IMPROVING AGGREGATION VALIDITY

by John E. Lee, Jr.*

In recent years, increasing concern has been expressed in the literature of our profession over the problems of aggregation in models of agricultural demand and supply (1, 3, 4, 5, 7, 8, 12). Most frequently the concern has been with the validity of aggregated supply estimates when micro supply situations are simulated in linear programming models.

The aggregation problem stems from the fact that while most economic activity originates with individual firms and households, economists, with increasing frequency, want to say something about the behavior, in the aggregate, of all or a group of such units. One obvious approach is to build up to the aggregate, firm by firm. This is not feasible in an atomistic economy such as agriculture. A second approach is to ignore the individual units by working with observed or postulated relationships among economic aggregates. A third alternative, especially for an atomistic or purely competitive economy, is to approach the aggregate through "representative" units or through the analysis of subaggregates.

The latter approach has been particularly popular in agricultural supply research. The so-called Regional Adjustment Studies (NC-54, S-42, etc.) have used linear programming models of representative farms to derive regional estimates of supply schedules for the major commodities. The basic units of analysis for a model of national agricultural production developed by the Economic Research Service are representative resource situation subaggregates. Linear programming models of these units are used to estimate aggregate supply response to economic stimuli. These are but two examples of the micro-oriented approach to estimation of aggregate behavior--an approach that has almost become conventional in farm management and production economics research.

* Agricultural Economics, Production Adjustments Branch, Economic Research Service. This version of the paper has benefited greatly from the suggestions of Lee M. Day. The author also appreciates the comments of Fred Abel, Glen T. Barton, and W. Neill Schaller. The views expressed are the author's and do not necessarily reflect the views of the U.S. Department of Agriculture.

Numbers in parentheses refer to Literature Cited at end of paper.
The concern for the validity of aggregate estimates of supply derived from the approach typified in these models is justified. There are at least three sources of invalidity in that approach:

1. The failure to be consistent with respect to time in the formulation and specification of supply models;

2. The failure to recognize that constraints on economic activity at higher levels of aggregation may be different than at micro levels; and

3. The failure to systematically group farms and construct units of analysis so as to recognize the interfarm variation in forces that shape economic behavior.

These failures are interrelated; thus, the distinctions between them are somewhat arbitrary. But the breakdown provides some "handles" on the aggregation problem and an organization framework for the discussion which follows. I will work from conceptual and theoretical benchmarks, but the emphasis in this paper will be on practical guidelines for improving aggregation validity in models of agricultural supply.

It seems appropriate to begin with a look at the aspect of the aggregation problem which has been discussed most frequently. The problem is that of exact aggregation; or simply, the problem of grouping farms so that analyses of aggregate behavior will be free from bias or error attributable to the grouping process itself.

**Grouping Microunits for Aggregation Validity**

**Recent Conceptual Developments**

Richard H. Day (3) has shown that conditions sufficient for exact aggregation are proportional variations of resources and behavioral "bounds," proportional variation of net return expectations among all farms in the aggregate, and finally, common technical coefficients which appear in the constraints on the farm's decisions. Tom Miller (7), searching for less restrictive conditions for exact aggregation, has shown conceptually that the responses of different farms to a given set of relative product prices will be proportional if the farms have homogeneous activity vectors and if the same activities appear in the linear programming solution vector for each farm. In other words, farms that have the same enterprises in the optimum enterprise mix can be grouped without bias or error. While Miller's conditions are less restrictive than Day's (groupings of farms under Day's conditions

-341-
would be subsets of groupings under Miller's conditions), they are of limited practical value because they are defined as requirements of the solutions to the individual farm problems, rather than as observable characteristics of the farms themselves.

In another recent paper, Sheehy and McAlexander (13) hypothesized that if farms could be sorted into groups having the same absolute restriction (limiting resource) on output, benchmark farms based on the averaging of resource levels within each group would give unbiased aggregate results. They proceeded to apply these grouping criteria only to single-product firms and not to the more general multiple-product, multiple-resource case where variations in relative prices must be recognized.

