Chapter V

THE PRODUCTION RELATION

Data Limitations

Certain aspects of the aggregate production relation are considered in this chapter along with results of statistical analysis of equation 4.1 and some alternative equations which were also studied. In view of this examination none of the equations studied appear to give a satisfactory empirical representation of the underlying relation, and the need for further study, possibly with different sources of data and different techniques, is strongly indicated. The results are presented in some detail, to indicate to other research workers the sort of attempts that have been made and the nature of the difficulties that arose. It is hoped that this will assist others in developing more fruitful avenues of attack in future studies. In addition to uncertainties about the theoretical specification (form of equation, way in which stochastic elements enter, etc.) and limitations of existing estimation procedures, two special difficulties with respect to the available data have already been cited and need to be borne in mind in considering the production relation. One is the difficulty in accurately measuring pasture consumption which constitutes about 35% of the total feed consumed by livestock.[1] The other is the lack of any observations on variables other than feed and animal numbers that might be expected to influence the relation significantly.

Pasture seemed too important to be omitted, and so such observations as could be obtained were used throughout the study. In Chapter II it was provisionally assumed that important unobserved factors had increased fairly smoothly over time and could be approximately taken into account by introducing a time variable in the production equation. One attempt to improve on this assumption is reported below (see pp. 85 to 87).

[1] There are, of course, serious difficulties in the measurement of other variables. However, the importance of pasture in this relation and the likelihood that the observations of pasture consumption are still less accurate than observations of other variables seem to justify citing it as a special problem.

Statistical Tests Employed

Fitted equations have been examined by considering the plausibility of estimates of coefficients, by applying statistical tests of some of the assumptions on which the estimates are based, and by checking the fitted equations against data for 1950, the only year for which observations were available that was not included in the sample. Where estimated coefficients seemed sufficiently implausible, other steps in the examination were omitted. Checks against 1950 data are reported in Chapter VIII. The two statistical tests applied in this chapter are the Durbin-Watson test[2] for serial independence of disturbances and the Rubin-Anderson test[3] of overidentifying restrictions. The latter is only appropriate when the equation in question is viewed as a member of a system of equations; hence, it is applied only to equations fitted by the limited-information method.

Durbin-Watson Test of Serial Independence

The Durbin-Watson test is based on the statistic d given by

$$d = \frac{\Delta^2}{S^2} \quad \text{where} \tag{5.1}$$

$$\Delta^2 = \sum_{t=2}^{T} (\tilde{u}_t - \tilde{u}_{t-1})^2 \qquad S^2 = \sum_{t=1}^{T} \tilde{u}_t^2 \tag{5.2}$$

$\tilde{u}_t$ is the residual of the fitted relation for time t. Durbin and Watson have shown that no exact critical region corresponding to a given level of significance can be found for d, but that upper and lower limits for each boundary of the critical region can be calculated. They have tabulated these limits, and their tables are used in the tests reported here. All of the serial-correlation tests in the present study are two-tailed tests (tests against both positive and negative serial correlation) at the 0.05 significance level (the relevant limits are thus contained in Durbin and Watson II, Table V, p. 174). If the calculated value of d for a particular equation is either less than the lower limit (d_L) for the boundary of the lower part of the critical region or greater than the upper limit ($4 - d_L$) for the

[2] J. Durbin and G. S. Watson, Testing for Serial Correlation in Least-Squares Regression I, *Biometrica*, Vol. 37, p. 409, and, Testing for Serial Correlation in Least-Squares Regression II, *Biometrica*, Vol. 38, p. 159.

