

DISCUSSION PAPER NO. 307

A Note on Aggregation and Disaggregation

by

Walter D. Fisher

November 1977

W. D. Fisher "A Note on Aggregation and Disaggregation"

E R R A T A

Page 3, equation (8). \tilde{P} should be \bar{P} .

Page 6, equation (14). $A = T'T^+$ should be $A = T'T'^+$.

Page 7, line after equation (16). Read "where each A_f is. . .and each P_f is.."

Page 8, line 6. Close parentheses after [4].

Page 8, equation (18), block-diagonal matrices on right. Insert three dots between upper left and lower right vectors in each matrix.

A NOTE ON AGGREGATION AND DISAGGREGATION

By Walter D. Fisher *

This note is motivated by some recent work by Chipman [2,3], Sondermann [6], and Tintner and Sondermann [7] that has generalized and extended the approach to clustering and aggregation pursued by Fisher [4,5]. The purpose of the note is to comment on the relationship between the concepts of simplification, aggregation, and disaggregation that appear in these various writings. The context is the multivariate regression model, which is assumed to be used by an investigator to make predictions of a set of endogenous variables with small quadratic loss. In section 3 a theorem states a correspondence between the use of a prescribed aggregation matrix and the use of a prescribed disaggregation matrix.

1. Model and Cost Function

Let the original or detailed model be

$$(1) \quad y = Px + v ,$$

where y is a vector of G endogenous variables, x a vector of H exogenous variables, P a known $G \times H$ coefficient matrix, and v a vector of random disturbances with zero expectation. It is assumed that the vector x is also random, but independent of v , with raw second moment matrix $E(xx') = M$, a known symmetric positive-definite matrix. The word "model" or "detailed model" will also be used to refer to the matrix P .

The basic purpose is assumed to be the prediction of y with small error. Another purpose is assumed to be to find a simplified model, \tilde{P} , of the same size as P , the manner of simplification to be described below, such that when

* Northwestern University

\tilde{P} is used to forecast y by the formula

$$\tilde{y} = \tilde{P}x ,$$

the expected loss

$$(3) \quad r = E(y - \tilde{y})'C(y - \tilde{y})$$

is minimized, where C is an arbitrary symmetric positive-definite matrix that weights the relative importance of forecast errors in the various endogenous variables and their interactions. Such a \tilde{P} will be referred to as optimal.

It is shown in Fisher [4, p. 27, Theorem 1] that $r = c + \text{constant}$, where c is the simplification cost.

$$(4) \quad c = \text{tr } C(\tilde{P} - P)M(\tilde{P} - P)' ,$$

and hence the simplification problem may be described as choosing a \tilde{P} so that c is minimized.¹

Throughout the rest of this note, with the exception of brief comments to be made in section 5, it will be assumed that the model P has full row rank (rank G) and that the simplified model is obtained by "one-way simplification on the left" -- that is by

$$(5) \quad \tilde{P} = AP ,$$

where A is a square matrix of rank F , $F < G$, and is called a simplifier. It will also be assumed that C and M are identity matrices. The loss of generality entailed by these assumptions will be only slight, will not bear on the points to be made, and hopefully will be more than compensated by the gain in simplicity of notation and exposition. More general formulations are available in the literature cited.

In the next three sections various types of constraint on the simplified model are treated.

2. Simplification under Rank Restriction

Say that it is desired to choose for the simplified model a \tilde{P} that is of rank no greater than F, where $F < G$, and such that the simplification cost

$$(6) \quad c = \text{tr } (\tilde{P} - P)(\tilde{P} - P)'$$

is a minimum. (Note that this representation of c uses the special assumptions $C = I, M = I$.) It has been shown [4, p 167, Lemma C.1] that the desired simplifier and simplified model are, respectively,

$$(7) \quad A = R_F' R_F \quad , \quad \tilde{P} = R_F' R_F P \quad ,$$

where R_F is the $F \times G$ matrix whose rows are the normalized characteristic vectors associated with the F largest characteristic roots of the matrix PP' . This solution may be called the characteristic vector solution and is unique if the characteristic roots of PP' are distinct², which we shall assume to be the case. This problem is a mild specialization of that considered by Tintner and Sondermann [7, p 522, Theorem 3], and becomes the same as theirs if, in their notation, we set $C = I, M = I, K = F$. Then the solution (7) above becomes identical to theirs with R_F equal to their Q_K' .

