A SIMULTANEOUS EQUATION ECONOMETRIC MODEL OF WINTER MAINTENANCE COST CATEGORIES

William J. Dunlay, Jr., Department of Civil Engineering, University of California, Berkeley

Little is known about the quantitative influence of factors affecting winter maintenance expenditures. This report proposes a simultaneous equation stochastic model to explain or predict county expenditures for snow and ice control based on selected measures of a county's need for such specific operations as spreading chemicals and abrasives and plowing snow. Emphasis is on methodology, with specific findings cited only to illustrate the procedure. A procedure for applying the estimates produced by the model to evaluate county performance is recommended. It is suggested that counties can best be evaluated by a residual analysis in which they are grouped according to how their actual costs compared with those estimated by the model to determine whether counties in a particular group share any similar practices, policies, or deficiencies.

•IN RECENT years expenditures for snow and ice control by state highway departments have increased greatly, but research expenditures for studies of snow and ice control have remained relatively insignificant—0.25 to 0.50 percent of the total highway research expenditures in 1967 (9). Rising expenditures can be attributed to increased mileage of multilane highways, huge increases in vehicle miles of travel, greater public demand for bare pavement maintenance, and inflation. Reasons for the relative lag in winter maintenance research, however, are difficult to pin down.

Winter maintenance operations are initiated as the need for them, in the form of snow and ice on the roads, occurs. This need is often of an emergency nature and is difficult to predict, or even explain, with a degree of accuracy usable for planning purposes. The severity, duration, location, and time of hazardous snow and ice conditions are just a few of the unknowns that complicate winter maintenance operations and have probably discouraged research in this field. Yet, it is these same uncertainties that make further research necessary to develop better and more economical techniques for snow and ice control.

SCOPE AND PURPOSE

This paper is directed primarily to maintenance managers with some knowledge of statistical techniques. Two-stage least-squares estimation procedures are covered in detail, however, because they are not as widely known as standard regression methods.

The purpose of this paper is to propose a method by which expenditures for winter maintenance operations may be predicted and understood. Specific results are cited only to illustrate the proposed methodology, and findings are not critically examined because they are only preliminary.

The model described here is a system of simultaneous equations, each designed to explain the variation of specific categories of winter maintenance expenditures among the basic operating units of a state highway department. The basic operating unit is assumed in this paper to be a county office of the department. Typical categories of winter maintenance expenditures are given in Table 1.

Sponsored by Committee on Maintenance and Operations Costs.

Testing and development of the model has thus far been accomplished using cross-section data from 66 counties for one winter season. Each equation has been constructed to explain the expenditure level for a particular operation using such influencing factors as (a) level of expenditures for other related operations, (b) temperature and frozen precipitation of a county, (c) state highway mileage and characteristics, (d) amount of traffic that must be accommodated, and (e) extent of the winter maintenance force operating within a county. The equations are of the general form

TABLE 1

TYPICAL CATEGORIES OF WINTER MAINTENANCE EXPENDITURES

Index	Operation
y_1	Purchase and stocking of abrasives
y_2	Purchase and stocking of de-icing chemicals
y_3	Spreading of abrasives and chemicals
y_4	Plowing snow, slush, and ice
y_5	Snow fence installation

$$y_i = a + b_1 y_1 + \dots + b_{i-1} y_{i-1} + b_{i+1} y_{i+1} + \dots + b_n y_n + c_1 x_1 + \dots + c_m x_m$$

where the y's are expenditures for such operations as given in Table 1, the x's are measured or observed characteristics of a county determined to have a predictable effect on y_i, and the coefficients (a, b's, and c's) are statistically estimated parameters that define the direction and significance of the relationships between y_i and the explanatory variables on the right-hand side of the equation. Variables not included in a particular equation can be considered to have a coefficient of zero.

For each equation one set of coefficients applies to all counties. Thus, it is the county-to-county variation that the model explains, and the coefficients must therefore be estimated using cross-section data. Procedures for using both time-series and cross-section data are available but are not widely understood. These procedures involve the added time-series problems of accounting for inflation and improved levels of service and nonuniform (among the counties) changes in these two items with time.

Endogenous variables are in units of dollars per mile of maintained highway within a county. These units were chosen because it is felt that such unit expenditures more accurately reflect differences in climate, extent of operation, and highway characteristics than do total expenditures.

OBJECTIVES

When constructing the equations one must select explanatory variables considered to have a predictable effect on each endogenous variable. The objective of this process, in addition to merely explaining the variation in existing expenditures, is to develop equations that reflect what a county's winter maintenance expenditures ought to be, based on characterics that seem inherently important in defining its true need for snow and ice control. Explanation of existing expenditures, however, is an important step in the development stage.