[3] T. W. Anderson and H. Rubin, Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equation., *op. cit.*, p. 56.

boundary of the upper part of the critical region, the null hypothesis of serially independent disturbances is rejected. If calculated d lies between the upper limit (d_U) for the boundary of the lower part of the critical region and the lower limit ($4 - d_U$) for the boundary of the upper part of the critical region, the hypothesis of serial independence is not rejected. If the calculated statistic falls between d_U and d_L or between $4 - d_U$ and $4 - d_L$, the result of the test is inconclusive. Durbin and Watson have presented an approximate test to be applied to inconclusive cases, but, since it would involve substantial extra calculations in our cases, and since they only assert that it is a good approximation when one has more than 40 observations, this has not been applied.

The test has been developed for the single-equation case with the independent variables regarded as fixed. Thus the Durbin-Watson limits are inexact when the test is applied to a least-squares equation containing a lagged dependent variable or when it is applied to an equation fitted by limited information. None of our least-squares equations contains the lagged value of the dependent variable, and so there is no difficulty as far as our least-squares equations are concerned. The use of the tabulated limits must be regarded as approximate when applied to an equation fitted by limited information. It was pointed out in Chapter IV that limited-information estimates of coefficients of the predetermined variables in an equation (the z_*) could be regarded as least-squares estimates of coefficients appearing in the regression of a linear combination (y^0) of the current endogenous variables (y_*) on the z_*. If the appropriate linear combination were exactly known, then no qualification would attach to the application of the Durbin-Watson test to residuals from limited-information equations. The fact that we only have estimates of coefficients of y^0 makes the test inexact. However, since we have no exact test, it was applied. The tabulated limits depend on the number (k') of independent variables in the equation. Because of the above least-squares interpretation of limited-information estimates, k' was taken to be equal to the number of predetermined variables in the equation when the test was applied to limited-information residuals.

For the limited-information version of (4.1), the calculated value of d is 1.11. The limits of the lower bound of the critical region for 30 observations and $k' = 3$ are $d_L = 1.12$, $d_U = 1.54$. Hence, the hypothesis of serial independence is rejected by this test. For the least-squares version of (4.1), the calculated value of d is 1.28. The limits for 30 observations and $k' = 5$ are $d_L - 0.98$, $d_U = 1.73$. The test is inconclusive in this case. However, in both cases, some doubt is cast on the validity of the assumption of serial independence of disturbances. The calculation of the statistic for both versions of (4.1) is shown in the appendix (pp.

146 and 149). For other equations the same procedure is used, but the details are not shown.

Rubin-Anderson Test of Overidentifying Restrictions

The Rubin-Anderson test was designed to test the assumption that certain coefficients are zero (certain variables do not appear) in a particular equation, given the validity of the remainder of the specification. It was observed in Chapter I that, in many applications to economics, the investigator is likely to have better grounds for the specification of which variables enter particular relations than for other aspects of his specification. In such cases the interpretation of the test is subject to doubt, but it remains true that an extreme value for the test statistic is an indication of difficulty somewhere in the statistical specification used.[4] The test statistic is $T \log (1 + 1/\lambda)$ where T is the number of observations and λ is the largest characteristic root of a determinantal equation used in the calculation of limited-information estimates.[5] The statistic has been shown to have a limiting χ^2 distribution, with degrees of freedom equal to the number of overidentifying restrictions used in the calculation of the estimates. In the notation of Chapter IV, the number of overidentifying restrictions used is the number of z_{**} employed in computing the estimates less $(H - 1)$, the number of z_{**} necessary for identification. Six z_{**} were used in calculating limited-information estimates of (4.1), and two were necessary for identification (the equation contained three current endogenous variables); hence the statistic has 4 degrees of freedom in this case. The calculated value for $T \log (1 + 1/\lambda)$ was 18.9. Since values greater than 13.3 would arise by chance only one time in one hundred, this is a strong rejection of the set of specifications on which limited-information estimates are based.[6]

[4] See Carl Christ, *A Test of an Econometric Model for the United States, 1921–47*, Conference on Business Cycles, National Bureau of Economic Research, 1951, pp. 35–129. Similar observations could be made about the Durbin-Watson test in the applications considered here. Apparent serial correlation of disturbances could, for example, be caused by the use of an inappropriate algebraic form or the omission of an important variable.

