Sufficient Conditions for Exact Aggregation in Linear Programming Models

By Thomas A. Miller

MANY GAPS exist in the economic analysis of broad problems of agricultural policy. One is the lack of a reliable method by which economists can generalize from analysis of individual farms. Current research methodology often includes scaling up the linear programming solution of a "representative" farm to generate information about the aggregate production behavior of the group or set of individual farms it represents. But this approach has a weakness. If the individual farms in the group do not respond alike to changes in economic stimuli, the estimates of aggregate output for the group will be biased. This article develops the conditions of similarity among individual farms which, if met, are sufficient to permit grouping farms so that a representative farm may be used to estimate the aggregate behavior of each group without bias.

The basic problem is how to obtain estimates of the total output of a given set of farms under various assumptions. One possible procedure would be to determine the optimum organization (and output) from every individual farm in the set and to sum them into the desired aggregate estimate. Although this procedure would result in a bias-free estimate, the limited resources available for study usually make it impractical. Alternative procedures involving greater abstractions become necessary to make the problem computationally feasible.

An often used alternative procedure is to define a "representative farm" within the set and to determine the optimum organization for this farm by linear programming techniques. The output of the set as a whole is then estimated by multiplying the solution of the representative farm by a weighting factor defined as the number of farms in the set. Since the resources of the representative farm are usually defined as the sum of all resources in the set divided by the total number of farms, a parallel method is to consider the total set and its resources as the representative farm and to determine the optimum solution for the entire set directly. These two procedures yield identical output estimates.

Inherent in these abstractions is the possibility of aggregation bias.[1] This is said to exist when the sum of the solutions for each of the individual farms in the set does not equal the estimate obtained by determining the optimum solution to the entire set directly (or the total obtained by weighting the solution for the representative farm).

The Aggregation Problem

At this point, it is desirable to make a more rigorous definition of aggregation bias and at the same time develop a notation for the discussion which follows. Consider the linear programming model representing the gth farm of the set of n farms, which is the problem of selecting a vector of production levels, X_g, such that profit is a maximum, resource limits are respected, and no production levels are negative. In the usual mathematical notation we solve for X_g such that

(1) $$\pi_g = Z_g X_g$$

is a maximum subject to

(2) $$B_g X_g = C_g$$

and

$$X_g \geq 0$$

[1] Lee M. Day, "Use of Representative Firms in Studies of Interregional Competition and Production Response," Jour. Farm Econ. 45:1438-1445, Dec. 1963.

where π_g = total net returns to the gth farm,

Z_g = the 1 by m vector of activity net returns for the gth farm,

X_g = the m by 1 vector of the activity levels to be chosen by the gth farm,

B_g = the k by m matrix of input-output coefficients for the gth farm, and

C_g = the k by 1 vector of available resources of the gth farm.

This is standard linear programming form with the necessary slack vectors included to reach equality of relations in equation (2).

If the optimum solutions are obtained for all n farms and totaled, the desired solution for the aggregate set of n farms becomes $\sum_{g=1}^{n} X_g$. This is the procedure referred to in the second paragraph of the introduction and the result is an estimate free of aggregation bias. Hence, it becomes a logical standard against which all other procedures may be judged.

As mentioned earlier, the alternatives often used are (1) to sum the total resources over all farms and to determine the optimum solution for the aggregate as a whole or (2) to weight results obtained for a representative farm. Since these procedures yield equivalent results, either one may be used in discussing the aggregation problem with no loss in generality. Choosing the former, the more abstract alternative then may be expressed in one problem of selecting a vector of aggregate area production levels, X*, such that

(3) $$\pi^* = Z^*X^*$$

is a maximum subject to

(4) $$B^*X^* = C^*$$

and

$$X^* \geq 0$$

The starred symbols represent the entire set of farms where the g subscripts represented the individual farms. The dimensions of the matrices are the same in both cases. Since the individual farm resources are summed to obtain the resources of the aggregate set,

$$C^* = \sum_{g=1}^{n} C_g.$$

Now we may define exact aggregation as the situation in which the levels of the various activities in the second formulation are exactly the same as that obtained by programming each farm separately and summing, that is

$$X^* = \sum_{g=1}^{n} X_g.$$

Conversely, aggregation bias is defined as the situation in which

$$X^* \neq \sum_{g=1}^{n} X_g.$$

The central question now becomes, given the set of n farms and the problem specified above, what conditions among the set of farms are sufficient to assure exact aggregation?

A Recent Contribution to Bias-Free Aggregation

In an article on aggregating linear programming models, Richard Day defines sufficient conditions for exact aggregation as the requirement of "proportional heterogeneity."[2] The conditions are that

(5) $$B_1 = B_2 = ..., = B_n = B^*$$

(6) $$Z_g = \gamma_g Z^*$$

where γ_g is a scalar greater than zero for all g and

(7) $$C_g = \lambda_g C^*$$

[2] Richard H. Day, "On Aggregating Linear Programming Models of Production," Jour. Farm Econ. 45: 797-813, Nov. 1963.

where λ_g, a scalar greater than zero and less than one for all g, represents the proportion of the sets' resources that the gth firm possesses. Condition (5) is that all firms must have identical matrices of input-output coefficients; condition (6) is that firms have only proportional variation in net return expectations; and condition (7) is that firms have only proportional variations in constraint vectors.