Thus, several writers have advanced sufficient conditions for exact aggregation. However, these have been rather restrictive conditions or have applied only to special cases. None of these efforts answered Lee Day's (1) call for research to determine the necessary conditions for exact aggregation and to determine the magnitude and direction of error at various prices associated with different levels of variance in the proportionality of resources. We turn to that task now.

The Necessary Conditions for Exact Aggregation

In an extension of Miller's analysis, this writer (6) showed how the "qualitatively homogeneous solution vector" conditions could be translated into observable characteristics of the farms themselves. This was done by solving the dual to Miller's primal problem. In so doing it was revealed that the shadow prices (marginal value products) of the resources were the same for all farms which Miller showed could be grouped without aggregation error. Further, these shadow prices were constant over the range of resource ratios represented by the aggregated farms. The task remaining was the practical one of determining the exact ranges of resource ratios over which the marginal value product was constant.

In the article to which I just referred, simple graphics were used to demonstrate that, in effect, the activity vectors in a linear programming situation represent marginal value product "borders." All farms with combinations of resources bounded by the same marginal value product "borders" or activity vectors have the same shadow prices in the dual, have the same activities in the primal solution vector, and can be aggregated without error. The number of cells or bounded areas represents the maximum number of groups of farms needed for exact aggregation.
Working with the dual counterpart to Miller's theorem rather than with the primal was a matter of convenience rather than necessity. All farms whose resource coordinates lie between the same two feasible activity vectors will maximize revenue with some combination of those same vectors, thus meeting the "qualitatively homogeneous solution vector" conditions of Miller's primal theorem.

From this exercise it was apparent that the key to determination of the resource ratios relevant to error-free grouping of farms is the relationship between the ratios in which resources are required by alternative activities and the ratios in which resources are available to farms. This observation has been used to develop an exact aggregation algorithm for multiple-product farms in a completely general price situation.

The difficulty with multiple-product farms is that the resources available to a given commodity (activity) are subject to change as prices (and thus the relative profitability of the activities) change. The answer to that difficulty, of course, is to recognize all possible orderings of the relative profitability of activities (that is, to recognize all the possible orders in which activities could come into the linear programming solution). This can be done in the following way. First, all the relevant combinations of two resources are determined. For example, if one or more activities use all three of resources A, B, and C, the relevant combinations of two resources would be AB, BC, and AC. Then the critical ratios of each of these combinations of two resources are determined from the technical coefficients of the activity vectors. In other words, the boundaries to the critical ranges of each resource ratio are the ratios in which the various activities use these resources. Then farms are grouped according to the critical resource ratio range in which each combination of two resources falls. Sequential sorting on the basis of all possible relevant combinations of two resources (assuming that prices of all outputs could vary from zero to infinity) produces a number of groups of farms free of aggregation error.

Since the same conditions must be met for each relevant resource ratio by the farms in each group, the order of the sequential sorting process will not affect the composition of the final groups of farms. Because for each farm in each group the output of each commodity will be limited by the same restraint at any given price, the shadow prices for the resources of each farm are identical and the conditions for zero errors are met.

In fact, this "critical resource ratios" grouping algorithm provides the minimum sufficient conditions (and therefore the necessary conditions)
for exact aggregation in a completely general model. If any farm in any one of the groups derived meets anything less than these conditions (that is, if any of its resource ratios fall outside the critical ranges which delineate that group) there will be aggregation error at some combination of prices.

The exact aggregation process can be made general with respect to technology and time as well as to prices. This can be done by expressing technological capacity as restraints, and variations in operating technology (crop and livestock practices) as alternative activities. In the case of multi-period models, all known technologies which are possible alternatives over the whole sequence of time periods can be included as vectors in the original grouping procedure.

The Concept of Error "Trade-off"

It should be obvious that in a problem of any size, the number of aggregation-error-free groups delineated by the critical-resource-ratios method just outlined will be very large. The natural question then is, "What effects on costs and quality of information would result from combining some of the exact aggregation groups?"