[5] See appendix, pp. 141–142, for an example of the determinantal equation and the calculation of the largest root.

[6] One of the assumptions on which the test is based is that the moment matrix of the predetermined variables converges to a fixed limit as the number of observations increases indefinitely. This is violated by our inclusion of time as a variable unless we assume that, beyond some point, the factors associated with time cease to increase linearly and approach a finite limit. Though this does not seem unreasonable, the questions raised emphasize both the undesirability of applying asymptotic theory to a fairly small sample and the complications that arise with the use of time as a variable.

The implausibility of the estimates of the coefficient of Z_{13} and the implications of statistical tests that have been performed indicate a need for other formulations of this relationship but do not indicate very precisely the directions in which one might usefully look for alternatives. In selecting alternatives to be examined in the present study, the authors limited themselves to equations that could be analyzed with the time-series data readily available and that could be readily incorporated into the economic model developed in Chapter II. More substantial revisions would have demanded more time and resources than could be allocated but may be necessary if real progress is to be made.

Essential Inputs Property of Logarithmic Relation

One feature of equations linear in the logarithms of observed variables that is of interest in connection with equation 4.1 is the following. If we imagine that all of the variables except two are held constant (at finite positive values) and one of these is arbitrarily set equal to zero, the other will be zero or infinite, depending on whether the signs of the coefficients of the logarithms of the two variables are different or the same. If we consider a production relation of this form, relating one output to several inputs, and if the coefficient of the log of output is positive and coefficients of the logs of inputs are negative, then, for output to be nonzero, all inputs must be positive. That is to say, at least some of every input is essential if any output at all is to be produced.

Clearly the appropriateness of assuming this property depends on the definitions of inputs employed. If the resources that may be used in production are classified into a large number of highly special inputs, it is likely that only a few will be absolutely essential to the production of some output. On the other hand, if only a few categories of inputs are set up, it is more likely that some amount from each category will be necessary for any production to be realized.

Reaggregation of Feed Inputs

In reference to the variables of equation 4.1, there is no question but that some output of livestock and livestock products could be produced with zero input of any of the types of feed, provided all three were not simultaneously zero. Thus there exists a discrepancy between our a priori knowledge and the implications of the form of equation adopted in (4.1). This discrepancy concerns a property of the equation in the large (i.e., it arises only if large variations in the variables are considered), and we are ordinarily more interested in the local properties of our equtions. However, the existence of the discrepancy, combined with other evidence that (4.1) is unsatisfactory, gives some inducement to consider

TABLE XI
ALTERNATIVE PRODUCTION RELATION

Method	Estimates of Coefficients			
	β_{18}	γ'_{11}	γ'_{12}	γ'_{10}
Limited information	−0.463(0.111)	−0.274(0.152)	−0.0045(0.0004)	−3.132(1.009)
Least squares	−0.565(0.107)	−0.170(0.148)	−0.0043(0.0004)	−3.274(0.987)

$R^2 = 0.978$

alternatives in which this property is modified. Since the feed variables are measured in common units (pounds of digestible nutrients), one way to remove this inconsistency between our a priori knowledge and the form of equation used is to introduce a variable representing aggregate feed fed to livestock into the production relation instead of the three separate feed variables of (4.1). In the notation of Chapter IV, we then have

$$y_{8t} = y_{6t} + y_{7t} + z_{3t} \tag{5.3}$$

$$Y_{1t} + \beta_{18}Y_{8t} + \gamma'_{11}Z_{1t} + \gamma'_{12}z_{2t} + \gamma'_{10} = U'_{1t} \tag{5.4}$$

where $Y_8 = \log y_{8t}$. Whether or not (5.4) permits a better approximation to the underlying relation than (4.1) can hardly be answered a priori. Whereas (4.1) did not seem to allow sufficiently for the substitutability of feeds, (5.4) treats them as perfect substitutes. Estimates of the coefficients of (5.4) are given in Table XI.[7]

