A Recursive Model of the Hog Industry

By Arthur A. Harlow

There are a number of estimation methods available to the econometrician. They range in complexity from least squares, which takes only a few minutes with pencil and paper for a simple problem, to full information maximum likelihood, which requires a large-capacity electronic computer for effective application. A list of the common methods, given in order of increasing complexity, includes least squares, two-stage least squares, instrumental variables, limited information, and full information. Recursive systems are fitted by least squares which makes estimation relatively simple, and they have other desirable statistical properties. Following a brief discusssion of the special characteristics of recursive systems, this paper presents an application of the recursive method to an analysis of the hog industry.

The author is indebted to Anthony S. Rojko and Hyman Weingarten of the Economic Research Service for their helpful suggestions during the preparation of this paper.

THE STATISTICAL properties of recursive models have been discussed for a number of years, a recent proponent being Herman Wold. But for all the theoretical examination of the method, only a few empirical applications have been made, notably those of Wold (*18*)[1] and Barger and Klein (*1*). Performance under actual conditions is a valuable criterion for evaluating theoretical systems.

Wold (*19*) has shown that under certain conditions recursive systems fitted by least squares produce maximum likelihood estimates of the structural coefficients. This is in addition to the unbiased estimates of the dependent variable that the least squares method normally gives. Maximum likelihood estimates have certain optimum properties that recommend them to the statistician. The estimates are consistent—it is unlikely that in large samples there would be any great difference between the estimate and the true value. They are also efficient in the sense that for large samples they have a smaller variance or standard error than any other estimate when the observations are from a normal population.

In order to give some basis for evaluating the performance of the method, a simple econometric model is described and the special characteristics of recursive systems briefly discussed.

Characteristics of a Recursive Model

A single equation regression model that assumes no errors of measurement in the observed variables can be written in the following form:

$$y_t = b_1 x_{1t} + b_2 x_{2t} + \ . \ . \ . \ + b_k x_{kt} + u_t.$$

The dependent variable y and the independent variables x_1, x_2 . . . x_k are observed; the disturbance u is an unobservable random variable. The index t runs from 1 to N—there are N observations in the sample. The b's are constant but unobservable coefficients of the independent variables. The usual procedure is to estimate b_1, b_2 . . . b_k, and to make confidence statements about the estimates from the given sample.

Similar definitions are used for multiple equation systems, although an additional problem arises when there is a simultaneous relationship between two or more variables in the system. This would occur, for example, when supply and price affect each other within the given time period. Under the condition of a simultaneous relationship between two variables, the use of a single regression equation fitted by least squares may result in a statistical bias in the estimation of the b coefficients as indicated by Haavelmo (*9*). To avoid this bias, a system of equations should be solved simultaneously. In these equation systems, the variables are commonly divided into two categories—endogenous variables and predetermined variables. The former are those variables whose relationships among themselves and with other variables are to be determined from knowledge of the predetermined variables whose values are assumed to be known. These two categories thus correspond with the dependent and inde-

[1] Italic numbers in parentheses refer to Literature Cited, page 11.

pendent classification in a single regression equation.

Certain assumptions must be made about the disturbances u_t in order to simplify computation and to obtain estimates of the coefficients with desirable statistical properties such as consistency and efficiency. The following assumptions are commonly made for most methods of estimation: u_t is a normally distributed random variable with zero mean and constant variance; u_t is distributed independently of the predetermined variables; and successive disturbances are distributed independently of one another, or in other words, u_t is not serially correlated.

A recursive system consists of a set of equations each containing a single endogenous variable other than those that have been treated as dependent in prior equations. The endogenous variables enter the system one by one, like links in an infinite chain where each link is explained in terms of earlier links. The explanation proceeds recursively, link by link, from one period to the next, and within each period in a specified order.

The general mathematical condition for the use of the recursive method is that the Jacobian of the transformation connecting the disturbances with the endogenous variables be triangular and equal to one. The Jacobian of the transformation is the matrix of partial derivatives of u_t with respect to y_t. This general condition may seem quite complicated, but there is a simple method of testing it. If the structural equations in a system are linear in the variables, the elements of the Jacobian are identical with the coefficients of the endogenous variables. So when these coefficients form a triangular matrix, the general condition is met.

In order for the least squares parameter estimate of the recursive system to be consistent and efficient, certain assumptions concerning the disturbances must be met. One of the assumptions is that they are independent of the predetermined variables in the equation. This can be met by assuming that the covariance matrix Σ of the residuals u_t is diagonal. This means that the covariance matrix has zeros everywhere except on the diagonal, i.e., u_t has constant nonzero variance and zero covariance. Thus the disturbance of one equation is not correlated with the disturbance in any other equation of the model in the same time period. This additional assumption permits the use of actual rather than calculated values in the estimating equations. It also leads to maximum likelihood estimates by the single equation least squares method in a recursive model.

However, if the disturbances in the several equations in a system are correlated with one another, they are also correlated with variables that were dependent in preceding equations of the system, since by definition a dependent variable in a particular equation is assumed to be correlated with the disturbances in that equation.

