THE USE OF REGRESSION TECHNIQUES WITH ECONOMIC DATA

by

R. L. Anderson
Institute of Statistics, North Carolina State College, Raleigh

1. The Method of Least Squares with a Single Multiple Regression Equation.

1.1. The Mathematical Model. In the traditional regression analysis we assume that we have a single dependent variate Y and several independent fixed variates, X₁, say n in number. These variates are connected by the following mathematical model

\[ Y = b_1X_1 + b_2X_2 + \ldots + b_nX_n + e \]

where the \( b_i \) are estimates of true regression coefficients, \( \beta_i \), and the residual \( e \) is the estimate of a true residual which is assumed to be normally and independently distributed with mean 0 and variance \( \sigma^2 \). Since the independent variates are assumed to be fixed (have no probability distribution), Y is normally and independently distributed with mean \( \sum_{i=1}^{n} \beta_iX_i \) and variance \( \sigma^2 \).

We consider a sample of n values of Y and the corresponding values of the Xᵢ. The assumption of fixed Xᵢ implies that in repeated samples of n from the population specified by (1), the Xᵢ remain fixed from sample to sample. The variates, Y and the Xᵢ may be polynomials, exponentials, logarithms, or any other mathematical functions which will meet the assumptions. With economic data, we frequently use logarithmic values in order to meet the requirement that the variance is constant from observation to observation.

The statistician's problem is to determine the values of the estimates \( b_i \) and an estimate of \( \sigma^2 \) from the sample of n. The method of least squares has certain desirable properties and produces the same estimates as the method of maximum likelihood for the type of distribution assumed for (1). After determining the \( b_i \), it is also desirable to have estimates of their variances and an over-all test of the adequacy of the regression model. The method of least squares minimizes the sum of squares of the sample residuals, e. The quantity to be minimized is

\[ (2) \quad SSE = \sum_{i=1}^{n} (Y - b_iX_i)^2 \]
where $S$ indicates summation over the sample of $x_i$. We have not used a second subscript to indicate the sample number, because it complicates the equations; instead it is understood that wherever $S$ is used, we are summing over the sample.

We form $r$ minimizing equations, such as for $b_1$:

$$
\frac{\partial (SSE)}{\partial b_1} = SX_Y - \left[ b_1 SX_1^2 + b_2 SX_1 X_2 + \cdots + b_r SX_1 X_r \right] = 0
$$

(3)

This is called the leading equation for $b_1$. Therefore we have the following set of $r$ equations in the $r$ unknowns, $b_i$:

$$
b_1 SX_1^2 + b_2 SX_1 X_2 + \cdots + b_r SX_1 X_r = SX_Y
$$

$$
b_1 SX_2^2 + b_2 SX_2 X_3 + \cdots + b_r SX_2 X_r = SX_Y
$$

(4)

$$
\vdots
$$

$$
b_1 SX_r^2 + b_2 SX_r X_{r+1} + \cdots + b_r SX_r^2 = SX_Y
$$

If these equations are solved by the matrix-inversion method, which utilizes either Fisher's $q$-values [4] or Snedecor's $k$-values [12], we have at once the necessary tools to determine an estimate of $\hat{\sigma}^2$ and the variance of each $b_i$, as well as a test of the adequacy of the entire model. For a short-cut matrix-inversion technique, one which we are using at North Carolina State College, I refer you to the abbreviated Doolittle method described by P. S. Dwyer [3]. I believe that the use of matrix-inversion methods in regression analysis needs a little promoting. Once a computer has mastered the matrix-inversion techniques, she can handle regression problems in very little more time than with the elimination method which has been generally used in the past. In addition the matrix-inversion method furnishes the requisite data with which to compute the variances of each $b_i$, something which cannot be obtained in any other manner than by the matrix-inversion method, when $r$ is greater than 2. Professor Peach of our staff has written up an excellent description of the abbreviated Doolittle method with an example. Copies can be obtained by writing to the Institute of Statistics, North Carolina State College.
After we have computed the \( b_1 \), the total reduction in \( SY^2 \) attributable to regression is \( SSR = \sum_{i=1}^{r} b_1(SX_1Y) \). Hence the residual (or error) sum of squares is

\[ SSE = SY^2 - \sum b_1(SX_1Y) = SY^2 - SSR. \]

SSE, under the assumptions made with (1), is distributed as \( \chi^2 \) with \( n - r \) degrees of freedom. Hence an unbiased estimate of \( \sigma^2 \) is \( s^2 = SSE/(n - r) \).

If we set up the null hypothesis that \( \{ \beta_1 = 0 \} \), then the reduction SSR is also distributed as \( \chi^2 \sigma^2 \), but with \( r \) degrees of freedom. Hence we can test this null hypothesis by means of

\[ F = \frac{SSR}{SSE} \cdot \frac{n - r}{r}, \]

with \( r \) and \( n - r \) degrees of freedom.

In most statistical investigations one of the regression coefficients, say \( b_1 \), estimates \( \mu \), the mean of \( Y \), which is not assumed to be 0 when testing the adequacy of the model (1). This is accomplished by setting \( X_1 = 1 \), and transforming all the other \( X_i \) to deviations from their means. Hence the model becomes

\[ Y = m + \sum_{i=2}^{r} b_i x_i + e \]

where \( x_1 = (X_1 - \bar{X}_1) \) and \( m \) is the estimate of \( \mu \). The least squares equations are the same as before except the first equation becomes simply

\[ m = \frac{\sum Y}{n} = \bar{Y}. \]

The other \( r - 1 \) equations use \( x_1 \) instead of \( X_1 \) with \( b_1 \) not in the equations. For example the \( b_2 \) equation is

\[ b_2Sx_2^2 + b_3Sx_2x_3 + \cdots + b_rSx_2x_r = Sx_2Y. \]

As everyone knows, we compute \( Sx_1x_j = SX_1X_j - \frac{SX_1Sx_j}{n} \). Under the null hypothesis, \( \{ \beta_1 = 0 \} \) the expected value of \( Y \) is \( \mu \), not assumed to be zero. The reduction in sum of squares due to regression is

\[ SSR = \sum_{i=2}^{r} b_1(Sx_1Y), \]
which is distributed as $\chi^2 \sigma_i^2$ with $r - 1$ degrees of freedom when \( \beta_i = 0 \),

\[ i = 2, 3, \ldots, r. \]

The error sum of squares is

\[ \text{SSE} = SY^2 - n\bar{Y}^2 - \text{SSR}, \]

which is distributed as $\chi^2 \sigma_i^2$, as before, with $n - r$ degrees of freedom.