In the calculation of the least-squares estimates, Y_1 was again treated as the dependent variable. To obtain the limited-information estimates the model given by equations 5.3, 5.4, and 4.2 to 4.7 was considered. Because (5.3) is not linear in the logarithms of the observed variables, this is a mixed linear, nonlinear system. Rubin and Anderson have shown[8] that the limited-information method yields consistent estimates of a linear subset (in this case 5.4 alone) of such a system. The estimates in the first row of Table XI were obtained, using Z_3, Z_4, Z_5, Z_7, Z_8, Z_9, Z_{10} as predetermined variables not appearing in the relation being estimated.

Both sets of estimates in Table XI appear plausible on first inspection.

[7] The production equation fitted by Lorie, *op. cit.*, Ch. IV, was of the same form as (5.4), with some differences in definitions of variables. In our notation, his estimates of coefficients, obtained by least squares, using data for the period 1910–44, were $\beta_{18} = -0.330$, $\hat{\gamma}'_{11} = -0.523$, $\hat{\gamma}'_{12} = -0.0021$.

[8] See T. W. Anderson and H. Rubin. The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations, *op. cit.* pp. 200–212.

The estimates of γ'_{12} imply an annual increase of about 1% per year in product obtained from given quantities of feed and animals. The sum of estimates of β_{18} and γ'_{11} is -0.74, using either limited-information or least-squares estimates. In the initial stages of the study, when the a priori impressions of a number of informed persons were sought on various aspects of livestock production and marketing, it was quite generally conjectured that, at levels of output that have prevailed in the past thirty years, livestock production is characterized by nearly constant returns to scale. It was somewhat less generally stated that numbers of animals fed and amounts of feed used probably dominated production to such an extent that output would be almost proportional to changes in these two inputs (it being recognized that some observed inputs would tend to change along with animal numbers and feed).

Test of Importance of Other Inputs

If both these conjectures are true, then, except for sampling fluctuations, the sum of the estimates of β_{18} and γ'_{11} should be -1. If the first conjecture is accepted a priori but not the second, a test of the hypothesis that $\beta_{18} + \gamma'_{11} = -1$ may be regarded as a test of the importance of unobserved factors that did not consistently vary with animal numbers or feed. We are inclined to interpret the test in this way, but we recognize that other a priori assumptions would lead to other interpretations. Since our limited-information procedures do not include tests of linear hypotheses, the test is based on the least-squares calculations. The test statistic has the F distribution with 1 and $T - k' - 1$ degrees of freedom where k', as before, is the number of independent variables in the equation.[9] The calculated value of F in this case is 6.97, whereas the value corresponding to the 0.05 significance level is 4.22. Thus, by our interpretation, the test offers substantial evidence of the importance of omitted variables.

Plausibility of Marginal Responses

Marginal response of output to small increments of feed or animals can be calculated for particular values of the inputs by solving the equa-

[9] Let the regression equation be given by $y_t + \alpha x'_t = u_t$ where y = dependent variable, x_t = vector of independent variables, α = vector of coefficients. Let the null hypothesis be given by $b\alpha' = \beta_0$. Then

$$F = \frac{(T - k' - 1)(b\hat{\alpha}' - \beta_0)^2}{T\hat{\sigma}^2 b M_{xx}^{-1} b'}$$

has the F distribution with 1 and $T - k' - 1$ degrees of freedom. $M_{xx} = T\Sigma_{t=1}^{T} x'_t x_t$, $\hat{\alpha}$ is the least-squares estimate of α, $\hat{\sigma}^2 = \frac{1}{T}\Sigma_{t=1}^{T} (y_t + \hat{\alpha}x'_t)$. For a general discussion of tests of linear hypotheses see A. M. Mood, *op. cit.*, p. 305.