Previously, this question of a "trade-off" between analytical costs and aggregation error has been thought of in terms of representing a heterogenous population with a single representative farm and then proceeding to increase the number of representative farms one by one, noting while doing so the marginal rate of substitution between increased costs and reduced aggregation errors. However, now that a procedure is available for exact aggregation of firms, it seems logical to move in the opposite direction, that is, to gradually reduce the number of aggregates to be programmed by successive combining of the basic groups which are free of aggregation error.

As a part of research recently completed, this approach was developed in three stages. First, a measure of aggregation error in multiple-product, price-general models was developed. That measure was based on the value of the productive capacity of resources mis-allocated as a result of combining the original error-free groups of farms. Then an algorithm was developed for successively combining the two groups which minimized the maximum marginal aggregation error. The final stage or result was the plotting of a schedule of accumulated maximum aggregation error for successively smaller numbers of groups of farms (from the original number of error-free groups to one). Conceptually, a cost schedule could be superimposed on the accumulated error schedule to develop a decision framework for an acceptable "balance" between the amount of aggregation error one is willing to live with and the costs one is willing to incur to further reduce error.
The point I want to make with this discussion is that the necessary conditions for exact aggregation are known. These conditions imply burdensome numbers of groups of farms. Offsetting this, proposals are being offered for efficient "trade-offs" between numbers of groups of farms for analysis and the potential aggregation error.

Recent Empirical Developments

Very few empirical analyses have explicitly addressed the problem of aggregation error. Two articles (4, 13) in a recent issue of the Journal of Farm Economics reported applications of Sheehy's and McAlexander's "homogeneous restriction method" for grouping farms. In both cases, use of this method resulted in less aggregation error than the more conventional grouping methods. However, the sample populations in both cases were farms on which dairying was the single dominant activity. Tom Miller has worked with samples of multiple-product farms at Iowa State, searching for practical guidelines for reducing aggregation error. The results of that work have not yet been published.

To evaluate the "critical resource ratios" method for exact aggregation and the associated error "trade-off" concept which were discussed in the previous section of this paper, a sample of 100 farms in southwest Georgia was selected. These were multiple-product farms; most of them produced some combination of cotton, peanuts, wheat, corn, oats, soybeans, beef cattle, and hogs. Survey data for each of the farms were used to construct separate linear programming models. The model for each farm was solved and the solutions to all farm models were added to derive the "true" aggregate supply schedule for each commodity.

In the models, the actual variations in activity coefficients for each commodity, as revealed by the survey data, were retained to provide maximum realism. In addition, there were as many as 18 potentially limiting resource restraints. For the population of 100 farms, exact aggregation required 100 groups. Each group obviously contained a single farm.

These results were not at all surprising. Exact aggregation is a demanding requirement. For the population of farms included in the experiment and considering the restraints included in the models, it is possible that a slightly larger sample would have produced some groups with more than one farm per group. That is, at a certain size of sample the ratio of number of error-free groups to farms in the sample would have dropped below one to one, and beyond that the ratio would have fallen at an increasing rate as the size of sample increased.
The error "trade-off" concept was tested with a subsample of 21 farms, for which exact aggregation required 21 groups. A completely general price situation was assumed. The 21 groups were sequentially reaggregated; first to 20 groups; then to 19 groups; and so on until finally all farms were back together in a single group. Each step consisted of combining two groups of farms so that the potential aggregation error was minimized. The resulting "trade-off" curve is shown in Figure 20.1. That curve shows the relationship between the number of groups of farms and the maximum relative aggregation error from misallocating the productive capacity of the resources involved.

The magnitude of the possible error depicted in Figure 20.1 is at first distressing. But it must be remembered that the situation in which all prices can vary from zero to infinity is an extreme case. Usually we are concerned with the magnitude of error for more modest price variations.