If the matrix inversion method of determining the $b_i$ is used, the estimated variance of each $b_i$ is simply $c_{ii}s^2$, where $c_{ii}$ is the diagonal element in the inverted matrix. This variance can be used to attach confidence limits to the estimate of $b_i$ or to make a test of significance by means of

\[ t = \frac{b_i - \hat{b}_i}{s \sqrt{c_{ii}}} \]

with $n - r$ degrees of freedom.

Finally it should be emphasized that the analysis of variance is a special type of a linear regression, in which the $X_i$ are definitely fixed numbers. I have discussed the use of the analysis of variance with economic data emphasizing the assumptions needed, in an article on the analysis of hog prices \[1\].

1.2. An Example of a Dairy Cost Study: For reasons which will be explained in the next section, much economic data do not meet the requirements for the multiple regression analysis. However certain types of data do seem to meet these requirements, such as data acquired from a set of farms, business places, etc., at a given period of time in which certain variables can be considered to be more or less fixed and can be used to estimate some dependent variable. One such study was a dairy cost study of 89 dairy farms in central North Carolina in 1941 \[6\].

Data were available for each farm on the amounts and the costs of concentrates, of silage, and of roughage per cow, the cost of pasture per cow, the amount of milk sold per cow, the number of cows, the net income per cow, the labor income per cow, the hours of labor per cow, and the total digestible nutrients (TDN) per cow. It was decided to use milk sold per cow as the dependent variable since labor income was so variable because of varying allowances for the hours of labor used. A preliminary analysis using labor income as the dependent variable produced no significant regression coefficients.
I shall demonstrate the multiple regression method with the amount of milk sold per cow as the dependent variable \( (Y) \) and the amount of concentrates per cow \( (X_2) \), the amount of silage per cow \( (X_3) \), the amount of roughage per cow \( (X_4) \), and pasture cost per cow \( (X_5) \) as the independent variables. The amounts have been quoted in hundreds of pounds per cow and the pasture cost in dollars per cow. The regression model is:

\[
Y = \beta_0 + \sum_{i=2}^{5} \beta_i X_i + e
\]

with \( x_i = X_i - \bar{X}_i \). The computations are presented in Table 1, using the matrix-inversion method.

**Table 1:** Computations for the Regression of the Amount of Milk Sold on the Amounts of Food Used per Cow from 89 Dairy Herds.

<table>
<thead>
<tr>
<th>Y(milk sold)</th>
<th>X2(concentrates)</th>
<th>X3(silage)</th>
<th>X4(Roughage)</th>
<th>X5(Pasture-cost)</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.994</td>
<td>29.4310</td>
<td>39.0647</td>
<td>36.0326</td>
<td>11.6426</td>
</tr>
</tbody>
</table>

**Sums of Squares and Cross-products**

\[
x_2 \quad x_3 \quad x_4 \quad x_5 \quad Y
\]

\[
\begin{array}{ccccc}
5051.54 & -616.17 & -931.7732 & -48.289 & 3679.74 \\
96710.77 & 3244.25 & 1358.95 & 999.432 & 7028.15 \\
19230.53 & -1251.95 & 999.432 & 7028.15 & \\
1254.57 & 999.432 & 7028.15 & & \\
\end{array}
\]

\[
S_Y^2 - n \bar{Y}^2 = 11358.72
\]

\[
10^{-4} x_i
\]

\[
\begin{bmatrix}
2.236903 & 0.1413543 & 0.0356220 & 0.7459220 \\
0.1413543 & 0.1152455 & -0.8249484 & -0.0951647 \\
0.0356220 & -0.0249484 & 0.5636525 & 0.6032603 \\
0.7459220 & -0.0951647 & 0.6032603 & 8.963883
\end{bmatrix}
\]

**Inverse or \( a \)-matrix**

**Regression Coefficients \( \pm \) Standard Errors**

\[
b_2 = \pm 0.0343 \quad b_3 = \pm 0.0787 \quad b_4 = \pm 0.1021 \quad b_5 = \pm 0.9276
\]

\[
SSR = \sum_{i=2}^{5} b_i (x_i Y) = 4535.08
\]

\[
SSE = 11358.72 - SSR = 6823.64
\]

\[
s^2 = SSE/84 = 81.2338
\]
Some of the computations may require an explanation. In order to illustrate how to compute the values of the $b_i$, consider $b_2$

$$b_2 = 10^{-4} \left[ (2.236903)(3679.74) + (0.1413543)(3905.56) + \cdots + (0.7459220)(702.815) \right] = .9343$$

Then the estimated standard error of $b_2 = s \sqrt{c_{22}} = 0.1348$ and similarly for the other $b_i$, with a standard error of $s \sqrt{c_{ii}}$.

Of these regression coefficients, all but $b_4$ were highly significant. Since in 1941 the cost of concentrates was $1.30 per cwt., of silage $0.27 per cwt., and of roughage $0.85 per cwt. and milk sold for $3.20 per cwt., we would conclude that added concentrates and pasture were profitable, added silage about paid for itself, and added roughage was unprofitable. It should be stated that in a regression analysis, any regression coefficient indicates the average change to be expected in the dependent variable for a unit change in the particular independent variable, with the other independent variables assumed to be held constant. For example in the above analysis, $b_2$ measures the added cwt. of milk expected from an addition of one cwt. of concentrates to the ration, while leaving the amounts of the other foods unchanged. I wish to warn against using such results for increases far beyond the means, because we did not have a sufficient range of the independent variables to study the diminishing returns to scale. Apparently these cows were being fed too much roughage. Even the average amount of roughage for all herds, 3600 pounds per cow, was said to be about all a cow could handle when she was on pasture as long as most cows in this region are. Hence added roughage would be expected to be wasted. If the pasture costs reflect the true cost of pasture, it appears that increased pasture is really the most profitable method of producing milk. Pasture might be put in an even more favorable light if the effects on soil improvement also were considered.