tions for y_1 and differentiating with respect to each input. This yields

$$\frac{\partial y_1}{\partial y_8} = -\beta_{18}(\text{antilog} - \gamma'_{10})y_8^{-\beta_{18}-1}z_1^{-\gamma'_{1}}10^{-\gamma'_{12}z_2} \tag{5.5}$$

$$\frac{\partial y_1}{\partial z_1} = -\gamma'_{11}(\text{antilog} - \gamma'_{10})y_8^{-\beta_{18}}z_1^{-\gamma'_{11}-1}10^{-\gamma'_{12}z_2}. \tag{5.6}$$

Whichever set of coefficients is used, marginal response to one input tends to decrease with increases in that input and to increase with increases in the other.

Marginal returns to feed and inventory have been calculated for an "average," year considered for this purpose to be a year in which the variables y_8, z_1, z_2 take on their average values over the period of the observations. These average values are 392 million for y_8, 6.09 billion for z_1, and 15.5 for z_2. The units are those used in obtaining the estimates given in Table XI, namely thousands of pounds of TDN for y_8, dollars worth of estimated potential production for z_1, and years for z_2. If we use these average values for the variables and the limited-information coefficients given in Table XI, the calculated marginal return for feed is 8.67, and the marginal return to livestock inventory is 0.32. If we use least-squares estimates of coefficients, the calculated marginal return to feed is 10.47 and to inventory 0.20.

The aggregate production relation of our model depends on the production functions of individual producers and on relations determining the distribution of inputs among individual producers. For a small increase in the aggregate of one input, the rate of response of aggregate production will be a weighted average of the marginal productivities of the input for individual firms, and the weights will be proportional to the amounts by which the individual firms increase their use of the input in question.[10] To the extent that individual producers succeed in equating marginal productivity of an input to the input-output price ratio, we should expect the marginal response of aggregate output also to approximate this ratio. In general, we would expect the approximation to

[10] Let $y = f(a, b, c)$ be the aggregate production relation with y as aggregate output and a, b, c as aggregates of inputs. Let y_i, a_i, b_i, c_i, be the output and inputs of an individual firm with production function $y_i = f_i(a_i, b_i, c_i)$. We have $y = \Sigma_i y_i$, $a = \Sigma_i a_i$, etc. Suppose the process by which the input a is allocated among firms can be summarized in a set of relations $a_i = g_i(a)$. For a small increase in a, with b and c fixed, we have $\frac{\partial y}{\partial a} = \Sigma_i \frac{\partial y_i}{\partial a_i}\frac{\partial a_i}{\partial a}$. It would be useful to consider in detail the implications of various assumptions about the distribution relations and the process underlying them. This, however, is outside the scope of the present study.

be closer for current endogenous inputs than for predetermined inputs, since the latter cannot be adjusted in response to current information.

In any case we do not have data on the cost of increases in inventory. We can verify that the general order of magnitude of the calculated marginal returns to the inventory variable are reasonable, by noting that average returns are about 1.2 and that average returns should decline if the size of the herd is increased while feed consumption remains constant. This implies that marginal returns are below average returns.

Data are also lacking on the cost of roughage. However, since roughage is predetermined, marginal adjustments in quantity of feed consumed take place largely through variations in grains fed. We might expect, therefore, a rough correspondence between marginal returns to grain and price of grain. Our calculated marginal returns above, 8.67 using limited-information coefficients and 10.47 using least-squares, are in dollars worth of product per thousand pounds of TDN. The average price of grain per thousand pounds of TDN during our sample period was \$19.40.[11] Thus there is a substantial discrepancy in this comparison.[12]

An explanation that readily suggests itself is that digestible nutrients obtained through roughage may cost substantially less than digestible nutrients obtained through grains. This is confirmed if we compare average production with average quantity of feed fed. Average production was \$7.4 billion at average prices. If the average quantity of feed, 392 billion lb of TDN, is evaluated at \$19.40 per 1000 lb, then the feed fed to animals was worth about \$7.6 billion. At the same time that this provides an explanation for the discrepancy noted above, it also makes highly questionable the assumption (utilized in 5.4) that digestible nutrients in one type of feed are perfect substitutes for digestible nutrients in another type of feed.