Once the equations have been fully developed they can be used for several purposes:

1. To apportion funds appropriated for winter maintenance among the counties according to their estimated needs for particular types of operation;
2. To predict the consequences of changes in policy on the different expenditure categories (e.g., what would be the effect of eliminating snow fences on plowing and spreading costs?); the consequences of such policy changes can thus be evaluated before they are implemented; and
3. To identify counties whose expenditures deviate significantly from those predicted by the model.

THEORY AND TERMINOLOGY

In single-equation multiple-regression analysis, variables are classified as either independent (explanatory) or dependent (to be explained). However, this classification is inadequate when considering a simultaneous system of the following form (which is the general form of all equations in this paper):

$$y_i = f_i\,(y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_n;\ x_1,\ldots x_m) + e_i \qquad i = 1,\ldots,n \tag{1}$$

Note that a variable explained by one equation may itself be an explanatory variable in other equations. For example, expenditures for erecting snow fences are expected to both influence and be influenced by expenditures for snow removal. Variables of this type that are affected by and affect other explained variables are called endogenous variables.

Some variables have an effect on the expenditures to be explained but are not in turn affected by those expenditures. Such variables are not explained by the model; their values are determined by forces outside the model. Total annual snowfall, for example, affects certain winter maintenance expenditures but is obviously not affected by such expenditures. Variables having these characteristics are known as exogenous variables and are represented by the x's in Eq. 1. Descriptions of suggested exogenous and endogenous variables are contained in this paper.

To illustrate the theory and terminology of this section, the following two-equation model will be discussed:

$$y_1 = a_1y_2 + b_1x_1 + e_1$$

$$y_2 = a_2y_1 + b_2x_2 + e_2$$

Note that y_1 is explained by the first equation and used as an explanatory variable in the second. Thus, y_1 is an endogenous variable. The term dependent variable is used when referring to the expenditure variable explained by a particular equation.

To many readers the most striking difference between these equations and the more familiar multiple-regression equation is probably the inclusion of the y's as explanatory factors on the right-hand side of the equations. Although this complicates the process of obtaining unbiased estimators for the coefficients, it is very important to retain the y's as explanatory variables. Because the y's appear in more than one equation of the system, their interrelationships cannot be determined by examining any single equation, as in ordinary multiple-regression analysis; instead, a simultaneous analysis of all the equations of the system is necessary. These y's represent unit winter maintenance expenditures for such operations as plowing snow and erecting snow fences, which by their very nature must affect each other. For example, the erection of snow fences in appropriate locations will retard the formation of drifts, thus reducing the requirements for plowing. On the other hand, the areas subject to much drifting and therefore to the need for considerable snow removal will be the areas in which the expenditures for snow fences will be highest. The inclusion of the expenditure variables as explanatory factors reflects this type of mutual dependence.

The form and content (variables included) of each equation is called the structure of that equation, and the equations themselves are called structural equations. Construction of the model requires that hypotheses be made concerning the relationships between explanatory variables and various categories of winter maintenance costs. This is necessary in order to choose the variables that should most logically explain each cost category. These hypothetical relationships are represented by the content of the structural equations.

Estimating Procedures

The use of conventional least-squares procedures for estimating parameters of single multiple-regression equations of the form

$$y = a_1 + b_1x_1 + b_2x_2 + \ldots + b_n x_n + e$$

involves several assumptions concerning the disturbance or error terms, denoted by the symbol e in the preceding equation. It is assumed that this error term is a stochastic variable that represents the aggregate effects on the dependent variable of explanatory factors not included in an equation. The exogenous variables are assumed to be known without error. Variables not included are unknown, considered unimportant, or not quantifiable. It is further assumed in conventional least-squares estimation that the disturbance term is a random variable with an expected value of zero, a constant variance, and zero covariance with (stochastically independent of) each of the explanatory variables.

Consider the two-equation simultaneous system

$$y_1 = a_1y_2 + b_1x_1 + e_1 \quad (2)$$

$$y_2 = a_2y_1 + b_2x_2 + e_2 \quad (3)$$

If these equations are solved for y_1 and y_2, the solutions will be affected by both e_1 and e_2 in the manner shown in Eqs. 4 and 5. Thus, in Eq. 2, y_2 cannot be assumed independent of e_1; and similarly in Eq. 3, y_1 cannot be assumed independent of e_2. Under these conditions, it has been proved that the use of conventional least-squares techniques leads to biased estimates of the parameters of the system. However, a modification of these techniques known as two-stage least squares may be used to obtain unbiased estimates.

