the optimized grouping:

Table 2: The Optimized Grouping

Group 1
– Petroleum products
– Iron and steel
– Nonferrous metals
– Precision and optical goods, clocks etc.
– Chemical 47 (Pharmaceuticals)
– Fine ceramics
– Paper and paperboard products
– Plastic products
– Leather
– Food and beverages
Group 2
– Iron, steel, and malleable cast iron
– Chemical 45 (Chemical fibres)
– Wood products
– Leatherware (including travelware)
– Footware
– Textiles
– Tobacco products
Group 3
– Steelworking Products
– Machinery (including farm tractors)
– Office machinery & data-processing eq.
– Musical instruments, toys, film etc.
– Chemical 46 (Colouring matter etc.)
– Chemical 49 (Other)
– Wood pulp, cellulose, paper etc.

Group 4
– Quarrying products
– Prod. of drawing & cold-rolling mills
– Chemical 43 (Fertilizer)
– Printed and duplicated matter
Group 5
– Coal
– Structural-steel prods. & rolling stock
– Ironware, sheet-metal ware, & hardware
– Chemical 41 (Inorganic)
– Glass and glassware
Group 6
– Agricultural, forestry, & fish. prod.
– Crude oil and natural gas
– Road vehicles (excluding farm tractors)
– Electrical products
– Chemical 42 (Organic)
– Chemical 44 (Plastic)
– Sawn timber, plywood etc.
– Rubber products
– Apparel

In contrast to our statement on "vertical groupings" **Group 1** in our

optimized grouping contains three commodities classified as basic materials in the official grouping together with a set of final consumption goods. However, already in **Group 2** a vertical grouping can be detected from chemical fibres to textiles and footware. It seems reasonable to find leatherware in the same group. Not much can be said about iron, steel, and iron products and wood products, while tobacco products as very small commodity class does not influence the objective function to a large extent. Hence, its place in the final grouping should not be given much importance.

In **Group 3** we find commodities related to machinery at different stages of production together with basic materials for the printing and paperboard industry. A strong vertical component is the one between quarrying products and fertilizer in **Group 4**. The same holds true for the grouping of coal together with structural–steel products and ironware in **Group 5**. In 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 with some commodities of the steel and iron industry on the one side and oil together with some chemical products on the other side in distinct groups.

The last group includes the important automobile industry together with rubber products which enter into automotive manufacturing in an obvious way. Related is the vertical grouping of crude oil, organic chemicals, and plastic.

The forecast error $\tilde{\phi}$ for the above optimized grouping amounts to 20409.5 compared with 57678.0 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.

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 $\tilde{\phi}$, is

an element of the discrete space $\mathcal{H}$,[25] and $\tilde{\phi}$ can be thought of as a function

$$\mathcal{H} \times \mathcal{X} \longrightarrow \mathbb{R}\,, \tag{19}$$

where $\mathcal{X}$ denotes the space of data sets (X, Y). $\tilde{\phi}$ is uniformly continuous in the second argument, as $\mathcal{H}$ is finite. Consequently, in the general case

$$H \neq \tilde{H} \Longrightarrow \tilde{\phi}(H) \neq \tilde{\phi}(\tilde{H}) \tag{20}$$

and therefore

$$\delta \equiv \min_{H,\tilde{H}\in\mathcal{H},H\neq\tilde{H}} |\,\tilde{\phi}(H) - \tilde{\phi}(\tilde{H})\,| > 0\,. \tag{21}$$

As $\tilde{\phi}$ is continuous in (X, Y) there exists a small positive ε such that a perturbation of the data by less then ε in the Euclidean norm will lead to a deviation in the values of $\tilde{\phi}$ 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 parameters of the algorithm will lead to completely different outcomes with regard to the values of the objective function $\tilde{\phi}$ and to the main features of the resulting groupings.

To begin with the optimization parameters, we tried a huge bundle of different threshold sequences, used different numbers of iterations from 10,000 to 1,000,000 and many different initial values for the random-number generator. The general impression is a negative correlation between the number of iterations and the achieved values for $\tilde{\phi}$, a rather weak influence of different forms for the threshold sequence—as long as the thresholds are not too

[25] As 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.

small—, and optimal values for $\tilde{\phi}$ nearly always in the order of magnitude of 20000. The run with 10 trials leading to the optimal grouping presented above gave a mean value of $\tilde{\phi}$ of 21807 with standard deviation of 999.1.

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

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 intractable, 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, a dynamic modelling as discussed in appendix A could cope with time lags in the price adjustment process. Secondly, we plan to apply our methods to different data sets, in particular to Swedish and Dutch price-index data.

Appendix

A Extension to a Dynamic Distributed-Lag Model

A.1 Dynamic formulation

The formulation in section 3 is purely static: each domestic price is assumed to be affected by external prices only in the contemporaneous month. However, it is well known that while the transmission of some international price changes may be very rapid, changing world prices of raw materials may take some time to be manifested in domestic prices of finished products. A distributed-lag model is therefore justified; retaining the linearity assumption, this will be of the form

$$(22) \qquad y_{tj} = \sum_{i=1}^{K} \sum_{l=0}^{L} z_{t-l,i} \gamma_{ij}(l) + \varepsilon_{tj}, \quad \mathcal{E}\varepsilon_{tj} = 0, \quad \mathcal{E}\varepsilon_{tj}\varepsilon_{t'j'} = \sigma_{jj'} v_{tt'}, \quad (j = 1, 2, \ldots, m)$$

where $t = 1, 2, \ldots, n$. The expectations are taken to be conditional on the $z_{t-l,i}$, and the $v_{tt'}$ are assumed to be known. The periods covered in the sample are $1 - L, 2 - L, \ldots, 0, 1, 2, \ldots, n$, numbering $T = n + L$.