Tests of Production Relation with Aggregate Feed Variable

The statistical tests described earlier in the chapter and applied to equation 4.1 also tend to cast doubt on the appropriateness of (5.4) as a formulation of the production relation. The Durbin-Watson statistic calculated for the residuals from the least-squares version of (5.4) is 0.634, whereas the lower limit (d_L) of the lower boundary of the critical region for 30 observations and 3 independent variables is 1.12. For the limited-information equation, the calculated statistic d is 0.596 and d_L for $k' = 2$ is 1.18. In both cases the test significantly rejects the assumption of serially uncorrelated disturbances. The value of $T \log(1 + 1/\lambda)$

[11] This is a simple average of the figures in col. 1, Table V, on p. 63.

[12] Comparisons, using data for scattered individual years instead of averages, are similar.

calculated for this relation is 41.7, which is also highly significant, since a value as large as 16.8 would occur only one time in 100 if all of the assumptions on which estimation was based were valid.

Assumptions Underlying Third Form of Production Relation

Another form of the production relation for which estimates were obtained involved different treatment of the feed variables and omission of the time variable. An attempt was made to learn of possible sources of the secular increase in feeding efficiency in livestock enterprises and to consider to what extent these might be related to observable variables that could be utilized in studying the relation. The principal explanatory factors that seemed to the authors and others to be of importance were improved breeds, improved feeding practices, and improved care of animals. If observations were available on all the inputs used in livestock production, it would seem that improved care and improved feeding practices could be related, to a considerable extent, to increased use of labor and/or equipment, and to changes in the composition of livestock rations. Different breeds of animals could be regarded as different inputs if sufficient data were available. Although the data necessary to study various of these factors are sometimes available in the analysis of feeding experiments and in intensive farm-management studies of a few firms, the only aggregate data that seemed promising for incorporation in the present study related to the amounts of particular feeds fed to livestock during the sample period.

Several advisers who were consulted suggested that one of the important historical changes during the period of observation was the increased recognition of the value of protein in animal diets and the increased use of high-protein feeds. Although the advisers disagreed somewhat in their speculation about the relative importance of this and other factors, it did seem worth while to consider a form of the production relation in which protein was distinguished from other digestible nutrients. It was possible to estimate the number of pounds of protein (DP) and the number of pounds of other digestible nutrients (ODN) fed to animals from the estimates of protein and total nutrient content of individual feeds given by Morrison and Jennings. The use of their estimates is explained in Chapter III (see especially footnote 43), and the composition figures actually used in this study are summarized in Table II, pp. 51–52.

Estimated Coefficients for Revised Relation

Let y_{9t} be the total number of pounds of digestible protein contained in feed fed to livestock in period t, and let y_{10t} be the total number of pounds of other digestible nutrients. $Y_{9t} = \log y_{9t}$, $Y_{10t} = \log y_{10t}$. (5.4) was modified by taking out z_2 and by substituting Y_9, Y_{10} for Y_8. Coefficients in the resulting equation were estimated by limited information and by least squares with these results:

$$(5.7) \qquad Y_1 - \underset{(0.91)}{3.98}Y_9 + \underset{(1.10)}{3.65}Y_{10} - \underset{(0.36)}{0.67}Z_1 + \underset{(2.32)}{3.62} = 0$$

$$(5.8) \qquad Y_1 - \underset{(0.40)}{1.97}Y_9 + \underset{(0.48)}{1.17}Y_{10} - \underset{(0.24)}{0.28}Z_1 - \underset{(0.77)}{1.80} = 0 \qquad R^2 = 0.88.$$

The limited-information coefficients are given in (5.7), the least squares in (5.8). The time subscript has been omitted from each variable. Calculated standard errors of the estimated coefficients are given in parentheses below their respective coefficients.