The first stage of the two-stage process is to solve the structural equations simultaneously to obtain an expression for each dependent variable with only exogenous variables as explanatory factors. Solving the preceding two-equation example simultaneously results in

$$\hat{y}_1 = \{b_1/[1 - (a_1a_2)]\}\, x_1 + \{a_1b_2/[1 - (a_1a_2)]\}\, x_2 + \{(e_1 + a_1e_2)/[1 - (a_1a_2)]\} \quad (4)$$

$$\hat{y}_2 = \{a_2b_1/[1 - (a_1a_2)]\}\, x_1 + \{b_2/[1 - (a_1a_2)]\}\, x_2 + \{(e_2 + a_2e_1)/[1 - (a_1a_2)]\} \quad (5)$$

This is called the reduced form of the model. Note that in the reduced form y_1 is a function of x_2 and x_1, whereas in structural Eqs. 2 and 3 it is a function of y_2 and x_1. This is the mechanism by which the effect of the endogenous variable (y_2) on y_1 is indirectly accounted for in the reduced form. That is, even though y_2 is not included in the reduced form equation for y_1, a linear function of the exogenous variable on which y_2 depends (x_2) is included.

The disturbance terms of the reduced form equations are linear functions of e_1 and e_2 and are therefore independent of the explanatory variables x_1 and x_2. Therefore, application of standard least-squares techniques results in unbiased estimators for the reduced form coefficient (the expressions in parentheses) and, thereby, unbiased estimates of the dependent variables $\hat{y}_1$ and $\hat{y}_2$.

In the second stage of the two-stage least-squares procedure, the estimates of the dependent variables are substituted for the actual values of the endogenous variables on the right-hand side of the structural equations as in the following:

$$y_1 = a_1\hat{y}_2 + b_1x_1 + v_1$$

$$y_2 = a_2\hat{y}_1 + b_2x_2 + v_2$$

A second regression is then performed on each of these equations to obtain unbiased estimators for the structural coefficients. Ordinary least-squares regression may now be used because the estimated endogenous variables are functions of only the exogenous variables and are therefore independent of the error terms.

Alternative Forms of the Model

Equation 1 illustrates the general form of the equations of this study. The f_i symbol of that equation represents the functional form of the ith structural equation. Three functional forms—linear, quadratic, and log-linear—have been considered.

In the linear form the equations are expressed as the weighted sum of a set of endogenous and exogenous variables raised to the first power:

$$y_i = a_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} b_{ij}y_j + \sum_{k=1}^{m} c_{ik}x_k + e_i \qquad i = 1, \ldots, n$$

The a_i's, b_i's, c_i's, and d_i's are the weights, or parameters, of the equations. The linear form is widely used because it is the simplest to specify and interpret. The sign of a coefficient (e.g., c_{ik}) represents completely the estimated direction of the relationship between x_k and y_i. The magnitude of the coefficient, along with the magnitude of the variable itself, is a measure of the sensitivity of y_i to changes in x_k, with other factors being held constant.

The quadratic form is similar to the linear form, except that the squares of the endogenous variables are included as additional explanatory factors. Quadratic equations have the form

$$y_i = a_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} b_{ij} y_j + \sum_{k=1}^{m} (c_{ik} x_k + d_{ik} x_k^2) + e_i \qquad i = 1, \ldots, n$$

In this form the change in the dependent variable (y_i) associated with a unit change in an exogenous variable (x_k), with other factors constant, is measured by the partial derivative of y_i with respect to x_k, which is $c_{ik} + 2d_{ik}x_k$. Thus the change in y_i resulting from a unit change in x_k is seen to be a function of x_k itself. Assuming that x_k is positive, the direction and strength of the relationship indicated by the sign and magnitude of this partial derivative depends on the magnitude of x_k and the signs and magnitudes of c_{ik} and c_{i2k}. To evaluate the average sign and strength of the predicted relationships, each partial derivative should be evaluated at the mean value of the associated exogenous variable.

The last functional form considered was the log-linear form. The equations in this form are expressed as the weighted sum of the logarithms of the explanatory variables,

$$\log y_i = \log a_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} b_{ij} \log y_j + \sum_{k=1}^{m} c_{ik} \log x_k + \log e_i \qquad i = 1, \ldots, n$$

Taking the antilog of both sides yields

$$y_i = a_i \prod_{\substack{j=1 \\ j \neq i}}^{n} y_j^{b_{ij}} \prod_{k=1}^{m} x_k^{c_{ik}} e_i$$

The coefficient of an explanatory variable in the log-linear form is equal to the ratio of the percentage of change in the dependent variable to the percentage of change in an explanatory variable. To verify this consider the two-variable example

$$\log y = \log a + b \log x$$

Taking the differential of both sides of this equation with respect to x, we get

$$dy/y = b\ dx/x$$

or

$$b = (dy/y)/(dx/x)$$

By interpreting the differentials as finite differences and by multiplying numerator and denominator by 100, we obtain

$$b = [\Delta y/y\ (100)]/[\Delta x/x\ (100)]$$

Thus, b is the percentage of change in y corresponding to a 1 percent change in x.

