PERFECTING METHODS FOR PREDICTING THE COURSE OF RURAL AREA DEVELOPMENT

PART II

Forecasting Income for Selected Rural Areas in Kentucky

By

Thomas H. Klindt, Garnett L. Bradford and Bruce R. Beattie

RESEARCH REPORT 13 : August 1972

University of Kentucky :: College of Agriculture
Agricultural Experiment Station :: Department of Agricultural Economics
Lexington

PREFACE

This is the second of two reports relating results of research conducted under Kentucky Hatch Project 89, "Development of Procedures for Quantifying and Assessing the Economic Well-Being of Rural Areas in Kentucky." The research reported in this and its companion publication is based in large part on the Ph.D. research of Thomas H. Klindt. This thesis research represented the initial contribution to Hatch 89.

The principal thrust of Hatch 89 is to perfect methods for predicting the course of rural area economic development. Specifically, five objectives are involved—(1) definition of economic development, (2) delineation of criteria and procedures for evaluating model efficacy, (3) construction of alternative models for predicting important components of economic development, (4) comparative tests of alternative models in accordance with criteria established in objective two, and (5) use of "best" models based on results obtained in carrying out objective four to predict the course of economic development for selected rural areas of Kentucky.

This publication presents empirical results pertaining to objectives three, four, and five. Specifically, single-equation income forecasting models are constructed, tested, and used to predict total and per capital personal income in selected rural areas of Kentucky. An overview of the total research effort—a definition of economic development and criteria for evaluating model efficacy—is reported in a companion piece to this report: *Perfecting Methods for Predicting the Course of Rural Area Development: Part I—Toward a Definition of Economic Development and A Framework for Evaluating Model Efficacy* (Research Report 11, University of Kentucky Department of Agricultural Economics, 1972).

The authors gratefully acknowledge the assistance and cooperation of Dr. Harold K. Charlesworth and staff of the Office of Development Services and Business Research at the University of Kentucky, Mr. William G. Herzel and staff at the Program and Research Staff, Kentucky Department of Revenue, and the Kentucky Department of Economic Security for providing much of the data for this study. Also, the criticisms and comments of our colleagues in the Department of Agricultural Economics are gratefully appreciated.

CONTENTS

	Page
PREFACE	3
LIST OF TABLES	6
LIST OF FIGURES	6
OBJECTIVES	7
MODELS: GENERAL DISCUSSION	8
SIMPLE FORECAST MODELS	11
POLICY-ORIENTED FORECAST MODELS	13
COMPARATIVE MODEL TESTING	16
1971 FORECASTS	21
SUMMARY AND CONCLUSIONS	22
APPENDIX A	25
APPENDIX B	26

LIST OF TABLES

Table | Page

1. Types of Income Forecast Models Studied 8
2. Selected Regression Results, Lagged Income Model 13
3. Selected Regression Results, Economic Sector Model 15
4. Predictive Accuracy Statistics, Per Capita Personal Income, in Dollars 19
5. Predictive Accuracy Statistics, Total Personal Income, in Dollars 20
6. Predicted Total and Per Capita Personal Income for 1971, in Dollars 21
7. Predicted Total and Per Capita Income for Selected Kentucky Area Development Districts, 1971, in Dollars 21

LIST OF APPENDIX B TABLES

1. Observed and Predicted Levels of Area Income, Income Extrapolation Model, 1968, in Dollars 26
2. Observed and Predicted Levels of Area Income, Lagged Income Model, 1968, in Dollars 27
3. Observed and Predicted Levels of Area Income, Economic Sector, Resource-Based Model, 1968, in Dollars 28

LIST OF FIGURES

Figure

1. Development Districts and Areas 10
2. Components of Model Efficacy 17

PERFECTING METHODS FOR PREDICTING THE COURSE OF RURAL AREA DEVELOPMENT

Part II

FORECASTING INCOME FOR SELECTED RURAL AREAS IN KENTUCKY

by

Thomas H. Klindt, Garnett L. Bradford and Bruce R. Beattie*

Rural development research is certainly nothing new, yet there has been a new surge of interest in certain aspects of this research, especially since 1967. This interest was exhibited in the findings of the President's National Advisory Commission on Rural Poverty, where it was pointed out that rural sectors of the United States have a disproportionate share of the nation's poverty [6]. Further, the President's Task Force on Rural Development indicated that economic problems in rural areas do not remain confined to rural areas. Disadvantaged persons from rural areas "in both hope and desperation...have turned to the cities to find an answer–many to find jobs, but a disproportionate number to fine welfare and slum housing." [7, p. 4].

Consistent with the above-mentioned stance, public policy makers have exhibited a renewed interest in rural development; however, to act judiciously they must have information to evaluate alternative courses of action. Lyle [5, pp. 7-8] contended that minimum needs for evaluation include:

> The future condition expected to prevail in the absence of the planned effects of new programs or projects proposed. A curve drawn to describe this condition over time may be called the "normal" growth path. The normal growth path... refers to expected future growth in the absence of newly planned or unanticipated programs, projects, technological developments, and so on.

This "no-change" information, together with expected effects from a planned program or project, yields the information required to determine the actual impact of a proposed program or project.

Objectives

An overview of some research along these (above-mentioned) lines recently conducted at Kentucky [4] is presented in this report. The primary emphasis was upon constructing simple, yet effective models to forecast total and per capita personal income for selected rural areas of Kentucky under the

*Assistant Professor of Agricultural Economics, Louisiana State University; Associate Professor of Agricultural Economics and Assistant Professor of Agricultural Economics, University of Kentucky, respectively.

assumption of Lyle's "normal growth path."[1] Resulting forecasts (predictions) may be used by policy makers to evaluate the impact of proposed programs or projects to determine which, if any, should be implemented. Further, such forecasts may give clues concerning which rural areas will be in greatest need of public assistance if no action is taken, thereby allowing a degree of forewarning to policy makers.

Specifically, objectives of the study were:

(1) to define economic development,
(2) to delineate operational criteria and procedures for evaluating model efficacy,
(3) to construct alternative models for predicting important components (levels of variables) of economic development in selected rural areas,
(4) to comparatively test the alternative models in accordance with the operational criteria, and
(5) to use "best" models, as determined by the testing procedure, to forecast the course of economic development for selected rural areas of Kentucky.

An earlier companion report [1] dealt with the first two objectives. This report focuses upon the latter three objectives.