In our particular application, y_{tj} is the value of the jth producer price index ($j = 1, 2, \ldots, m$) in month $t = 1, 2, \ldots, n$, where month 1 is January 1969 and month $n = 316$ is April 1995.[26] The various price-index series are linked together and expressed relative to the year 1991 as a base, and multiplied by their 1991 weights. There are $m = 42$ of these indices; this differs from the number $m = 42$ used in the static application by reason of the omission of "other mining products" (for lack of available data), the merging

[26]The data set used for the dynamic model in this appendix was generated based on publications by the Statistisches Bundesamt Wiesbaden in printed form and on its CD-ROM StatisBund. We are indebted to Ms. Knauer from the Statistisches Bundesamt for helping us in filling gaps caused by unpublished price indices or weights.

of categories 291 and 295 into the single category 29 (foundry products)—again for lack of data—and the replacement of the chemicals category (40) by its eight component subcategories.[27]

The exogenous variable $z_{t-l,i}$, where $l = 1, 2, \ldots, L = 12$ and $i = 1, 2, \ldots, K = 2m$, is the value of the ith import price index ($i = 1, 2, \ldots, m = 42$) or the jth export price index ($i = 42 + j, j = 1, 2, \ldots, 42$), also expressed relative to the year 1991 as a base and multiplied by the 1991 weight, in month $t - l$, where $t = 1, 2, \ldots, n = 316$ and $l = 0, 1, \ldots, L = 12$. The earliest month covered is thus month $1 - 12 = -11$ which corresponds to January 1968. Thus there are $K = 2m = 84$ exogenous variables. The 42 commodities and their weight s in the import, export, and producer price indices are listed in Table 4 in Appendix C.

The $(n + L) \times K$ and $(n + L) \times m$ raw data matrices $\breve{Z}$ and $\breve{Y}$ for this problem may be constructed as

$$(23)\quad \breve{Z} = [z_{ti}]_{\substack{t=1-L,2-L,\ldots,0,1,2,\ldots n \\ i=1,2,\ldots,K}} \quad \text{and} \quad \breve{Y} = [y_{ti}]_{\substack{t=1-L,2-L,\ldots,0,1,2,\ldots n \\ i=1,2,\ldots,m}} .$$

From the $(n + L) \times m$ matrix $\breve{Y}$ we shall need only the bottom n rows. We therefore define the $n \times m$ matrix Y as $Y = [y_{ti}]_{\substack{t=1,2,\ldots,n \\ i=1,2,\ldots,m}}$. The required transformation of the matrix $\breve{Z}$ is considerably more complicated. From (22) we see that the contribution of the ith exogenous variable to the jth endogenous variable at time t is given by

$$\sum_{l=0}^{L} z_{t-l,i}\gamma_{ij}(l) = (z_{ti}, z_{t-1,i}, \cdots, z_{t-L,i}) \begin{bmatrix} \gamma_{ij}(0) \\ \gamma_{ij}(1) \\ \vdots \\ \gamma_{ij}(L) \end{bmatrix} .$$

[27] Data for two of these chemical subcategories—categories 47 (pharmaceuticals) and 49 (other chemicals)—were unavailable for the export price index over the period January 1980 to December 1984; however, data for the combined category were available, as well as data for SITC category 54 (medicinal and pharmaceutical products) whose weight in the 1976-base series (12.38) differed only slightly from the weight of GP category 47 (12.90); consequently, SITC category 54 was taken as a proxy for GP category 47 over this period, and the residual figures computed by subtracting this proxy from the aggregate series were used as a proxy for GP category 49. Another compromise made was to use the 1970-base import price index for crude petroleum (GP 2121) over the period January 1968 to December 1969 as a proxy for the export price index for natural gas (GP 2122) over the same period.

Its contribution to the jth column of Y is therefore given by

$$\begin{bmatrix} z_{1i} & z_{0i} & z_{-1,i} & \cdots & z_{1-L,i} \\ z_{2i} & z_{1i} & z_{0i} & \cdots & z_{2-L,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{ni} & z_{n-1,i} & z_{n-2,i} & \cdots & z_{n-L,i} \end{bmatrix} \begin{bmatrix} \gamma_{ij}(0) \\ \gamma_{ij}(1) \\ \vdots \\ \gamma_{ij}(L) \end{bmatrix}.$$

Thus, defining the $n \times (L+1)$ matrix Z_i by

$$Z_i = [z_{t-l,i}]_{\substack{t=1,2,\ldots,n \\ l=0,1,\ldots,L}} = \begin{bmatrix} z_{1i} & z_{0i} & z_{-1,i} & \cdots & z_{1-L,i} \\ z_{2i} & z_{1i} & z_{0i} & \cdots & z_{2-L,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{ni} & z_{n-1,i} & z_{n-2,i} & \cdots & z_{n-L,i} \end{bmatrix} \quad (i = 1, 2, \ldots, K)$$