The same farm models were resolved with the prices of all commodities except cotton held constant at 1962 prices. The price of cotton was varied from $0.16 per pound of lint (including a return for cotton seed) to $0.40 a pound. The resulting error for alternative numbers of groups of farms is shown in Table 20.1. Data from Table 20.1 were used to construct the "maximum actual error" and the "average actual error" curves in Figure 20.1.

The models for the groups of farms shown in Table 20.1 were solved with all the activities of the original farms included. For the three-group and one-group models, "modal" activities were estimated and the models rerun. The results are compared in the first four columns of Table 20.2. The use of modal activities reduced the aggregation error considerably. In fact, constructing modal activities for the single aggregate farm reduced error more than breaking the aggregate into three groups, each of which retained the original activities.

For comparison of the minimum-error-grouping procedure with more conventional procedures, the same farms were grouped according to size. The size categories were those used for the same region for the S-42 Regional Adjustment Study: Under 150 acres of open land; 150 through 249 acres of open land; and 250 acres and more open land. The results are shown in the two right-hand columns of Table 20.2.

The absolute results of this simple experiment must be used judiciously and may not hold for other regions or farming situations. Certainly the complexity of the aggregation problem and the magnitudes of error involved in the "trade-off" process will be smaller in regions where the production of one major commodity, such as wheat in parts of the Great Plains, dominates alternatives over a wide range of prices.
Figure 19.1. The relationship between number of groups of farms and aggregation error, 21-farm sample, southwest Georgia.
Table 19.1. Aggregation error as a percentage of "true" aggregate value of production for specified numbers of groups of farms and at selected prices of cotton

<table>
<thead>
<tr>
<th>Cotton price ($/lb.)</th>
<th>Error for alternative number of groups of farms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
</tr>
<tr>
<td>0.16</td>
<td>11</td>
</tr>
<tr>
<td>0.20</td>
<td>11</td>
</tr>
<tr>
<td>0.22</td>
<td>11</td>
</tr>
<tr>
<td>0.24</td>
<td>11</td>
</tr>
<tr>
<td>0.26</td>
<td>12</td>
</tr>
<tr>
<td>0.28</td>
<td>12</td>
</tr>
<tr>
<td>0.30</td>
<td>11</td>
</tr>
<tr>
<td>0.32</td>
<td>8</td>
</tr>
<tr>
<td>0.34</td>
<td>9</td>
</tr>
<tr>
<td>0.36</td>
<td>5</td>
</tr>
<tr>
<td>0.38</td>
<td>11</td>
</tr>
<tr>
<td>0.40</td>
<td>10</td>
</tr>
<tr>
<td>Weighted average error</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 19.2. Aggregation error as a percentage of "true" aggregate value of production for specified groups of farms and at selected prices of cotton