Models: General Discussion

Four types of models were constructed, fitted and comparatively tested (Table 1). The models were, by design, single-equation relationships. Part of the reason for emphasizing simplicity was the commonly expressed assertion, especially among rural development decision makers, that economists have achieved a level of model building sophistication that far surpasses the capability of action groups to apply such models in the solution of their problems. More important, however, was the practical expediency of first working with relatively simple models, then hopefully moving on to more complex ones such as simultaneous equation systems, simulation, etc. Economic and statistical theory underlying single-equation systems is much better developed; and such systems readily lend themselves to the "proven" statistical

[1] Total and per capita area incomes were predicted, realizing that they are not the sole determinants of welfare or development but do represent important components thereof; see [1].

Table 1.--Types of Income Forecast Models Studied

A. Simple Forecast Models

1. Time trend extrapolation (1 variation)

2. Lagged income models (3 variations)[a]

B. Policy-oriented Forecast Models

1. Economic sector models--resource based, independent variables in physical terms (4 variations)[b]

2. Economic sector models--investment based, independent variables in dollar terms (4 variations)[b]

[a] Variations involved alternative model specifications.

[b] Variations involved alternative methods of forecasting parameter values.

technique of ordinary least squares regression. In general, this approach was designed to enhance the use of definite criteria for comparatively testing model efficacy. Testing of the models primarily involved comparative accuracy in predicting future levels of total personal income and per capita personal income in five development districts located in Kentucky. These districts were subdivided into 19 two- or three-county areas which served as the basic observational units (Fig. 1).[2] Areas were delineated with the aim of achieving a reasonable degree of intra-area homogeneity and greater inter-area heterogeneity than if the five multi-county districts had been used as the basic units of observation. Data were obtained from the Kentucky State Department of Revenue, the Office of Development Services (University of Kentucky), and other secondary sources [2]. Since personal income data were available for only six years—1959, 1962, 1965, 1966, 1967 and 1968—division of the development districts into smaller areas enabled the fitting of regression models using a cross-sectional approach.

Two of the four models (Table 1) are referred to as simple forecast models in that they were constructed strictly for forecast purposes. The other two models are referred to as policy-oriented forecast models in that they were constructed with explanatory variables amenable to policy (program) changes. Each model had certain common structural characteristics. Specifically, a functional relationship was hypothesized between the dependent variable (total or per capita personal income) and a set of independent variables. The set of independent variables differs with each model.

The first simple forecast model was a simple linear extrapolation relationship based on time; that is, future income was hypothesized to be linearly related to an index of time. This model was developed to serve as a benchmark. Certainly other models should *predict* at least as accurately as a single linear extrapolation if they are to be seriously considered as an information source for decision making.[3]

The second simple forecast model was a single-equation lagged-income model with income hypothesized to be a function of income in some base year plus a measure of income change in the years immediately preceding the base year.[4] Hypothesizing income to be a function of past income, and a measure of the change in past income seems reasonable, i.e., given the assumption of no change in policy coupled with the fact that resources and resource-use usually tend to change gradually in rural areas. To predict with this model, a prediction equation was formulated in which the dependent variable represented income during the year to which predictions were to be made. Parameter projections of the prediction equation were based on parameter estimates obtained by fitting similar models for previous time periods. Since this model contained only lagged independent variables, all observations needed for the prediction equation were, of course, known.

Policy-oriented models as the label implies were constructed to provide more policy information than simple forecast models while retaining emphasis on predictive

[2] For a more detailed enumeration of districts and areas, see Appendix A. Nineteen areas are shown in Fig. 1; however, observations for Area 19 (Christian and Todd Counties) were very suspect and so were not employed in fitting the models.

[3] Throughout this report the words "predict" and "forecast" are used interchangeably. We are aware that some economists use the word predict to mean what will happen *if* certain other events happen—a type of conditional forecast. In this study the predictions are largely unconditional and so can be called forecasts.

[4] The base year is defined as the last year for which observable data were used for making a given prediction.

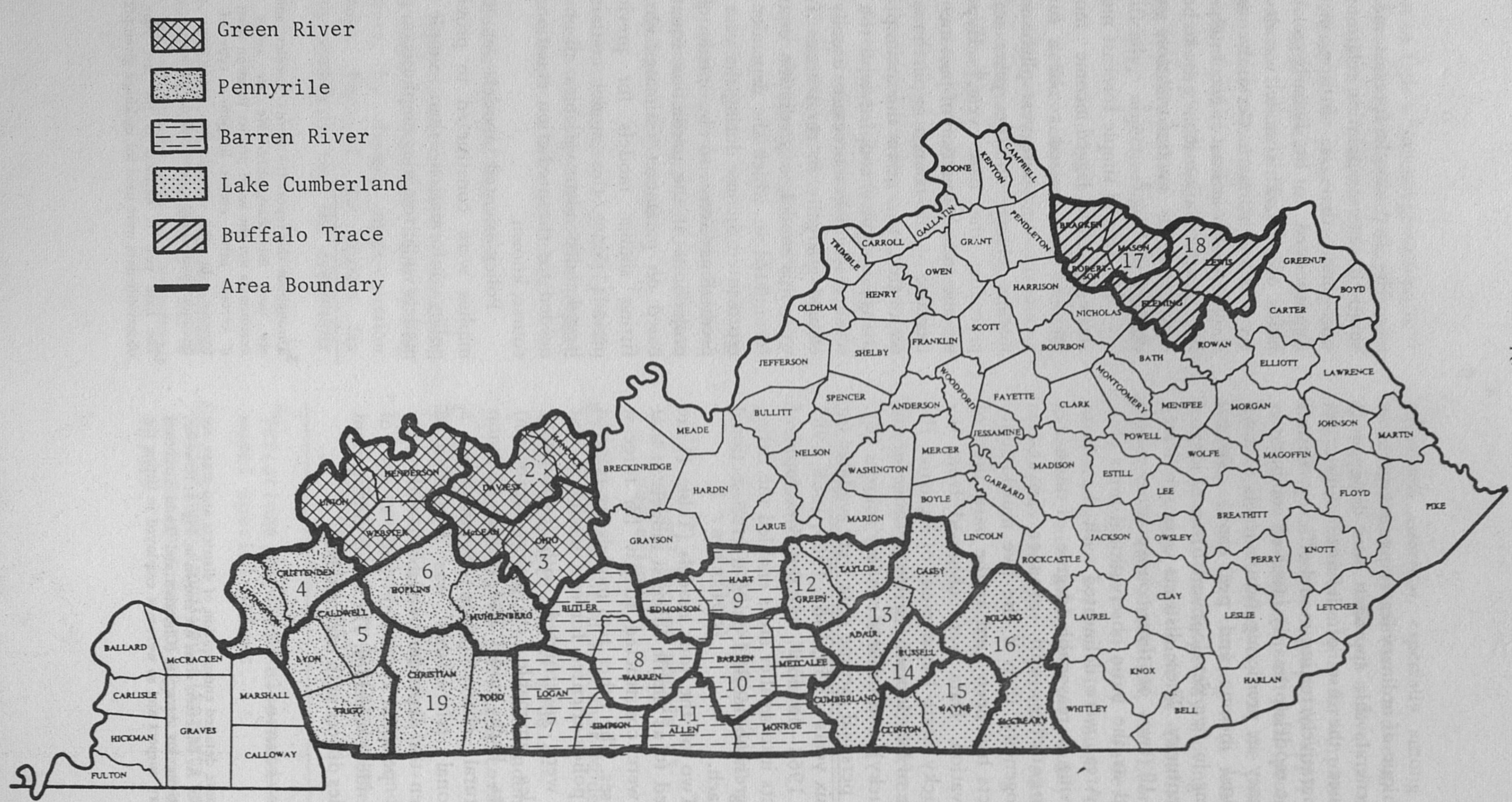

Figure 1. Development Districts and areas.

