# IMPLICATIONS OF AGGREGATION BIAS FOR THE CONSTRUCTION OF STATIC AND DYNAMIC LINEAR PROGRAMMING SUPPLY MODELS*

**Allan E. Buckwell and Peter B. R. Hazell**
*Agricultural Adjustment Unit, University of Newcastle upon Tyne*

*Aggregation bias is acknowledged to be one of the more serious problems confronting the Linear Programming approach to supply analysis. Whilst the literature abounds with theoretical solutions to the problem there is a notable lack of ideas on how these solutions might be made operational. This conclusion holds* a fortiori *for the dynamic case. This paper discusses a practical methodology for the classification of farms in order to minimise aggregation bias, and also the implications of avoiding bias for the specification of dynamic linear models.*

Most studies of agricultural supply response for the United Kingdom have been based on econometric analysis of aggregate time series data. However, at a time when the Government is considering radical changes in agricultural policy, such as the adoption of the E.E.C. Common Agricultural Policy, there are definite advantages in adopting an alternative approach based on a microeconomic analysis using mathematical programming techniques. These advantages can be summarised as follows:

(i) Microeconomic models provide a wealth of information at the farm and regional levels, as well as at the national level. This is extremely useful in the evaluation of the impact of policy on many problems of farm management, rural development and regional income distribution.

(ii) A mathematical programming model necessarily embodies a complete causal system of the functioning of the individual farm and its interrelationships with all other sections of the industry. It is therefore not so susceptible to the problems which arise when the policies to be evaluated involve extrapolation of explanatory variables beyond the range of past experience.

(iii) A mathematical programming model can take formal account of the fact that most farms produce many products using many resources (i.e. multiproduct/multiresource firms), and hence is well suited to examining the total impact of changes in relative prices on the supply of individual products.

These advantages must be weighed against the immense data requirements of any comprehensive microeconomic model. A further difficulty is the alleged normative nature of the approach which arises because farmers are assumed to

* We are grateful for the helpful comments of Professor S. John Rogers, Ian M. Sturgess and Brian H. Davey on an earlier draft of this paper. Any remaining errors are entirely our responsibility.

be sufficiently rational in their decision-making process to permit approximation by a maximising model.* Normative models are generally more helpful in the formulation of policy than in the evaluation thereof. To the extent that mathematical programming models are normative, this does reduce their usefulness for evaluating current agricultural policy problems.

At Newcastle University we consider that the advantages of a microeconomic approach using mathematical programming techniques far outweigh the disadvantages at the present time, and we are now completing the construction of a comprehensive model of agricultural supply response for the United Kingdom using linear programming.† In the course of this work we have to face the problems of aggregation bias and model specification errors. These problems have not yet been satisfactorily resolved, and must be treated as the programming counterpart of the statistical problems which arise in econometric work.

Problems of aggregation bias have been studied in some depth in the context of static linear models [2,4,8,11–15] while problems of model specification, which remain more elusive, have been little more than recognised.[13,15] Both sets of problems are even less well understood for dynamic models, and yet dynamic models are probably the most relevant for the analysis of present problems in United Kingdom agricultural policy. This is because static models are only useful for a comparative analysis of equilibrium situations, and only then when these situations are sufficiently far apart in time to allow equilibrium to be attained after a host of existing dynamic constraints have worked themselves out. These constraints include problems of capital accumulation necessary to finance lumpy farm investments, risk aversion or other technical characteristics which inhibit farmers from making very rapid adjustments to their farm plans or technology in any one year, and the sluggish response of various institutional and marketing arrangements to changes in the economic environment. Dynamic models can incorporate these features‡ and thus ensure that feasibility is maintained over relatively short forecast periods. Dynamic models also provide detailed information about the adjustment path of supply response in proceeding from one set of policy variables to another. Such information is likely to be of far greater use for policy evaluation than simple knowledge of the supplies forthcoming under equilibrium situations.

These considerations are particularly important in studying the impact of a switch to the Common Agricultural Policy, because it is important to consider both the final 'in E.E.C.' prices and the sets of prices which will obtain during the transition period up to 1978.

This paper is concerned with problems of aggregation bias in dynamic linear models of agricultural supply response, and in particular, with the implications of avoiding bias for both farm classification procedures and model specification. We begin with a review of the problem of aggregation bias in static linear models, and of existing theorems pertaining to criteria for classifying farms to avoid bias in such models. Given that aggregation bias is unlikely to be totally avoided in any empirical model, we propose a method of approximating the classification criteria which has proved practical in our work at Newcastle. Theorems for avoiding aggregation bias are then extended to dynamic linear models, and their implications pursued for farm classification procedures and

*Day[6] has argued that *rational* choice need not necessarily imply *normative* choice, because decisions are bounded by the limited extent of the individual's knowledge. Simulating the decision environment in a microeconomic model only leads to the 'best that can be done' under the circumstances and not what 'necessarily ought to be done'.

† The current study grew out of earlier work by Davey and Weightman.[3]

‡ For example, through a recursive linear programming specification.[5]

for model specification. In particular, it is shown that aggregation bias can only be avoided if certain construction constraints are observed in the farm models, and that recursive linear programming models are generally not consistent with these requirements.

## 1. Aggregation bias

One way of avoiding aggregation bias would be to construct a farm model for each individual farm, and to solve these models taking into account various inter-dependencies between farms, such as the movement of intermediate goods of production and competition for common scarce resources. A good model might also be closed with respect to price to permit some degree of price determination as well as to guarantee feasibility with respect to final demand.

Let $\dot{x}_{kt}$ denote the $n \times 1$ linear programming solution vector of the $k^{\text{th}}$ farm model in the $t^{\text{th}}$ year, where $n$ denotes the total number of activities over all farms, and zeros are inserted where appropriate for individual farms. The aggregate supply vector from all farms is then

$$\sum_k \dot{x}_{kt}.$$

This value may or may not equal $\sum_k \bar{x}_{kt}$, where $\bar{x}_{kt}$ denotes the actual output vector of the $k^{\text{th}}$ farm in the $t^{\text{th}}$ year. Discrepancies may arise because of:

(i) Model specification error,[13,15]
(ii) Inaccuracies in the data,
(iii) The incomplete specification of inter-dependencies between farms,
(iv) The normative nature of the model.