Historical Changes in Proteins and Other Nutrients

The result that output is decreased by increases in other digestible nutrients is quite implausible, as is the implication that production could be so greatly increased by small increases in protein fed. Several possible explanations for the implausibility of the results were considered. It was noted with mild surprise that the simple correlation coefficient between Y_9 and Y_{10} was 0.98 for our sample period. The ratio of protein to other digestible nutrients fed was 0.166 for the first five years of observation (1920–24), compared with 0.177 for the last five years (1945–49). The prior impressions of the investigators had led them to expect a fairly substantial increase in protein, relative to other nutrients. Because others may have similar impressions, Table XII, accounting for protein and other nutrients by sources, may be of some interest.

Though the use of protein concentrates did expand greatly during our period of observation, a fairly modest expansion in the feeding of grains relative to forage was sufficient to keep the ratio of digestible protein to other digestible nutrients almost constant. To the extent that high correlation between Y_9 and Y_{10} contributed to the inaccuracy of the estimates of coefficients in (5.7) and (5.8), we should expect the calculated standard errors of the estimated coefficients to be large. Though the calculated standard errors are, in fact, large, they are small enough to suggest that this was not the only source of difficulty with this equation.

Other Forms of Relation

Two other production equations were fitted. An alternative algebraic form involving the same variables as (5.7), (5.8) is given in (5.9).

TABLE XII

HISTORICAL CHANGE IN LIVESTOCK RATIONS

Figures in Upper Six Rows Are in Tens of Billions of Pounds

	Sources of Nutrients			
	Grain	Protein concentrates	Roughage	Total
1920–24				
TDN	62.70	3.61	118.93	185.24
DP	6.44	1.50	18.49	26.42
ODN	56.26	2.11	100.45	158.82
1945–49				
TDN	82.67	8.17	137.59	228.43
DP	8.62	3.47	22.32	34.41
ODN	74.05	4.70	115.28	194.02
Per cent increase				
TDN	32	126	16	23
DP	34	131	21	30
ODN	32	123	15	22

TDN = total digestible nutrients.
DP = digestible protein.
ODN = other digestible nutrients.
The increase in animal units fed during this interval was about 16%.

$$(5.9)\quad y_1/z_1 - \underset{(1.83)}{2.7}y_9/z_1 + \underset{(26.92)}{57.5}y_{10}/z_1 - \underset{(299.9)}{384.3}y_9y_{10}/z_1^2 - \underset{(1.507)}{2.1} = 0 \qquad R^2 = 0.81$$

The coefficients were estimated by least squares. The nutrient variables were deflated by the inventory variable. This implies constant returns to scale which was conjectured by several of our consultants. The cross product of the deflated feed variables was introduced to allow for possible interaction between the nutrients. The positive coefficient of y_{10}/z_1 is sufficiently large to make $\partial y_1/\partial y_{10}$ negative over the ranges of the observed values of the variables. This was regarded as sufficiently implausible that no further statistical analysis was performed on this equation.

Another equation that was considered was obtained by introducing time into the equation estimated in (5.7), (5.8). The resulting equation was estimated by both methods and the results are indicated in (5.10) (limited information) and (5.11).

$$(5.10)\quad Y_1 + \underset{(2.07)}{4.25}Y_9 - \underset{(2.04)}{4.66}Y_{10} + \underset{(0.27)}{0.09}Z_1 - \underset{(0.003)}{0.01}z_2 - \underset{(1.42)}{3.90} = 0$$

$$(5.11) \quad \underset{}{Y_1} + \underset{(0.47)}{1.32Y_9} - \underset{(0.46)}{1.84Y_{10}} - \underset{(0.13)}{0.09Z_1} - \underset{(0.008)}{0.006z_2} - \underset{(0.84)}{3.54} = 0 \qquad R^2 = 0.91$$

In these cases the implausibility of the coefficients of Y_{10} made it seem not worth while to apply further analysis.