usefulness. In this report, only the first of these two types of models—the economic sector models—is discussed. In these models additional policy information is obtained by hypothesizing area income to be a function of several independent variables representing economic activities in various sectors of the economy. The model yielded information concerning the relative income effects of various broad economic sectors. These sectors were represented by physical activity variables such as employment and acres harvested.

Simple Forecast Models

Linear extrapolation is probably the most common technique used to forecast economic variables. The extrapolation model which was fitted may be specified as

$$Y_t = \alpha_o + \alpha_1 T + U_1 \qquad (1)$$

where

Y_t = total (or per capita) personal income of an area in year t;

T = a year-identification integer (index) corresponding directly to the subscript for Y, e.g., 1959 = 1, 1962 = 4, 1965 = 7 and 1968 = 10;

α_o and α_1 = model parameters; and

U_1 = the error term with the conventional assumptions being made.[5]

This model was fitted for each of the 18 areas using income data for three of the six years available (1959, 1962, and 1965) to permit a test of predictive accuracy, assuming a three-year prediction interval.[6] Thus, in effect it was assumed that 1965 was the last year for which income was known.[7] Parameter estimates ($\hat{\alpha}_o$ and $\hat{\alpha}_1$) for each area were then used to forecast income in the most recent year for which income observations actually were available—in this case, 1968. Consequently the test-forecast equation may be written as

$$\tilde{Y}_{68} = \tilde{\alpha}_o + \tilde{\alpha}_1 T \qquad (2)$$

where

$\tilde{\alpha}_o = \hat{\alpha}_o$, $\tilde{\alpha}_1 = \hat{\alpha}_1$, and T = 10.

The superscript notation ($\tilde{}$) is employed to identify forecasted values of the parameters, α_o and α_1. In subsequent discussion, an explanation is advanced why no error term is included in the test-forecast equation. If these income forecasts were found to be reasonably accurate [see the subsequent section on comparison of models], then the model could be refitted using all available income data i.e., refit the model specified in equation (1) using six years—1959, 1962, 1965 through 1968, and the revised parameter estimates employed to forecast income in some *actual* future year, say 1973.

This simple procedure was not employed in order to predict the past, even though 1968 obviously is now history. Rather, the model (equation 1) was fitted assuming that 1965

[5]See Johnston [3, pp. 3-11].

[6]Work involving longer prediction intervals would have been desirable, but data restrictions limited predictions to three years. Two-year predictions were made but results are not presented here. Note, however, that all models and test procedures are constructed such that an additional year's data will allow an additional year's prediction and test.

[7]Obviously, it would have been desirable to have more complete time series data. However, it is the authors' contention that in formulative research of this type it is better to have research results which have been constrained by lack of data than a research project which is still awaiting "complete" data.

was the present, then forecasts were made from 1965 into what was assumed to be the future [1968 in equation (2)]. This amounts to setting up an *ex post facto experiment* to test the predictive capabilities of the model. A rigorous test of a model's predictive capability is more complex than one might suspect at first glance. It is one thing to construct, using all information available, a model that adequately explains past events; it is quite another to demonstrate that a model based on experience and information of the distant past is capable of *predicting* less distant past events. The latter approach was chosen for this research. Surely, if a model has performed adequately "in the past," then the same type of model can be employed (or recommended) with more confidence than if past performance is not impressive, or worse yet, unknown. This is the essence of the methodology which undergirded this research, and hopefully its full meaning will become more evident as other models are discussed and comparatively evaluated.

Lag models are also frequently used for forecast purposes. The lagged income model which ultimately was adopted[8] may be specified as

$$Y_t = \beta_o + \beta_1 Y_{t-i} + \beta_2[Y_{t-i} - Y_{t-j}] + U_2 \quad (3)$$

where

Y_t = total (or per capita) personal income of an area in year t;

Y_{t-i} = income of the area in the year t-i, i denoting a discrete lag interval;

Y_{t-j} = income of the area in the year t-j, j denoting a greater discrete lag interval than i;

β_o, β_1 and β_2 = model parameters; and

U_2 = the error term.

Procedures followed in fitting this model, deriving the test-forecast equation, and forecasting, were nearly identical to those described for the linear extrapolation model. The only substantive procedural change was dictated by the fact that income data were available only for six years (1959, 1962, 1965 through 1968). Consequently, when 1965 was assumed to be the most recent (or known) year the model was specified as

$$Y_{65} = \beta_o + \beta_1 Y_{62} + \beta_2 [Y_{62} - Y_{59}] + U_3 \quad (4)$$

Fitting this model for each area obviously would more than exhaust the degrees of freedom. Hence, even though we were cognizant of the hazards of a cross-sectional approach, a common set of parameters for the 18 areas was estimated. The reader should recall that 18 sets of parameters were estimated with the extrapolation model—one set for each multi-county area. Selected regression statistics corresponding to equation (4) are shown in Table 2.

The simple forecast models were constructed to serve the sole purpose of predicting (forecasting) levels of area total and per capita income. Presumably, models of this type can be used for forecasting levels of any number of other development variables, e.g., total agricultural income *or* employment in agriculture and other sectors of the areas' economies. Statistical theory calls attention to certain hazards of employing such models. For example, the problems of biased estimates in lagged models are discussed by Johnston [3, pp. 211-213]. This could be especially serious when the forecasts are long-term; indeed, there was a tendency (as will be illustrated subsequently) for the lagged model to "undershoot" actual 1968 income when forecasting from 1965, a mere

[8] Two other slightly different algebraic forms were also specified and tested. All three gave essentially the same results, therefore this, the simplest, form was used.