One way to overcome this correlation is to use calculated values for endogenous variables serving as independent variables in succeeding equations. Calculated values for a dependent variable in an equation are known to be uncorrelated with the disturbances in that equation because the residuals are ignored in the computation. Hence calculated values for an endogenous variable obtained from one equation are uncorrelated with the disturbances in another equation within the same system.

With triangularity of the Jacobian but no restrictions on the Σ matrix, single equation least squares estimates are not necessarily also maximum likelihood estimates. Successive application of least squares methods to separate equations using calculated values of previously dependent variables leads to consistent but not efficient estimates in this case. This method is not recommended for large systems with a long chain of substitution of results from some equations into others. The accumulation of errors may eventually become prohibitive.

In many cases it does not appear justified to assume the diagonality of the covariance matrix. This assumption is clearly inapplicable in the following situations: (1) In a given year, all observations were subject to larger than usual errors because the funds available for the collection of statistics were less than normal. (2) One variable, such as national income, is affected by errors of aggregation (perhaps due to shifts in distribution), and this variable enters several equations of the model. (3) Omitted variables are known to affect two or more equations—for instance, weather affects all crop production. On the other hand, if two crops are grown in widely separated areas, the inclusion of weather in the random disturbances is permissible because relatively independent drawings of weather affect the two crops.

But unless there is a specific reason for assuming that the Σ matrix is not diagonal, it is usually worthwhile to make this assumption because this can reduce computations for maximum likelihood estimates by a factor of 2 or 3 for a model of 3 equations and by a much greater factor for larger systems.

Formulation of the Model

In formulating an econometric model, the nature of the economic system to be analyzed should determine the type of equations to be used and the method used in fitting them.

The initial task in analysis is thus to determine the nature of the economic system under consideration. Wold (*17*) believes that many economic systems are of the recursive type, whereas certain other econometricians believe that systems of simultaneous relations are more typical. But regardless of which type is more prevalent, cobweb models are considered to fall into a recursive pattern, and it is generally agreed that hog production is a fairly good example of a cobweb phenomenon.

As an extremely simplified example, hog production can be described as a cobweb model where price in one period affects production in the next, which in turn affects price, and so on. The equations for a simple cobweb model where consumption equals production are

$$Q_t = aP_{t-1} + u_t$$
$$P_t = bQ_t + v_t$$

These two equations are linear in the variables and the matrix of the coefficients of the endogenous variables is triangular, so it satisfies the conditions for a recursive system. Of course, a more complete model would include more variables and more equations. Such a model is presented later in this paper.

The time devisions used in the analysis have considerable effect on the simultaneity of the determination of variables. With short time units, a recursive or causal chain system might be more appropriate than a simultaneous equation system. But when substantial units of time, such as a year, are used, more variables have time to interact and hence the simultaneous system is perhaps preferable.

Thus, in a recursive model, quarterly data are often preferred to annual observations. However, in quarterly models serial correlation in the disturbance terms tends to become more of a problem than when annual data are used. Barger and Klein (*1*) have attempted to solve this difficulty by randomizing the disturbance terms through autoregressive schemes. But the use of such schemes amounts to an admission that an economic explanation for a systematic part of the movement in the dependent variable has not been found.

The omission of relevant variables is an important reason for the existence of serial correlation in the residuals. Economic variables tend to be serially correlated. When some relevant economic variables are omitted from a structural equation, their effects on the dependent variable are left with the unexplained residuals. As a result, the residuals are also serially correlated. An attempt to deal with this difficulty would thus require the inclusion of more explanatory variables in the structural relationships. These additional variables may be either current or lagged.

Nerlove and Addison (*13*) have found that the inclusion of the lagged quantity of a given commodity in its demand and supply functions is not only theoretically sound but that it also reduces serial correlation in the estimated residuals to insignificant levels. In the supply functions of agricultural products, lagged variables probably play a more important part than do current variables. In the case of demand functions, however, a number of current variables, such as current prices of competitive and complementary products, may be at least as important as lagged quantity as explanatory variables. And they may also significantly reduce serial correlation in the residuals.

The following six equations, fitted by least squares to quarterly data for 1949 through 1959, form a recursive system based on an expansion of the simple cobweb model given earlier. Although there are now additional variables and more equations, production is still basically a function of lagged price, and price is a function of present production. As intermediate steps, there are now equations for the number of sows farrowing, hogs slaughtered, cold storage holdings of pork, and farm price of hogs.

$$(1)\quad F = -1163.260 + \underset{(.032)}{0.994} F_{t-4} + \underset{(20.100)}{28.716} P_{H_{t-1}} - \underset{(336.756)}{714.202} P_{C(4)y-1} - \underset{(11.275)}{19.509} P_{B(4)y-1}$$

$$+15.282G_{y-1}+84.350P_{H(4)y-1}$$
$$(7.678) \quad (21.002)$$
$$R^2=0.965$$
$$d=1.840^{\Delta}$$

(2) $$H=10583.172+3.101F_{t-2}$$
$$(.170)$$
$$-9010.958D_{(3)}+58.125T$$
$$(477.373) \quad (15.134)$$
$$R^2=0.927$$
$$d=1.287^{*}$$

(3) $$Q_P=313.377+0.128H-2.584R$$
$$(.002) \quad (1.213)$$
$$R^2=0.994$$
$$d=1.278^{\circ}$$