Although the log-linear form shows promise for application in this study, problems of interpreting the resulting residuals have delayed an accurate evaluation of this form; therefore, no further discussion of it is included in this paper.

VARIABLES OF THE MODEL

Endogenous Variables

The endogenous variables are winter maintenance expenditure categories such as those given in Table 1. The exact form of these variables depends on how maintenance costs are categorized for accounting purposes. If the expenditures are not broken down by individual functional operations, a model of this type is of little use because it is only these operations that can be related to conditions encountered in the field. As stated earlier, expenditures are in units of dollars per mile.

It is essential that the endogenous variables be mutually exclusive. For example, the cost of chemicals and abrasives must be excluded from spreading costs because there are separate endogenous variables for abrasive expenditures and chemical expenditures. Thus, only the labor and equipment costs associated with spreading are included in y_3. By the same requirement, endogenous variables for labor and equipment costs may not be included in the same model as the variables in Table 1. If equations for labor and equipment costs are desired, a separate model must be constructed.

Exogenous Variables

All characteristics of a county that could possibly affect some aspect of winter maintenance must be enumerated and their significance evaluated. These characteristics will at first be general in nature, such as highway characteristics, climate, traffic demand, and intensity of operations. It is then necessary to select or define some specific measure or index of each of the characteristics.

Examples of characteristics and associated specific measures that have been used in the background study of this paper are given in Table 2.

Data for calculating the highway, traffic, and operation indexes of Table 2 should be available from records kept by most highway departments (1, 2, 3, 4, 5). The climate variables can be determined from the climatic summaries of the U. S. Weather Bureau (11, 12, 13). Data for the suggested measure of highway ruggedness (x_2) can be compiled from 1:24,000 U. S. G. S. quadrangle maps or 1:250,000 Army Map Service maps.

Altitude may be an important determinant of winter maintenance requirements, but, because of the difficulty of obtaining a single index to represent the altitude of an entire county, this index has not yet been tested. The average elevation of a county's highways is suggested for initial consideration.

TABLE 2

EXOGENOUS VARIABLES

County Characteristic	Specific Measure or Index
1. Highway characteristics	
Multilane highways	x_1 = ratio of mileage of four or more lanes to total mileage
Ruggedness	x_2 = mean number of contour lines crossed per unit length of mapped road
2. Climate	
Coldness	x_3 = mean temperature, November through March
	x_4 = degree days below 32 F
Frozen precipitation	x_5 = total annual snowfall
	x_6 = annual number of days with ≥ 1 in. of snow on ground
3. Traffic demand	x_7 = population density, persons per square mile
	x_8 = motor-vehicle registration receipts in dollars per mile
4. Operation density	x_9 = number of trucks and graders per mile
	x_{10} = number of stockpiles per mile

TABLE 3

CORRELATION COEFFICIENT FOR THE ENDOGENOUS VARIABLES

y_1		y_2		y_3		y_4		y_5	
y_3	0.80	y_3	0.70	y_1	0.75	y_3	0.72	y_4	0.50
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
x_4	0.20	x_3	-0.22	x_1	0.25	x_8	0.21	x_9	0.21

CORRELATION ANALYSIS

To aid in the selection of specific explanatory variables to be included in each structural equation, the sample correlation coefficients between each endogenous variable and the other endogenous and exogenous variables of the model should be obtained and analyzed. The sample correlation coefficient is an estimate of the strength of the linear association between a pair of variables. Thus, the correlation coefficient may be close to zero when there is actually a strong but nonlinear relationship between two variables. The square of the correlation coefficient, known as the coefficient of determination, is a measure of the proportion of the variation in one variable that may be attributed to differences in another variable.

A significant correlation does not necessarily imply a causal relationship between two variables. Both may depend on some common third factor. Thus, a statistical relationship may exist between two variables even though to infer a causal relationship would be absurd.

To facilitate examination of the results of the analysis, the correlation coefficients between an endogenous variable and all other variables are tabulated, in descending order of significance, under that endogenous variable, as in Table 3.

Preliminary regression equations can also be constructed to determine the joint effect of various combinations of variables and to further analyze the significance of certain relationships. A step-wise least-squares procedure is suggested wherein those explanatory variables that do not have regression coefficients significantly different from zero (at some predetermined level of significance as measured by a t-value) are dropped from the equation one at a time, until only variables with significant coefficients remain.