This problem is also closely related to another problem. Define an aggregator as a full rank $F \times G$ matrix, S, that premultiplies into P, yielding the $F \times H$ aggregated model

$$(8) \quad \tilde{P} = SP \quad .$$

Define a disaggregator as a full rank $G \times F$ matrix, T', that premultiplies into some aggregated model \bar{P} , yielding the simplified model

$$(9) \quad \tilde{P} = T' \bar{P} = T' SP \quad ,$$

the term "simplified model" and the symbol \tilde{P} in (9) being justified by (9)

being in the form of (5) with $A = T'S$. When the simplifier A is in this form, we shall also say that the pair (S, T') is an aggregation-disaggregation sequence, or more simply, a sequence. Consider now the problem of finding a sequence (S, T') and an associated aggregated model, \bar{P} , such that the resulting simplified model \tilde{P} in (9) is optimal -- that is, such that the simplification cost in (6) is a minimum. Such a sequence and aggregated model will also be called optimal.

Since any \tilde{P} in the form of (9) must have rank no greater than F , an optimal \tilde{P} in that form -- i.e. subject to the restriction $\text{rank}(\tilde{P}) \leq F$ -- must satisfy the characteristic vector solution (7). That is, a solution to the present problem is the optimal sequence and aggregated model, respectively,

$$(10) \quad (S, T') = (R_F, R'_F) \quad , \quad \bar{P} = R_F P \quad .$$

But note that the solution is not unique. An infinity of solutions may be generated by, say,

$$(11) \quad (S, T') = (ZR_F, R'_F Z^{-1}) \quad , \quad \bar{P} = ZR_F P \quad ,$$

where Z is an arbitrary $F \times F$ nonsingular matrix, and all of these solutions will yield the same simplifier $A = R'_F R_F$ and same simplified model \tilde{P} .

After presenting their solution to the problem of simplification under rank restriction on \tilde{P} Tintner and Sondermann state that the solution "provides only a partial answer to the problem of optimal aggregation with free choice" of the disaggregator, and

"Since we are interested in the aggregation problem, our main concern is the aggregated model \bar{P} instead of the simplified model \tilde{P} , which here serves only to measure the aggregation bias. The problem of simultaneous optimal determination of \bar{P} , T and S is still unsolved."³

They then advance a conjecture which, in the special case considered in the present note, is that the optimal aggregated model is $\bar{P} = R_F P$, as in (10).

While recognizing the importance of the aggregation problem, one may question the appropriateness of the definite article "the" in the phrase "the optimal aggregated model" [ibid] since the conjectured solution is, without further restrictions, one of an infinite number, as shown above. One may also perhaps question the "main concern" of finding a unique aggregated model in these circumstances, unless convincing criteria can be provided for choosing a preferred one from the infinite number.

3. Simplification under Prescribed Aggregator or Disaggregator

Sometimes it is expedient to consider sequences where either the aggregator or the disaggregator is specified a priori. Sometimes a particular structure such as grouping is desired.⁴

DEFINITION. An aggregator-prescribed sequence is a sequence (S, T') where S is given. A disaggregator-prescribed sequence is a sequence (S, T') where T' is given.