Table 2.--Selected Regression Results, Lagged Income Model[a]

Statistic	Total Personal Income[b]	Per Capita Personal Income[b]
R^2	0.994	0.957
β_o	-2989	31.09
β_1	1.28 (39.57)	1.24 (18.09)
β_2	0.12 (0.34)	-0.41 (-1.78)

[a]For t = 1965, the last year for which income observations were assumed to be available.

[b]Calculated t values are shown in parentheses below β values (15 degrees of freedom).

three-year prediction interval. Yet, on balance, these simple models should prove useful for development action groups in that:

> Meaningful planning requires information regarding the future condition expected to prevail in the absence of the planned effects of new programs... [5, p. 7].

Certainly the simple forecast models are consistent with this "no-change" situation.

Policy-Oriented Forecast Models

Two policy-oriented models were fitted: (1) a model having four independent variables which corresponded to certain rather broad economic sectors (economic sector model—resource based), and (2) a model having five independent variables which also corresponded to certain economic sectors, but with the variables specified in dollar units (economic sector model—investment based). These models were constructed to provide more policy related information than the simple forecast models, while retaining emphasis on predictive usefulness. Since the two types were similar in statistical structure and the investment-based model yielded inferior results as gauged by the testing techniques used in the subsequent section on comparative model testing, the discussion in this section is limited to variations of the resource-based model.[9]

The resource-based economic sector model which ultimately was selected may be specified as

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + U_4 \qquad (5)$$

where

Y = total (or per capita) personal income of an area;

X_1 = burley tobacco acres harvested;

X_2 = cropland acres (exclusive of burley tobacco) harvested;

X_3 = average monthly employment in nonmanufacturing industries;

X_4 = average monthly employment in manufacturing industries;

β_o, β_1, β_2, β_3 and β_4 = model parameters; and

U_4 = the error term.[10]

[9] County tax data were utilized in developing the investment based economic sector model. These results are presented in Klindt [4].

[10] Three other independent variables were considered in earlier formulations of the model, viz., (1) value of intangible personal property (such as annuities and bank shares), (2) transfer payments from public (governmental) institutions and (3) value of breeding livestock. They were statistically nonsignificant at the 0.05 level.

This model was fitted cross-sectionally for each of five years—1959, 1962, 1965, 1966 and 1967. Some of the regression estimates—R^2, $\hat{\beta}$ and t values—are shown in Table 3. Note, that practically all of the variation in total personal income could be attributed to variation in the four independent variables and that a sizable but lower proportion of the variation in per capita personal income could be similarly attributed to the independent variables. This was true for all years. Presumably, the difference (in R^2 values) was because per capita income is a ratio of two variables (total income and population) and, in this case, was subject to much more variation. Lower t values for the respective regression coefficients of the per capita results may be similarly explained.

Derivation of the test-forecast equation involved forecasting new levels (values) for the independent variables and for the parameters. In most other respects, the methodology was identical to that employed for the simple forecast models. Assuming 1965 was the last year for which data (observations for Y and X_1 through X_4) were available and that a three-year income forecast was to be made (into 1968), the test-forecast equation could be specified as

$$\tilde{Y}_{68} = \tilde{\beta}_o + \tilde{\beta}_1\tilde{X}_1 + \tilde{\beta}_2\tilde{X}_2 + \tilde{\beta}_3\tilde{X}_3 + \tilde{\beta}_4\tilde{X}_4 \quad (6)$$

where the notation (coefficients and variables) has similar economic meaning as for the stochastic form of the model specified in equation (5), except the superscript notation (~) is employed to denote forecasted values. (Note that the $\tilde{\beta}_i$ and $\tilde{X}_i$ are 1968 forecasted values and these values, in turn, yield a prediction of income in 1968, $\tilde{Y}_{68}$). Since the forecast of each coefficient and variable is single-valued, $\tilde{Y}$ is, in a sense, deterministic and thus the error term was deleted.[11]

Forecasts for X_3 and X_4 ($\tilde{X}_3$ and $\tilde{X}_4$) were based on a fit of the following exponential model for each of the 18 areas:

$$X = \Theta \, \psi^t e \quad (7)$$

where

$X = X_3$ or X_4 in years 1954 through 1965;

t = an index of the year, i.e., 1954 = 1, 1955 = 2...,1965 = 12;

Θ and ψ = model parameters; and

e = the error term.

Estimates for Θ and ψ were then "plugged into" an equation of the same form as (7), and forecasts of X_3 and X_4 for 1968 ($\tilde{X}_3$ and $\tilde{X}_4$) were obtained for each area by setting t = 15.

Trend analyses also were made for the other two independent variables, X_1 and X_2. However, no significant trends could be established for X_2 (cropland acres harvested). It was decided that even though values for X_1 (burley tobacco acreage) have been tending downward they are so subject to decisions made vis'-a-vis' the price-support program, it would be presumptous to forecast this variable. In short, it was assumed that values for X_1 and X_2 would remain at the last "known" level which was the 1965-level in this case.[12]

---

[11] All this raises questions about whether Y is a statistic and, if so, then what would be its likely frequency distribution? We take the position that it must be a statistic, but are unprepared to argue for a particular frequency distribution.

[12] One should note at this point that our procedure for forecasting levels of independent variables is not sacrosanct. However, we believe in the absence of additional information that trend extrapolation may be as good as one can do, particularly when trying to predict consequences of "no-change" or "status quo" development strategies. Certainly one would utilize different procedures if an overt attempt was to be made to alter an independent variable level or if on the basis of prior information an existing trend was not expected to continue.

Table 3.--Selected Regression Results, Economic Sector Model

Statistic	1959	1962	1965	1967
		Total Personal Income[a]		
R^2	0.993	0.996	0.995	0.995
$\hat{\beta}_0$	1,957	5,791	4,169	5,913
$\hat{\beta}_1$	1.007 (2.36)	1.102 (3.42)	1.921 (3.51)	1.883 (2.55)
$\hat{\beta}_2$	0.062 (2.72)	0.090 (3.84)	0.148 (5.48)	0.100 (3.40)
$\hat{\beta}_3$	9.006 (19.46)	10.373 (25.83)	11.958 (20.82)	13.416 (20.42)
$\hat{\beta}_4$	6.656 (10.15)	5.691 (10.48)	4.884 (6.73)	5.297 (7.12)
		Per Capita Personal Income[a]		
R^2	0.840	0.711	0.730	0.727
$\hat{\beta}_0$	642	870	1,022	1,277
$\hat{\beta}_1$	0.03604 (1.82)	0.04130 (1.89)	0.05692 (1.72)	0.04992 (1.19)
$\hat{\beta}_2$	0.00132 (1.25)	0.00202 (1.27)	0.00313 (1.91)	0.00348 (2.08)
$\hat{\beta}_3$	0.06333 (2.95)	0.05486 (1.69)	0.02623 (0.75)	0.01871 (0.50)
$\hat{\beta}_4$	0.05850 (1.93)	0.04544 (1.24)	0.06537 (1.49)	0.07069 (1.66)

[a]Total personal income is estimated in thousands of dollars, per capita income in actual dollars. Calculated t values are shown in parentheses directly below each β value (13 degrees of freedom).