To evaluate which of several alternative exogenous variables best relates a particular county characteristic to each endogenous variable, the correlation coefficients between a particular cost variable and the measures of that characteristic (e.g., climate) can be listed as in Table 4.

We see from Table 4 that x_3 serves as a better measure of "coldness" in explaining y_1 than does x_4. When constructing the equations, the measure of a characteristic that best relates to an endogenous variable as determined from Table 4 is the one used in the structural equation for that variable.

HYPOTHETICAL RELATIONSHIPS

To establish the structure of an equation, it is necessary to postulate the nature of the relationships between the dependent variable of that equation and the explanatory variables included. This involves making assumptions concerning the direction of the relationships and specifying how, on the basis of judgment and experience, each chosen characteristic is thought to affect the dependent variable in question, independent of any statistical correlations. The specific variables to represent the included characteristics are then selected on the basis of correlation and preliminary regression studies.

As an example of this process consider the equation for y_1, expenditures for abrasives, whose structure is given in Table 5.

TABLE 4

CORRELATION COEFFICIENTS OF CLIMATE INDEXES

Endogenous Variables	Exogenous Variables			
	x_3	x_4	x_5	x_6
y_1	-0.30	0.20	0.25	0.40
y_2	-0.25	0.30	0.45	0.40
y_3	-0.10	0.20	0.40	0.35
y_4	-0.40	0.30	0.60	0.50
y_5	-0.30	0.40	0.20	0.10

TABLE 5
VARIABLES IN THE EQUATION FOR y_1

Endogenous	Exogenous
y_2 = purchase of chemicals	x_1 = ratio of mileage of four or more lanes to total mileage
y_3 = spreading expenditures	x_2 = mean number of contour lines crossed per unit length of mapped highway
	x_3 = mean temperature, November through March
	x_6 = annual number of days with ≥1 in. of snow on the ground
	x_{10} = number of stockpiles per mile

Expenditures for purchasing chemicals (y_2) should affect abrasive expenditures because the same number of stockpiles is usually supplied and the same network of roads is treated from each stockpile. It is expected that this relationship will be negative because the greater the amount of chemicals used, the less should be the need for abrasives. The cost of purchasing abrasives is directly affected by spreading expenditures (y_3) because abrasive purchases are for the replenishment of stocks and spreading is the process by which stocks are depleted.

The exogenous measure of multilaning (x_1) is included because the larger the value of this variable is, the greater is the area of pavement per mile to be treated with abrasives. The index for highway ruggedness (x_2) reflects the fact that grades require more intense, immediate, and frequent treatment with anti-skid materials than level sections.

Climate is measured by mean temperature (x_3) and by the number of days with more than 1 in. of snow on the ground (x_6). Both meausres of climate are included because conditions that require abrasives are a function of both the amount of frozen precipitation experienced and the length of time that the snow or ice might be expected to linger on the pavement, which, in turn, is a function of coldness. Frozen precipitation and coldness are best correlated with y_1 by x_3 and x_6, respectively, as indicated in Table 4.

The number of stockpiles per mile (x_{10}) is the last exogenous factor. It seems reasonable to assume that counties with a greater number of stockpiles per mile will buy and stock more abrasives per mile. The structure of all other equations is established in a similar manner (14).

Because of the manner in which they are selected, the explanatory variables in an equation represent factors that are presumed to determine a particular winter maintenance expenditure. It may happen that a hypothesis is not supported by subsequent statistical analysis. An estimated coefficient may not be statistically significant, or it may have an algebraic sign opposite to that expected. This would indicate either a weakness in the equation structure or a true discrepancy between what has been assumed to be true and what actually occurs.

A weakness in structure often takes the form of an excessive interdependence among the explanatory variables, which causes one of two or more interrelated variables to have an unreasonable sign or coefficient. This condition, known as multicollinearity, can be tested by computing the correlation coefficients or by omitting the variable with a reasonable sign and coefficient in a subsequent analysis to see if the undesirable condition remains. When two or more explanatory variables are highly correlated it turns out that some linear function of these variables does the explaining that any one of them could have done alone. This function may be such that the estimated coefficients of the individual variables are considerably distorted.

EQUATION SOLUTIONS

After the equations are constructed they are estimated using the two-stage least-squares technique available in many computer library programs for statistical analysis (7). The resulting solutions must then be analyzed to determine if the hypothetical relationships have been confirmed.