1.3. Production Functions for Iowa Farms. Using data secured in a random sample of 738 Iowa farms in 1939, a study was made of the production functions [9].

The regression model used was:
\[(14) \quad Y = \alpha + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + \varepsilon,\]

where the \(x_i\) are deviations from the means of the \(X_i\). \(Y\) and \(X_i\) are given in terms of logarithms for the following variables:

- \(Y\) = total product: cash sales, home consumption and inventory increases ($)
- \(X_1\) = value of real estate ($)
- \(X_2\) = months of labor (family and hired)
- \(X_3\) = value of machinery and equipment ($)
- \(X_4\) = value of livestock on hand and livestock expense, including feed ($)
- \(X_5\) = cash operating expenses ($)

Each regression coefficient in (14) is a measure of elasticity, indicating the average proportional change in total product which would result from a change in the corresponding independent variate. For example, \(b_1\) represents the elasticity of product with respect to the value of real estate. We did not use a logarithmic relationship for the dairy study \([6]\) because we did not believe that there was enough of a range in the variables studied to justify using it. The logarithmic relationship is necessary for data of this kind if the range in the variables is very great in order that the assumptions of normality and equal variances for the residuals should hold. Also the logarithmic relationship more adequately measures the diminishing returns to scale.

The data were analyzed separately for each of 5 type of farming areas, for each of 5 types of farms for large and small farms, and for all farms. The over-all regression coefficients were:

\[(14a) \quad b_1 = 0.2316, \quad b_2 = 0.0282, \quad b_3 = 0.0844, \quad b_4 = 0.4767, \quad b_5 = 0.0317,\]

with a multiple correlation coefficient of 0.3832. All but \(b_2\) were significant at the .01 probability level. The returns to labor were not significant as was found in our dairy study, probably because of the extreme variability in stating the total labor used on each farm. Since the sum of the coefficients is less than unity, we conclude that there was a diminishing return to scale, since a 1% increase in the input of all resources would not be followed by a 1% increase in output. Using the estimates of the elasticities (14a) and the geometric means of the variables, estimates of the marginal productivities were also secured, with confidence limits.

1.4. The Use of the Analysis of Variance for the Dairy Cost Study. The analysis of variance can also be used to study the relationship between the amount of
milk sold per cow and the average amount of silage, roughage, and concentrates fed to each cow. I have omitted pasture from this analysis because the problem is already quite complicated for an example. We divided the silage feeding into three classes - 0, Low, High; and the concentrates and roughage each into Low and High amounts per cow. The Low-High division was made so that about the same number of herds would be included in each category. The dividing point for silage was 6000 pounds per cow, for concentrates 2900 pounds per cow, and for roughage 3500 pounds per cow. The average milk sold per cow and the number of herds for each of the 3 x 2 x 2 = 12 classes are presented in Table 2.

Table 2: Average cwt.s. of milk sold for different amounts of feed per cow.

<table>
<thead>
<tr>
<th>Silage</th>
<th>Concentrates</th>
<th>Low</th>
<th>High</th>
<th>Low</th>
<th>High</th>
<th>Low</th>
<th>High</th>
<th>Ave.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>53.06(6)</td>
<td>61.92(8)</td>
<td>53.03(10)</td>
<td>60.13(7)</td>
<td>55.42(10)</td>
<td>60.62(3)</td>
<td>56.84(44)</td>
</tr>
<tr>
<td>Roughage</td>
<td>Low</td>
<td>46.61(6)</td>
<td>60.12(10)</td>
<td>64.07(3)</td>
<td>59.99(9)</td>
<td>55.03(9)</td>
<td>62.87(8)</td>
<td>57.95(45)</td>
</tr>
<tr>
<td></td>
<td>Ave.</td>
<td>56.49(30)</td>
<td>47.92(29)</td>
<td>47.81(30)</td>
<td>57.40(89)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The figures in parenthesis are numbers of cows in each class.

We shall consider here the following mathematical model:

(15) \[ Y = \alpha + b_{11}X_{11} + b_{12}X_{12} + b_{20}X_{20} + b_{21}X_{21} + b_{22}X_{22} + b_{31}X_{31} + b_{32}X_{32} + e, \]

where \( X_{11} \) is 1 for the low level of concentrates and 0 elsewhere, \( X_{12} \) is 1 for the high level of concentrates and 0 elsewhere, \( X_{20} \) is 1 for no silage and 0 elsewhere, etc. (15) assumes that the interactions are non-existent. This assumption can be tested by computing the total sum of squares attributable to the 12 classes with 11 degrees of freedom and subtracting the amount due to the main effects with 1 + 2 + 1 = 4 degrees of freedom, leaving 2 interaction degrees of freedom.

The analysis of variance is somewhat complicated for the data given in Table 2 because of the disproportionate frequencies from one class to another. Snedecor presents several techniques of handling this disproportionate problem for
two way tables, but the techniques for 3-way tables are not so simple.

If the averages alone in Table 2 were analyzed and the within mean square adjusted by the harmonic mean of the frequencies for the different classes, we have the following analysis (called the method of unweighted means):

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of freedom</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentrates (C)</td>
<td>1</td>
<td>120.5**</td>
</tr>
<tr>
<td>Silage (S)</td>
<td>2</td>
<td>16.1</td>
</tr>
<tr>
<td>Roughage (R)</td>
<td>1</td>
<td>1.4</td>
</tr>
<tr>
<td>C x S</td>
<td>2</td>
<td>24.4</td>
</tr>
<tr>
<td>C x R</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>S x R</td>
<td>2</td>
<td>22.0</td>
</tr>
<tr>
<td>C x S x R</td>
<td>2</td>
<td>19.0</td>
</tr>
<tr>
<td>Within Classes</td>
<td>77</td>
<td>19.76</td>
</tr>
</tbody>
</table>

The harmonic mean of the frequencies was 12/1.91508 = 6.26606. Since the within mean square was 123.8, the error mean square for this analysis was 123.8/6.26606 = 19.76. Note that only C is significant. Silage does not become significant in this analysis as it did in the regression equation (13). This method of unweighted means, in general, is quite good if the class frequencies are not too unequal.