From Chipman [2, p 692, Theorem 2.2] in an aggregator-prescribed sequence conditional on the aggregator S , the optimal disaggregator, simplifier, and simplified model are, respectively, in the present case

$$(12) \quad T' = S_W^+ \quad , \quad A = S_W^+ S \quad , \quad \tilde{P}_1 = S_W^+ S P \quad ,$$

where $S_W^+ = WS'(SWS')^{-1}$ and $W = PP'$, which yields the simplification cost

$$(13) \quad \tilde{c}_1 = \text{tr } W - \text{tr } WS'(SWS')^{-1} SW.$$

Chipman calls this result "best approximate disaggregation". We have applied his formulas, which deal with a more general case, to our special assumptions,

and under these assumptions \tilde{P}_1 is unique.⁵

From Fisher [4, p 34, Theorem 2] in a disaggregator-prescribed sequence conditional on the disaggregator T' , the optimal aggregator, simplifier, and simplified model are, respectively, in the present case

$$(14) \quad S = T'^+ , \quad A = T'T^+ , \quad \tilde{P}_2 = T'T'^+P ,$$

where $T'^+ = (TT')^{-1}T$ (the Moore-Penrose generalized inverse of T'), which yields the simplification cost

$$(15) \quad \tilde{c}_2 = \text{tr } W - \text{tr } WT'(TT')^{-1}T .$$

The simplified model \tilde{P}_2 is found to be unique.

A correspondence between the two approaches is provided by the following theorem.

THEOREM. For any optimal aggregator-prescribed sequence there exists an optimal disaggregator-prescribed sequence that entails the same simplification cost, and vice versa.

PROOF. When S is a prescribed aggregator, the optimal simplifier is, from (12), S_W^+S , and the cost is \tilde{c}_1 from (13). Let $T' = W^{1/2}S'$ be the prescribed disaggregator. Then the optimal simplifier conditional on that T' is, from (14), $T'T'^+ = W^{1/2}S'(SWS')^{-1}SW^{1/2}$, and the cost is obtained from \tilde{c}_2 in (15). Then it is found, after substituting $W^{1/2}S'$ for T' in (15) and using properties of the trace operator, that $\tilde{c}_1 = \tilde{c}_2$. When we start with some arbitrary T' as a prescribed disaggregator, the optimal simplifier is, from (14), $T'T'^+$, and the cost is \tilde{c}_2 from (15). Then, letting $S = TW^{-1/2}$ be the corresponding prescribed aggregator, the corresponding optimal simplifier and cost in terms of

T can be found from (11) and (13), and again the two costs are found to be the same.

In general the corresponding simplifiers S_W^+ and $T'T'^+$, as well as the corresponding simplified models \tilde{P}_1 and \tilde{P}_2 , are themselves different. In an important special case they are the same.

COROLLARY. If the prescribed aggregator is R_F , or if the prescribed disaggregator is R_F' , then the two simplifiers are the same and equal to $R_F'R_F$.

The proof is immediate from the fact that $R_F'R_F$ is the unique optimal unconditional simplifier of rank F, and that the optimal conditional simplifiers must be consistent with the optimal unconditional one. The corollary can also be established by direct substitution into (12) and (14), using the properties of normalized characteristic vectors.

4. Simplification by Grouping

An important special case of the simplification problem arises frequently in economics and other fields when it is desired to group the elements of the model P into disjoint subsets. Continuing with the special case of simplification on the left only, say that the rows of P are grouped into F disjoint and exhaustive subsets, as are also the rows of \tilde{P} correspondingly, and it is required that any row of \tilde{P} depend only on the subset of P corresponding to its own subset. That is, it is required that \tilde{P} be of the form

$$(16) \quad \tilde{P} = \begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_F \end{bmatrix} \begin{bmatrix} P_1 \\ \vdots \\ P_F \end{bmatrix},$$

where A_f is a square matrix of less than full rank and P_f is a submatrix of P according to some partition of rows. This case is called simplification by grouping. The partition may be prescribed a priori from economic-theoretical

consideration,⁵ or the problem of simplification may include the finding of an optimal partition, based on the structure of the detailed model. In the latter case the optimal \tilde{P} may be found in two stages: first finding a minimal cost conditional on a partition, and second, searching over the domain of relevant partitions for the global minimal cost. The search problem (see, for example, [3] and [4]), will not be discussed here.