In practice it is not possible to programme all the individual farms. For the United Kingdom, for example, this would involve programming some 200,000 individual farms. Some degree of farm aggregation is necessary. This can be done with the use of representative farms, or by using aggregate regional or area models.

The representative farm approach, which has been used widely in the United States and is the approach adopted in the Newcastle work, involves classifying the universe of farms into a smaller number of homogenous groups, and constructing a model for a 'representative' farm for each group. Aggregate supply estimates are obtained through appropriate weighting of the representative farm solution vector according to the number of farms in the group.

Let $\dot{x}_{ht}$ denote the solution vector of the representative farm for the $h^{\text{th}}$ group in the $t^{\text{th}}$ year, and let $k_{ht}$ denote the number of farms in the group. Aggregate supply for the group is then $k_{ht} \,.\, \dot{x}_{ht}$,* and aggregate estimated supply over all groups for the $t^{\text{th}}$ year is:

$$\sum_h k_{ht} \,.\, \dot{x}_{ht}$$

Because we assume identical model specifications and the same data for the representative farms as would be used for programming each individual farm, this estimate should be compared with the ideal estimate:

$$\sum_k \dot{x}_{kt}$$

Any discrepancy that arises between these two estimates is due to aggregation bias.

* This weighting procedure is only correct if the representative farm is the mean farm, or a median or modal farm when resources are symetrically distributed amongst farms.

The aggregate farm approach involves aggregating the resources of a homogenous region or area (not necessarily involving adjacent land) and programming the aggregated variables as a single farm.

Let $\dot{x}_{rt}$ denote the solution vector for the $r^{th}$ region in the $t^{th}$ year, then aggregation bias exists if

$$\dot{x}_{rt} \neq \sum_{k \epsilon r} \dot{x}_{kt}$$

The regional farm approach is generally adopted for recursive linear programming and spatial equilibrium models.

## 2. Avoiding aggregation bias in static linear models

### 2.1 *Day's requirements*

Aggregation bias can only be avoided if farms are classified into groups or regions which are defined according to rigid theoretical requirements of homogeneity. The most comprehensive set of sufficient conditions for multiproduct/multiresource farms have been established by Day.(4)

Let the linear programming model of the $k^{th}$ farm be written in matrix notation as:

$$\begin{aligned} &\text{Max} && c'_k\, x_k \\ &\text{Subject to} && A_k\, x_k \leq b_k \\ & && x_k \geq 0 \end{aligned}$$

where $c_k$ = the vector of activity return expectations
$A_k$ = the input–output coefficient matrix
$b_k$ = the vector of resource and constraint levels
$x_k$ = the vector of activity levels to be determined.

Similarly, let the linear programming model for the aggregate group farm be written in the same way without the $k$ subscripts. That is:

$$\begin{aligned} &\text{Max} && c'\, x \\ &\text{Subject to} && Ax \leq b \\ & && x \geq 0 \end{aligned}$$

where

$$c = \frac{1}{k} \sum_k c_k{}^* \text{ and } b = \sum_k b_k .$$

Day has shown that only farms satisfying the following three conditions may be grouped together if aggregation bias is to be avoided:

(i) $A_k = A$
(ii) $c_k = \gamma_k\, c$
(iii) $b_k = \lambda_k\, b$ where $\gamma_k, \lambda_k \geq 0$.

Day termed the first requirement 'technological homogeneity', and it requires that each farm has the same production possibilities, the same type of resources and constraints, the same levels of technology and the same level of managerial ability.

The second requirement demands that individual farmers in a group hold expectations about unit activity returns which are proportional to average expectations. This requirement is termed 'pecunious' (sic) proportionality.

The third requirement is that the constraint vector of the programming model for each individual farm should be proportional to the aggregate constraint vector. This is termed 'institutional' proportionality. This requirement is strictly only necessary for those constraints which will be binding in the model solution,

---

* $c$ may also be a weighted average.(4, p. 802)

but since these can rarely be identified beforehand, it is usual to demand institutional proportionality in all constraints.

Day's conditions are sufficient to guarantee unbiased supply estimates according to the definitions presented earlier for both the representative and regional farm approaches. The theorem follows because the conditions ensure $\dot{x}_k=\lambda_k \dot{x}$ for all $k$, a result which will be used later in the paper.

A direct consequence of Day's requirements is that each individual farm should be pecuniously and institutionally proportional to the arithmetic mean farm for its group. This means that when using the representative farm approach, the representative farm might as well be defined as the arithmetic mean farm, for while any other farm is proportional and could be used, it would require additional knowledge about the distribution of the proportionality coefficients $\lambda_k$ to permit estimation of group supply.

If the average farm is selected as the representative farm, then the representative farm approach simply reduces to the aggregate farm approach, because the average farm is $1/k$ times the aggregate farm, where $k$ is the number of farms in the group.

### 2.2 *Alternative requirements*

Day's requirements are very demanding, and alternative and less demanding sufficient conditions have been developed to replace the requirement of institutional proportionality.

Sheehy and McAlexander[14] suggested that farms should be classified on the basis of the most limiting resource for each product.* While this approach is most directly applicable to single product farms, it may be applied to multi-product farms through repeated reclassifications.[2, p. 706] Using an example of single product farms, Sheehy and McAlexander showed that use of this 'homogeneous restriction method' led to less aggregation bias than conventional grouping procedures based simply on size and type of farm. This result was subsequently reinforced in a paper by Frick and Andrews.[8]

A closely related set of conditions were later developed by Miller.[12] These are based on the conception that Day's institutional proportionality requirement can be relaxed to allow non-proportional resource variation for an individual farm, providing only that they are not sufficiently large to lead to a different solution basis from that of the average or aggregate farm. More formally, he suggests grouping farms which have (a) identical technology matrices, and (b) homogeneous solution vectors (i.e. the individual farm solution vectors should contain the same activities, though not necessarily at the same levels).