The first step in the analysis of solutions is to determine which functional form, linear or quadratic, yields the best estimates of the actual expenditures. The following is an example of this comparison:

Dependent Variable	R^2	Linear Form, Sum of Squared Errors	R^2	Quadratic Form, Sum of Squared Errors
y_1	0.72	151,000	0.76	130,000

In the linear form 72 percent of the variation in y_1 is accounted for by differences in the explanatory variables whereas the quadratic form accounts for 76 percent, which implies that the sum of the squared differences between estimated and actual expenditures is less in the quadratic form. Because the primary objective of this study is not merely to explain existing expenditures, choosing the form that gives the best estimates is valid only if the structure of the equation is reasonable. All equations in the model must be compared to ensure that the best form for the model as a whole is selected.

In Table 6 the solution of the equation for y_1 is given in tabular form. The significance of the coefficients, as indicated by the t-ratios, should be considered first because if a variable's coefficient is not significantly different from zero that variable may just as well be omitted. Since, at 53 degrees of freedom, a t-value of 2.67 indicates significance at the 1 percent level, the hypothesis that a coefficient is equal to zero is rejected for all variables in Table 6 at the 1 percent level of significance.

Next consider the regression coefficients of the endogenous explanatory variables. The minus sign of the coefficient associated with y_2, expenditures for purchasing chemicals, indicates that the model has predicted a negative relationship between y_1 and y_2, which agrees with the hypothesis of the last section that the greater the expenditures for chemicals, the less the need for abrasives is. The positive coefficient of y_3, spreading costs, supports the hypothesis of a direct relationship between abrasive costs and spreading.

As described earlier, to obtain an estimate of the average direction of the relationship with each exogenous variable, the sign of the partial derivative with respect to the variable evaluated at its mean must be considered. These partials are shown in the last column of Table 6. In this example it turns out that three of the indicated directions are as hypothesized, and two (x_1, multilaning, and x_6, snowfall) are not. Further study of x_1 and x_6 is required to ascertain the reasons for the discrepancy in the signs of their partials.

ERROR ANALYSIS

A significant difference between the actual expenditure by a county for a particular operation and the corresponding estimate produced by the model may itself be revealing. If a county's expenditure is very much lower than that predicted by the model, one or more of the following causes may apply:

1. The county may perform the operation in a highly economical manner;
2. Greater emphasis may be placed by the county than by the state as a whole on related operations in treating a particular condition;
3. The extent of operation is not sufficient to provide an adequate level of service; and
4. Other explanatory variables not previously identified are required, in addition to or in place of existing variables, to make the structure more realistically reflect a county's true need for the operation.

TABLE 6

SOLUTION OF QUADRATIC EQUATION FOR y_1, UNIT EXPENDITURE FOR ABRASIVES

Explanatory Variable	Regression Coefficient	Student t-Ratio	Partial Derivative Evaluated at Mean of Variable
y_2	-0.08	106.7	
y_3	0.55	780.2	
x_1	262.18	55.2	-199.6
x_1^2	-6,124.77	148.6	
x_2	1.55	60.0	2.3
x_2^2	0.03	45.7	
x_3	-113.01	190.6	-11.5
x_3^2	1.62	171.3	
x_6	0.09	3.5	-1.1
x_6^2	-0.01	50.6	
x_{10}	4,697.00	245.2	2,911.7
x_{10}^2	-43,649.48	55.0	

Note: Degrees of freedom = 53
Explained variation = 76 percent.

Corresponding but opposite indications exist for large positive estimation errors.

The fourth possible cause must be resolved first because the other three are not valid unless the model reasonably represents the winter maintenance needs of a county. The third possibility must be evaluated outside the model because the model does not include any measure of the level of service provided. The model does, however, point to the possibility of a deficient level of service when actual expenditures are less than predicted.

To facilitate the evaluation of possibilities 2 and 4, bar graphs of the type shown in Figure 1 can be constructed. Percentage of error is used rather than the actual magnitude of the error to place all counties on a comparable scale. From Figure 1 it appears, because of the large errors of opposite sign, that county Y emphasizes some operations and deemphasizes others relative to the practices in the state as a whole. For example, expenditures for purchasing abrasives (y_1) are considerably greater than the model predicts, whereas chemical purchases (y_2) are less. County Z, on the other hand, has very small errors of opposite sign. In county X, the actual expenditures were all significantly greater than estimated by the model. The actual reasons for these differences must be ascertained outside the model.

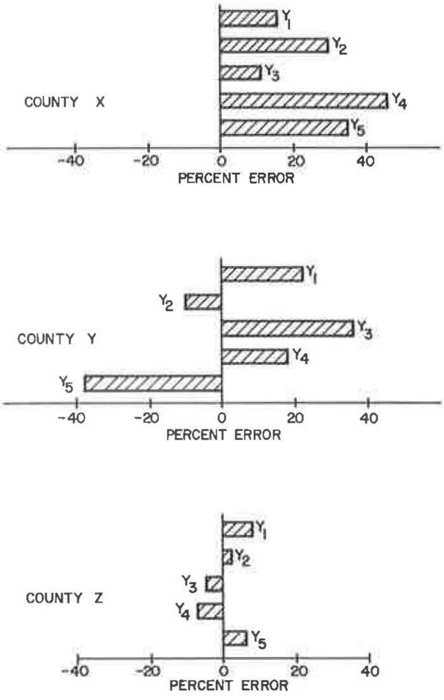

Figure 1. Percent errors in estimates of y's for three counties; percentage of error [(y_i = yi)/yi] 100.