We might make a complete least-squares solution of the problem, whereby we compute the values of the $b_i$ in (15) by the methods given in Table 1. Since the total milk sold summed over any of these classifications (concentrates, silage, or roughage), is the same as the grand total, we actually have only 4 independent $b$'s in (15). The simplest method of taking account of this difficulty is to let each $b_{12} = 0$ and consider the following regression model.

$$Y = n' + b_1'x_{11} + b_{20}'x_{20} + b_{21}'x_{21} + b_{31}'x_{31} + e,$$

where the $b_i'$ indicates that we have changed the variables to meet the dependency conditions. The coefficients of the $b_i'$ and $m'$ and the right hand-side of the equations (4) are

<table>
<thead>
<tr>
<th>m'</th>
<th>b_{11}'</th>
<th>b_{20}'</th>
<th>b_{21}'</th>
<th>b_{31}'</th>
<th>SX_{i}Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>89</td>
<td>44</td>
<td>30</td>
<td>29</td>
<td>44</td>
<td>5108.55</td>
</tr>
<tr>
<td>44</td>
<td>44</td>
<td>12</td>
<td>13</td>
<td>26</td>
<td>2370.61</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>30</td>
<td>0</td>
<td>14</td>
<td>1694.60</td>
</tr>
<tr>
<td>29</td>
<td>13</td>
<td>9</td>
<td>22</td>
<td>17</td>
<td>1679.66</td>
</tr>
<tr>
<td>44</td>
<td>26</td>
<td>14</td>
<td>17</td>
<td>44</td>
<td>2500.98</td>
</tr>
</tbody>
</table>

The variable for which each is the leading equation is underlined.
We next determine the estimates $b_i$, which are as follows:

\[(22) \quad m^i = 62.4205; \quad b_{11}^i = -7.53957; \quad b_{20}^i = -3.09490; \quad b_{21}^i = -1.34355; \quad b_{31}^i = 0.378992\]

The reduction due to these 5 variates is

\[(23) \quad SSR = \sum b_i^i x_i y + 5108.55m^i = 294,456\]

The reduction due to the 4 feed variates alone is

\[(24) \quad SSR - n\bar{y}^2 = 1228.\]

Hence we can build up the following analysis of variance table to test the interactions.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concentrates (adj.)</td>
<td>1</td>
<td>1159</td>
<td>1159**</td>
</tr>
<tr>
<td>Silage (adj.)</td>
<td>2</td>
<td>139</td>
<td>69.5</td>
</tr>
<tr>
<td>Roughage</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Interactions</td>
<td>7</td>
<td>600</td>
<td>85.7</td>
</tr>
<tr>
<td>Within Classes</td>
<td>77</td>
<td>9530</td>
<td>123.8</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>11358</td>
<td></td>
</tr>
</tbody>
</table>

**Highly significant (probability level of .01)**

As before, we note that there are no significant over-all interaction effects, since the interaction mean square is less than the within classes (error) mean square. Next we would like to break down the main effects into the three separate effects. If any effect is to be tested, that regression coefficient (or coefficients) is omitted from the regression equation and matrix \[(22)\] and \[(23)\] and the reduction in total sum of squares due to the remaining effects is computed. The difference between this reduction and $SSR = 294,456$ is the sum of squares attributable to the omitted effect, adjusted for the other effects. For example, consider the concentrates (adj.) sum of squares. When we omit $b_{11}^i$ from \[(22)\] and omit the second row and column from \[(23)\], the values of the other regression coefficients are:

\[(25) \quad m^i = 58.3246, \quad b_{20}^i = -1.28342, \quad b_{21}^i = 0.29131, \quad b_{31}^i = -1.18837\]

The reduction in sum of squares due to these variables is 293,297. Hence the concentrate (adjusted) sum of squares is
This is the only significant effect. Note that the sum of the main effects sums of squares is greater when they are considered separately than when together. The mathematical statistician recognizes this as evidence of a certain correlation between the effects because of the disproportionate frequencies. However, the frequencies in this case were not variant enough to produce much of a discrepancy between the two sums of squares.

1.5. Summary. I have presented an example of the use of the multiple regression and the analysis of variance models in determining the best use of feed for producing milk. The analysis of variance model is not quite so sensitive to differences in this case, because we did lose some information in grouping the independent variables as we did in the analysis of variance model. The analysis of variance is generally used when the independent variables do not vary from place to place. For example, we might have used 0, 4000, and 8000 pounds of silage; 2000 and 4000 pounds of concentrates; and 3000 and 6000 pounds of roughage per cow. Then we would have divided our herds into 12 groups, at random, and fed each group one of the 12 rations. In this case the analysis of variance model (15) would have been an exact representation of the experiment, and when an experiment is so planned, we know that the assumption of fixed $X_i$ is met exactly.

In my article on the analysis of hog prices $[1]$, I had data on the ratios between the prices received at Cincinnati and Louisville for each of two weight classes on each of the five complete market days of the week averaged for each month over five years. Since we had complete data, the disproportionate frequencies problem was avoided. In this case the simple analysis of variance techniques for multiple classifications could be used.

These examples have been presented to show that there are certain types of economic data which meet the requirements of the multiple regression model (1).

These requirements are: (i) Fixed independent variates from sample to sample, (ii)
normal and independent residuals with the same variance, and (iii) the independent effects are additive. The economist finds that (i) is often not met unless he can plan the experiment as suggested above for the dairy study or he can regard the $X_i$ as rather typical of the situation which will continue to prevail. Transformations of the type explained by M. S. Bartlett $[2]$ can usually be used to handle non-normality, unequal variances, and lack of additivity. For example multiplicity of the effects is reduced to additivity by the use of logarithms. The problem of lack of independence of the residuals and correlation between residuals and the main effects has not as yet been satisfactorily solved. Some use of the tests for serial correlation may be made here. These problems have been discussed in detail in the hog price article $[1]$.