Forecasts of the coefficients ($\tilde{\beta}_o$ through $\tilde{\beta}_4$) were derived using each of the following four decision-rules (model variations):[13]

Variation A--select value from most recent year (1965);

Variation B—analyze the array of estimates (e.g., $\hat{\beta}_i$'s) for a trend and if a significant trend was detected, extrapolate to the test-forecast year (1968). If no significant trend was detected, the value for the most recent year (1965) was used.

Variation C—same as Variation B, except if no trend was detected, then a simple average of the array of $\hat{\beta}$'s was used.

Variation D—use a dummy variable model to pool observations from all known years (1959, 1962 and 1965 in this case) and, thus, simultaneously generate a single set of parameter estimates for β_1, β_2, β_3 and β_4. An array of β_o values was generated, and if a singificant trend was detected, then the trend was extrapolated to the test-forecast year. That is, the model specified in (5) was modified by adding two zero-one dummy variables to detect intercept differences (viz., 1962 versus 1965 and 1959 versus 1965).

Since no *a priori* rationale seemed appropriate for knowing which of the variations would most accurately predict income, all four variations were employed. Coefficients which were used to test-forecast per capita income from 1965 to 1968 are shown as follows:

	$\tilde{\beta}_o$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$
Variation A	1,022	.0569	.0031	.0262	.0654
Variation B	1,223	.0657	.0040	.0262	.0654
Variation C	1,223	.0657	.0040	.0451	.0564
Variation D	1,310	.0438	.0022	.0459	.0556

Similarly, four sets of parameter forecasts were derived for total personal income.

Altogether, 54 test-forecast equations were formulated—(a) one time extrapolation equation for each of 18 areas, (b) one lagged income equation, and (c) four each of the resource-based and investment-based economic sector models, where each type of model [(a), (b) and (c)] was developed for both per capita and total personal income for the three-year prediction interval. With the use of these test-forecast equations, personal income (total and per capita) was forecasted into 1968 for each of the 18 observational units. The predicted Y's were then compared with actual 1968 values to form the basis for an empirical test of the predictive accuracy of each type of model and/or procedure.[14]

Comparative Model Testing

Comparison of models for efficacy always involves test criteria. In many studies the criteria are implicit or unknown. A primary objective of this study was to delineate explicitly such criteria, and this (as noted in the section on objectives) was a major concern of an earlier companion report [1]. Suffice it to reiterate that a framework for assessing model efficacy was structured and its components were discussed in some detail. They are reproduced in Fig. 2. The emphasis is placed upon one subset of this framework, viz., *predictive accuracy*, which in turn is depicted as being a subset of the "fruitfulness" component.

Assume that the only purpose of the models is to predict area income, given a "no-change" development strategy. Further assume that all the models are logically valid.

[13] Details on each of these procedures are given in Klindt [4, pp. 67-70].

[14] Predicted and actual Y's for each of the three test-forecast equations constructed in this report are shown in Appendix B Tables 1-3.

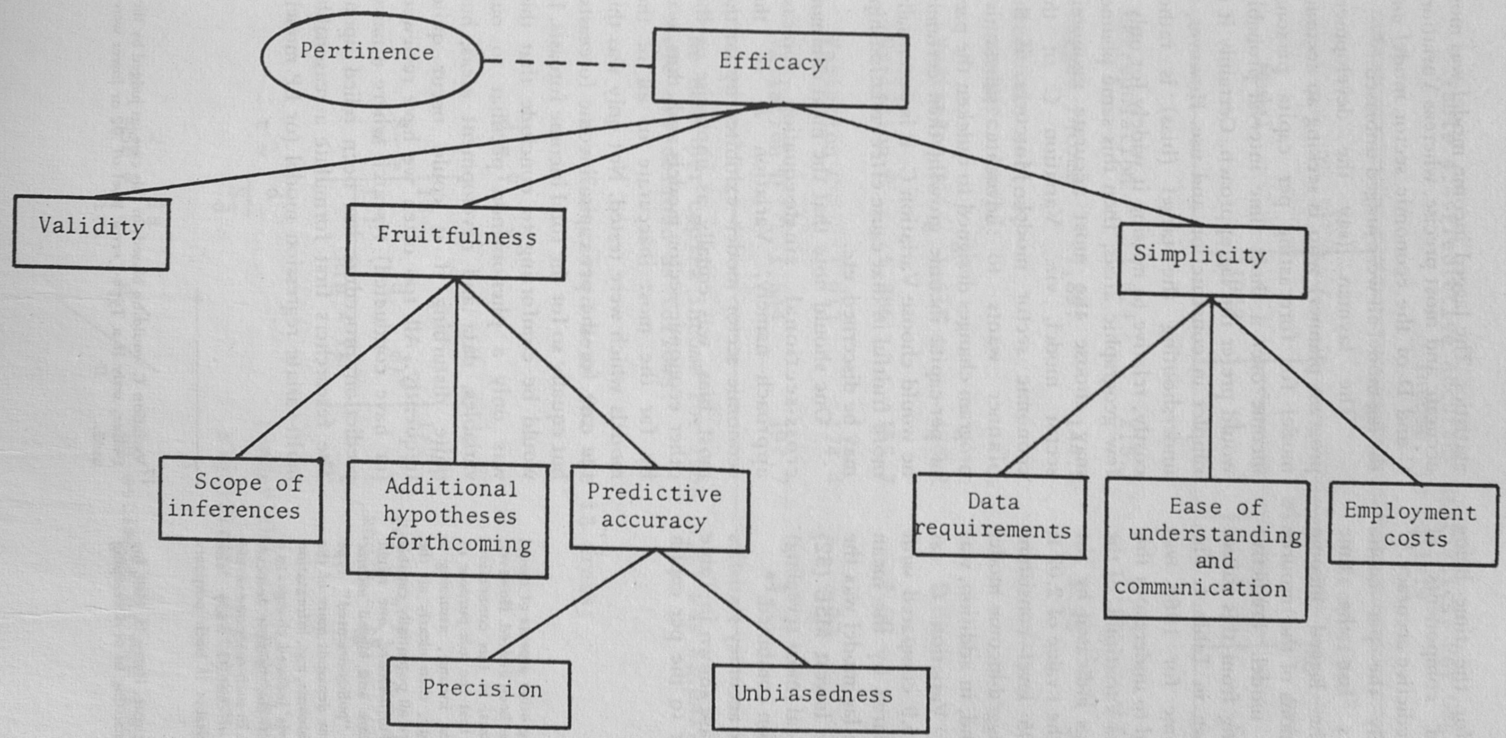

Figure 2. Components of model efficacy.