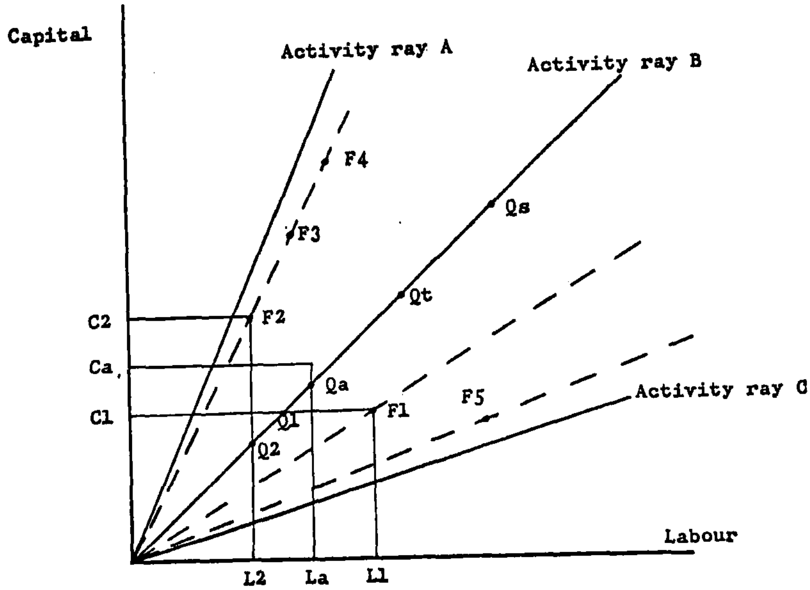
Miller's requirements unfortunately depend upon successful anticipation of the solution vectors of all the individual farm linear programming models, but Lee[11] has demonstrated that by extending Miller's analysis to the dual problem, a more practicable set of conditions can be obtained. These are that all farms in a group should satisfy the condition (a) above, and in addition (b) they should have identical net revenue expectations, and (c) the range of resource ratios should be such that the dual solution vector is the same for all farms in the group.

It is instructive to demonstrate geometrically the relationship between the sets of conditions suggested by Day and Lee.† Consider a simple situation where five farms have in common three productive processes using two resources, capital and labour. This situation is illustrated in Figure 1. Considering for the moment

* It was assumed that this would apply to farms which were technologically homogenous.

† The following is a composite from Schaller[13] and Lee.[11]

**Figure 1**




just two of the farms, F1 and F2 with resources (C1, L1) and (C2, L2) respectively, producing only one product on the production ray B, then farm 1 would produce $Q1$, farm 2, $Q2$, and the aggregate supply would be $Qt=Q1+Q2$. If the two farms were grouped together and represented by the average of the two, then the resulting farm would have Ca capital and La labour, and its output would be $Qa$, with aggregate supply $Qs=2Qa$. This conveniently demonstrates the aggregation bias as $Qs-Qt$. The problem is to group these farms in such a way as to minimise or eliminate the bias. In using the sets of conditions of both Day and Lee, we assume that net revenue expectations are identical for all the farms. Given that all our five farms are facing the identical production functions (represented by the three rays A, B and C in resource space in Figure 1) we need only consider the third conditions of institutional proportionality and identical dual solution vectors respectively.

Farms satisfying Day's institutional proportionality condition will lie along a ray through the origin (Figure 1). Thus in the current example three groups are required to eliminate aggregation bias; farms F2, F3 and F4 in one group and F1 and F5 each in separate groups. Lee's third condition requires that all farms be grouped together which have resource ratios of capital to labour lying between the ratios implied by two adjacent activity rays. Following this condition only two groups of farms are required for exact aggregation; F2, F3 and F4 in one group and F1 and F5 in the other. In general Day's condition of exact proportionality is unnecessarily severe, and adherence to his rules leads to defining more groups than are necessary for exact aggregation. In fact, Lee's requirements are not only less severe than Day's but actually become identical to Day's as the number of activities tends to infinity.

It can also be demonstrated that a formal specification of the discussion by Sheehy and McAlexander results in a grouping of farms which is always identical to that produced from Lee's requirements.* Hence the three sets of conditions proposed by Sheehy and McAlexander, Miller and Lee are identical, and furthermore, are directly related to Day's more demanding set of requirements.

A radically different approach to aggregation bias in farm supply analysis has been made by Fisher.(7) Recognising that *elimination* of aggregation bias is not a practical objective, Fisher uses a systematic cluster analysis technique to discover the grouping which *minimises* the bias.

One of the principal problems which confronts the application of both this and all the above sets of conditions is that full fore-knowledge of the technology matrices of individual farms is required. Given that one of the main *raison d'être* of representative farm programming is to avoid the need for full data on the individual units, this problem is far from trivial. Without exception, the literature cited above offers no help in this problem: any empirical work reported at all refers to farm populations of less than 100 (and usually farms of the same type). There is, therefore, a considerable hurdle to be cleared in applying these conditions to the 200,000 or so farms currently being considered in our supply analysis of the U.K. Ignorance of input–output coefficients for individual farms is most critical in applying Fisher's method and the homogeneous restriction method, where the actual coefficients are required for the analysis. However, in applying either Day's or Lee's sets of conditions, one possible solution is to delineate a grouping of farms based on the other requirements, and then to disaggregate the resulting groups where there is evidence of differences in technology.