(4) $$S=-750.755+0.218Q_{P_{t-1}}+0.890S_{t-1}$$
$$(.028) \quad (.058)$$
$$+6.448P_{P_{t-1}}-1.358W_{t-1}$$
$$(2.048) \quad (.378)$$
$$R^2=0.906$$
$$d=2.006^{\Delta}$$

(5) $$P_P=103.127-1.249\frac{Q_P}{N}-2.841\frac{S}{N}$$
$$(.155) \quad (.461)$$
$$-0.419\frac{Q_B}{N}-0.678\frac{Q_R}{N}-0.017\frac{I}{N}$$
$$(.194) \quad (.431) \quad (.017)$$
$$-6.613D_{(55)}-0.026W$$
$$(1.126) \quad (.013)$$
$$R^2=0.873$$
$$d=1.298^{*}$$

(6) $$P_H=-8.933+0.602P_P-0.044M-0.009W$$
$$(.029) \quad (.012) \quad (.003)$$
$$R^2=0.945$$
$$d=1.332^{*}$$

$^{\Delta}$No serial correlation in residuals.
*Inconclusive test for serial correlation in residuals.
°Positive serial correlation in residuals.

The Durbin-Watson statistic (d) provides a test for the presence of serial correlation in the residuals, i.e., the degree of correlation of each residual with the residual for the previous year. The table for the Durbin-Watson test extends only to five independent variables. For equations with more independent variables, a linear extrapolation was used. The degree of serial correlation is indicated by the symbols given above.

In general, the mnemonic device of denoting each variable by its first letter is followed, althou[gh] there are some exceptions to prevent duplication. All price variables are measured in constant prices by deflating them by the consumer price index. Unless otherwise specified, quarterly data are used. For lagged variables, $t-1$, $t-2$, and so on, indicate the number of quarters lagged, and $y-1$ indicates a yearly lag. For instance, $P_{C(4)y-1}$ refers to the price of corn in the fourth quarter of the previous year.

Variables are defined as follows:

F=sows farrowing (1,000 head)
P_H=deflated price received by farmers for hogs (dol. per cwt.)
P_C=deflated price received by farmers for corn (dol. per bu.)
P_B=deflated price received by farmers for beef cattle (dol. per cwt.)
G=annual production of barley, oats, and grain sorghum (mil. tons)
H=hogs slaughtered (1,000 head)
D=dummy variable (0 or 1 in different periods)
T=time (1, 2, 3, . . . in successive quarters)
Q_P=quantity of pork produced (mil. lbs.)
S=cold storage holdings of pork, beginnin[g] of quarter (mil. lbs.)
P_P=deflated retail price of pork (cents per lb.)
W=seasonal index, based on mean quarterly temperatures
Q_B=quantity of beef produced (mil. lbs.)
R=ratio of pigs saved fall $y-1$ to spring y
Q_R=quantity of broilers produced (mil. lbs.)
N=U.S. population (mil.)
I=deflated discretionary income (mil. dol.)
M=Index numbers of unit marketing charges (1947–49=100)

Discussion of the Variables

The basic supply equation of the simple cobweb model has been expanded into three equations to reflect some of the steps in the production process. There are no data on sows bred but there are estimates of the number of sows farrowing, and most bred sows are held for farrowing. Factors that influence the number of sows farrowing include the facilities available on farms for raising hogs, the expected price of hogs at marketing time, the price and supply of feed, and the relative profitability of alternative uses of farm resources used in hog production, mainly feed and labor. Since this analysis uses quarterly data, some provision for the seasonal pattern of farrowings is also needed.

One way of accounting for the equipment available for hog production as well as the seasonal pattern is to include the farrowings for the same quarter of the previous year as an explanatory variable. Various studies have indicated that farmers make their basic decisions concerning next year's hog production in the fall. Thus the prices of hogs, corn, and beef cattle—the latter being the main competitor of hogs for feed and labor—for the fourth quarter of the previous year are included. But conditions may change during the year so the prices of these three commodities lagged one quarter were also tried as explanatory variables. Since the gestation period for hogs from breeding to farrowing is about 4 months, a 3-month lag is not precise, but it is the closest approximation possible using quarterly data. The previous year's production of small grains (oats, barley, and sorghum) affects farrowings during the first two quarters, while the present year's production influences third and fourth quarter farrowings.

Intercorrelation resulted in wrong signs and insignificant coefficients for corn and beef prices lagged one quarter, hence these variables were dropped from the equation. The effects of intercorrelation (that is, correlation between independent variables) upon correlation and regression coefficients are discussed by Fox and Cooney (*6*). In general, high intercorrelation tends to mean lowered reliability for individual regression coefficients.

All regression coefficients in equation (1) have the correct sign and all are larger than their standard errors, although they are not all significantly different from zero at the 5 percent level. This is the standard significance level used in this study—it means that a coefficient has a 95 percent probability of being different from zero and five chances out of a hundred that it is not different from zero. The *d* statistic for the Durbin-Watson test for serial correlation in the residuals shows no serial correlation.