(24)

and the $(L+1) \times m$ matrix Γ_i by

$$\Gamma_i = [\gamma_{ij}(l)]_{\substack{l=0,1,\ldots,L \\ j=1,2,\ldots,m}} = \begin{bmatrix} \gamma_{i1}(0) & \gamma_{i2}(0) & \cdots & \gamma_{im}(0) \\ \gamma_{i1}(1) & \gamma_{i2}(1) & \cdots & \gamma_{im}(1) \\ \gamma_{i1}(2) & \gamma_{i2}(2) & \cdots & \gamma_{im}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{i1}(L) & \gamma_{i2}(L) & \cdots & \gamma_{im}(L) \end{bmatrix} \quad (i = 1, 2, \ldots, K),$$

(25)

the contribution of the ith exogenous variable to all m columns of Y is $Z_i\Gamma_i$. Writing

(26) $$Z = [Z_1, Z_2, \ldots, Z_K] \quad \text{and} \quad \Gamma' = [\Gamma_1', \Gamma_2', \ldots, \Gamma_K']$$

we may express (22) in the form:

(27) $$Y = Z\Gamma + \mathrm{E} = \sum_{i=1}^{K} Z_i\Gamma_i + \mathrm{E}$$

where $\mathrm{E} = [\varepsilon_{tj}]$ is defined analogously to Y. Thus it is necessary to form the $K = 2m = 84$ matrices Z_i from the matrix $\breve{Z}$. Each $n \times (L+1)$ matrix Z_i is constructed from the ith column of the $(n+L) \times K$ matrix $\breve{Z}$. The first column of Z_i consists of the last n elements of the ith column of $\breve{Z}$; the second column of Z_i consists of the next-to-last n elements of the ith column of $\breve{Z}$; and finally, the last column of Z_i consists of the first n elements of the ith column of $\breve{Z}$.

B Fitting the lag distributions by spline functions

The $n \times (L+1)K = 316 \times 1092$ matrix Z of (24) having been constructed from (23) as just indicated, we now need to transform it in order to fit the distributed lags by cubic spline functions. Letting the contemporaneous lag coefficients $\gamma_{ij}(0)$ be estimated freely, and letting the remaining L lag coefficients $\gamma_{ij}(l), l = 1, 2, \ldots, L$ be assumed to lie on a natural cubic spline with N knots, denoted

$$1 \leq \kappa_1 \leq \kappa_2, \ldots, \kappa_N \leq L, \tag{28}$$

it is known from the theory of natural cubic splines (cf. Greville, 1969) that there exist $N+2$ coefficients $a_{ij}(0), a_{ij}(1), c_{ij}(1), c_{ij}(2), \ldots, c_{ij}(N)$ satisfying the following conditions:

$$\gamma_{ij}(l) = a_{ij}(0) + a_{ij}(1)l + \sum_{h=1}^{N} c_{ij}(h)(l - \kappa_h)_+^3 \quad (l = 1, 2, \ldots, L), \tag{29}$$

where $d_+ \equiv \max(d, 0)$, and where the coefficients $c_{ij}(h)$ satisfy the following two constraints:

$$\sum_{h=1}^{N} c_{ij}(h) = 0 \quad \text{and} \quad \sum_{h=1}^{N} c_{ij}(h)\kappa_h = 0. \tag{30}$$

Conditions (29) and (30) may be expressed in matrix form as

$$\begin{bmatrix} \gamma_{ij}(1) \\ \gamma_{ij}(2) \\ \vdots \\ \gamma_{ij}(L) \\ \cdots \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \vdots & (1-\kappa_1)_+^3 & (1-\kappa_2)_+^3 & \cdots & (1-\kappa_N)_+^3 \\ 1 & 2 & \vdots & (2-\kappa_1)_+^3 & (2-\kappa_2)_+^3 & \cdots & (2-\kappa_N)_+^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & L & \vdots & (L-\kappa_1)_+^3 & (L-\kappa_2)_+^3 & \cdots & (L-\kappa_N)_+^3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 1 & 1 & \cdots & 1 \\ 0 & 0 & \vdots & \kappa_1 & \kappa_2 & \cdots & \kappa_N \end{bmatrix} \begin{bmatrix} a_{ij}(0) \\ a_{ij}(1) \\ \cdots \\ c_{ij}(1) \\ c_{ij}(2) \\ \vdots \\ c_{ij}(N) \end{bmatrix}. \tag{31}$$

Restriction of the relations (29) to the lag coefficients $\gamma_{ij}(l)$ at the knots