Possible Modifications in Approach

Reconsideration of our approach to the production relation is clearly in order. A variety of alternative approaches are in principle possible, and we have not investigated them thoroughly enough to try to make strong recommendations at this time. By way of illustration, two possible ways of proceeding are briefly outlined.

On the assumption that an important weakness in the approaches already tried is the lack of observations on inputs other than feed and animals, one might consider trying to specify the missing inputs and the relations that determine their values, in the hope that the other variables in these relations would be observable. If so, one could eliminate the unobservable inputs from these equations, thus obtaining a smaller number of equations that would not contain the unobserved variables. The new equations would be called partially reduced form equations and would have to be reinterpreted in light of the way in which they were derived. Consider the following hypothetical example. Let (5.12), (5.13) be two equations of a larger structural system.

$$(5.12) \qquad y_1 + \beta_{12}y_2 + \gamma_{11}z_1 + \gamma_{12}z_2 + \gamma_{10} = u_1$$

$$(5.13) \qquad y_2 + \beta_{23}y_3 + \gamma_{22}z_2 + \gamma_{23}z_3 + \gamma_{20} = u_2$$

Suppose that y_2 is not observable. By solving (5.13) for y_2 and substituting for it in (5.12), the following relation is obtained.

$$(5.14) \quad y_1 - \beta_{12}\beta_{23}y_3 + \gamma_{11}z_1 + (\gamma_{12} - \beta_{12}\gamma_{22})z_2 - \beta_{12}\gamma_{23}z_3 = (u_1 + \beta_{12}u_2)$$

In general, statistical specifications commonly made in analyzing linear equations will hold for (5.14) if they are valid for (5.12), (5.13). If u_1, u_2 are normal, independent of the z's and serially independent, for example, the new disturbance $(u_1 + \beta_{12}u_2)$ will also have these properties. (5.14) will be less autonomous[13] than (5.12), (5.13) and typically more difficult to interpret, but its substitution for (5.12), (5.13) does not affect the validity of the model.

Another possibility is to seek sampling data on the operations of a number of individual producers and to use these as the basis for the analysis. To draw reliable inferences about the relation among aggregate

[13] For a discussion of the autonomy of economic relations see Trygve Haavelmo. The probability Approach in Econometrics, *op. cit.* pp. 26–39.

variables, the sample data would have to be rather comprehensive, and the model would have to specify (perhaps only in part) relations between individual and aggregate observable variables and between individual and aggregate parameters. Tobin's use of sample data to estimate income elasticity in an aggregate demand relation is an example of this approach.[14] As sampling data continue to become more available the possibilities for successfully developing this approach should improve.[15] The possibility of making use of data provided by technical experiments in this connection is also worth exploring. It is possible that models like those used in *Activity Analysis*[16] could be developed to provide a basis for combining experimental data with sampling data and time series.

[14] See James Tobin., *op. cit.*

[15] Some preliminary speculation about problems that might be encountered in this development are given in Clifford Hildreth, Combining Cross Section Data and Time Series, *Cowles Commission Discussion Paper, Statistics 347*. Cowles Commission for Research in Economics, University of Chicago, Chicago, May 15, 1950. See also

Stephen G. Allen Jr., Estimation of a Single Equation in a Complete System of Stochastic Equations with Cross-Section, Time-Series Data, *Cowles Commission Discussion Paper, Statistics 366*, Cowles Commission for Research in Economics, University of Chicago, Chicago, Oct. 8, 1951.

Lawrence R. Klein. Sample Surveys of Households: a New Tool in Econometrics *Econometrica*, Vol. 19, p. 345, 1951.

[16] See *Activity Analysis of Production and Allocation*, T. C. Koopmans, editor, Cowles Commission Monograph 13, John Wiley & Sons, New York, 1951.