The number of hogs slaughtered depends upon the number of sows farrowing 6 to 8 months previously and the number of pigs raised per sow. The age of barrows and gilts at marketing varies from 5 to 10 months; therefore it is difficult to achieve a precise determination of the lag between farrowing and marketing. But most hogs are slaughtered at 6 to 8 months, so a two-quarter lag was used. As the number of pigs raised per sow has been increasing in recent years at a fairly steady rate, time was used as a variable to account for this trend. Graphic analysis of the data revealed that estimates of third quarter slaughter based on lagged farrowings were considerably higher than actual slaughter. Many gilts are retained for breeding purposes in the early fall months, and consequently slaughter is reduced. Also, the predominance of March farrowings in the first quarter and the tendency to feed spring pigs longer than fall pigs before marketing mean that many pigs farrowed in the first quarter are not marketed until the fourth. To account for this, a dummy variable was introduced having a value of 1 for the third quarter where there were fewer marketings, and 0 for other quarters when conditions were normal.

All three regression coefficients are significant and have the expected signs. The *d* statistic is within the inconclusive range.

The quantity of pork produced depends upon the number of hogs slaughtered, their average slaughter weight, and their dressing yield. The latter is virtually stable and therefore does not create changes in pork production. The average slaughter weight varies seasonally and from year to year.

The year-to-year variation in slaughter weights results from different quantities fed, and this may be caused by hog prices or grain supplies. Also, spring farrowed pigs are usually fed longer and to heavier weights than fall pigs. Thus in a year when the number of spring pigs slaughtered is proportionately larger, the average slaughter weight would be expected to increase, and vice versa. Pigs farrowed in the fall are marketed during the following spring and summer, whereas most spring pigs are marketed in the fall of the same year. A ratio of pigs saved in the fall of the preceding year to the number of pigs saved in the spring is therefore included as a variable. The negative sign for this variable means that as the proportion of fall pigs marketed during the year increases, the quantity of pork produced decreases when the number of hogs slaughtered remains constant.

No satisfactory method was found for accounting for the seasonal variation in slaughter weights. The absence of any measure for this variable in the equation is one likely reason for the positive serial correlation in the residuals.

Since it was found that cold storage holdings of pork at the beginning of a quarter have an effect on the retail price of pork during that quarter, a storage estimating equation was necessary. A certain quantity of pork is always in cold storage owing to curing processes and the normal delay between slaughter and consumer purchase. Cold storage holdings lagged one quarter are used to account for these factors. The quantity of pork produced during the previous quarter also influences the quantity in storage, as do price and the season of the year.

The seasonal variation may perhaps be accounted for by using a seasonal index of temperature variation. This is simply the average temperature for each quarter as a percentage of the "normal" annual temperature. Estimates of average U.S. temperatures are used in computing the index, admittedly a rough approximation, yet one that gives some indication of variation due to the changing seasons.

All of these factors are significant and the *d* statistic indicates no serial correlation in the residuals. The sign of the price coefficient is somewhat surprising as it indicates that storage holdings increase as price increases. We might expect that storage would be reduced when prices were high. However, there is probably some interaction between price and storage within a quarter, and with other variables held constant, it may be that a price increase accompanies an increase in stocks, because when stocks increase the quantity marketed decreases.

On the demand side, price analysis for agricultural commodities is frequently centered at the retail level. In the short run, prices of the more basic commodity are regarded as derived primarily from the price as determined for the retail product. As Thomsen (*15*, p. 52) states it: "The demand for products at the farm end of the marketing system consists of consumer demand (i.e., prices which consumers will pay for different quantities) minus a schedule of marketing charges . . . ."

Using this approach, the retail price is estimated first and the farm price is derived from it. The retail price of pork is affected by the supply of pork available for consumption, the supply of competing meats, and consumer income. These variables are put on a per capita basis to allow for population increase. The supply of pork consists of the quantity in storage at the beginning of a quarter plus production during the quarter. The major competing meats are beef and broilers. As storage of these meats is small, the available supply is approximately equal to production. Deflated discretionary income is used as a measure of consumer income because it is not so highly correlated with the other predetermined variables as is disposable income.[2] As was mentioned earlier, high intercorrelation among the predetermined variables makes interpretation of their coefficients difficult because it increases standard errors and sometimes changes the signs of the coefficients. Graphic analysis of the data shows a shift in demand for pork occurring in 1955, so a dummy variable having a value of 0 from 1949 through 1954 and a value of 1 from 1955 to 1959 was used to account for this shift. Stanton (*14*) has shown that there is a seasonal difference in the demand for pork, therefore the previously described seasonal index is included as a variable.

All coefficients are larger than their standard errors except that for discretionary income. The sign for the income coefficient indicates that as income rises, the price of pork falls. But since the coefficient is not significantly different from zero, the sign means very little, if anything. All other signs are as expected. The sign for the dummy variable indicates a downward shift in the demand for pork. The seasonal factor shows a decrease in price during the warmer spring and summer quarters, other things constant. The coefficient for broiler production is not significant.

---

[2] Discretionary income is derived from disposable income by subtracting the essential expenses associated with the relatively high standard of living achieved in the postwar years. The items deducted from disposable income include major fixed commitments such as rent, taxes, home mortgage and installment debt repayments, and essential expenditures including medical expenses, household utilities, local transportation, and the maintenance of a given level of food and clothing consumption. For a full explanation of the concept of discretionary income and the procedures for its calculation, see Franklin (7).