<table>
<thead>
<tr>
<th>Cotton price ($/lb.)</th>
<th>One group, inclusive activities</th>
<th>One group, modal activities</th>
<th>Three groups, inclusive activities</th>
<th>Three groups, modal activities</th>
<th>Three sizes, inclusive activities</th>
<th>Three sizes, modal activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>Percent</td>
<td>Percent</td>
<td>Percent</td>
<td>Percent</td>
<td>Percent</td>
<td>Percent</td>
</tr>
<tr>
<td>0.16</td>
<td>56</td>
<td>36</td>
<td>41</td>
<td>16</td>
<td>64</td>
<td>24</td>
</tr>
<tr>
<td>0.20</td>
<td>56</td>
<td>36</td>
<td>41</td>
<td>16</td>
<td>65</td>
<td>24</td>
</tr>
<tr>
<td>0.22</td>
<td>56</td>
<td>36</td>
<td>41</td>
<td>16</td>
<td>66</td>
<td>24</td>
</tr>
<tr>
<td>0.24</td>
<td>67</td>
<td>36</td>
<td>43</td>
<td>19</td>
<td>66</td>
<td>26</td>
</tr>
<tr>
<td>0.26</td>
<td>67</td>
<td>37</td>
<td>44</td>
<td>20</td>
<td>68</td>
<td>25</td>
</tr>
<tr>
<td>0.28</td>
<td>61</td>
<td>38</td>
<td>49</td>
<td>22</td>
<td>84</td>
<td>31</td>
</tr>
<tr>
<td>0.30</td>
<td>59</td>
<td>34</td>
<td>48</td>
<td>23</td>
<td>82</td>
<td>27</td>
</tr>
<tr>
<td>0.32</td>
<td>44</td>
<td>32</td>
<td>42</td>
<td>15</td>
<td>66</td>
<td>20</td>
</tr>
<tr>
<td>0.34</td>
<td>47</td>
<td>38</td>
<td>39</td>
<td>12</td>
<td>60</td>
<td>20</td>
</tr>
<tr>
<td>0.36</td>
<td>44</td>
<td>49</td>
<td>39</td>
<td>14</td>
<td>52</td>
<td>18</td>
</tr>
<tr>
<td>0.38</td>
<td>44</td>
<td>36</td>
<td>36</td>
<td>15</td>
<td>46</td>
<td>19</td>
</tr>
<tr>
<td>0.40</td>
<td>44</td>
<td>35</td>
<td>35</td>
<td>14</td>
<td>44</td>
<td>19</td>
</tr>
<tr>
<td>Weighted average error</td>
<td>54</td>
<td>37</td>
<td>41</td>
<td>17</td>
<td>62</td>
<td>23</td>
</tr>
</tbody>
</table>
However, these empirical experiments do serve to illustrate concepts which in turn may provide practical guidelines for improving aggregation validity.

**Implications for Constructing the Basic Units of Analysis**

Despite the theoretical developments reported in this paper, it is a practical reality that, for the time being, programming models of regional and national agricultural production as well as models of resource use and interregional competition must contain a relatively small number of basic analytical units (each representing a group of farms). That being the case, the greatest returns in terms of reduced aggregation error may come from doing a better job of constructing these units of analysis.

**Weighted Average Activities**

The empirical experiment with the sample of Georgia farms demonstrated that the use of "modal" activities resulted in less error than including all the activities of each farm in models of groups of farms. Theoretically, at least, error from variations in coefficients for a given activity can be totally eliminated by using coefficients that are weighted averages of the coefficients for that activity for all farms in a group. Using weighted averages is necessary because different farms furnish different proportions of the total potentially limiting resources and the influence of the technical coefficients of each farm is proportional to that farm's share of the total output.

Weighted average activities may be obtained by weighting the reciprocal of the technical coefficient of each farm (for a given activity) by the proportion of the potentially limiting resource contributed to the aggregate by that farm and then obtaining the reciprocal of the result. In other words, the output of given commodities for individual farms of a group of \( j \) farms could be expressed as

\[
Q_{ij} = a_{ij} X_1 + b_{ij} X_2 + c_{ij} X_3
\]

where \( i \) represents the commodity or activity \((1...M)\), \( j \) is the farm \((1...n)\), and \( a, b, \) and \( c \) are the technical coefficients for resources \( X_1, X_2, \) and \( X_3, \) respectively. The total output of each commodity for the group of farms could be expressed as

\[
Q^T_i = A_i X_1 + B_i X_2 + C_i X_3
\]
where \( A_1 = \sum_{j=1}^{n} \left( \frac{1}{a_{ij}} \right)^{-1} \), and where \( \Theta \) is the proportion of \( X_1 \) contributed by the \( j \)th farm. The weighted average coefficients in equation (2) would constitute the activity vectors in models of farm aggregates.

Perhaps a small sample of the farms represented by a model would be sufficient to establish both the variations in coefficients for a given activity and the weights for each of those variants. The only difference in the data required to estimate weighted average coefficients and that required to estimate "modal" coefficients is that the former requires resource quantities data for each farm.