Given this general approach, Day's conditions are the most manageable for the multiproduct/multiresource case, especially as Lee's conditions are not readily generalised for the multidimensional case.†

---

* This may be proved by noting that in essence Lee's third condition compares the ratio of resources with the slopes of the activity rays. Thus in the two activity, two resource case (using Lee's notation) $\frac{b1}{b2}$ is compared to $\frac{a11}{a21}$ and to $\frac{a12}{a22}$. There are three possible results:

(i) $\frac{b1}{b2} > \max\left(\frac{a11}{a21}, \frac{a12}{a22}\right)$

(ii) $\mathrm{Min}\left(\frac{a11}{a21}, \frac{a12}{a22}\right) < \frac{b1}{b2} < \max\left(\frac{a11}{a21}, \frac{a12}{a22}\right)$

(iii) $\frac{b1}{b2} < \min\left(\frac{a11}{a21}, \frac{a12}{a22}\right)$

Suppose for a particular group of farms which satisfy Lee's three conditions the outcome was the second case, and that $\frac{a11}{a21} > \frac{a12}{a22}$.

Then $\frac{a12}{a22} < \frac{b1}{b2} < \frac{a11}{a21}$

holds for each of the farms. Consider the second two terms of this inequality, $\frac{b1}{a11} < \frac{b2}{a21}$, i.e. for these farms resource 1 is more restricting in the production of activity 1 than resource 2. Similarly looking at the first two terms, $\frac{b1}{a12} > \frac{b2}{a22}$ resource 2 is most restricting for activity 2. Thus for the particular group of farms to which these results apply, each activity has the same most restricting resource, and Lee's conditions are identical to those of Sheehy and McAlexander. This result generalises for cases (i) and (iii) above, and for any number of activities and resources. Thus the two methods are exactly equivalent.

† In his paper Lee did promise that such a generalisation was being undertaken, but no results of such work have yet appeared.

## 3. Implications for farm classification in the U.K.

### 3.1 *Adaptation of Day's conditions*

The conclusion of the last section was that the requirements as set out by Day were the most operationally convenient for farm classification and that the application could proceed in two stages.

(i) The classification of farms on the basis of requirements (ii) and (iii) (Section 2.1).

(ii) The disaggregation of the resulting groups in order to approximate more closely requirement (i).

The second stage does not readily lend itself to any general systematic approach. An example will illustrate the principles involved. Assume results of stage 1 included a group consisting of dairy farms in the South West of England. In stage 2 the task would be to discover if there are radical differences in the technology of dairying within this particular subset of farms. It might be decided that the principal determinants of differences in technology in dairying are breed type and herd size. Therefore, on the basis of data on the distributions of dairy breeds and herd sizes for the South Western dairy farms, it might be decided to disaggregate the group into four sub groups with large herds and small herds of Friesians and non-Friesian breeds. Similarly, one might further disaggregate according to management standard, land ownership status and so on as far as data could be found to show differences in technology coefficients. Finding such data is likely to prove difficult and so steps should be taken to reduce the need for disaggregating groups in stage 2. One such step is to introduce a regional grouping in stage 1, in this way differences in technology based on climate, soil type and altitude may be taken into account.

Since the general function of a supply model is to predict the output response to a given set of prices and costs, adoption of this approach should also assist in the attainment of pecunious proportionality. This is further reinforced by the high degree of governmental interference in prices which enables us to assume fairly uniform price expectations for large areas of the country. The remaining components of pecunious proportionality are yields. To the extent that yields are explained by regional differences, an appropriate breakdown will ensure reasonable uniformity of yields within a group. Also, any disaggregations in stage 2 taking into account management level may help to ensure greater uniformity of yields within a group.

### 3.2 *Grouping farms for institutional proprotionality*

We are now left with the problem of deriving a grouping of farms in which institutional proportionality holds within the groups. It was demonstrated in Figure 1 that Day's third condition could be expressed geometrically as the requirement that all farms in the same group should lie along a ray through the origin when the farms are plotted in resource space. Thus is we could plot the characteristics of our population of farms in resource space, then by inspection we could discover the groups of farms which lie approximately along rays through the origin. In practice a multivariate technique is needed for finding such groups. One such technique is known as cluster analysis.* It can be used to discover (a) the number of clusters or groups of farms required, and (b) the members of each group. The analysis cannot, however, be carried out on the points plotted in resource space directly. Figure 2a shows 13 farms plotted in three-dimensional resource space where the farms fall approximately along two distinct rays through

* This term will be explained in section 3.3.

the origin, i.e., the 13 farms fall into two groups according to the requirement of institutional proportionality. If these 13 points were to be grouped in order to minimise some measure of within group 'distance', then points 1, 2, 3, 7, 8 and 9 would tend to fall into the same group, and points 4, 5, 6 and 10, 11, 12, 13 would form two other groups. In this case the proportionality requirement would be violated for the group of small farms. To avoid this problem, the cluster analysis is carried out not on the original points, but on the co-ordinates of the points after they have been projected onto a plane* as shown in Figure 2b (a hyperplane when we have more than three original dimensions). In this way the points which lie closely along a ray will tend to be grouped together as desired.

Figure 2A

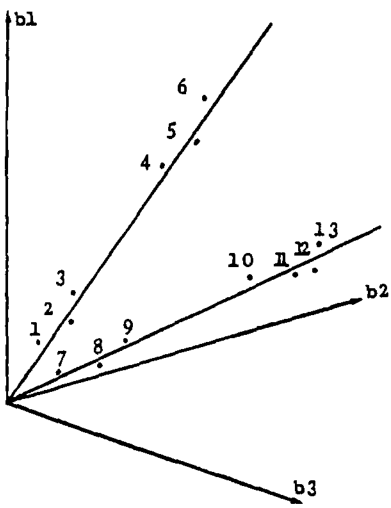
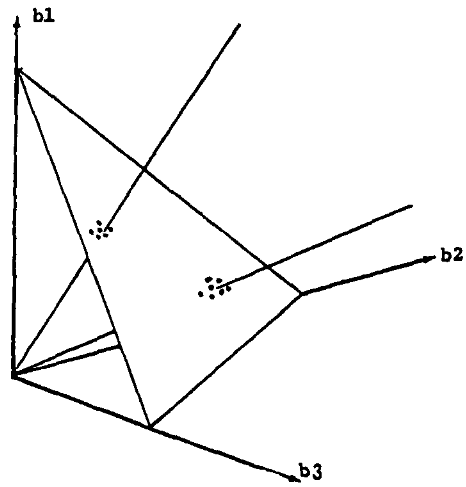



Figure 2b




### 3.3 *Cluster analysis*

Because these techniques have not yet received much attention from agricultural economists, and because there is considerable confusion and disagreement over terminology in the literature in which these techniques are used, this section briefly outlines a definition of cluster analysis and one possible usage of it for farm classification work.