The rest of the coefficients (except that for income) are more than twice their standard errors and are significant at accepted probability levels. The d statistic is in the inconclusive range.

The farm price of hogs is derived from the retail price and a marketing margin. Index numbers of unit marketing charges are used as a measure of this margin. The margin varies seasonally, increasing during the summer. All three variables have the expected sign and are significantly different from zero at the 5 percent level. The test for serial correlation of the residuals is again inconclusive.

Evaluation of the Model

This system of six equations is not linear in the variables because in equation (5) pork production and storage have been converted to per capita bases after being used as total figures in other equations. Therefore, the matrix of the coefficients of the endogenous variables does not equal the Jacobian, and the latter will have to be computed in order to determine whether the general condition for recursive treatment is met. The Jacobian is

$$|J| = \frac{\partial(u_1, u_2, u_3, u_4, u_5, u_6)}{\partial(F, H, Q_P, S, P_P, P_H)} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.128 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-1.249}{N} & \frac{-2.841}{N} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0.602 & 1 \end{vmatrix} = 1,$$

so the general condition is fulfilled.

The Jacobian contains unknown values of the variable N, but since it is a triangular matrix, its determinantal value is the product of the elements along the main diagonal which equals unity. Thus in this case, there is no need to develop linear approximations of the nonlinear variables in order to evaluate the Jacobian.

Residuals are estimates of the unobservable disturbances u_t. The assumptions made earlier about the u_t do not necessarily apply to the residuals $\hat{u}_t$. If the equations have been properly formulated, however, the residuals should have the same properties as were assumed for the disturbances.

Testing for serial correlation in the residuals has become fairly common in time series analysis—it has been shown that serious errors and biases may result if residuals are autocorrelated. Nonindependence of residual terms from an equation fitted by least squares leads to loss of statistical efficiency, underestimation of the true variance by the usual formula, and invalid $t-$ and $F-$ tests; it does not lead to statistical bias or inconsistency (Foote, *5*). For equations utilizing lagged dependent variables as explanatory variables, Hurwicz (*12*) has demonstrated that least squares estimates of coefficients are biased in small samples. But provided the residuals are not autocorrelated, these estimates of the parameters are consistent. If the residuals for these equations are autocorrelated, least squares estimates of the parameters may be seriously biased even in large samples (Griliches, *8*).

A completely satisfactory test for serial correlation in residuals has not yet been devised. A simple correlation between $\hat{u}_t$ and $\hat{u}_{t-1}$ too frequently accepts the hypothesis that they are independent. In other words, the correlation coefficient will sometimes be nonsignificant when the sample has been drawn from a serially correlated population. Durbin and Watson (*2*) derived some approximate tests for autocorrelation in successive residuals of least squares regression models. The tests do not apply to models that use lagged values of the dependent variable as independent variables (*2*, p. 410). The d statistic computed from the residuals of such a regression often accepts the false hypothesis of nonautocorrelated errors. The test also has a considerable range of indeterminancy.

Durbin and Watson (*3*) present an approximate procedure for use when their bounds test is inconclusive. This method involves transforming (¼)d to a Beta distribution and should be sufficiently accurate with a large number of degrees of freedom, say greater than 40. As the transformation requires lengthy computations and gives only a rough approximation for small samples, it is rarely used.

Although two of the six equations in the model use lagged values of the dependent variable for independent variables (equations (1) and (4)), the Durbin-Watson test is used to give some indication of the degree of serial correlation in the residuals for these analyses. The test indicates that the condition of serial independence of the residuals is relatively well met for a quarterly system.

There is only one case of positive serial correlation; there are three cases where the test is inconclusive and two cases of no autocorrelation. The two cases of nonautocorrelation are suspect, however, because these equations include lagged values of the dependent variable.

A common method of testing for independence between two distributions is to compute the correlation coefficient. Absence of linear correlation, however, is not proof of independence (*4*, p. 222). But for want of a more precise test for the independence of the residuals from different equations, a correlation matrix for the $\hat{u}$'s was used to measure the degree of independence. The diagonality of the covariance matrix Σ can be determined from this matrix since

$$r_{12}=\frac{\text{covariance } x_1 x_2}{\sqrt{(\text{variance } x_1)\ (\text{variance } x_2)}}$$

If there is no correlation between the $\hat{u}$'s, the $r_{ij}=0$, and the covariances are also zero, making the Σ matrix diagonal. The correlation matrix for the residuals of the six equations is:

	$\hat{u}_1$	$\hat{u}_2$	$\hat{u}_3$	$\hat{u}_4$	$\hat{u}_5$	$\hat{u}_6$
$\hat{u}_1$	1. 000	0. 055	0. 027	0. 103	0. 161	0. 061
$\hat{u}_2$	-----	1. 000	. 124	. 083	. 013	. 130
$\hat{u}_3$	-----	-----	1. 000	. 141	. 037	. 194
$\hat{u}_4$	-----	-----	-----	1. 000	. 002	. 019
$\hat{u}_5$	-----	-----	-----	-----	1. 000	. 183
$\hat{u}_6$	-----	-----	-----	-----	-----	1. 000

The residuals from different equations may be assumed to be independent because none of the correlation coefficients is significantly different from zero at the 5 percent level. The assumption of a diagonal covariance matrix and the use of actual values of endogenous variables serving as predetermined variables in subsequent equations appear to be justified.