Also, as noted above, for the time being ignore simplicity and components of fruitfulness other than predictive accuracy.[15]

First, consider only the per capita income test-forecasts for the time extrapolation model, the lagged income model and the four variations of the resource based economic sector model. Predictive accuracy statistics resulting from this subset of test-forecasts are shown in Table 4. The average per capita income for 1968 was $2,041. All models tended to undershoot this mean. The least biased was Variation C of the economic sector model, as indicated by the lowest $\bar{d}$ (124.4). In fact, the t value of 2.02 is nonsignificant at the .05 level—consistent with unbiasedness. The lagged-income model was only slightly biased and, in addition, was far more precise than Variation C (as indicated by the S_d of 25.9 compared with 61.2). Moreover, as measured by the mean square error criterion, the lag model was the most accurate—it had the lowest $\hat{MSE}$ (32) which was used as the critical (most revealing) measure of bias and precision combined.[16]

Comparable predictive accuracy statistics for total personal income are shown in Table 5. Results are very similar to the per capita statistics. The lagged income model was most accurate and most precise, whereas Variations C and D of the economic sector model *and* the lag model all were judged unbiased.[17]

The layman (say the development program planner) who is seeking an accurate model for forecasting per capita personal income over a short time interval probably would prefer the lag approach. Certainly it is simpler in construction and use. However, if undershooting the target (bias) is rather costly, relative to missing it widely for only a few geographic areas, then this same planner may choose the most accurate economic sector model, viz., Variation C of the economic sector model. Moreover, if the planner wants to advocate substantive program changes designed to quicken the pace of per-capita income growth, then certainly he would choose Variation C—it is potentially more fruitful in that cause-effect relationships may be discerned, etc.

One should note that the most common cross-sectional single-equation forecast approach—namely, Variation A of the economic sector model—exhibited by far the most bias, was equally as imprecise as the other economic sector models and, thus, was by far the most inaccurate of all of the models which were tested. Not only was this the case for the per capita income forecasts, but equally so for the total income forcasts. It would be comforting to conclude that this was only a phenomenon peculiar to our variables, data and development areas, but quite disturbing if it could occur quite frequently. All too often we have reviewed (or have conducted) research where a naive prediction procedure has been relied upon. The researchers first formulate a reasonable multi-variate regression model (or LP model,

---

[15] The effort to date in quantifying other aspects of model efficacy (as outlined in Fig. 2) has been limited. However, a few rather obvious observations, in this connection, might be noted. If one assumes that the sole purpose of the models is to forecast area income, assuming a "no-change" development strategy, then clearly all the models are pertinent, i.e., they are all potentially capable of yielding such information. Furthermore, one might argue that the time extrapolation and lagged income models are simpler than the "policy-oriented" type models. On the other hand, if one demands more of the models than a mere prediction of income, e.g., information concerning income changes vis'-a-vis' induced changes in a particular economic sector, then the "simple forecast models" clearly are not pertinent. In such cases one would reject these models out-of-hand; only the "policy-oriented" models are suitable if both purposes must be served.

[16] We wish to acknowledge our colleague, Harry H. Hall, for pointing out the usefulness and assisting us in developing the mean square error criterion.

[17] Variation C would be biased, in this case, as judged by the t-value, only if a Type-I error level of .02 or lower were used.

Table 4.--Predictive Accuracy Statistics, Per Capita Personal Income, in Dollars

Model	Statistics[a]				
	$\bar{\bar{Y}}_{1968}$	Bias		Precision	Accuracy
		$\bar{d}$	t	$S_{\bar{d}}$	$\hat{MSE}$
					(000)
Extrapolation	1,824	218.1	9.22	23.7	57
Lagged income	1,897	144.8	5.58	25.9	32
Economic sector, resource based:					
Variation A	1,615	426.2	7.00	60.9	245
Variation B	1,886	155.7	2.57	60.8	87
Variation C	1,917	124.4	2.02	61.2	79
Variation D	1,844	197.7	3.21	61.6	104

[a]The mean per capita income for 1968 ($\bar{Y}_{68}$) was $2,041. Statistics shown here are defined as follows:

$\bar{\bar{Y}}$ = forecasted mean for the 18 areas;

$\bar{d} = \bar{Y} - \bar{\bar{Y}}$;

$$S_{\bar{d}} = \left[\frac{\sum_{i=1}^{18} (d_i - \bar{d})^2}{n(n-1)} \right]^{1/2}$$ for i = 1, 2,...18, and n = 18;

$$t = \frac{\bar{d}}{S_{\bar{d}}}$$ (under the null hypothesis, H_o: there is no significant difference between $\bar{Y}$ and $\bar{\bar{Y}}$); and

$$\hat{MSE} = \sum_{i=1}^{18} d_i^2 / n$$ (an estimate of the mean square error).

Table 5.--Predictive Accuracy Statistics, Total Personal Income, in Dollars

Model	Statistics[a]				
		Bias		Precision	Accuracy
	$\bar{\bar{Y}}_{1968}$	$\bar{d}$	t	$S_{\bar{d}}$	$\hat{MSE}$
Extrapolation	71,215	8,638	4.01	2,154	153,476
Lagged income	78,425	2,027	1.15	1,763	57,002
Economic sector resource based:					
Variation A	65,608	14,853	4.69	3,167	390,304
Variation B	73,670	6,791	3.44	1,974	136,643
Variation C	75,484	4,977	2.52	1,975	107,425
Variation D	76,720	3,741	1.11	3,370	220,035

[a]Statistics shown here are defined in Table 4. Income and $\hat{MSE}$ are in thousands of dollars. The actual mean income for 1968 ($\bar{Y}_{68}$) was approximately $80.5 million.

Table 6.--Predicted Total and Per Capita Personal Income, 1971, in Dollars

Area	Total Income[a]	Per Capita Income
1	205,648	3,149
2	284,203	3,367
3	69,562	2,404
4	41,498	2,516
5	79,821	2,781
6	194,021	2,791
7	110,686	2,100
8	180,369	2,782
9	52,970	2,333
10	103,694	2,612
11	54,301	2,133
12	83,824	3,071
13	55,284	2,089
14	45,327	2,352
15	39,917	1,712
16	101,620	2,201
17	83,851	2,814
18	54,720	2,206

[a]Total income predictions are in thousands of dollars.

for that matter), estimate parameters using cross-section data from the most recent year (or time-series data from recent years), and then proceed to crank out dependent variable predictions by utilizing these "historical" parameters (coefficients) and perhaps more up-to-date observations for the explanatory variables. Their only saving grace, in our opinion, is to actually expose any or all proposed predictive models to a rigorous set of test criteria and if they must actually forecast steps should be taken (criteria developed) to at least use the "least bad" of the models which have been tested.