2.1 Defects of the Multiple Regression Method. Much economic data do not meet the requirements for using the multiple regression model (1). There are four main difficulties.

(i) Many of the $X$-variates cannot be considered any more constant than the $Y$-variante. As Professor Marschak has stated $[11]$: 

The economist has no independent variables at his disposal because he has to take the values of all variables as they come, produced by a mechanism outside his control. This mechanism is expressed by a system of simultaneous equations, as many of them as are variables. The experimenter can isolate one such equation, substituting his own action for all the other equations. The economist cannot. (page 143).

For example, the economist may be given a series of average prices and quantities sold at these prices. But these series cannot be summarized in a single equation which reflects how prices and quantities (and other economic variables) interacted to produce the final results. We might draw up a single regression equation expressing price as a function of quantity, time and any other pertinent variables. However this prediction equation probably would tell us nothing of the price which would have been obtained at a given point of time if some of the other variables had taken on different values. This prediction
equation would be merely a historical record of the price-quantity relationship through time. T. Haavelmo was one of the first econometricians to recognize that if a system of equations is needed to describe a given economic situation, the estimates of the $\beta_1$ for any single equation will be biased unless the entire system is considered simultaneously [7]. In order to estimate the parameters in a system of equations, we must use the general method of maximum likelihood. The mathematical complexities in many cases are almost insurmountable. To date, methods of estimation have been developed but no adequate tests of significance.

(ii) Another major problem in economic regression models is the "identification" of each regression equation. For example, we may want to estimate a demand and a supply equation or a production function and a profit maximizing equation. It is necessary, if the system of equations is to be used as a guide for economic policy, that the analyst be able to "identify" each equation after the estimation has been completed.

(iii) When the data used in estimating the parameters in an econometric system are derived from a time series, we run into another defect in the usual regression model—the residuals are not independently distributed. Techniques of handling these difficulties are being developed, but these techniques merely add to the already overburdened complexities of the estimation procedures. As noted before, aberrations from the other assumptions regarding the distribution of the residuals—normality and the same variance—can usually be taken care of by the use of transformations.

(iv) As noted before, it is also necessary that the effects be additive.

The problem of setting up systems of economic equations has been under consideration by the Cowles Commission for Research in Economics at the University of Chicago for almost a decade. The research has been directed by Professor J. Marschak with the able assistance of such men as T. W. Anderson, W. H. Andrews,
M. A. Girshick, T. Haavelmo, L. Hurwicz, L. R. Klein, T. Koopmans, H. B. Mann, H. Rubin, G. Tintner, and A. Wald, O. Reiersol, M. G. Kendall, H. Wold and J. R. N. Stone, to mention a few, have also made important contributions to the problems involved. A list of some of the pertinent publications is given at the end of this article. An exhaustive and rigorous treatment of the theoretical results so far obtained will be published soon in Statistical Inference in Dynamic Economic Models, Cowles Commission, Monograph No. 10. One of the best articles which presents the basic concepts of this new approach to the use of regression equations in economics is The Probability Approach to Econometrics by T. Haavelmo [8].

2.2. An Example of the Simultaneous Equation Model. As an example of the techniques used by these econometricians, I shall use the results presented by Girshick and Haavelmo in Statistical Analysis of the Demand for Food [5]. It is shown that the demand equation cannot be estimated unless the supply equation is either known or estimated simultaneously with the demand equation. A system of five equations was set up.\(^1\)

\[
\begin{align*}
Y_1(t) &= a_{10} + a_{12}Y_2(t) + a_{13}Y_3(t) + b_{18}X_8(t) + b_{19}X_9(t) + e_1(t) \\
Y_1(t) &= a_{20} + a_{22}Y_2(t) + a_{24}Y_4(t) + b_{28}X_8(t) + e_2(t) \\
Y_3(t) &= a_{30} + b_{37}X_7(t) + b_{39}X_9(t) + e_3(t) \\
Y_4(t) &= a_{40} + a_{45}Y_5(t) + b_{46}X_6(t) + b_{48}X_8(t) + e_4(t) \\
Y_5(t) &= a_{50} + a_{52}Y_2(t) + b_{58}X_8(t) + e_5(t)
\end{align*}
\]

(27)

In (27) \(Y_1\) = consumption of food per capita, \(Y_2\) = retail price of food products divided by the cost of living, \(Y_3\) = disposable income per capita divided by the cost of living, \(Y_4\) = production of agricultural food products per capita, \(Y_5\) = prices received by farmers for food products, divided by the cost of living, \(X_6 = Y_5\) for the previous year, \(X_7\) = net investments per capita divided by the cost of living, \(X_8\) = time (1922 = 1 through 1941 = 20), \(X_9\) = \(Y_3\) for the previous year.

\(^1\)Note that \(a_{10}\) are not means but \(y\)-intercepts. For example, \(a_{30} = Y_3 - b_{37}X_7 - b_{39}X_9\). Hence we actually use the model corresponding to (7) and (9).
Table 3: Data used in Girshick-Haavelmo Study.