Inclusion of Additional Variables

In an attempt to reduce the serial correlation in the residuals, the model was altered slightly by including more explanatory variables in the equations. The method of determining the appropriate variables to include in the equations was to compute a correlation matrix of the residuals and the predetermined variables used in the system. An examination of this matrix showed that no predetermined variables are correlated significantly with $\hat{u}_1$. The residuals of the other five equations, however, are correlated with at least one of the predetermined variables used in other equations of the system. After selecting the correlated variable that made the most economic sense in each case, equations (2) through (6) were refitted using these additional variables.

Four predetermined variables are significantly correlated with $\hat{u}_2$. As would be expected, they are highly intercorrelated since they are each correlated with the same variable. Of the four, the price of hogs for the previous quarter was the most appropriate to include as an explanatory variable for the number of hogs slaughtered.

Since slaughter hogs can vary in age from 5 to 10 months, some response to price would be expected. A high price during the previous quarter would induce farmers to sell their hogs at that time. This would reduce the number of hogs sold in the present quarter. Conversely, there would be an increase in hogs slaughtered in the present quarter with low prices during the preceding quarter. Thus a negative regression coefficient would be expected.

The inclusion of this variable in equation (2) increases the multiple coefficient of determination to 0.94 and eliminates serial correlation in the residuals as measured by the Durbin-Watson statistic. The sign of the regression coefficient is negative as expected.

The residuals of equation (3) are significantly correlated with only one predetermined variable, the price of corn in the fourth quarter of the preceding year. It seems reasonable that the price of feed would influence the quantity of pork produced. A low corn price would lead to increased feeding and heavier hogs. When this variable is included in the equation, its coefficient is significantly different from zero with the expected negative sign. Serial correlation in the residuals is reduced so that the *d* statistic now falls in the inconclusive range.

The residuals of equation (4), which estimates the storage of pork at the beginning of a quarter, are correlated with four predetermined variables. The estimates for third-quarter beginning storage were consistently low, so the dummy variable for the third quarter was included as an additional variable. The other three correlated variables could not reasonably be included in the equation on economic grounds.

One reason for the higher than estimated storage for the beginning of the third quarter may be that packers and wholesalers increase storage holdings at this time in anticipation of smaller supplies and higher pork prices during the summer. In any event, the addition of this variable increases the R^2 to 0.94 and the d statistic still indicates that there is no serial correlation in the residuals.

Of the two variables correlated with $\hat{u}_5$, the price of hogs during the previous quarter seemed most appropriate. The introduction of this variable into the equation for the retail price of pork means that pork prices are not completely determined by consumer demand as was previously hypothesized. Retail prices are somewhat "sticky" and for short periods may reflect the prices retailers paid for the pork now on their counters. Thus if the price of hogs was high the previous quarter, the price of pork would tend to be higher than it would be otherwise.

The addition of this variable changes the sign of the income coefficient to positive, but the coefficient is still not significantly different from zero. It also increases the R^2 to 0.92 and reduces serial correlation in the residuals. The d statistic is now at the upper end of the inconclusive range. The coefficient of the variable itself is significant with a positive sign.

Five predetermined variables are correlated with $\hat{u}_6$. Of the five, per capita pork storage seems the most reasonable to affect the price of hogs. A large quantity of pork in storage at the beginning of a quarter would be expected to decrease the price of hogs. But when this variable is included in the equation, its coefficient has a positive sign and the marketing margin coefficient becomes insignificant.

There is a fairly strong negative correlation between marketing margins and per capita cold storage holdings of pork. Thus with other factors held constant, when storage increases, margins decrease, and when margins decrease, hog prices increase. By this indirect route, an increase in cold storage holdings of pork may raise the price of hogs.

The five equations that were refitted using additional variables are presented below. In each case the additional variable is significantly different from zero and serial correlation in the residuals is reduced.

(2a) $$H = 15{,}188.994 + \underset{(.154)}{3.127}F_{t-2} - \underset{(433.451)}{8905.125}D_{(3)} + \underset{(17.486)}{24.212}\,T - \underset{(77.757)}{243.264}P_{H_{t-1}}$$

$R^2 = 0.941$
$d = 1.792^{\Delta}$

(3a) $$Q_P = 474.592 + \underset{(.002)}{0.128}H - \underset{(1.254)}{3.657}R - \underset{(36.720)}{81.884}P_{C(4)_{y-1}}$$

$R^2 = 0.995$
$d = 1.526^{*}$

(4a) $$S = -865.271 + \underset{(.023)}{0.238}Q_{P_{t-1}} + \underset{(1.721)}{7.918}P_{P_{t-1}} - \underset{(0.312)}{1.456}W_{t-1} + \underset{(.050)}{0.825}S_{t-1} + \underset{(18.318)}{80.655}D_{(3)}$$

$R^2 = 0.937$
$d = 1.718^{\Delta}$

(5a) $$P_P = 73.489 - \underset{(.133)}{1.068}\frac{Q_P}{N} - \underset{(.437)}{1.875}\frac{S}{N} - \underset{(.160)}{0.332}\frac{Q_B}{N} - \underset{(.355)}{0.460}\frac{Q_R}{N} + \underset{(.015)}{0.003}\frac{I}{N} - \underset{(1.073)}{4.193}D_{55} - \underset{(.010)}{0.025}W + \underset{(.151)}{0.658}P_{H_{t-1}}$$

$R^2 = 0.918$
$d = 1.831^{*}$

(6a) $$P_H = -14.826 + \underset{(.024)}{0.621}P_P - \underset{(.012)}{0.008}M - \underset{(.003)}{0.015}W + \underset{(.110)}{0.512}\frac{S}{N}$$

$R^2 = 0.965$
$d = 1.498^{*}$

$^{\Delta}$No serial correlation in residuals.
*Inconclusive test for serial correlation in residuals.