Improved Restraints

Even when good aggregate restraint data are available, we get error in models of production because of the variation in the relative scarcity of specific restraints from one farm to the next. It is common, for example, in some regions that some farms must hire additional labor while others have idle labor. In most cases the idle labor on one farm is not available to farms with labor shortages. In addition, some restraints (and their coefficients) are somewhat elastic and the restraint quantities are hard to define. In critical periods, farmers may work harder, faster, and more hours per day. In a crisis, other family members, not normally a part of the farm labor force, may pitch in and help. Machine capacity may be expanded by using machinery more hours a day.

Two suggestions may be appropriate for improving estimates of restraints for models of farm aggregates. Both should help reduce aggregation error. The first is the use of the "inclusive" method for estimating restraints. For example, in short-run supply models, it is not uncommon to use census of agriculture estimates of total land available in a region. Sometimes this estimate is adjusted by subtracting out land used by enterprises and types of farms not included in the model. But this process is often not taken far enough. It may be better to start by estimating the land used by activities specifically included as model alternatives. From that benchmark, estimates could be made of (1) additional quantities of land that could be available to the included activities, and (2) the economic conditions which might make those quantities available.

The second suggestion is a variation of the inclusive approach and may be particularly useful for shorter-run predictive models. It may be called the engineering method of estimating restraints. Weighted average coefficients for, say, labor in period \( t-1 \) could be multiplied
by actual activity levels in period \( t-1 \). The product is a first approxi-
mation of the labor restraint in period \( t \). This estimate could be ad-
justed if there are reasons and data for doing so. Then the model
could be run to see what pressure points develop with respect to labor.
These points could then be examined and the reasonableness of the
restraints further evaluated. This approach helps to reduce the problem
of attributing positive productive capacity in the aggregate to resources
which on individual farms may in fact be in surplus. Again, the use-
fulness of this suggestion depends on the research objectives.

Formulating Models to Improve Aggregation Validity

Recent attempts to evaluate some of the regional studies of
agricultural production and adjustments prompt me to say something
about the relation of assumptions about time, resource mobility,
and research objectives to the validity and usefulness of the results
of supply models.

Model Realism and Aggregation Error

The original assignment to write this paper called for emphasis
on realism in aggregation. But the realism of the aggregation pro-
blem is a function of the realism of our models, which are themselves
abstractions to the extent that they do not perfectly simulate the real
world. I have not been able to understand the surprise and dismay of
researchers who, when they conduct optimizing analyses with highly
unrestrained models, built on assumptions of advanced technology,
and weight these results to aggregate levels, get supply-price relation-
sips highly unlike those they observe in the real world. It would be
grossly inaccurate to attribute all the differences between the model
results and the real world to aggregation error. If we want results to
compare with the real world we have to build models of the real world--
including all the relationships, uncertainties, and subjective and
objective goals and restraints of the real world.

The nature of the aggregation problem varies with the "realism" of
our economic models. In highly abstract models in which all resources
are assumed variable, knowledge is assumed perfect, and all farmers
are assumed to have perfectly adjusted to a common "best" technology,
the problem of aggregation bias or error from grouping farms with unlike
resource ratios, technical coefficients or expectations does not exist.
If, however, any of the real world's asset fixity and imperfect know-
ledge are permitted, the possibility of aggregation error in the results
of a given model must be recognized. In general, the intra-model
aggregation problem becomes more complicated as the realism of the
models increase; that is, as we move from static to dynamic assumptions
and from more "normative" to more "predictive" assumptions.
Potential aggregation error may be suppressed in more predictive models where so-called "behavioral" or "flexibility" restraints are included. These restraints, which generally take the form of upper and lower bounds on the solution levels of specified enterprises, indirectly account for the many forces that cause lags in adjustment or temper profit-maximizing behavior in the real world. Thus, this particular method of adding realism to supply models may reduce aggregation error as a by-product.