<table>
<thead>
<tr>
<th>Year</th>
<th>Y_1</th>
<th>Y_2</th>
<th>Y_3</th>
<th>Y_4</th>
<th>Y_5</th>
<th>Y_6</th>
<th>Y_7</th>
<th>Y_8</th>
<th>Y_9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1922</td>
<td>98.6</td>
<td>100.2</td>
<td>87.4</td>
<td>108.5</td>
<td>99.1</td>
<td>98.0</td>
<td>92.9</td>
<td>1</td>
<td>77.4</td>
</tr>
<tr>
<td>1923</td>
<td>101.2</td>
<td>101.6</td>
<td>97.6</td>
<td>110.1</td>
<td>99.1</td>
<td>99.1</td>
<td>142.9</td>
<td>2</td>
<td>87.4</td>
</tr>
<tr>
<td>1924</td>
<td>102.4</td>
<td>100.5</td>
<td>96.7</td>
<td>110.4</td>
<td>98.9</td>
<td>99.1</td>
<td>100.0</td>
<td>3</td>
<td>97.6</td>
</tr>
<tr>
<td>1925</td>
<td>100.9</td>
<td>106.0</td>
<td>98.2</td>
<td>104.3</td>
<td>110.8</td>
<td>98.9</td>
<td>123.8</td>
<td>4</td>
<td>96.7</td>
</tr>
<tr>
<td>1926</td>
<td>102.3</td>
<td>108.7</td>
<td>99.8</td>
<td>107.2</td>
<td>108.2</td>
<td>110.8</td>
<td>111.9</td>
<td>5</td>
<td>98.2</td>
</tr>
<tr>
<td>1927</td>
<td>101.5</td>
<td>106.7</td>
<td>100.5</td>
<td>105.8</td>
<td>105.6</td>
<td>108.2</td>
<td>121.4</td>
<td>6</td>
<td>99.8</td>
</tr>
<tr>
<td>1928</td>
<td>101.6</td>
<td>106.7</td>
<td>103.2</td>
<td>107.8</td>
<td>109.8</td>
<td>105.6</td>
<td>107.1</td>
<td>7</td>
<td>100.5</td>
</tr>
<tr>
<td>1929</td>
<td>101.6</td>
<td>108.2</td>
<td>107.8</td>
<td>103.4</td>
<td>103.7</td>
<td>109.3</td>
<td>142.9</td>
<td>8</td>
<td>103.2</td>
</tr>
<tr>
<td>1930</td>
<td>99.8</td>
<td>105.5</td>
<td>96.6</td>
<td>102.7</td>
<td>100.6</td>
<td>108.7</td>
<td>92.9</td>
<td>9</td>
<td>107.8</td>
</tr>
<tr>
<td>1931</td>
<td>100.3</td>
<td>95.6</td>
<td>88.9</td>
<td>104.1</td>
<td>81.0</td>
<td>100.6</td>
<td>97.6</td>
<td>10</td>
<td>96.6</td>
</tr>
<tr>
<td>1932</td>
<td>97.6</td>
<td>88.6</td>
<td>75.1</td>
<td>99.2</td>
<td>68.6</td>
<td>81.0</td>
<td>52.4</td>
<td>11</td>
<td>88.9</td>
</tr>
<tr>
<td>1933</td>
<td>97.2</td>
<td>91.0</td>
<td>76.9</td>
<td>99.7</td>
<td>70.9</td>
<td>68.6</td>
<td>40.5</td>
<td>12</td>
<td>75.1</td>
</tr>
<tr>
<td>1934</td>
<td>97.3</td>
<td>97.9</td>
<td>84.6</td>
<td>102.0</td>
<td>81.4</td>
<td>70.9</td>
<td>64.3</td>
<td>13</td>
<td>76.9</td>
</tr>
<tr>
<td>1935</td>
<td>98.0</td>
<td>102.3</td>
<td>90.6</td>
<td>94.3</td>
<td>102.3</td>
<td>81.4</td>
<td>78.6</td>
<td>14</td>
<td>84.6</td>
</tr>
<tr>
<td>1936</td>
<td>99.2</td>
<td>102.2</td>
<td>103.1</td>
<td>97.7</td>
<td>105.0</td>
<td>102.3</td>
<td>114.3</td>
<td>15</td>
<td>90.6</td>
</tr>
<tr>
<td>1937</td>
<td>100.3</td>
<td>102.5</td>
<td>101.1</td>
<td>101.1</td>
<td>110.5</td>
<td>105.0</td>
<td>121.4</td>
<td>16</td>
<td>103.1</td>
</tr>
<tr>
<td>1938</td>
<td>100.2</td>
<td>97.0</td>
<td>96.4</td>
<td>102.3</td>
<td>92.5</td>
<td>110.5</td>
<td>78.6</td>
<td>17</td>
<td>105.1</td>
</tr>
<tr>
<td>1939</td>
<td>104.1</td>
<td>95.8</td>
<td>104.4</td>
<td>104.4</td>
<td>89.3</td>
<td>92.5</td>
<td>109.5</td>
<td>18</td>
<td>96.4</td>
</tr>
<tr>
<td>1940</td>
<td>105.3</td>
<td>96.4</td>
<td>110.7</td>
<td>108.5</td>
<td>93.0</td>
<td>89.3</td>
<td>128.6</td>
<td>19</td>
<td>104.4</td>
</tr>
<tr>
<td>1941</td>
<td>107.6</td>
<td>100.3</td>
<td>127.1</td>
<td>111.3</td>
<td>106.6</td>
<td>93.0</td>
<td>238.1</td>
<td>20</td>
<td>110.7</td>
</tr>
</tbody>
</table>

Mean 100.8 100.7 97.5 104.2 97.1 96.7 108.0 10.5 95.0

Sums of Squares and Products

<table>
<thead>
<tr>
<th>x_6</th>
<th>x_7</th>
<th>x_8</th>
<th>x_9</th>
<th>x_1</th>
<th>x_2</th>
<th>x_3</th>
<th>x_4</th>
<th>x_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3071.72</td>
<td>3963.81</td>
<td>-415.25</td>
<td>1714.92</td>
<td>257.79</td>
<td>920.30</td>
<td>1407.01</td>
<td>364.65</td>
<td>2169.42</td>
</tr>
<tr>
<td>32367.31</td>
<td>658.15</td>
<td>4956.04</td>
<td>1793.92</td>
<td>1870.02</td>
<td>8385.36</td>
<td>3657.55</td>
<td>6297.52</td>
<td></td>
</tr>
<tr>
<td>665.00</td>
<td>317.70</td>
<td>72.25</td>
<td>-257.75</td>
<td>436.85</td>
<td>-172.70</td>
<td>-306.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2067.07</td>
<td>401.62</td>
<td>430.58</td>
<td>1764.46</td>
<td>307.19</td>
<td>1290.24</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All data are in terms of indexes with 1935-39 = 100. These data are presented in Table 3. The Y(t) variables are considered to be jointly dependent variables, while the X(t) variables are regarded as fixed or predetermined variables. The first equation in (27) represents the demand for food by the consumer, the second the supply of food in the retail market, the third the income equation, the fourth the supply of foodstuffs by farmers, and the fifth the demand for farm food products by the commercial sector (for sale, export, and maintenance of stocks). Other variables might be included in (or excluded from) these five equations, but this nodal was decided upon as one which would present all of the computational problems to be faced in the estimation of the parameters in a system of equations.
Incidentally, one important simplification would have been to regard \( Y_4(t) \) as a fixed variable so far as the retail market was concerned. This would have reduced the number of required equations to four.