$l = \kappa_h$ leads to the $(N+2) \times (N+2)$ system

$$(32) \qquad \begin{bmatrix} \gamma_{ij}(\kappa_1) \\ \gamma_{ij}(\kappa_2) \\ \gamma_{ij}(\kappa_3) \\ \vdots \\ \gamma_{ij}(\kappa_{N-1}) \\ \gamma_{ij}(\kappa_N) \\ 0 \\ 0 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & \kappa_1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & \kappa_2 & (\kappa_2-\kappa_1)^3 & 0 & \cdots & 0 & 0 \\ 1 & \kappa_3 & (\kappa_3-\kappa_1)^3 & (\kappa_3-\kappa_2)^3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \kappa_{N-1} & (\kappa_{N-1}-\kappa_1)^3 & (\kappa_{N-1}-\kappa_2)^3 & \cdots & 0 & 0 \\ 1 & \kappa_N & (\kappa_N-\kappa_1)^3 & (\kappa_N-\kappa_2)^3 & \cdots & (\kappa_N-\kappa_{N-1})^3 & 0 \\ 0 & 0 & 1 & 1 & \cdots & 1 & 1 \\ 0 & 0 & \kappa_1 & \kappa_2 & \cdots & \kappa_{N-1} & \kappa_N \end{bmatrix} \begin{bmatrix} a_{ij}(0) \\ a_{ij}(1) \\ c_{ij}(1) \\ c_{ij}(2) \\ \vdots \\ c_{ij}(N-1) \\ c_{ij}(N) \end{bmatrix}.$$

Defining the $L \times 2$ and $L \times L$ matrices P and Q by

$$(33) \quad P = [p_{lr}] = [l^{r-1}]_{\substack{l=1,2,\ldots,L \\ r=1,2}} \quad \text{and} \quad Q = [q_{ll'}] = [(l-l')_+^3]_{l,l'=1,2,\ldots,L},$$

we may write (31) as

$$(34) \qquad \begin{bmatrix} \gamma_{ij}^1 \\ 0 \end{bmatrix} = \begin{bmatrix} P & QJ' \\ 0 & P'J' \end{bmatrix} \begin{bmatrix} a_{ij} \\ c_{ij} \end{bmatrix}.$$

Likewise, defining the $N \times L$ matrix J by

$$(35) \qquad J = [\delta_{\kappa_h,l}]_{\substack{h=1,2,\ldots,N \\ l=1,2,\ldots,L}}$$

where $\delta_{\kappa,l}$ is the Kronecker delta, the $N \times 1$ vector whose components are the values of the spline function at the knots, $\beta_{ij}(h) \equiv \gamma_{ij}(\kappa_h)$, $h = 1, 2, \ldots, N$, may be denoted

$$(36) \qquad \beta_{ij}^1 = J\gamma_{ij}^1.$$

Corresponding to (32) we then have

$$(37) \qquad \begin{bmatrix} \beta_{ij}^1 \\ 0 \end{bmatrix} = \begin{bmatrix} JP & JQJ' \\ 0 & P'J' \end{bmatrix} \begin{bmatrix} a_{ij} \\ c_{ij} \end{bmatrix} \equiv A \begin{bmatrix} a_{ij} \\ c_{ij} \end{bmatrix},$$

which defines the $(N+2) \times (N+2)$ matrix A.

Now it is implicit that $\gamma_{ij}(l) = 0$ for $l > L$, so to ensure that the function $\gamma_{ij}(l)$ is continuous and differentiable for $l \geq 1$ we shall require that $\gamma_{ij}(l)$ and its first derivative vanish at $l = L$. Thus we impose the end-point constraints

$$\gamma_{ij}(L) = a_{ij}(0) + a_{ij}(1)L + \sum_{h=1}^{N} c_{ij}(h)(L - \kappa_h)_+^3 = 0; \tag{38}$$

$$\gamma'_{ij}(L) = a_{ij}(1) + 3\sum_{h=1}^{N} c_{ij}(h)(L - \kappa_h)_+^2 = 0. \tag{39}$$

Assuming $\kappa_1 = 1$ and $\kappa_N = L$ in (28), these are realized by setting $\beta_{ij}(N) = \gamma_{ij}(\kappa_N) = \gamma_{ij}(L) = 0$, and replacing $\gamma_{ij}(\kappa_{N-1})$ by 0 in (32) and, conformably, replacing row $N-1$ of the matrix A by the row

$$pa_{ij} + qc_{ij} = (p,\ q) \begin{bmatrix} a_{ij} \\ c_{ij} \end{bmatrix}$$

where

$$p = (0, 1) \quad \text{and} \quad q = 3\Big((L-1)^2, (L-\kappa_2)^2, \ldots, (L-\kappa_{N-1})^2, 0\Big).$$

We denote this modified A matrix by $\bar{A}$.

In place of (37) we now have

$$\begin{bmatrix} \beta_{ij}^1 \\ 0 \end{bmatrix} = \bar{A} \begin{bmatrix} a_{ij} \\ c_{ij} \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} a_{ij} \\ c_{ij} \end{bmatrix} = \bar{A}^{-1} \begin{bmatrix} \beta_{ij}^1 \\ 0 \end{bmatrix}, \tag{40}$$

assuming $\bar{A}$ to be invertible.[28] Substituting the second equation of (40) in (34), we obtain the relation between the L restricted lag coefficients and the $N-2$ values of these lag coefficients at the knots:

$$\gamma_{ij}^1 = \Phi^1 \beta_{ij}^1 \quad \text{where} \quad \Phi^1 = [P \mid QJ']\bar{A}^{-1} \begin{bmatrix} I_{N-2} \\ 0 \end{bmatrix}. \tag{41}$$