Comparisons such as that of Figure 2 are recommended as an additional aid in identifying explanatory characteristics previously overlooked, or in evaluating the relative economy of each county's operation. Counties with higher rank (those in the top portion of Figure 2) may share some practice or policy that leads to relatively lower costs for the operation (purchasing abrasives in this example). Such practices can best be discovered by first grouping the counties according to how their cost compares with the model estimate and then examining counties in similiar groups. A ranking of this type also facilitates the identification of new characteristics whose values depend on the rank of a county, i.e., on the magnitude and sign of its estimation error. This can be accomplished by performing a contingency table or regression analysis to determine which new characteristics are most strongly related to the rank or residuals of the counties. Characteristics so identified are then converted to new or modified exogenous variables and included in subsequent solutions of the model. This type of feedback is essential to the development of the best structure for each equation.

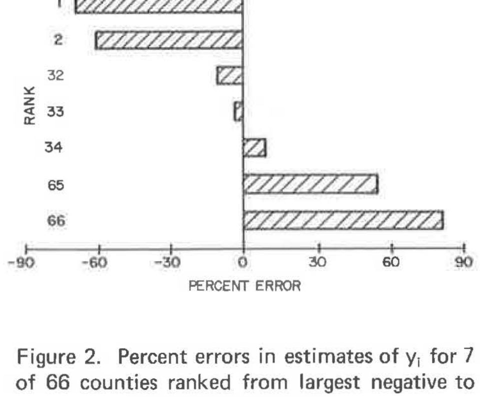

Figure 2. Percent errors in estimates of y_i for 7 of 66 counties ranked from largest negative to largest positive error.

CAUTIONS IN USING THE MODEL

It must be emphasized that the choice of variables and the types of functions to be fitted follow from one's familiarity with the maintenance operations involved. Intimate knowledge of the data is especially critical in a simultaneous equation model because a "wrong" variable in one equation may distort the coefficients of other equations. A "right" variable subject to large measurement errors may have similar consequences. Because explanatory variables are assumed in the analysis to be known without error, the precision with which each variable is measured may have important implications in applying the model.

The simultaneous equation method is appropriate in cases where the values of two or more variables are jointly dependent. The number of structural equations in a model must equal the number of endogenous variables, but if an individual equation contains only one endogenous variable, the dependent variable, that equation should be estimated separately as a single equation model using ordinary least-squares.

SUMMARY

This paper has described a model of simultaneous equations designed to predict county expenditures in Pennsylvania for certain basic winter maintenance operations. The flow chart shown in Figure 3 summarizes the procedures associated with this model.