The number of parameters to be estimated in (27) are 19 regression coefficients plus the variances and covariances of the residuals. As explained before, the data were yearly indexes for each of the twenty years, 1922-1941. It is impossible at this time to estimate how many degrees of freedom are represented in 5 series of jointly dependent variables (Y) and 4 series of predetermined variables (X), nor can any statement be made as to the number of degrees of freedom used in estimating the parameters. As I stated above, no tests of significance have been developed to determine the over-all adequacy of the model and the usefulness of various terms in the individual equations. The complicating feature of a possible serial correlation in the series of indexes has also been neglected in the analysis presented in this article.

I do not intend to outline the computing procedures used in estimating the parameters. These have been adequately presented in the article itself, and are too highly mathematical to be condensed into a short summary. I might mention that equation 3 in (27) can be handled by the method of least squares, since it contains but one dependent variable, the other variables being fixed. The problem of identifiability can be illustrated by a comparison of the first two equations in (27). If we let \( H \) be the total number of \( Y \)'s in a given equation and \( K \) the number of \( X \)'s not used in this equation, then an equation is said to be completely identified if \( K = H - 1 \). It is said to be over-identified if \( K > H - 1 \). And if \( K < H - 1 \), some restrictions must be imposed on the variance-covariance matrix of the residuals. Hence for identification purposes, \( K \geq H - 1 \). It should be emphasized that these are mathematical limitations on the model. We see that in the first equation \( K = H - 1 = 2 \) and in the second equation \( K = 3 \) and \( H - 1 = 2 \). Hence both equations are identified, but the second is over-identified. The
difference between complete and over-identification is merely a difference in computational procedures used in estimating the regression coefficients. From a practical standpoint, identification involves the necessary features in an equation so that it is distinguishable from all other equations in the system. Comparing the first two equations, we note that the demand equation has two income variables (one fixed) and no production variable, while the supply equation has a production variable and no income variables.

The estimates of the regression coefficients found in this study were as follows:

\[
\begin{align*}
  a_{10} &= 97.677, & a_{12} &= -0.246, & b_{18} &= -0.104 \\
  a_{20} &= 13.319, & a_{22} &= 0.157, & b_{28} &= 0.399 \\
  a_{30} &= 40.731, & a_{52} &= 2.883, & b_{48} &= -0.190 \\
  a_{40} &= 81.250, & a_{13} &= 0.247, & b_{58} &= 0.656 \\
  a_{50} &= -200.068, & a_{24} &= 0.653, & b_{19} &= 0.051 \\
 & & a_{45} &= 0.556, & b_{39} &= 0.367 \\
 & & & b_{37} &= 0.203 \\
 & & & b_{46} &= -0.300
\end{align*}
\]

The \( a_{12} \) regression coefficients measure the influence of the retail price of food, respectively, on consumer demand, on retail supply, and on commercial demand. We note that the values of \( a_{12} \) correspond to the economic theory that consumer demand should decrease as the price increases and that the other two should increase. Similarly we note that the demand increases with an increase in the disposable income (\( a_{13} \)), that the supply of retail goods increases with an increase in the agricultural production (\( a_{24} \)), and that agricultural production increases with an increase in the current price level of farm products (\( a_{45} \)). Also we note the depressing influence of the previous year's farm prices on the current farm production (\( b_{46} \)), which I cannot explain. Also the uneven influence of the time variable (\( X_8 \)) needs some explanations. Income for the previous year appears to have had little influence on the consumer demand in the current year (\( b_{19} \)) but naturally was closely related to the current year's income (\( b_{39} \)). And finally we note the apparent sizeable increase in income when net investment was
increased (b37). It is unfortunate that it is not possible to place confidence limits on these estimates.

If we were merely interested in forming predictive equations for the dependent variables \( Y \) in terms of some or all of the predetermined variables \( X \), we could have used the sums of squares and cross-products given in Table 3 and estimated, by the single equation method, the regression coefficients in these equations:

\[
\hat{Y}_i = a_i + b_{16}^i X_6 + b_{17}^i X_7 + b_{18}^i X_8 + b_{19}^i X_9 \quad (i = 1, 2, 3, 4, 5),
\]

where \( \hat{Y}_i \) is the predicted value of \( Y_i \), neglecting the residual. These equations are called "reduced" equations and can also be determined by solving the equations in (27), omitting the \( e_i \), simultaneously for the \( Y_i \)'s in terms of the \( X \)'s. We note that equations 3. in (27) and in (29) are the same, because \( Y_3 \) was estimated only from predetermined variables. Equations (29) in general do not represent fundamental economic relationships, such as demand or supply functions. As Koopmans states [10].

We consider the problem of estimation of parameters like elasticities of demand or supply, marginal propensities to consume, etc., not the problem of prediction of the value of economic variables like production, prices, etc. It is true that the use to which we put estimates of the parameters of relations between economic variables is again one of prediction. But often this is a different type of prediction, in which the effects of certain presumed acts of policy like price regulation, or instituting compensatory public works, or influencing savings habits, etc., are to be predicted. In such cases we are dealing with a type of prediction in which one or more of the relations found to govern the past are altered, and which is therefore not a straight forecast assuming continuation of all past relationships. (Page 451).

In this article, which presents an excellent condensation of the simultaneous equation method, Koopmans points out that under certain restrictive circumstances it is possible to use the single equation estimates, \( b_i^1 \), in (29) to determine all of the structural coefficients in (27). However, these conditions are seldom met in practice, so that it is necessary to use the more complicated procedures which have been developed by the Cowles Commission Research Group. Koopmans also points out that the use of time-lags often allows the statistician to use single equation estimation with a system of equations [10].
In order to demonstrate how well the four predetermined variates \((X)\) could be used as predictors of each of the dependent variates \((Y)\), we present in Table 4 the regression coefficients for the reduced equations (29) with their standard errors, based on single-equation estimation. Also included is a measure of the adequacy of prediction in each case, the squared multiple correlation coefficient \(R^2\),

\[
R^2 = \frac{SSR}{Sy^2}, \quad \text{where} \quad Sy^2 = SY^2 - \overline{e}^2.
\]

Table 4: Reduced Form Regression Coefficients and the Standard Errors for the Study of the Demand for Food.