Defining the $(L+1) \times (N-1)$ matrix

$$\Phi = \begin{bmatrix} 1 & 0 \\ 0 & \Phi^1 \end{bmatrix}, \tag{42}$$

[28] That A is invertible follows from the uniqueness theorem of natural splines (Greville, 1969). A similar argument could be used in the case of $\bar{A}$, but is unnecessary since in the present case invertibility is verified numerically.

as well as the vectors

$$\gamma_{ij} = \begin{bmatrix} \gamma_{ij}^0 \\ \\ \gamma_{ij}^1 \end{bmatrix} = \begin{bmatrix} \gamma_{ij}(0) \\ \gamma_{ij}(1) \\ \gamma_{ij}(2) \\ \vdots \\ \gamma_{ij}(L) \end{bmatrix}, \quad \text{and} \quad \beta_{ij} = \begin{bmatrix} \beta_{ij}^0 \\ \\ \beta_{ij}^1 \end{bmatrix} = \begin{bmatrix} \beta_{ij}(0) \\ \beta_{ij}(1) \\ \beta_{ij}(2) \\ \vdots \\ \beta_{ij}(N-2) \end{bmatrix},$$

(43)

we have

$$\gamma_{ij} = \Phi\beta_{ij} \quad (i = 1, 2, \ldots, K; j = 1, 2, \ldots, m). \tag{44}$$

Since γ_{ij} is the jth column of the matrix Γ_i of (25), defining the $(N-1) \times m$ matrix B_i as the matrix consisting of the m columns β_{ij} for $j = 1, 2, \ldots, m$, (44) implies

$$\Gamma_i = \Phi\mathrm{B}_i \quad (i = 1, 2, \ldots, K). \tag{45}$$

Defining the $(N-1)K \times m$ matrix B by

$$\mathrm{B} = \begin{bmatrix} \mathrm{B}_1 \\ \mathrm{B}_2 \\ \vdots \\ \mathrm{B}_K \end{bmatrix} \quad \text{where } \mathrm{B}_i = \begin{bmatrix} \beta_{i1}(0) & \beta_{i2}(0) & \cdots & \beta_{im}(0) \\ \beta_{i1}(1) & \beta_{i2}(1) & \cdots & \beta_{im}(1) \\ \beta_{i1}(2) & \beta_{i2}(2) & \cdots & \beta_{im}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{i1}(N-2) & \beta_{i2}(N-2) & \cdots & \beta_{im}(N-2) \end{bmatrix},$$

(46)

it follows from this and (25) that

$$\Gamma = (I_K \otimes \Phi)\mathrm{B}. \tag{47}$$

Thus, (27) becomes

$$\begin{aligned} Y &= \sum_{i=1}^{K} Z_i\Phi\mathrm{B}_i + \mathrm{E} = \sum_{i=1}^{K} X_i\mathrm{B}_i + \mathrm{E} \\ &= Z(I_K \otimes \Phi)\mathrm{B} + \mathrm{E} = X\mathrm{B} + \mathrm{E}, \end{aligned} \tag{48}$$

where

$$X_i = Z_i\Phi \quad \text{and} \quad X = [X_1, X_2, \ldots, X_K] = Z(I_K \otimes \Phi). \tag{49}$$

We may denote

$$X_i = [x_{tih}]_{\substack{t=1,2,\ldots,n \\ h=0,1,\ldots,N-2}} = \begin{bmatrix} x_{1i0} & x_{1i1} & \cdots & x_{1i,N-2} \\ x_{2i0} & x_{2i1} & \cdots & x_{2i,N-2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{ni0} & x_{ni1} & \cdots & x_{ni,N-2} \end{bmatrix} \quad (i = 1, 2, \ldots, K).$$

(50)

The matrix Z of the original system (27) is of order $n \times (L+1)K$, while the matrix X of the transformed system (48) is of order $n \times (N-1)K$, representing a gain in degrees of freedom of $(L-N+2)K$. Denoting $k = (N-1)K$, (48) has the general form of our multivariate model.

In our particular application in which $L = 12$, $N = 4$, and the four knots are placed at $\kappa_1 = 1, \kappa_2 = 4, \kappa_3 = 8$, and $\kappa_4 = 12$, the first equation of (40) becomes

$$(51)\quad \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 4 & 3^3 & 0 & 0 & 0 \\ 0 & 1 & 3\cdot 11^2 & 3\cdot 8^2 & 3\cdot 4^2 & 0 \\ 1 & 12 & 11^3 & 8^3 & 4^3 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 4 & 8 & 12 \end{bmatrix} \begin{bmatrix} a_{ij}(0) \\ a_{ij}(1) \\ c_{ij}(1) \\ c_{ij}(2) \\ c_{ij}(3) \\ c_{ij}(4) \end{bmatrix} = \begin{bmatrix} \gamma_{ij}(1) \\ \gamma_{ij}(4) \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The inverse of the matrix in (51), which we have denoted $\bar{A}$, is

$$(52)\quad \bar{A}^{-1} = \frac{1}{8064}\begin{bmatrix} 11328 & -3696 & -1728 & 432 & 165888 & -13824 \\ -3264 & 3696 & 1728 & -432 & -165888 & 13824 \\ 64 & -112 & -192 & 48 & 18432 & -1536 \\ -120 & 231 & 528 & -111 & -54720 & 4560 \\ 64 & -154 & -528 & 90 & 82944 & -6912 \\ -8 & 35 & 192 & -27 & -38592 & 3888 \end{bmatrix}.$$