Consider now the further specialization where each A_f is specified to be of rank 1, which implies that \tilde{P} is to be of rank F. Since the determination of various optimal \tilde{P}_f are independent of each other, the characteristic vector solution (7) may be applied to each P_f separately, using only the largest root in each problem. That is, we have

$$(17) \quad A_f = R_1^f R_1^f, \quad f = 1, \dots, F,$$

where R_1^f is the normalized characteristic row vector corresponding to the largest characteristic root of $P_f P_f'$. Noting that A_f is in the form of a column vector postmultiplied by a row vector, we have the simplifier

$$(18) \quad A = T'S = \begin{bmatrix} t_1' & & \\ & & \\ & & t_F' \end{bmatrix} \begin{bmatrix} s_1 \\ & & \\ & & s_F \end{bmatrix},$$

where t_f' are column vectors and s_f are row vectors. Thus the optimal simplifier is an aggregation-disaggregation sequence where both aggregator and disaggregator are grouping matrices, that is, matrices in the form of T' or S in (18), where the weights (non-zero elements) have been determined as part of the solution.⁶

Next consider the optimal simplifier when the aggregator or disaggregator is prescribed to be a particular grouping matrix, with both the partition and the weights fixed a priori, perhaps with the weights positive. With

the aggregator S so prescribed, (12) applies; with the disaggregator T' so prescribed, (14) can be used. But note that in this case the simplifier A in (14) separates into the block diagonal form of A in (18), while such does not occur with the simplifier A in (12) because of the intervention of the general positive-definite matrix W in the formula for S_W^+ . So, simplification by grouping can be attained when the disaggregator is prescribed to be a grouping matrix, but not when the aggregator is so prescribed. Too much should not be made of this result since it depends crucially on our initial assumption that C (or M in the parallel case of columnwise simplification) is a diagonal matrix, an untenable assumption in many applications.⁷ It is, nevertheless, of some interest.

As Chipman points out [2 p. 710], an investigator may wish to emphasize the aggregated model rather than the simplified one, and in that case may wish to prescribe the aggregator as a grouping matrix so that the aggregated model will possess a grouping structure, even if the simplified model does not. While this may certainly be true, it would seem that for those other problems where disaggregation is also necessary or desirable, there may be still further advantages in having the final simplified model also possess a group structure. When this result is possible (such as when C is a diagonal matrix), it seems reasonable to formulate the problem in a way (such as prescribing a grouping disaggregator T') that will attain it.

5. Two-way simplification

This treatment has been restricted to simplification "row-wise" or "on the left" of the coefficient matrix P in equation (1), ignoring the possibilities of simplification "column-wise" or "on the right". An entirely

analogous treatment could be made of the columnwise case and is made in [4] and [5] by interchanging the role of rows and columns of P , \bar{P} , \tilde{P} , and in an aggregation-disaggregation sequence placing the disaggregator to the right of the aggregator. Then M plays the role of C . Or the simplification by rows and columns may be considered simultaneously. In that case, when rank restrictions are considered, the simplifier with the lower rank is the effective one.

In commenting on Fisher's formulation in the case of simultaneous row-wise and column-wise simplification Chipman remarks that it "involves an asymmetry in the treatment of independent and dependent variables, an aggregation operator being given in the former case but a disaggregation operator in the latter" [2, p 710]. This characterization of asymmetry results from looking at the vectors of variables in (1) rather than at the coefficient matrix P . From the viewpoint of the coefficient matrix Fisher's formulation involves a given disaggregation operator both over rows and over columns of P , while it is Chipman's formulation that is assymmetric. One might also say that in (1) the role of the independent and dependent variables are themselves in a asymmetric relationship to each other, so that it is a moot point which formulation is fundamentally the more symmetric one.