<table>
<thead>
<tr>
<th>Dependent Variate</th>
<th>Constant (a)</th>
<th>(X_1)</th>
<th>(X_2)</th>
<th>(X_3)</th>
<th>(X_4)</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_1)</td>
<td>87.932</td>
<td>-0.059</td>
<td>0.040**</td>
<td>-0.041</td>
<td>0.154*</td>
<td>0.7595**</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.011)</td>
<td>(0.085)</td>
<td>(0.070)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_2)</td>
<td>80.560</td>
<td>0.241</td>
<td>0.041</td>
<td>-0.253</td>
<td>-0.052</td>
<td>0.5865**</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.028)</td>
<td>(0.219)</td>
<td>(0.179)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_3)</td>
<td>40.731</td>
<td>0.203**</td>
<td></td>
<td>0.367**</td>
<td></td>
<td>0.8833**</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td></td>
<td></td>
<td>(0.038)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_4)</td>
<td>97.923</td>
<td>-0.128</td>
<td>0.062*</td>
<td>-0.487*</td>
<td>0.180</td>
<td>0.5651*</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.024)</td>
<td>(0.184)</td>
<td>(0.150)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Y_5)</td>
<td>45.072</td>
<td>0.649*</td>
<td>0.161*</td>
<td>-0.078</td>
<td>-0.287</td>
<td>0.6549**</td>
</tr>
<tr>
<td></td>
<td>(0.277)</td>
<td>(0.058)</td>
<td>(0.452)</td>
<td>(0.370)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at .05 probability level; **at the .01 level.

The regression coefficients which are given in the first line for each dependent variable refer to the reduced equations (29) in the text. The values in the second line (in parentheses) are the standard errors, computed on the basis of single-equation estimation.

From Table 4 we can draw some inferences about the reliability of the estimates of the reduced form regression coefficients.

(i) The regression of \(Y_2\) on the \(X\)'s is rather peculiar in that no single regression coefficient is significantly greater than its standard error, yet the overall reduction is highly significant.

(ii) We note that in all cases the dependent variables declined through time \((X_3)\); however, the decreases for \(Y_1\) (food consumption per capita) and \(Y_5\) (prices received by farmers) were smaller than their standard errors. The only real reduction was in \(Y_4\) (agricultural production) probably because of the AAA program after 1932.
Remember that this shows a reduction in $Y_5$, after adjusting for the effects of the other independent variates.

(iii) Prices received by farmers during the previous year ($X_6$) exerted a stimulating effect on the two dependent price variables ($Y_2$ and $Y_5$), especially $Y_5$ which should be highly related to its previous year's data. The unexplainable negative influence of $X_6$ on the current agricultural production ($Y_4$) is not significant but still disturbing.

(iv) Net investment ($X_7$) stimulated all of the dependent variables, and usually by a large amount.

(v) The influence of the previous year's income ($X_9$) was definitely positive on consumption ($Y_1$) and of course on the current year's income ($Y_3$). The unexplained negative effects on prices were smaller than their standard errors.

If we could solve directly for the structural regression coefficients (27) in terms of the reduced form ones (29), then we might attach tentative standard errors to the structural coefficients using the standard errors given in Table 4.

It should be reaffirmed that we cannot use equations (29) for policy formulations because they do not represent real economic relationships. For example, the $Y_1$ and $Y_2$ equations (29) are mixed-up supply and demand equations. $Y_1$ cannot be called a demand equation, because it does not consider the influence of current prices on the consumption; similarly $Y_2$ does not consider the influence of consumption or production on retail prices. In other words, we must consider some of the dependent variables along with the independent variables if a real economic relationship is to be estimated.

Comparing the various $R^2$, we note that the variable requiring only independent variables for estimation ($Y_3$) has by far the largest $R^2$. This rather sustains a statement made at the close of the Girshick-Haavelmo article:
We should like to point out, however, that when one is searching for structural economic relations one cannot in general expect to find as high correlations as those obtained from a mechanical application of the method of multiple correlation to the variables in the structural equations. For the correlations obtained by the latter procedure are due not only to the acceptance of the same predetermined variables in the equation of the reduced form, but also to the intercorrelations between the residuals in these equations. The method of multiple correlations would produce high correlation at the expense of a bias in the estimates of the structural coefficients involved. (Page 110).

Girshick and Havelmo have used the term "multiple correlation" in the same sense that I have used "multiple regression" and probably refer to the higher correlations which would be obtained when some of the dependent variables are included with the present independent ones in single equation estimation. Note that this inclusion would not be used for \( Y_3 \), which could be estimated from \( X \)'s alone.

As all these writers have indicated, this analysis using a system of equations cannot be said to be complete until we can determine some means of attaching confidence intervals to the estimates of the structural regression coefficients. Tintner has developed a suggested plan of determining how many equations are needed in the system and also gives some approximate confidence intervals for the regression coefficients \( [13] \). However he attacks the problem by assuming the errors occur in the measurement of the variables and not in the equations themselves. The Cowles Commission Group has neglected errors of measurement and has considered only residual errors in the equations.

2.2. **Summary.** I have attempted to present some results of an analysis of food prices using a system of equations, rather than a simple multiple regression equation. This analysis was used because the estimates of the parameters in a single structural equation will be biased if there is actually a relationship between this equation and other unspecified equations. Also it is impossible to identify a single equation as either a demand, supply or income equation unless all of these equations are included in the system. And the statistician must be very careful when constructing a system of equations that he can identify each
equation when the estimation has been completed. The main drawback in the use of this model of a system of equations is that no tests of significance are available to determine either the adequacy of the model or the reliability of the estimates of the individual parameters. However, I anticipate that this gap in the theory will soon be filled.

REFERENCES CITED


OTHER REFERENCES