Following the lead of Lance and Williams,(10) the term cluster analysis will be reserved for the purpose of this paper as a numerical technique which sorts a number of objects into a predetermined number of exhaustive and mutually exclusive groups which maximise some criterion of intra group homogeneity. In contrast are those numerical processes which optimise a hierarchical route from individual elements to the population considered as a whole. In our use of these techniques for farm classification there are in fact elements of both approaches. This arises because we use a hierarchial agglomerative process initially to help describe the nature of the data and more specifically to enable a decision to be made on the number of groups required. Having decided on the

* The simplest and most convenient projection to use is to replace the original co-ordinates of each point in $n$-dimensional space $b_1, b_2, b_3 \ldots, b_n$ by $b'_1, b'_2, b'_3 \ldots, b'_n$ where;

$$b_i = b_i / \sum_n b_i \text{ for all } i = 1, n.$$

We are grateful to K. J. Thomson for help on this point.

number of groups, a cluster analysis enables the optimum grouping for that number of cells to be obtained.

The hierarchial method itself involves two components, namely a measure of the similarity or likeness of the objects which is used to judge the successive stages in the grouping process, and a strategy by which the process takes place. There are numerous examples of both criteria and strategies; Ball[1] and Freidman and Rubin[9] discuss many of the commonly found criteria of similarity, and Lance and Williams,[10] and Williams and Dale[16] discuss strategies by which these criteria are used to discover optimal hierarchies. In our own farm classification work at Newcastle we have chosen the error sum of squares (E.S.S.) as a criterion of within group similarity, and centroid sorting as an appropriate strategy.

Given our particular criterion and strategy, the computational procedure is an iterative process involving basically two steps. Firstly the similarity coefficients between all the objects are calculated, and the two objects having the least 'distance' between them (i.e. the greatest similarity) are merged. Secondly the group formed by the merging of the two objects is now represented by its centroid.* These two steps above are repeated and successive fusions eventually involve groups of objects merging so that finally there is just one group containing all the objects. One problem with this hierarchial procedure is that once objects have been merged with a group, they cannot leave the group. This problem is overcome by allowing objects to change groups in the next stage of the analysis. As the fusions occur the within group error sum of squares is recorded. This statistic is initially zero when all objects are in individual groups, and it increases as objects are fused. It reaches a maximum of $n$ (the number of objects being considered)† when all the objects are in the same group. By examining the value of this statistic, and changes in it as fusions occur it is possible to make a judgement of the number of groups inherent in the data. This decision marks the end of the first part of the analysis. The second part is to use the methods of cluster analysis proper to optimise the grouping of objects for the decided number of groups.

The essence of the cluster analysis is that the similarity of each object with the centroid of each group is measured, and if the within groups error sum (E.S.S.) can be reduced by moving an object into a different group then this is done. The relocation process is repeated until no further improvement of the E.S.S. occurs. Hence the rigidity of the hierarchical procedure mentioned above may be overcome.

A package computer programme 'Clustan'[17] is available which enables the above procedures to be followed, and has been successfully used at Newcastle in the construction of an initial U.K. farm classification.

## 4. Avoiding aggregation bias in dynamic linear models

### 4.1 *Aggregation theorems*

We now return to Day's theorem and examine its applicability to dynamic models. Day has shown that the theorem does apply for a recursive linear programming model.[5] However, the implications deserve detailed exploration because they are rather more stringent than generally recognised.

---

* That is the vector of co-ordinates of the centre of gravity of the group.

† This is, of course, only true of normalised data, i.e. where the data matrix $X$ of $p$ observations on $n$ objects is transformed to $X^*$ where the vector of observation averages is the zero vector, and the observation variances are all unity.

In a dynamic model aggregate supply estimates are obtained for each year of the supply forecast period. If each of these estimates is to be unbiased then farms must be classified so that within each group:

$$\left.\begin{array}{lll}\text{(i)} & A_{kt} = A_t & , \\ \text{(ii)} & c_{kt} = \gamma_{kt} c_t & , \\ \text{(iii)} & b_{kt} = \lambda_{kt} b_t & , \end{array}\right\} \text{for all } t$$

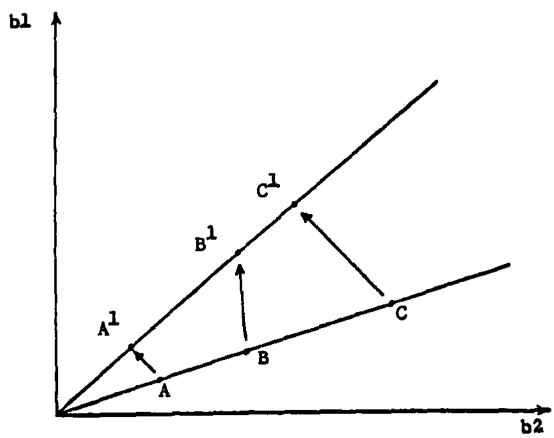
where again $k$ subscripts refer to the $k^{\text{th}}$ farm, and non-subscripted variables refer to the aggregate farm. These conditions are sufficient to ensure $\dot{x}_{kt}=\lambda_{kt}\,\dot{x}_t$ for all $t$.