Assumptions about Time

Too often the validity of the supply schedules generated by models is impaired by inconsistent assumptions about time. This may be called an aggregation-over-time problem. For example, in one major study: (1) The time period used for yields was assumed to be of sufficient length for the full effect of improved practices on crop yields to materialize; (2) the time period used for restraints was assumed to be long enough for intermediate-term capital investments in buildings, farm machinery, equipment, livestock, and pasture improvements to be considered as variable costs (yet operator labor and management were considered fixed); (3) farm numbers were those projected for 1975; and (4) the price and cost projections represented "... the level of prices that may be expected to prevail over an extended period of years under assumptions of relatively high employment, a trend toward peace, continued population and economic growth, and a stable general price level." One wonders whether the time periods of (1), (2), (3), and (4) were consistent. Failure to ensure that all the assumptions about time are consistent could result in mongrel supply schedules that have meaning for no time period.

As researchers, we have some license to play around with any assumptions we like. It may be useful to know the nature of "equilibrium" under alternative assumptions even though we imply ceteris paribus assumptions about the rest of the world and ignore the number of iterations of our model needed to reach equilibrium. But we get into trouble if we try to ascribe realism to our model results when we have failed to translate economic time into calendar time. Imagine trying to explain to a policymaker how he should interpret the "long-run" supply schedules I described a moment ago. Some economists tend to hold the policymaker's lack of appreciation of long-run research in contempt. But economists who have had occasion to work closely

2/ For an excellent discussion of the meaning and estimation of "proxy" restraints, see Schaller (9). For examples of their use, see (2) and (11).

3/ See pp. 8-10, (14).
with policy problems share the policymaker's disappointment at the failure of supply models to provide results to which he can attach specific dates under real world conditions.

I do not mean to imply that "long-run" research is not useful. But we should be as rigorous and consistent in our long-run models as we have to be in short-run predictive models.

Assumptions about Resource Mobility

There is an inter-model as well as an intra-model dimension to the aggregation problem. This is the question of the validity of deriving aggregate supply schedules by horizontal summation of representative farm supply schedules. Sharples, Miller, and Day (12) have recognized this problem in their evaluation of the NC-54 study. But the process of horizontal summation of supply schedules, per se, is valid. The invalidity lies in the failure to account for the aggregate limitations on availability of resources that are assumed to be variable or unlimited at the representative farm level. This failure has been particularly serious in the Regional Adjustment Studies because of the longer-run assumptions about the variability of resources. It is less serious in the shorter-run predictive models wherein most resources are fixed.

Individual farms or representative farms may buy land, but the aggregate supply of land is fixed. Labor or capital may be assumed unlimited for representative farms, but the aggregate quantities of the resources are certainly limited. When these aggregate limitations are not included as constraints on model results it is no wonder that the representative farm supply schedules and thus the summed schedules are biased upwards.

One possibility for reducing the distortion of supply estimates resulting from assumption of unlimited resources for representative farms is to include the matrices for a number of such farms as diagonal blocks within a regional matrix. Thus, some resources could be unlimited for individual representative farms but fixed for the aggregate of all such farms. The final solutions for each representative farm or resource situation subaggregate would reflect the competition for the fixed amount of resources available at the regional level. The validity of the supply schedules so derived would depend on the extent to which resources were mobile among regions represented by separate matrices. For example, the supply of land may be fixed for a region but labor and capital may flow among regions.

For year-to-year predictive models the failure to consider the aggregate effects of the decisions of individual producers on input and output prices may not be particularly serious. For many agricultural
enterprises, the nature of production is such that resources are committed to production in a planning period and the aggregate consequences of individual behavior are not relevant until the next planning period. This lagged response to the interaction of aggregate supply and demand means that individual regions and representative situations can be analyzed independently of others within a given production period without the kind of simultaneity considerations which would be necessary in the analysis of an oligopolistic industry. However, the aggregate price implications must be considered in models of multi-period enterprises such as livestock or in longer-run models in which the passage of time is implicit.