Of the six equations (including equation (1), which was not refitted), the Durbin-Watson test shows no serial correlation for three equations and for the other three the d statistic is at the upper end of the inconclusive range. The fit for the system as a whole is quite good. The coefficient of multiple determination (R^2) is above 0.90 for each equation.

Evaluation of the Revised Model

In an attempt to resolve the indeterminancy of the Durbin-Watson test, another test for serial correlation was tried. The von Neumann-Hart (*16*, *10*) test is designed to test for autocorrelation in an observed sequence of random variables. When it is applied to residuals of fitted relations, it accepts the independence hypothesis too frequently. It does not take account of the added correlation of the estimated residuals resulting from the necessity of estimating the regression coefficients; this defect becomes worse as the number of independent variables increases. However, Hildreth and Lu (*11*) applied this test to the residuals of a number of least squares equations with good results.

The von Neumann-Hart test is based on the ratio of the mean square successive difference to the variance. The ratio is:

$$\frac{\delta^2}{s^2}=\frac{\dfrac{\sum_{t=2}^{N}(\hat{u}_t-\hat{u}_{t-1})^2}{N-1}}{\dfrac{\sum_{t=1}^{N}(\hat{u}_t-\bar{u})^2}{N}}$$

The relation to the Durbin-Watson statistic, d, is readily apparent where

$$d=\frac{\sum_{t=2}^{N}(\hat{u}_t-\hat{u}_{t-1})^2}{\sum_{t=1}^{N}\hat{u}_t^2}$$

The mean of least squares residuals is zero so that

$$\frac{\delta^2}{s^2}=\frac{N}{N-1}\,d.$$

The two test statistics for each equation are given in table 1.

TABLE 1.—*Tests for autocorrelation in residuals*

Equation No.	Durbin-Watson statistic	von Neumann-Hart statistic
(1)	1. 840△	1. 883△
(2a)	1. 792△	1. 834△
(3a)	1. 526*	1. 561△
(4a)	1. 718△	1. 758△
(5a)	1. 831*	1. 874△
(6a)	1. 498*	1. 533△

△The hypothesis of zero autocorrelation is not rejected at the 5-percent level.
*The test value falls within the indeterminate range.

The von Neumann-Hart test shows that in no equation is the independence hypothesis rejected at the 5 percent level. The value of $\frac{\delta^2}{s^2}$ for each equation is considerably above the critical value of 1.455 for a two-tailed test at this significance level. This test indicates that there is only a remote possibility of serial correlation in the residuals.

The inaccuracy of this test when applied to the residuals of a regression equation containing a large number of independent variables is shown by the results for equation (5a). Although this equation has a comparatively high d value, it contains eight independent variables so the Durbin-Watson statistic falls within the indeterminate range. The von Neumann-Hart statistic, on the other hand, gives the same significance level for equation (5a) as for equation (1). The latter shows no serial correlation by the Durbin-Watson test. The residuals of equation (5a) may be serially independent but not to the same level of significance as those for equation (1).

The three equations in which the Durbin-Watson test is inconclusive do not contain lagged dependent variables, and each has at least 35 degrees of freedom. The procedure presented by Durbin and Watson (*3*) for use when their test is inclusive is therefore appropriate. The approximate critical values of Fisher's z for 5 and 1 percent tests against positive serial correlation were computed for each of the three equations. These values and the observed value of z for each equation are presented in table 2. When the critical value is less than the observed value, there is significant serial correlation.

The observed value of z is greater than the 5 percent critical value for equations (3a) and (6a) so the residuals of these two equations are serially correlated at accepted significance levels. The residuals for equation (5a) show no significant correlation by this test.

The correlation matrix of the residuals of the refitted equations was computed to give an approximation to the covariance matrix. The matrix for the residuals of the six equations still showed no significant correlation at the 5 percent level, hence the covariance matrix is still diagonal by this approximate test.

TABLE 2.—*Test for autocorrelation in residuals when d is inconclusive*

Equation No.	Critical value of z		Observed value of z
	5 percent	1 percent	
3a	0. 2613	0. 3712	0. 2992
5a	. 4167	. 5954	. 3924
6a	. 2461	. 3495	. 3183

This recursive model adequately explains the variation in six major variables in the hog industry. The coefficient of multiple determination (R^2) is above 0.90 for each aquation. Of the 30 regression coefficients, only 7 are not significantly different from zero at the 5 percent level. The assumptions made about the disturbances are comparatively well met. An examination of the residuals shows only two cases of significant serial correlation as measured by the available tests. The residuals for each equation are uncorrelated with those from other equations.