As with static models, the first condition requires the same activity possibilities, the same types of constraints, the same technology and managerial ability, and in addition identical rates of technological innovation over time.

The second condition requires proportional activity return expectations each year, thus requiring that any changes in expectations must be proportional, although the proportionality coefficient, $\lambda_{kt}$, need not be constant over time for the individual farm.

The third requirement implies that any changes in the constraint ratios of the aggregate (or average) farm are matched by proportional variations by all the individual farms. That is, in the base year all the individual farms should lie on a single scale ray (of their common production function), and if the average farm moves onto a new ray, then so must *all* the individual farms. The two-dimensional case is illustrated in Figure 3, where $b_1$ and $b_1$ denote two resources, and two

**Figure 3**




scale rays are $t$ and $t+1$. Here, the average farm in year $t$ is denoted by C, and is located on the scale ray $t$ together with all the individual farms. By the following year however, the average farm has changed its resource ratio, and has moved to $C^1$ on scale ray $t+1$. The third condition requires that all the individual farms act in a similar way by also moving to ray $t+1$.

Exactly how the individual farms can move relative to the average farm will depend on whether or not $\lambda_{kt}$ is a constant over time. If $\lambda_{kt}$ is a constant, then the $k^{\text{th}}$ farm must retain its relative scale relationship to the average farm. For example, farm A in the diagram moves from A to $A^1$ and $OA/OC=OA^1/OC^1$. Constant $\lambda_{kt}$ also means that if the average farm were to expand along the same scale ray, the individual farms must also move along the ray to retain their relative scale sizes.

If $\lambda_{kt}$ is not constant over time then the $k^{th}$ farm can adopt any scale size relative to the average farm. Thus, for example, farm B moves from B to $B^1$ and OB/OC is larger than $OB^1/OC^1$. If the average farm were to expand up the same scale ray, however, it is no longer necessary for all the individual farms to follow suit, and some may even accelerate their rate of expansion relative to the average farm.

Consider now the practical implications of these two possibilities.

If $\lambda_{kt}$ is a constant over time for the $k^{th}$ farm, then any change in the constraint levels of the average farm, denoted by $\Delta b_{it}$ for the $i^{th}$ constraint in the $t^{th}$ year, must be matched by a change of $\lambda_k \ \Delta b_{it}$ by the $k^{th}$ farm. Further, because $\dot{x}_{kt}=\lambda_k \dot{x}_t$, this also means that any change in the activity levels of the average farm, denoted by $\Delta \dot{x}_{jt}$ for the $j^{th}$ activity in the $t^{th}$ year, must be matched by a change of $\lambda_k \ \Delta \dot{x}_{jt}$ by the $k^{th}$ farm in the same year. For example, if the average farm increased its dairy herd size in the $t^{th}$ year, then all the individual farms must increase their herd sizes in the same year, and furthermore, do so by fixed proportional amounts. This is quite a demanding requirement when the lumpy nature of most farm investments is considered. The magnitude of changes in dairy herd size over time for an individual farm, for example, are largely determined by the need to achieve certain practical size units which are compatible with integer values of men and the discrete nature of machinery and building investments. Farms would probably have to be almost identical in their initial resource and enterprise structure if dairy herd size changes were to truly satisfy the fixed proportionality requirement.

This problem is particularly acute in recursive linear programming models, where flexibility constraints representing average or aggregate farm behaviour are imposed, and which tend to constrain changes in activity levels to small incremental units at the average farm level, thus necessarily imposing even smaller changes for the smaller than average farms. In fact the question arises as to whether recursive constraints can exist under the present assumptions for those activities which involve lumpy investments. Lumpy investments tend to require lenthy periods of capital accumulation for individual farms, so that if all farms which expand at the same time were grouped, the observed average farm would not change smoothly over time. Rather, the average farm's adjustments would follow a step function. The argument can also be extended to resources. If an individual farmer buys land he usually obtains a discrete package, and assuming he increases all other inputs to maintain proportionality with the average farm, his discrete change in resource structure will define a definite leap or bound in his farm plan. Again, if all the individual farms were to obtain additional lumps of land at the same time, a smooth recursive relationship would not be observed at the aggregate level for changes in activity levels.

If $\lambda_{kt}$ are constant over time, and if farms are classified to satisfy the requirements for avoiding aggregation bias, it would seem that not only are recursive linear programming formulations suspect, but that the actual average or aggregate farm model would have to take account of the lumpy nature of farm investments and activity levels, presumably through integer programming specifications.

It is apparently very desirable that $\lambda_{kt}$ should be allowed to vary over time, so that individual farms are free to make small, or even zero, changes in their resource structure and farm plans in some years (providing proportionality is maintained if the average farm changes its constraint ratios), whilst making substantial changes periodically as capital accumulation, or availability of resources permits. Such a possibility allows a more heterogenous group of farms with respect to dynamic behaviour, without necessarily introducing aggregation bias. This possibility also solves the problem of lumpy investments and activity levels, for while an individual farm model would require an integer programming specification, this is no longer necessary for the average or aggregate farm model where the 'lumps' smooth out.

One remaining difficulty, which cannot be resolved through variable $\lambda_{kt}$, is that leaps and bounds around the average farm are not acceptable when a change in the solution basis occurs. Thus if the average farm adopts a new activity in year $t$, so must all the individual farms. While expansion of existing activities can proceed at varying rates, decisions to adopt or abandon activities must be made in the same year. This problem is particularly difficult with Day's formulation of a recursive linear programming model, because real activities can never leave the solution basis, but only tend to asymptotic limits of zero. This means that if the average farm has a low level of a lumpy activity, it is necessarily assumed that the majority of farms have disbanded the activity, and only a few have retained it. Proportionality cannot be maintained under these conditions, and aggregation bias must exist.