Table 7.--Predicted Total and Per Capita Income for Selected Kentucky Area Development Districts, 1971, in Dollars

District	Total Income[a]	Per Capita Income
Green River	559,413	3,156
Pennyrile[b]	315,340	2,748
Barren River	502,020	2,474
Lake Cumberland	325,972	2,274
Buffalo Trace	138,571	2,531

[a]Thousands of dollars.

[b]Excludes Christian and Todd Counties.

1971 Forecasts[18]

The lagged income model, judged to be the most accurate predictor of both total and per capita personal income for the test interval, was used to forecast income for a three-year interval past the last year for which

[18]It should be noted that at the time this study was conducted 1971 was the future; and at the date of this writing 1971 personal income data still were not available. These data probably will not be available until sometime in 1973, so in a real sense the predictions to 1971 are forecasts.

income data were available, i.e., from 1968 to 1971. If one uses observed levels of income from 1962, 1965 and 1968 to determine forecast equation coefficients, the forecast equation for 1971 total income is

$$\tilde{Y}_{71} = 5936 + 1.1398Y_{68} + 0.2623(Y_{68} - Y_{65}) \quad (8)$$

while the forecast equation for 1971 per capita income is

$$\tilde{Y}_{71} = 393 + 0.9394Y_{68} + 0.5930\,(Y_{68} - Y_{65}) \quad (9)$$

Area income forecasts (predictions) for 1971, obtained by using equations (8) and (9) are presented in Table 6.

Recall that the areas with which this study dealt were multicounty subdivisions of Development Districts. Therefore, it was possible to sum forecasted total area incomes to the District level. These District level predictions are presented in Table 7.

District level per capita income forecasts for 1971 are also presented in Table 7. These forecasts were derived by determining the weighted average of the per capita incomes of the areas encompassed by a district. The "weights" in this case were the 1968 populations in respective areas.

SUMMARY AND CONCLUSIONS

The primary goal of this study was to devise a relatively simple and low cost method for forecasting income in rural areas. In meeting this goal, four separate models for predicting area income were constructed and tested. Each of these stochastic models was single equation in structure and employed data from secondary sources.

After the stochastic models were fitted, nonstochastic forms of the same equations were structured, with the dependent variable being the income level to be forecasted. The nonstochastic equations were referred to as test-forecast equations. Coefficients for these equations were generated by using least squares regression techniques for a time interval in which all variables were observable.

After the test-forecast equations were developed, they were used to "forecast" total and per capita area income for 1968 using only data available during and prior to 1965. These test-forecasts were made for each of 18 rural areas of Kentucky so that the models could be compared in terms of predictive accuracy.

Although all models yielded forecasts which generally were biased on the low side, the lagged income model was judged to be more accurate in predictions of both total and per capita income. For that reason it was used to predict total and per capita income levels from 1968 (the last year for which data were available) to 1971. Then, total income predictions were aggregated to the Development District level.

Even the "policy-oriented" models of this study were designed primarily to provide income forecasts and do not necessarily indicate clearly the sources of area income. It is possible that more complex models could be used to predict area income and also provide information concerning the sources of that income. Further, it would be desirable to construct models which might be used to forecast incomes more accurately than does the lagged income model. The extrapolation and lagged income models were constructed as a necessary first step in perfecting methods of predicting area income. Because of their simplicity, they should have value as a benchmark with which to compare other methods.

There are a number of potentially productive research avenues suggested by this preliminary effort in "quantifying and assessing the course of rural economic development." One important research topic is to develop alternative single equation policy-oriented type models emphasizing a single economic sector, probably agriculture and/or certain other subsectors. In connection with this effort, input-output (interindustry) analysis should be a useful tool for projecting predicted changes in area agricultural income throughout an area's economy in order to assess the "total" impact of alternative development strategies. Furthermore, the equity or income redistributive implications of alternative development strategies may be analyzed by using such techniques.[19]

[19] In our work concerning the "definition" of economic development, we noted that equity (as well as other considerations) in addition to "efficiency" (aggregate income effects) is an important dimension in the rural area development decision-making matrix [1].

There is an apparent need for additional research aimed at perfecting and quantifying criteria for evaluating all aspects of model efficacy. Current research has just begun to scratch the surface; high payoff would be expected from further efforts.

A final potentially productive area for additional research would be to extend this same sort of general framework, particularly the model testing aspects, using "more sophisticated" analytical techniques, e.g., programming models, simulation, etc. There is considerable speculation in the profession concerning the relative merits and demerits of "sophisticated" mathematical models vis'-a-vis' "less sophisticated" models. Seemingly, this speculation should be confronted with some empirical evidence.

REFERENCES

(1) Beattie, Bruce R., T. H. Klindt and G. L. Bradford, *Perfecting Methods for Predicting the Course of Rural Area Development: Part I, Toward a Definition of Economic Development and a Framework for Evaluating Model Efficacy*, Department of Agricultural Economics, University of Kentucky, Research Report 11, 1972.

(2) Charlesworth, Harold K. and William W. Davis, *Kentucky Personal Income: State and County for Selected Years, 1929-1968*, Lexington, Kentucky: Office of Development Services and Business Research, College of Business and Economics, University of Kentucky, 1969.

(3) Johnston, J., *Econometric Methods*, New York, McGraw-Hill Book Company, Inc., 1963.

(4) Klindt, Thomas H., "Development of Procedures for Quantifying and Assessing the Economic Well-Being of Rural Areas," unpublished Ph.D. dissertation, Lexington, Kentucky, University of Kentucky, 1971.

(5) Lyle, C. V., *Proposal for Meeting the Basic Requirements for Resource Planning*, Atlanta, Georgia, Environmental Protection Agency, Federal Water Quality Administration, Southeast Region, 1971.

(6) President's National Advisory Commission on Rural Poverty, *The People Left Behind*, Washington, D. C.: U. S. Government Printing Office, 1967.

(7) President's Task Force on Rural Development, *A New Life for the Country*, Washington, D. C.: U. S. Government Printing Office, 1970.