The relationship (41) then becomes:

$$\begin{bmatrix} \gamma_{ij}(1) \\ \gamma_{ij}(2) \\ \gamma_{ij}(3) \\ \gamma_{ij}(4) \\ \gamma_{ij}(5) \\ \gamma_{ij}(6) \\ \gamma_{ij}(7) \\ \gamma_{ij}(8) \\ \gamma_{ij}(9) \\ \gamma_{ij}(10) \\ \gamma_{ij}(11) \\ \gamma_{ij}(12) \end{bmatrix} = \frac{1}{8064}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 3 & 2^3 & 0 & 0 & 0 \\ 1 & 4 & 3^3 & 0 & 0 & 0 \\ 1 & 5 & 4^3 & 1 & 0 & 0 \\ 1 & 6 & 5^3 & 2^3 & 0 & 0 \\ 1 & 7 & 6^3 & 3^3 & 0 & 0 \\ 1 & 8 & 7^3 & 4^3 & 0 & 0 \\ 1 & 9 & 8^3 & 5^3 & 1 & 0 \\ 1 & 10 & 9^3 & 6^3 & 2^3 & 0 \\ 1 & 11 & 10^3 & 7^3 & 3^3 & 0 \\ 1 & 12 & 11^3 & 8^3 & 4^3 & 0 \end{bmatrix} \begin{bmatrix} 11328 & -3696 \\ -3264 & 3696 \\ 64 & -112 \\ -120 & 231 \\ 64 & -154 \\ -8 & 35 \end{bmatrix} \begin{bmatrix} \gamma_{ij}(1) \\ \gamma_{ij}(4) \end{bmatrix}$$

$$
\begin{aligned}
(53) \qquad &= \frac{1}{8064}
\begin{bmatrix}
8064 & 0 \\
4864 & 3584 \\
2048 & 6496 \\
0 & 8064 \\
-1016 & 7847 \\
-1216 & 6328 \\
-936 & 4221 \\
-512 & 2240 \\
-216 & 945 \\
-64 & 280 \\
-8 & 35 \\
0 & 0
\end{bmatrix}
\begin{bmatrix} \gamma_{ij}(1) \\ \gamma_{ij}(4) \end{bmatrix}.
\end{aligned}
$$

The 12×2 matrix in (53) is Φ^1, and β^1 is the 2×1 vector of lag coefficients $\beta_{ij}(1) = \gamma_{ij}(1) = \gamma_{ij}(\kappa_1)$ and $\beta_{ij}(2) = \gamma_{ij}(4) = \gamma_{ij}(\kappa_2)$ at knots 1 and 2. We define the 1×3 vector

$$x_{ti} = (x_{ti0}, x_{ti1}, x_{ti2}) = (z_{ti}, z_{t-1,i}, z_{t-2,i}, \ldots, z_{t-12,i})\Phi.$$

Our new variables are, from (53),

$$
\begin{aligned}
x_{ti0} &= z_{ti}; \\
x_{ti1} &= z_{t-1,i} + 0.603174 z_{t-2,i} + 0.253968 z_{t-3,i} - 0.125992 z_{t-5,i} \\
&\quad - 0.150794 z_{t-6,i} - 0.116071 z_{t-7,i} - 0.063492 z_{t-8,i} - 0.026786 z_{t-9,i} \\
(54) \qquad &\quad - 0.007937 z_{t-10,i} - 0.000992 z_{t-11,i}; \\
x_{ti2} &= z_{t-4,i} + 0.444444 z_{t-2,i} + 0.805556 z_{t-3,i} + 0.973090 z_{t-5,i} \\
&\quad + 0.784722 z_{t-6,i} + 0.523438 z_{t-7,i} + 0.277778 z_{t-8,i} + 0.117188 z_{t-9,i} \\
&\quad + 0.034722 z_{t-10,i} + 0.004340 z_{t-11,i}.
\end{aligned}
$$

B.1 The aggregation process

Aggregation according to a system of industrial classification is a process of partitioning the industries into groups and summing relevant entries in each group. In the case of the matrix Y of observations on the endogenous variables, this is most simply effected by postmultiplying the $n \times m$ matrix Y by a (proper) *grouping matrix* H of order $m \times m^*$ to obtain the $n \times m^*$ matrix $Y^* = YH$. The same applies, of course, to the $T \times m$ raw data matrix $\breve{Y}$ of (23). In the case of the raw data matrix $\breve{Z}$ of observations on the $K = 2m$ exogenous variables, it is natural to classify the commodities appearing in

the import and export price indices in the same way as the corresponding domestic variables are classified. The corresponding grouping matrix is then

$$\breve{G} = \begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix} = I_2 \otimes H. \tag{55}$$

However, given the rearrangement of $\breve{Z}$ that has led to the matrix Z, this grouping must be applied to the *matrices* Z_i. Thus our aggregated matrix takes the form of the $n \times K^*(L+1) = 316 \times 156$ matrix

$$Z^* = [Z_1, Z_2, \ldots, Z_K] \begin{bmatrix} \breve{g}_{11} I_{L+1} & \breve{g}_{12} I_{L+1} & \cdots & \breve{g}_{1K^*} I_{L+1} \\ \breve{g}_{21} I_{L+1} & \breve{g}_{22} I_{L+1} & \cdots & \breve{g}_{2K^*} I_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ \breve{g}_{K1} I_{L+1} & \breve{g}_{K2} I_{L+1} & \cdots & \breve{g}_{KK^*} I_{L+1} \end{bmatrix} = Z(\breve{G} \otimes I_{L+1}). \tag{56}$$