Some Additional Aspects of the Aggregation Problem

Aggregation Validity and the Representative Farm Concept

The "representative farm" is a useful tool for budgeting and farm management purposes. But the conventional representative farm may not be the appropriate unit of analysis for aggregate supply studies. There is a tendency to construct such units on the basis of "look alike" criteria such as acres of land, size of labor force, or current enterprise combinations. These criteria may not be those which minimize aggregation error in estimates of aggregate supply.

The representative farm concept as it has evolved has chained researchers to thinking in terms of "typical" identifiable entities which are instantly recognizable (small farms, part-time farms, peanut-cotton farms, etc.). The purposes of aggregate supply analysis may be best served if the researcher is not constrained by the idea that he has to be able to typify the characteristics of the farms in a group within the context of a single "representative" farm. It may well be that when farms are grouped in a way which minimizes aggregation error, the farms in a given group may not "look alike" in the standard sense. For example, when the Georgia farms in the sample described earlier were grouped according to criteria which minimized aggregation error, the intra-group variance in size (acres of land) was greater than the differences in the mean sizes of the groups.

It may be useful to think in terms of treating whole groups of farms as single decision-making entities (while ensuring that the decision "rules" are appropriate for the individual farms involved). This is the aggregate representative resource situation approach. There is no apparent need for researchers who take this approach to "scale down" the aggregate to the average or representative farm to see how that micro unit looks.
If it is really necessary to interpret the results of aggregate supply analysis for identifiable situations, it may be necessary to group farms into those situations first. Each group could then be subdivided for the explicit purpose of minimizing aggregation error.

The Need for "Open-Structured" Data 4/

In practice, the approach to aggregation may be dictated more by the nature of the data available than by the beauty of the relevant theory. To deal with the aggregation problem, more "open-structured" as opposed to "pre-structured" data are needed. Much of the survey and census data available are already processed and "canned" before the user gets them. Perhaps for general descriptive uses this form of presentation must be retained. But aggregates and averages for counties, states, and other units cover up the micro variations that are the very source of the aggregation problem. With today's modern data processing facilities it is becoming more feasible to preserve, on tapes, cards, and discs original sample data for micro units. With proper programs this data could be retrieved in unstructured form or restructured to meet the needs of users. The same data that would permit grouping of farms to minimize aggregation error would also be the best data for estimating the parameters of the analytical models of those groups.

Until such data are available, builders of models of aggregate supply can use sample surveys designed for their own needs or try to utilize data from surveys conducted for other purposes. There are always likely to be data gaps from the latter approach and these may be critical.

A Look to the Future

The changes in the structure of agricultural production and marketing which have been discussed in these papers will affect the problem of aggregation validity. On the marketing side there is likely to be less price uncertainty as more commodities are produced under contracts or on a custom basis for large buyers. This reduces the variation in net return expectations. The same marketing forces will call for increased specification of product qualities, grades, sizes, and dates of delivery. Contractors may even specify certain types of pesticides, levels of fertilization and harvesting methods. The result will be greater uniformity of products and ways of producing them.

---

4/ See (10) for a more elaborate discussion of data needs for aggregate supply models.
There will still be variations in weather and soil qualities, but even these may be more subject to control through cloud seeding, irrigation, and more scientific feeding of nutrients to the soil.

Changes in both marketing and production will continue to provide incentives for further specialization. This may tend to reduce the number of enterprises on individual farms and increase the specialization of machinery, labor, and the overall bundle of resources. There will thus be increased validity in grouping farms on the basis of their specialization because of the increased fixity of specialized resources and because farms within these groups will have increasingly similar resource ratios and technical coefficients.

Even management may become more homogeneous. As Professor Leontief once remarked to a Harvard class: "Every year more managers think like economists say the economic man should think because every year more managers have been trained by economists to think like the economic man."

It is likely that some of the emerging structural changes will complicate the aggregation problem. But on the whole, the kinds of changes I have described will simplify the aggregation problem and improve aggregation validity. This, in turn, will increase both the accuracy and the relevance of the more predictive models of economic behavior.
LITERATURE CITED