Summary

A recursive system as described by Wold has many desirable properties. As it is fitted by least squares, the computation is relatively simple. And provided certain assumptions are met concerning the covariance matrix of the residuals, it produces maximum likelihood estimates of the parameters.

The hog production process has been used many times as an example of a cobweb phenomenon. In a cobweb model, price in one period affects production in the succeeding period which in turn affects price, and so on. The link by link progression of price-quantity-price-quantity marks the hog industry as suitable for analysis by a recursive system of equations. If there were considerable interaction between price and quantity within a given period, a simultaneous equation system would be preferable. To minimize this interaction, quarterly data were used in fitting the equations.

The price-quantity relation given in the cobweb is an extremely simplified model of hog production. There are many factors that affect variation in the quantity of pork produced and the price received for it. The number of sows farrowing, the pigs raised per sow, the number of hogs slaughtered and their average weight all affect the quantity of pork produced. Storage holdings of pork and pork production affect the price of pork, and this in turn affects the price of hogs, which affects the number of sows farrowing in following periods. All these variables within the hog industry are influenced by numerous factors outside the industry such as the prices and supplies of raw materials (feed grains) and competing goods (beef, poultry).

In order to measure the effects of these variables, a recursive system of six equations was fitted by least squares. The dependent variables in these equations were: (1) The number of sows farrowing; (2) number of hogs slaughtered; (3) quantity of pork produced; (4) cold storage holdings of pork; (5) retail price of pork; and (6) farm price of hogs.

The residuals of these six equations were tested to determine whether they fulfilled the assumptions made about the disturbances in formulating the model. In an attempt to reduce serial correlation in the residuals, five of the equations were refitted using additional independent variables.

The final set of equations furnish a creditable model for describing the hog economy, judging by the variation explained and the significance of coefficients. According to the available tests, the residuals meet most of the requirements necessary for maximum likelihood estimates. There is little evidence of serial correlation in the residuals and no evidence of correlation between residuals of different equations.

With the correct selection of commodities and time intervals, recursive systems fitted by least squares are appropriate. The results of this study indicate that an adequate analysis of the hog industry can be made by using a quarterly recursive model.

Literature Cited

(*1*) BARGER, H. AND KLEIN, L.
1954. A QUARTERLY MODEL FOR THE UNITED STATES ECONOMY. Jour. Amer. Statis. Assoc. 49: 413–437.

(*2*) DURBIN, J. AND WATSON, G. S.
1950. TESTING FOR SERIAL CORRELATION IN LEAST SQUARES REGRESSION. I. Biometrika. 37: 409–428.

(3) ——— and Watson, G. S.
1951. testing for serial correlation in least squares regression. II. Biometrika. 38 : 159–178.

(4) Feller, William
1957. an introduction to probability theory and its applications. Vol. I. Second Edition. New York, John Wiley & Sons, Inc.

(5) Foote, R. J.
1958. analytical tools for studying demand and price structures. U.S. Dept. Agr., Agr. Handb. 146.

(6) Fox A., and Cooney, James F.
1954. effects of intercorrelation upon multiple correlation and regression measures. U.S. Agr. Marketing Service. AMS–341.

(7) Franklin, W. B.
1958. discretionary income. National Industrial Conference Board, Technical Paper No. 6.

(8) Griliches, Zvi.
1961. a note on serial correlation bias in estimates of distributed lags. Econometrica. 29 : 65–73.

(9) Haavelmo Trygve.
1943. the statistical implications of a system of simultaneous equations. Econometrica. 11 : 1–12.

(10) Hart, B. I.
1942. tabulation of the probabilities for the ratio of the mean square successive difference to the variance. Annals of Math. Statis. 13 : 207–14.

(11) Hildreth, C. and Lu, J. Y.
1960. demand relations with autocorrelated disturbances. Mich. Agr. Expt. Sta. Tech. Bul. 276.

(12) Hurwicz, L.
1950. least squares bias in time series. *In* Statistical Inference in Dynamic Economic Models. T. C. Koopmanns, ed. New York, John Wiley & Sons, Inc., pp. 365–383.

(13) Nerlove, Marc and Addison, W.
1958. statistical estimation of long-run elasticities of supply and demand. Jour. Farm Econ. 40 : 861–880.

(14) Stanton, B. F.
1961. seasonal demand for beef, pork, and broilers. Agr. Econ. Res. 13 : 1–14.

(15) Thomsen, F. L. and Foote, R. J.
1952. agricultural prices. New York, McGraw-Hill Book Co., 509 pp.

(16) von Neumann, J.
1941. distribution of the ratio of the mean square successive difference to the variance. Annals of Math. Statis. 12 : 367–395.

(17) Wold, H.
1956. causal inference from observational data : a review of ends and means. Jour. Royal Statis. Soc. 119 : 28–61.

(18) ———.
1959. ends and means in econometric model building : basic considerations reviewed. Paper *in* Probability and Statistics, The Harold Cramér volume, U. Grenander, ed. Stockholm : Almquist and Wiksells, pp. 355–434.

(19) ——— and Jureen, L.
1953. demand analysis. New York, John Wiley & Sons, Inc., 358 p.