But now we will wish to apply the same spline transformation to the distributed lags of the aggregated model as we did to the disaggregated model. Defining from (56)

$$Z_j^* = \sum_{i=1}^{K} \breve{g}_{ij} Z_i \quad (j = 1, 2, \ldots, K^*),$$

the analogue of (49) is $X_j^* = Z_j^* \Phi$, or

$$X^* = [X_1^*, X_2^*, \ldots, X_{K^*}^*] = [Z_1^*, Z_2^*, \ldots, Z_{K^*}^*] \begin{bmatrix} \Phi & 0 & \cdots & 0 \\ 0 & \Phi & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi \end{bmatrix} = Z^*(I_{K^*} \otimes \Phi). \tag{57}$$

We may now put all these results together as follows:

$$\begin{array}{rcll} X^* & = & Z^*(I_{K^*} \otimes \Phi) & \text{from (57)} \\ & = & Z(\breve{G} \otimes I_{L+1})(I_{K^*} \otimes \Phi) & \text{from (56)} \\ & = & Z(\breve{G} \otimes \Phi) & \text{by Kronecker multiplication} \\ & = & Z(I_K \otimes \Phi)(\breve{G} \otimes I_3) & \text{by Kronecker multiplication} \\ & = & X(\breve{G} \otimes I_3) & \text{from (49).} \end{array} \tag{58}$$

We shall denote the $k \times k^* = 3K \times 3K^* = 252 \times 36$ matrix

$$G = \breve{G} \otimes I_3 = I_2 \otimes H \otimes I_3. \tag{59}$$

Defining the $n \times m^* = 316 \times 6$ and $n \times k^* = 316 \times 36$ matrices Y^* and X^* by

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

the (false) aggregative model is

$$Y^* = X^*\mathrm{B}^* + \mathrm{E}^*.$$

The model (22) having been translated into the multivariate model (48), the symbols Y, X, V, H, and G have exactly the same meanings as in section 3, hence the objective function ϕ of (13) is computed in straightforward fashion.

B.2 Numerical results

The dynamic distributed-lag model described in this appendix increases the number of explanatory variables by the factor three. As the calculation of the objective function ϕ requires some matrix inversion, computing time increases by a factor of about 27. Therefore, we are only able to present some preliminary results for the dynamic setting based on a limited number of iterations for the TA algorithm.

As in section 5, in a first step we calculated the value of ϕ for the official grouping of the 42 categories into only six groups as given by the publications of the Statistisches Bundesamt. The official grouping is the same as the one presented in section 5 and results in a value of ϕ of 898.1. The best grouping achieved so far exhibits a mean-square forecast error as measured by ϕ of solely 270.4. The optimized grouping is given in the following table:

Although the complexity of the problem does not allow for finding the global optimum with probability one, some additional runs of the algorithm for different parameter constellations might improve both the result and our confidence in it.

Table 3: The Optimized Grouping for the Dynamic Model

Group 1
– Structural-steel prod. & rolling stock
– Chemical 45 (Chemical fibres)
– Chemical 47 (Pharmaceuticals)
– Glass and glassware
– Leatherware (including travelware)
– Textiles
Group 2
– Coal
– Crude oil and natural gas
– Quarrying products
– Electrical products
– Chemical 42 (Organic)
– Paper and paperboard products
– Plastic products
– Leather
– Tobacco products
Group 3
– Chemical 46 (Colouring matter etc.)
– Fine ceramics
– Wood products
– Apparel

Group 4
– Petroleum products
– Nonferrous metals
– Iron, steel, and malleable cast iron
– Prods. of drawing & cold-rolling mills
– Steelworking Products
– Ironware, sheet-metal ware, & hardware
– Chemical 44 (Plastic)
– Office machinery & data-processing eq.
– Printed and duplicated matter
– Rubber products
– Footware
– Food and beverages
Group 5
– Machinery (including farm tractors)
– Musical instruments, toys, film etc.
– Chemical 41 (Inorganic)
– Chemical 49 (Other)
– Sawn timber, plywood etc.
Group 6
– Agricultural, forestry, & fish. prod.
– Iron and steel
– Road vehicles (excluding farm tractors)
– Precision and optical goods, clocks etc.
– Chemical 43 (Fertilizer)
– Wood pulp, cellulose, paper etc.