### 4.2 *Implications for dynamic model specification*

If $\lambda_{kt}$ are variable over time this permits a grouping of more heterogenous farms on the basis of dynamic behaviour. However, in order to avoid aggregation bias it is necessary that the model specifications also be compatible with variable $\lambda_{kt}$. It is now shown that in order to ensure this, there are restrictions on model specification for both single and multiperiod models. These requirements are quite stringent, and if they cannot be met whilst maintaining a satisfactory description of the farms then one must either accept aggregation bias or be willing to strive for more homogenous groupings of farms. The latter option also requires that formal account be taken of the lumpy nature of farm investments and activity levels, and that a recursive linear programming specification is avoided which is then invalid.

*The requirement for single period models* is that the model cannot have any non-zero constraint elements that are either equal to activity levels in the previous year or equal to activity levels plus one or more constraint levels in the previous year.

As already seen Day's theorem is sufficient to ensure that

$$\dot{x}_{kt} = \lambda_{kt}\ \dot{x}_t$$

for all $t$. One of the requirements was institutional proportionality, namely

$$b_{kt} = \lambda_{kt}\ b_t.$$

Consider now what happens if some element of $b_t$, say the $i^{\text{th}}$, is determined by the $j^{\text{th}}$ activity level in year $t-1$, together with the actual constraint level in $t-1$. For example, we might be considering a resource level, the value of which in year $t$ is dependent on the supply in year $t-1$ together with the amount purchased through a resource buying activity in year $t-1$. This can be written as

$$b_{it} = b_{it-1} + x_{jt-1}.$$

For the $k^{\text{th}}$ individual farm this means

$$b_{ikt} = b_{ik\ t-1} + x_{jkt-1}$$

or

$$b_{ikt} = \lambda_{k\ t-1}\,(b_{i\ t-1} + x_{j\ t-1}).$$

It inevitably follows that the solution to the $k^{\text{th}}$ farm problem in year $t$ must be proportional to the solution of the average farm by the same proportionality factor, $\lambda_{k\ t-1}$, as in the previous year. That is, $x_{ikt} = \lambda_{k\ t-1}\ x_{it}$. This means that $\lambda_{kt}$ cannot vary over time so that leaps and bounds are not possible. The same result follows if $b_{it} = x_{j\ t-1}$, as sometimes occurs with control rows in a linear programming matrix, or with the more general situation involving several activity or constraint levels.

There appear to be only two specifications which avoid this problem. Either the offending constraint must be arranged so that $b_{tt}=0$ (because any proportionality factor holds against zero), or else it is necessary to have an exogenous component in the constraint level, the magnitude of which is determined entirely 'outside' the model. Leaps and bounds must then involve changes in the exogenous components.

An example of the latter might arise with a dairy cow building constraint. In any one year the building constraint could be specified to increase through a building investment activity, but an exogenous component might also arise if the individual farms periodically attach additional lumps of land to their farm, which are likely to include additional cow housing facilities. The decision to invest in new land is presumed to be determined independently of the model, but this is quite reasonable in that it would take a highly sophisticated (and probably non-linear) model to satisfactorily incorporate an explanatory system of land purchasing activities. A leap for the individual farm involving dairy expansion would then be tied to an increase in farm acreage size. However, a leap must also involve simultaneous changes in all the other exogenous constraint variables in the model, and these changes must retain proportionality to the average farm. Unless such exogenous variables can be satisfactorily introduced, it is unlikely that model specification can comply with variable $\lambda_{kt}$ coefficients.

*The requirement for multiperiod models* is that the specification must include exogenous components for all non-zero constraints, and the model must be solved anew each year with updated values for these exogenous components. This follows because the multiperiod solution vector for the individual farm is known to be proportional to the multiperiod solution vector of the average farm by a single factor of proportionality. Obviously this implies $\lambda_{kt}$ do not vary over time. $\lambda_{kt}$ can only change each year if the model is re-solved each year with revised exogenous components in the constraint levels. In practice, the requirement essentially simulates farmers who plan their farms with definite multiperiod goals in mind, but who also reconsider their decisions as far as possible each year according to the outcome of various events not described in the model. For example, decisions are reconsidered in the event of success in acquiring an additional lump of adjacent land.*

Both the requirements discussed above are made conceptually more difficult if leaps and bounds involve changes in resource ratios, for then those farms making leaps in any given year change the resource ratios of the average farm, and this requires adjustments in the exogenous components of all other farms in the same year to maintain institutional proportionality. This problem also arises in recursive linear programming models when the flexibility coefficients for different activities are different, and hence the constraints grow at different rates.

The requirements for model specification developed above require that an important part of a farmer's decision problems remain exogenous to the model, and worse, that these decisions comply with requirements for continued institutional proportionality over time. A possible exception arises with single period models if the farm models can be constructed so that all offending constraint values are zero. However, our experience in constructing models at Newcastle suggests that realistic farm models for the United Kingdom cannot be constructed under this condition.

We conclude that farms ought to be classified on the basis of constant $\lambda_{kt}$ over time, and hence farm models ought to be constructed which (a) avoid a

* Decisions might also be reconsidered on the grounds of changes in activity return expectations, but it is changes in the exogenous resource components that are necessary to meet the requirement.

recursive linear programming specification, and (b) take formal account of the lumpy nature of farm investments and activity levels through integer programming specifications. Unfortunately, in constructing a model of the size required for the United Kingdom it is hard to achieve these ends. This is particularly so when there are other and competing objectives in the construction of a model such as an adequate description of the dynamic behavioural and institutional constraints which act on supply adjustments. Consequently, it may be necessary to build models which necessarily imply some degree of aggregation bias.