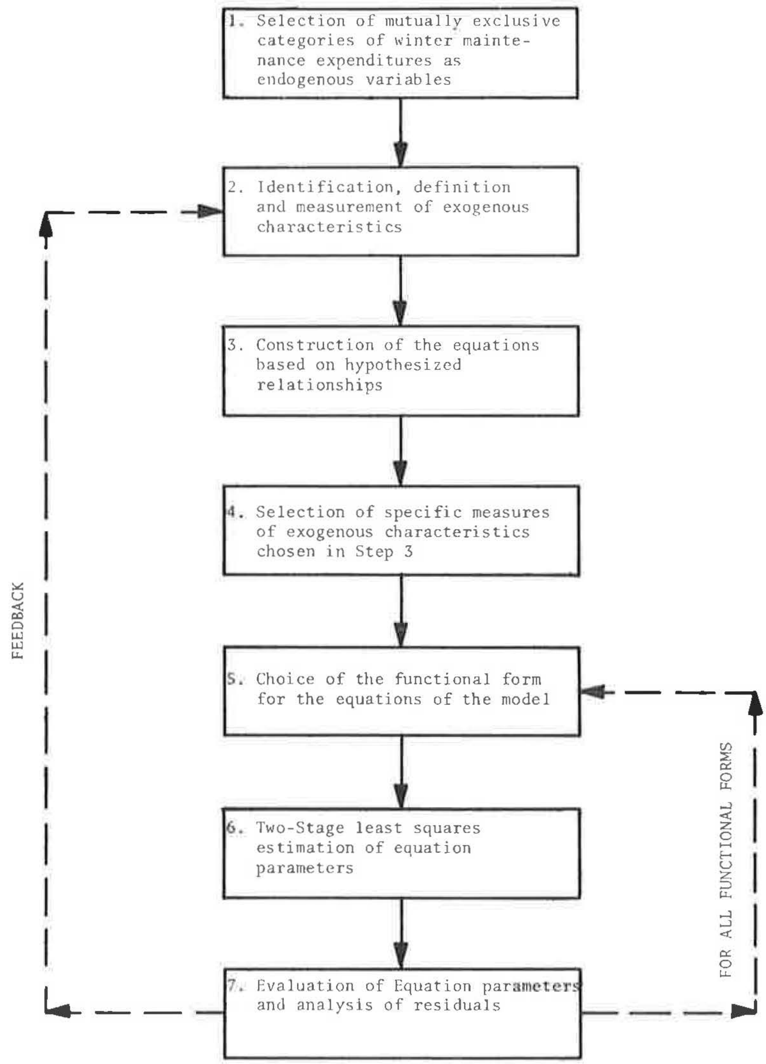

Figure 3. Summary of steps in model construction and evaluation.

The variables in each equation of the model were selected based on an examination of current practices and available data for describing characteristics of counties that affect their winter maintenance problems. A statistical analysis of the variables is used to measure the effect of certain county characteristics on expenditures for snow removal and ice control. Climate, traffic demand, highway width and grade, and operation intensity are the characteristics found to be most important in explaining expenditures.

Hypotheses concerning the nature of the relationships between expenditure categories and selected measures of the preceding characteristics serve as the basis for constructing the equations of the model. These hypothetical relationships are evaluated by examining the estimated parameters of the equations.

Finally, a procedure is recommended for applying the estimates produced by the model to evaluate county performance. It is suggested that counties can best be evaluated by grouping them according to how their actual costs compared with those estimated by the model and seeing if counties in a particular group share any practices or policies.

It is emphasized that only methodology is covered in this paper; details concerning specific variables and findings may be found elsewhere (14).

ACKNOWLEDGMENTS

This model was developed as part of a study of the cost-effectiveness of winter maintenance practices in Pennsylvania and was sponsored by the Pennsylvania Department of Highways and the Federal Highway Administration. The findings and conclusions are those of the author and not necessarily those of the sponsoring agencies. The author gratefully acknowledges the guidance of Owen H. Sauerlender in the development of the model. Appreciation is due other members of the staff of the Transportation and Traffic Safety Center for their invaluable assistance.

REFERENCES

1. Breakdown by Counties of Registration Receipts. Pennsylvania Department of Revenue, Harrisburg, 1968.
2. Correlation of Account Codes—Motor License and Highway Beautification Funds. Pennsylvania Department of Highways, Harrisburg, 1969.
3. Expenditure-Cost Report Number 511 by Organization Symbol, Allotment, and Object. Pennsylvania Department of Highways, Harrisburg, 1969.
4. Expenditure-Cost Report Number 561 by Organization Symbol, Allotment, and Function. Pennsylvania Department of Highways, Harrisburg, 1969.
5. Mileage by County of the State Highway System by Location in Townships, Boroughs, or Cities. Pennsylvania Department of Highways, Harrisburg, 1968.
6. Snow Removal and Ice Control. Highway Maintenance Manual. Pennsylvania Department of Highways, Harrisburg.
7. Halberg, M. C. Statistical Analysis of Systems of Simultaneous Linear Equations Using the Digital Computer. Pennsylvania State Univ., Agricultural Experiment Station, 1967.
8. Parman, W. J. A Pilot Study of Maintenance Costs of Idaho Highways. Univ. of Idaho, 1965.
9. Snow and Ice Control. Research Problem Statement No. 54, Highway Research Circular 62, March 1967, p. 48.
10. Twark, R. D. A Predictive Model of Economic Development at Non-Urban Interchange Sites on Pennsylvania Interstate Highways. Pennsylvania State Univ., 1967.
11. U. S. Weather Bureau. Climatic Summary of the United States—Supplement for 1931 through 1952—Pennsylvania. U. S. Govt. Printing Office.
12. U. S. Weather Bureau. Climatic Summary of the United States—Supplement for 1951 through 1960—Pennsylvania. U. S. Govt. Printing Office, 1964.
13. U. S. Weather Bureau. Climatological Data—Pennsylvania. U. S. Govt. Printing Office, Nov. 1968 through April 1969.
14. Dunlay, W. J. A Model of Simultaneous Stochastic Relations to Predict Winter Maintenance Costs in Pennsylvania. Pennsylvania State Univ., Pennsylvania Transportation and Traffic Safety Center, Report TTSC 6916, 1969.