C Commodity Groups and Their Weights

Table 4: Weights per thousand for German price indices 1991=100

Item	Category	Code	Imports	Exports	Prod.
1	Agricultural, forestry & fish. prod.	0000	56.07	12.10	‡52.48
	[Electricity, gas, and water]	1000	0.00	0.00	93.08
	[Mining products]	2100	63.93	5.93	54.45
2	Coal & lignite	2110	3.06	2.74	15.20
3	Crude petroleum & natural gas	2120	53.89	†0.69	*38.58
	[Crude petroleum]	2121	*37.55	0.00	0.25
	[Natural gas]	2122	16.34	†0.69	38.33
	[Other mining products]				
	(210 – 211 – 212)	–	†6.98	†3.19	†0.67
	[Iron ore]	2130	3.99	0.00	–
	[Nonferrous metal ore]	2140	2.24	†0.08	–
	[Potash salt]	2150	0.00	†1.68	0.37
	[Rock & pit salt]	2160	†0.13	†0.26	0.22
	[Fluorspar, heavy spar, graphite, etc.]	2170	†0.57	†0.08	–
	[Peat]	2180	†0.05	†0.40	–
4	Petroleum products	2200	28.21	5.94	34.14
	[Nuclear fuels]	2400	1.53	2.35	0.00
5	Quarrying products	2500	9.50	8.89	25.04
6	Iron & steel	2700	28.78	34.42	20.97
7	Nonferrous metals	2800	31.06	19.86	11.82
8	Foundry products	2900	2.92	4.56	8.57
	[Iron, steel & malleable cast-iron prod.]	2910	2.50	4.04	5.22
	[Nonferrous-metal foundry products]	2950	†0.42	†0.52	3.35
	[Drawing, cold-rolling, & steelworking products]	3000	9.77	14.18	20.45
9	Drawing & cold-rolling products	3010	4.01	5.10	5.88
10	Steelworking products	3020	5.76	9.08	14.57
11	Structural-steel products & rolling stock	3100	6.28	12.37	21.50
12	Machinery	3200	71.19	161.81	87.31
13	Road vehicles (excl. farm tractors)	3300	116.96	176.42	95.10
	[Ships & boats]	3400	0.00	0.00	0.13
	[Aircraft]	3500	0.00	0.00	0.00
14	Electrical products	3600	107.98	124.72	90.74
15	Precision & optical goods, clocks & watches	3700	19.01	21.22	9.77
16	Ironware, sheet-metal ware, & hardware	3800	23.17	29.87	35.86

(Continued on next page)

Table 4 (continued): Weights per thousand for German price indices 1991=100

Item	Category	Code	Imports	Exports	Prod.
17	Musical instruments, sporting goods, etc.	3900	11.73	8.19	4.61
	[Chemicals]	4000	93.50	135.24	70.18
18	Inorganic materials & chemicals	4100	5.43	8.39	4.66
19	Organic materials & chemicals	4200	25.08	31.06	10.91
20	Fertilizers & insecticides	4300	3.97	4.19	1.68
21	Plastic & artificial rubber	4400	20.13	25.58	12.08
22	Chemical fibers	4500	2.20	4.66	2.24
23	Dyes & paints	4600	4.83	13.93	7.79
24	Pharmaceuticals	4700	11.72	17.75	13.10
25	Other chemical products	4900	20.14	29.68	17.72
26	Office machinery & data-processing equipment	5000	41.84	22.63	9.52
27	Fine ceramics	5100	4.42	3.34	2.54
28	Glass	5200	6.57	6.81	6.71
29	Sawn timber, plywood, other worked wood	5300	8.06	3.29	6.99
30	Wood products	5400	14.43	10.26	25.84
31	Wood pulp, cellulose, paper & paperboard	5500	22.97	11.49	7.86
32	Paper products	5600	6.13	9.60	16.59
33	Printed & duplicated matter	5700	4.58	9.19	23.82
34	Plastic products	5800	21.13	28.20	34.68
35	Rubber products	5900	10.39	9.71	8.22
36	Leather	6100	2.23	1.81	0.73
	[Leatherware & footware]	6200	16.25	4.33	3.49
37	Leatherware (incl. travelware)	6220	3.77	1.28	1.13
38	Footware	6250	12.48	3.05	2.36
39	Textiles	6300	53.64	36.93	20.09
40	Apparel	6400	43.90	15.83	16.14
41	Food & beverages	6800	61.06	45.41	113.63
42	Tobacco products	6900	0.81	3.10	17.43
	[Prefabricated housing]	7000	0.00	0.00	2.00

Legend: Square brackets indicate items omitted from the data set. An asterisk (*) in front of a weight indicates data computed by aggregation. A dagger (†) indicates that published data are not available. A double dagger (‡) indicates a weight computed from 1991 data. The commodity codes used are those of the GP (Güterverzeichnis für die Produktionsstatistik).

References

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.

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., AND T. SCHEUER (1990): "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

FANG, K.–T., D. K. J. LIN, P. WINKER, AND Y. ZHANG. (2000): "Uniform Design: Theory and Application" *Technometrics*, forthcomming.

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

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.

GREVILLE, T. N. E. (1969): "Introduction to Spline Functions," in *Theory and Application of Spline Functions,* ed. by T. N. E. Greville. New York: Academic Press, 1–35.

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

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.

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.

SHEPHARD, 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.

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

WINKER, P. (1995): "Identification of Multivariate AR-Models by Threshold Accepting." *Computational Statistics and Data Analysis*, 20,9, 295–307.

WINKER, P. (1999): "Optimized Multivariate Lag Structure Selection." Computational Economics, forthcomming.

WINKER, P., AND K.–T. FANG (1997): "Application of Threshold Accepting to the Evaluation of the Discrepancy of Set of Points." *SIAM Journal on Numerical Analysis*, 34,5, 2028–2042.