Probably the usefulness of alternative formulations is the most important criterion for choice, and it will require time to tell which will be best chosen for which uses.

Footnotes

1. The simplification cost as defined in (4) has a close relationship to what may be called the matrix bias: $B = E(\tilde{y}-y)(\tilde{y}-y)' = (\tilde{P}-P)M(\tilde{P}-P)'$, which can be minimized in the matrix sense -- meaning that \tilde{P} is chosen so that the resulting B differs by a positive semi-definite matrix from the B resulting from any other choice of \tilde{P} . Indeed, it can be shown that such a minimization of the matrix bias is equivalent to the minimization of the scalar c for any positive semi-definite C. However, the choice of a particular C is necessary in the "second stage" of a minimization problem to be described below where a search is made over discrete alternatives (as is also pointed out by Chipman [2, pp. 677, 706]), so we elect to use the scalar criterion c throughout the present treatment. It should also be noted that B has a close relationship to Chipman's "absolute disaggregation bias" [2, p 690] and in the case of oneway simplification on the left, which is the case we shall discuss here, the two biases are precisely equal.
2. Proof: When the characteristic roots are distinct, all non-normalized characteristic vectors associated with a particular root are proportional (see, e.g. Bellman [1, p 42]), and hence the rows of R_F are unique except for possible multiplication by -1. Hence the product $R_F'R_F$ is unique.
3. [7, p 523] In the quoted statement we have changed the original \bar{A} and \tilde{A} to \bar{P} and \tilde{P} in order to correspond to our notation.
4. See section 4.
5. The uniqueness of \tilde{P}_1 results from the uniqueness in this case of Chipman's "generalized quasi-inverse" of S, which reduces to S_W^+ , sometimes referred to as an "oblique" generalized inverse of S with respect to W, and that satisfies all four properties of Chipman's conditions (\ddagger) [2, p 552] with $V = I$ and $U = W$.
6. Here the weights may possibly be negative (as would be the case, for example where two rows of P forming a subset by themselves were negatively correlated). The term grouping matrix or partitioning operator is sometimes used to include only the case of positive weights (Chipman [2, p 650], Fisher [4, p 28]), often prescribed as in the next paragraph. Note that it would be easy to generalize the case of P_f of rank one to allow ranks higher

than one.

7. See [4 pp. 51-52] for an example of the implausibility of a diagonal M matrix and [5, pp 759-761] for a discussion of the problem of attaining optimal \tilde{P} when C or M is nondiagonal.

References

- [1] Bellman R. Introduction to Matrix Analysis, McGraw Hill, New York, 2nd ed., 1970.
- [2] Chipman, J.S. "Estimation and Aggregation in Econometrics: an Application of the Theory of Generalized Inverses," Generalized Inverses and Applications, M.Z. Nasied, ed., 549-769, Academic Press, New York, 1976.
- [3] Chipman, J.S. "Optimal Aggregation in Large-scale Econometric Models," Sankhya, 37C (1975), 121-159.
- [4] Fisher, W.D. Clustering and Aggregation in Economics, Baltimore, The Johns Hopkins Press, 1969.
- [5] Fisher, W.D. "Optimal Aggregation in Multi-equation Prediction Models," Econometrica, 30 (1962), 744-769.
- [6] Sondermann, D. "Optimale Aggregation von Grossen Linearen Gleichungssystemen," Zeitschrift für Nationalökonomie, 33 (1973), 235-250.
- [7] Tintner, G. and D. Sondermann, "Statistical Aspects of Economic Aggregation," Mathematical Economics and Game Theory: Essays in Honor of Oskar Morgenstern, R. Henn and O. Moeschlin, eds., 516-535, Springer-Verlag, Berlin, 1977.