APPENDIX A

DELINEATION OF GEOGRAPHIC STUDY AREAS

The state of Kentucky has been divided into nearly identical sets of multi-county Development Districts by the Kentucky Cooperative Extension Service and the Office of Development Services and Business Research. This study utilized these existing boundaries to achieve a degree of consistency in development efforts. However, only 5 of 15 Development Districts were selected: Pennyrile, Green River, Barren River, Lake Cumberland, and Buffalo Trace (Fig. 1).

A primary consideration in selecting these five Districts was that they had economies based primarily upon commercial agriculture. This was done to insure a degree of economic homogeneity throughout the entire range of observational units. Urban areas and Appalachian regions were excluded, not because they were unimportant, but because they represent different economies and their inclusion would have unduly complicated the analysis.

From the five selected Districts, observational units, referred to as "areas," were delineated. Each area consisted of two or three counties selected with the aim of keeping them reasonably homogeneous internally. Further, areas were selected such that they would exhaust the geographic area of a District without intersecting its boundary. Delineating areas in this manner allowed variables to be aggregated to the District level.

From the 5 districts, 19 areas were delineated. The counties which comprised each area are as follows:

Area 1 – Henderson, Union, and Webster
Area 2 – Daviess and Hancock
Area 3 – McLean and Ohio
Area 4 – Crittenden and Livingston
Area 5 – Caldwell, Lyon, and Trigg
Area 6 – Hopkins and Muhlenberg
Area 7 – Logan and Simpson
Area 8 – Butler and Warren
Area 9 – Edmondson and Hart
Area 10 – Barren and Metcalfe
Area 11 – Allen and Monroe
Area 12 – Green and Taylor
Area 13 – Adair and Casey
Area 14 – Cumberland and Russell
Area 15 – Clinton and Wayne
Area 16 – McCreary and Pulaski
Area 17 – Bracken, Mason, and Robertson
Area 18 – Fleming and Lewis
Area 19 – Christian and Todd

Area 19 was later excluded because income data were atypical and suspect.

APPENDIX B

Appendix B, Table 1.--Observed and Predicted Levels of Area Income, Extrapolation Model, 1968, in Dollars

Area	Total Income[a]		Per Capita Income	
	Observed	Predicted	Observed	Predicted
1	166,100	133,524	2,583	2,271
2	236,300	224,428	2,885	2,616
3	52,500	42,017	1,868	1,670
4	29,100	22,856	1,902	1,535
5	60,900	50,778	2,207	1,945
6	156,800	132,251	2,256	1,922
7	87,000	79,534	2,500	2,270
8	145,200	129,159	2,248	2,002
9	39,200	35,363	1,806	1,632
10	81,800	76,735	2,087	1,936
11	40,600	38,637	1,657	1,596
12	65,200	60,334	2,479	2,191
13	41,000	36,109	1,559	1,381
14	32,400	26,644	1,770	1,488
15	28,400	26,145	1,224	1,104
16	78,500	62,486	1,625	1,308
17	66,600	68,015	2,387	2,401
18	40,700	36,861	1,703	1,559

[a]In thousands of dollars.

Appendix B--*Continued*

Appendix B, Table 2.--Observed and Predicted Levels of Area Income, Lagged Income Model, 1968, in Dollars

Area	Total Income[a]		Per Capita Income	
	Observed	Predicted	Observed	Predicted
1	166,100	162,192	2,583	2,349
2	236,300	260,638	2,885	2,852
3	52,500	44,003	1,868	1,708
4	29,100	22,969	1,902	1,590
5	60,900	54,174	2,207	1,954
6	156,800	153,691	2,256	2,169
7	87,000	82,383	2,500	2,277
8	145,200	142,073	2,248	2,134
9	39,200	36,289	1,806	1,660
10	81,800	81,127	2,087	1,966
11	40,600	39,302	1,657	1,639
12	65,200	63,919	2,479	2,209
13	41,000	37,123	1,559	1,407
14	32,400	26,844	1,770	1,526
15	28,400	25,740	1,224	1,167
16	78,500	68,112	1,625	1,406
17	66,600	73,089	2,387	2,518
18	40,700	37,980	1,703	1,609

[a]In thousands of dollars.

Appendix B--*Continued*

Appendix B, Table 3.--Observed and Predicted Levels of Area Income, Economic-Sector, Resource-Based Model, 1968, in Dollars

Area	Total Personal Income[a]				
	Observed 1968	Predicted, Variation A	Predicted, Variation B	Predicted, Variation C	Predicted, Variation D
1	166,100	124,595	141,082	144,145	140,097
2	236,300	196,482	216,350	224,773	242,702
3	52,500	40,577	46,427	46,857	50,832
4	29,100	20,405	22,709	22,282	33,325
5	60,900	37,422	42,402	42,523	52,245
6	156,800	119,719	134,153	135,368	126,204
7	87,000	83,490	91,112	96,797	116,692
8	145,200	119,724	135,463	139,375	147,420
9	39,200	30,773	34,257	34,792	43,475
10	81,800	72,322	80,557	82,434	88,053
11	40,600	42,589	47,135	48,632	60,858
12	65,200	54,566	59,526	62,393	79,865
13	41,000	33,725	37,988	38,190	44,374
14	32,400	22,333	24,399	25,167	37,957
15	28,400	29,354	31,919	33,212	47,631
16	78,500	58,803	64,424	66,479	77,228
17	66,600	58,983	65,162	67,187	78,368
18	40,700	34,959	38,595	29,774	50,201

(*Continued*)

[a]In thousands of dollars.

Appendix B--*Continued*

Appendix B, Table 3 *(Continued)*.--Observed and Predicted Levels of Area Income, Economic-Sector, Resource-Based Model, 1968, in Dollars

Area	Per Capita Personal Income				
	Observed 1968	Predicted, Variation A	Predicted, Variation B	Predicted, Variation C	Predicted, Variation D
1	2,583	2,102	2,480	2,558	2,261
2	2,885	2,530	2,859	2,958	2,753
3	1,868	1,484	1,771	1,789	1,682
4	1,902	1,165	1,391	1,403	1,437
5	2,207	1,356	1,607	1,627	1,603
6	2,256	1,599	1,864	2,002	1,953
7	2,500	1,957	2,255	2,238	2,103
8	2,243	1,840	2,107	2,194	2,131
9	1,806	1,399	1,650	1,661	1,624
10	2,087	1,781	2,074	2,111	1,978
11	1,657	1,454	1,705	1,717	1,689
12	2,479	1,648	1,911	1,911	1,847
13	1,559	1,422	1,683	1,702	1,645
14	1,770	1,300	1,536	1,533	1,538
15	1,224	1,332	1,566	1,568	1,577
16	1,625	1,475	1,713	1,745	1,743
17	2,387	1,728	2,008	2,025	1,918
18	1,703	1,502	1,763	1,765	1,705