4.3 *Implications for farm classification in the U.K.*

The dynamic version of Day's theorem requires that farms should be classified not only according to the criteria discussed in section 2.1, but also to ensure that they satisfy these criteria for each year of the supply forecast period. This does not mean that the technology matrix and the proportionality coefficients must remain constant from year to year, but only that a set of relationships exist as specified in section 4.1. Our conclusion that constant scale relationships (constant $\lambda_{kt}$) are required for realistic farm model specification increases the degree of farm homogeneity that is necessary. However, even this requirement can be partially avoided if farms are permitted to jump between groups over time.*

The clustering technique proposed for minimising aggregation bias in static models can easily be extended for use with dynamic models. This is achieved by testing the stability of the classification scheme derived under the static conditions over the recent past, and then disaggregating groups in which differences in behaviour are observed.

**Summary and conclusion**

This paper explored the implications of aggregation bias for farm classification and model specification in the types of microeconomic models that would be useful for evaluating current problems of United Kingdom agricultural policy. More specifically, the import of Day's[4] aggregation theorem has been explored for dynamic linear microeconomic models of supply response.

It has been shown that unbiased aggregate supply estimates are only attainable if very stringent homogeneity criteria are applied in the classification of farms. In addition to the well-known results for static models, it is also necessary that farms within a group grow in proportional ways and have identical rates of technological innovation. In an attempt to relax the classification criteria, the possibility of grouping farms which grow in proportional but erratic ways was pursued, since this is most compatible with the lumpy nature of many farm investments and activity levels. It was shown that if farms were classified in this way, the representative or aggregate farm model built would have to comply with certain restrictions in order to be consistent with the behaviour of individual farms. Experience at Newcastle has shown that these requirements do not always permit realistic farm models, and they often require assumptions that are hard to maintain. Failure to satisfy these requirements was shown to be particularly damaging for recursive linear programming model specification.

Because it is not possible to construct a microeconomic model for the United Kingdom which eliminates aggregation bias yet involves sufficiently few groups of farms to be computationally feasible, a classification procedure which uses

* In the Newcastle model forecasts of changes in the size and numbers of farms are predicted independently of the actual linear programming model. In estimating these forecasts a statistical model is used which takes direct account of movements of individual farms between classification groups. In this case we do have to assume that farms satisfy the proportionality requirements with the new group which they join.

cluster analysis was proposed which attempts only to minimise aggregation bias. This procedure has been used at Newcastle to classify farms according to Day's static requirements, but is applicable to the dynamic requirements if used in conjunction with time series data.

We conclude that, while in principle, aggregation bias could be eliminated, some degree of bias is inevitable in the actual construction of a dynamic micro-economic model of United Kingdom agriculture using linear programming techniques. While the most important source of bias is likely to arise through the impracticality of meeting the stringent classification criteria in classifying farms, bias is also likely to be introduced through specifying farm models which assume farms change in more homogeneous ways than is compatible with other features of the models. The magnitude of aggregation bias, and how it might change through the forecast period in a dynamic model, are not known, but in our view the advantages of the linear programming approach are so great at the present time that one can usefully proceed with the construction of microeconomic models.

References

1. **Ball, G. H.**: 'Data Analysis in the Social Sciences: What about the Details?' *Proceedings Fall Joint Computer Conference, 1965*, pp. 533–559.
2. **Barker, R. and Stanton, B. F.**: 'Estimation and Aggregation of Firm Supply Functions'. *Journal of Farm Economics*, Vol. 47, August 1965, pp. 701–712.
3. **Davey, B. H. and Weightman, P. W. H.**: 'A Microeconomic Approach to the Analysis of Supply Response in British Agriculture'. *Journal of Agricultural Economics*, Vol. XXII, September 1971, pp. 297–320.
4. **Day, R. H.**: 'On Aggregating Linear Programming Models of Production'. *Journal of Farm Economics*, Vol. 45, November 1963, pp. 797–813.
5. **Day, R. H.**: *Recursive Programming and Production Response*. North Holland, Amsterdam, 1963.
6. **Day, R. H.**: 'Dynamic Coupling, Optimizing and Regional Interdependence'. *Journal of Farm Economics*, Vol. 46, May 1964, pp. 442–451.
7. **Fisher, W. D.**: *Clustering and Aggregation in Economics*. John Hopkins Press, Baltimore, 1969, pp. 70–89.
8. **Frick, G. E. and Andrews, R. A.**: 'Aggregation Bias and Four Methods of Summing Farm Supply Functions'. *Journal of Farm Economics*, Vol. 47, August 1965, pp. 696–700.
9. **Friedman, H. P. and Rubin, J.**: 'On Some Invariant Criteria for Grouping Data'. *Journal American Statistical Association*, Vol. 62, December 1967, pp. 1159–1178.
10. **Lance, G. N. and Williams, W. T.**: 'A General Theory of Classificatory Sorting Strategies. 1. Hierarchical Systems'. *Computer Journal*, Vol. 9, 1967, pp. 373–380.
11. **Lee, J. E.**: 'Exact Aggregation – A discussion of Miller's Theorem'. *Agricultural Economics Res.*, Vol. 18, 1966, pp. 58–61.
12. **Miller, T. A.**: 'Sufficient Conditions for Exact Aggregation in Linear Programming Models'. *Agricultural Economics Res.*, Vol. 18, 1966, pp. 52–57.
13. **Schaller, W. N.**: 'Aggregation'. Paper presented at the Southern Regional Workshop on Simulation.
14. **Sheehy, S. J. and McAlexander, R. H.**: 'Selection of Representative Bench Mark Farms for Supply Estimation'. *Journal of Farm Economics*, Vol. 47, August 1965, pp. 681–695.
15. **Stovall, J. G.**: 'Sources of Error in Aggregate Supply Estimates'. *Journal of Farm Economics*, Vol. 48, May 1966, pp. 477, 479.
16. **Williams, W. T. and Dale, M. B.**: 'Fundamental Problems in Numerical Taxonomy'. *Advances in Botanical Research*, Vol. 2, 1965, pp. 35–68.
17. **Wishart, D.**: *Clustan I.A. User Manual*. Computing Laboratory, University of St. Andrews, Scotland, 1969.