Day presents proof of the sufficiency of these conditions through the duality theorem of linear programming. In addition to fulfilling the previously defined requirements of exact aggregation, he notes that the condition

$$R^* = \frac{1}{n} \sum_{g=1}^{n} R_g$$

would also be achieved in a set of firms conforming to equations (5), (6), and (7) where R^* is the "average marginal net revenue productivities" of the resources in the set and the R_g are the vectors of marginal net revenue productivities of resources of the individual firms.[3]

Day has an excellent discussion of the implications of the conditions of proportional heterogeneity from an operational standpoint. The interested reader is urged to refer to his article. His comments concerning conditions (5) and (6) are particularly detailed. In this area I find that many current research projects in agricultural economics are based upon the assumption of a given and specified level of management and hence identical input-output matrices for large groups of firms. The assumption of proportionality in the vectors of expected net returns is likewise easily met. In fact, an often used procedure is merely to assume all firms have the same net return expectations.

This article is primarily concerned with the restriction posed by condition (7), which allows only the variation in resources among firms that are usually expressed as differences of scale of operation. If two firms differ in one resource by a certain ratio, they must differ in all other resources by that ratio. This condition appears very restrictive from the operational standpoint. For example, in agriculture, there are nearly as many different cropland resource situations as there are farms while, on the other hand, a majority of these same farms may have labor supplied by only one operator. The implications should be obvious to researchers working in this area, and it does not appear desirable to go into them in any great detail. Instead, I have developed less binding sufficient conditions for exact aggregation.

[3] These values represent the solution of the dual linear programming problem.

Less Binding Requirements

The first step is to define the less binding requirements both intuitively and then somewhat more rigorously. A theorem and proof of the sufficiency of these requirements then follow.

An intuitive idea of the relaxed requirements is gained by considering the optimum solutions of a set of individual farms as determined by linear programming. Assume the set of farms under consideration is similar to the extent necessary for all individual optimum solutions to include identical sets of activities. Such a set of individual farms may then vary in both resource and net return vectors, so long as this variation is not great enough to cause a change in the set of optimum activities common to all farms in the group. The variation in net return and resource vectors among farms will, of course, cause differences among farms in optimum activity levels. The important thing is that the identity of the activities in the optimum solutions must be the same for all farms. Farms meeting this requirement will be defined as having qualitatively homogeneous output vectors.

To make a more rigorous specification of the new conditions, consider the optimum solution of each individual farm. The optimum solution for the gth farm may be expressed as a column vector

$$X_g = \begin{bmatrix} X_{1g} \\ X_{2g} \\ \cdot \\ \cdot \\ \cdot \\ X_{mg} \end{bmatrix}$$

Previously, m was defined as equal to the number of production processes considered by the farm plus the number of slack vectors necessary to permit nonuse of resources and k was defined as the number of resources or constraints. Observe that $m > k$ for this formulation since k is also the number of required slack vectors that are included in m to achieve equality in the restraints. For each optimum solution, X_g is made up of at most k activities that are greater than zero and at least m minus k activities that are equal to zero.[4]

Now consider a set of farms which have qualitatively homogeneous output vectors. For each of these we could express a streamlined output vector as

$$X'_g = \begin{bmatrix} X'_{1g} \\ X'_{2g} \\ \cdot \\ \cdot \\ \cdot \\ X'_{kg} \end{bmatrix}$$

by omitting the m minus k activities which are common to each and equal to zero. We note now that the X'_g (streamlined output vectors) for all farms will all consist of the same k basic activities. All such farms will have the same resources limiting, the same resources in disposal, and the same real processes in their final solution vectors. This leads to the following theorem.

Theorem. Sufficient conditions for exact aggregation are (1) that all farms have identical coefficient matrices, that is, that $B^* = B_g$ for all g, and (2) that all farms have qualitatively homogeneous output vectors.

Proof. For farms meeting conditions of the theorem, the original linear programming problem may be narrowed to the more trivial problem of solving a set of k equations in k unknowns

$$(8) \qquad B'^* X'_g = C_g$$

where $B'_g = B'^*$ is the k by k part of the coefficient matrix corresponding to the k activities in X'_g. Equation (8) is then equivalent to equation (2) with the unused activities of the coefficient matrix omitted and the zero elements of X_g omitted. This is no more than saying that if the identity of the final basis activities is known in advance, the linear programming problem may be solved simply as a set of simultaneous equations.

Similarly, the solution to the aggregate farm may be determined from the relation

$$(9) \qquad B'^* X'^* = C^*$$

which is developed in a similar fashion from equation (4).

Summing equation (8) over all n farms gives

$$B'^* \sum_{g=1}^{n} X'_g = \sum_{g=1}^{n} C_g$$

Since $\sum_{g=1}^{n} C_g = C^*$ by definition, it is obvious from equations (8) and (9) that $X'^* = \sum_{g=1}^{n} X'_g$. All that remains is to include the m minus k zero level elements to both vectors to complete the proof that $X^* = \sum_{g=1}^{n} X_g$. The conditions of the theorem are hence sufficient conditions for exact aggregation.[5]

---

[4] These k activities are often called the basic variables in the literature, while the remaining activities are called nonbasic variables. The theorem generally developed is that an optimum solution involves at most k unknowns at nonzero values (where k equals the number of equations). For example, see: R. D. Dorfman, P. A. Samuelson, and R. M. Solow, "Linear Programming and Economic Analyses," New York, McGraw-Hill Book Company, Inc., 1958, Theorem 2, p. 75.

[5] The conditions of the theorem are general in respect to the price or revenue vectors used; hence, the theorem covers variable-price programming. This is because the consideration of different prices has the effect of further restricting the groups of farms that have qualitatively homogeneous output vectors. To have exact aggregation under varying sets of prices, all farms in the group must merely meet the conditions of the theorem for every set of prices considered. In other words, the farms must all have qualitatively homogeneous output vectors for the first set of prices, have a set of possibly different but again qualitatively homogeneous output vectors for the second set of prices, and so on for all price ratios considered.

A parallel argument could be developed for aggregation of the dual solutions over the same set of n farms to obtain the "average marginal net revenue productivities" of the resources. Under conditions stated in the theorem, if $C^* = \sum_{g=1}^{n} C_g$, then $X^* = \sum_{g=1}^{n} X_g$. Likewise, for the dual solutions the same argument may be developed to show that if $Z^* = \frac{1}{n} \sum_{g=1}^{n} Z_g$, then $R^* = \frac{1}{n} \sum_{g=1}^{n} R_g$ where Z_g are (as defined earlier) the vectors of expected net returns per unit of the respective activities and the R_g are the vectors of desired "marginal net revenue productivities" of the resources in the optimum solutions of each individual farm. Hence, the conditions of the theorem also appear to be sufficient conditions for exact aggregation of the "marginal net revenue productivities" of the resources of the individual farms into the "average marginal net revenue productivities" of all resources in the aggregate.

Implications of New Conditions

The problem of aggregation bias is not trivial to research workers in agricultural economics. Large amounts of money are being allocated to projects which are utilizing the representative farm linear programming concept in developing aggregate area production estimates and area supply functions. One example is the NC-54 regional project, "Supply Response and Adjustments for Hog and Beef Cattle Production," in which 13 Corn Belt States are cooperating. Iowa's contribution to this project involves the use of 63 farms to represent all commercial farms in the State. Results from these 63 farms are used to develop State supply functions for hogs and beef cattle. At later stages, similar supply functions from all cooperating States will be combined.[6]

[6] For results of a recently completed study using similar methodology, see: W. B. Sundquist, et al., "Equilibrium Analysis of Income-Improving Adjustments on Farms in the Lake States Dairy Region, 1965," Minn. Agr. Expt. Sta. Tech. Bul. 246, Oct. 1963.

The question of how sufficient conditions for exact aggregation affect such current research studies is certainly important. The conditions developed in this paper are substantially less binding from an operational standpoint than the original ones developed by Day (see footnote 2). Some range of different resource situations and net return expectations can now be combined without incurring aggregation bias. Moreover, there is no restriction on the type of variation that may occur between farms, so long as all of the individual farms in the set have solutions made up of the same activities.

On the other hand, the new conditions are defined as a requirement of the solutions to the individual farms rather than a requirement of the farms themselves. Therefore, they provide less than an ideal solution to the problem of delineating representative farms. It may be difficult to anticipate the solutions of various individual farms with the accuracy necessary to stratify them into the separate classes required to avoid aggregation bias. Considerable prestratification analysis may be necessary. Nevertheless, the theorem still provides the researcher with a definite idea of the objective of the stratification. This is to delineate sets of individual farms in such a way that all farms within a respective set will meet the conditions of (1) identical input-output matrices, and (2) qualitatively homogeneous output vectors.

In addition, it may be desirable to answer such questions as: Given a representative farm, to what extremes may its coverage be extended? This question may arise after some preliminary linear programming work is done with the basic data, and a general idea is obtained of the types of optimum solutions involved and the effect of variance of different resources upon them. The answer can be obtained by parametric programming on the resource vector of the representative farm. The results will give the ranges of individual farm resource vectors which may be included in the set represented by that particular representative farm, without creating aggregation bias. This assumes, of course, that the appropriate adjustment will be made in the representative farm's resource vector.

Little can be added to these general statements. The problem of specifically how many representative farms are required to avoid aggregation bias in a given instance is left unsolved. The problem of how rapidly aggregation bias accumulates as we move away from the sufficient conditions stated in the theorem is also unsolved. Both of these are essentially empirical questions which must be answered by empirical means for each individual research problem.