Despite
living stan
as possible
both fiscal
level of l
for older
facilities.
to stay out
varied in
As it
were specif
measures.
conditional
age and gra
proportion o
geographic a
at the date
Phase
as predicted
effects of
these variab
Phase
allowed in t

ABSTRACT

CHARACTERISTICS OF LOW INCOME RURAL FAMILIES RELATED TO EXPENDITURE AND CONSUMPTION PATTERNS: AN ANALYSIS OF RURAL POVERTY FOR PUBLIC PROGRAM PURPOSES

by Pius Weisgerber

Despite current interest in measures which can gauge family level of living status, little empirical work has been done on either a criterion or possible predictors for low-income living levels. This study examines both family expenditures and the value of family consumption as possible level of living criteria; it then examines a number of family characteristics for their capacity to predict the consumption level of low-income rural families. A multivariate regression model is the chief technique employed to show the effect of several social, economic, and demographic characteristics in determining low income spending and consumption.

As dependent variables, "total expenditure" and "total consumption" were specified along with six subclassifications of these same two measures. As independent variables, each of the following factors were considered: money income, housing, family size, chief source of income, age and education of the homemaker, race, farm and tenure status, and the proportion of food produced at home by the family. Data were from three geographic areas in which low income was considered a problem. Analysis of the data was conducted in four phases:

Phase I -- Money income, housing index, and family size were tested as predictors of the living level. Initial assumptions were that the effects of the three variables were continuous and could be fitted as linear variables.

Phase II -- Additional variables were introduced and more freedom allowed in the way the independent variables were permitted to operate,

e.g., transformations were introduced so that specifications could conform to curvature; categorical variables were used to reflect discrete effects when continuous variables did not seem to fit. Dummy variables were specified for all characteristics other than housing index and family size.

Phase III -- Additional exploratory regressions were run allowing for the possibility that an "additivity" specification over the whole range of a continuous variable might not be warranted. This method, which was used to test the hypothesis of null interaction effects, used interaction terms in an extension of the dummy variable regression technique.

Phase IV -- A final model incorporating insights gained from preliminary analyses was tested. Joint variables were specified which presumed to represent a valid approximation to the underlying theoretical structure of consumption behavior among low-income rural families. The race, age, and major source of income factors were most salient for identifying and segregating the "best" set of family subgroups. "Proportion of home produced food" categories were also important in predicting more basic family consumption.

Housing index and log of family size were also important predictors in the final model but the slopes of the consumption functions for these two variables were known to vary for different categories of other family characteristics. The conclusion arrived at was that the true effect of the housing index and family size could best be determined by running separate computations for certain population groups such as (a) owners and renters, (b) whites and nonwhites, and (c) families earning most of their incomes from wages, salaries, professions or business, and all other families as a group.

CHARACTERISTICS OF LOW INCOME RURAL FAMILIES RELATED TO EXPENDITURE AND CONSUMPTION PATTERNS: AN ANALYSIS OF RURAL POVERTY FOR PUBLIC PROGRAM PURPOSES

By

Pius Weisgerber

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
1966

ACKNOWLEDGMENTS

Acknowledgment is made of the assistance and encouragement received throughout this study from Dr. James T. Bonnen, Michigan State University. Also, acknowledgment is made to Robert B. Reese, who by relieving the author of many other duties as an employee of the U. S. Department of Agriculture, afforded the time that was needed to do this study.

Special appreciation is extended to Faith Clark, Jean Pennock, and Mary Jane Ellis of the Consumer and Food Economics Research Division, ARS, USDA, who supplied the carefully collected data without which this study could not have been undertaken.

My thanks also to the secretarial, clerical, and computer programming staff of the Marketing Economics Division, ERS, USDA, for their vital assistance in bringing this study and manuscript into being. Special thanks are extended to Mrs. Rita B. Witten who did nearly all the desk calculator work which went into this study.

CHAPTER

I. [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

II. [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

III. [illegible]

A. [illegible]

[illegible]

[illegible]

IV. [illegible]

A. [illegible]

[illegible]

[illegible]

V. [illegible]

A. [illegible]

[illegible]

TABLE OF CONTENTS

CHAPTER PAGE

I. INTRODUCTION

Qualifications for Public Assistance 2

Objectives of the Study 4

Previous Work Done 5

The Primary Data 10

Design and Methodology 13

II. SPECIFYING THE MODEL

The Role of Proxy Variables 20

Specifying the Dependent Variables 22

Specifying the Independent Variables 26

Analytic Procedure 33

III. EAST TEXAS

A General Description of the Area 41

Description of the Survey Sample 42

The Analysis
Phase I 43
Phase II 55
Phase III 66
Phase IV 83

IV. BLACKLAND TEXAS

A General Description of the Area 106

Description of the Survey Sample 107

The Analysis
Phase I 108
Phase II 111
Phase III 114
Phase IV 121

V. SOUTH CENTRAL KENTUCKY

A General Description of the Area 135

Description of the Survey Sample 136

CHAPTER PAGE

The Analysis
Phase I 137
Phase II 140
Phase III 142
Phase IV 149

VI. THE STUDIES COMPARED

Population Comparisons 162

Sample Comparisons 162

Analytic Comparisons
Phase I 163
Phase II 164
Phase III 166
Phase IV 169

VII. CONCLUSIONS

Empirical and Theoretical Conclusions -- Dependent Variables . . .175

Empirical and Theoretical Conclusions -- Independent Variables . .177

Methodological Conclusions 183

VIII. APPENDIX 186

IX. BIBLIOGRAPHY 194

CHARACTERISTICS OF LOW INCOME RURAL FAMILIES RELATED TO EXPENDITURE AND CONSUMPTION PATTERNS: AN ANALYSIS OF RURAL POVERTY FOR PUBLIC PROGRAM PURPOSES

CHAPTER I

INTRODUCTION

The Economic Opportunity Acts of 1964 and 1965 draw increasing and renewed focus on a special segment of American society -- the poor and the impoverished. The Congress has enacted several public programs to help those in very low income households achieve a better level of living.

Most antipoverty programs attempt to help people rise above the poverty level by providing them the means to improve their employability, or by giving them better opportunities for greater self-help. Some other programs are designed to alleviate the effects of poverty, but do not move the recipient poor out of the ranks of the impoverished on any self-sustaining basis. The aged, the ill, and the incapacitated are essentially isolated from the productive world around them, and they can be aided only by direct income transfers or specialized programs. A direct attack on their poverty status can be made with such assistance as improved low-cost housing or supplemental food. The purpose is to upgrade the living conditions of those who are in poverty. It is with this latter type of program that the present thesis is chiefly concerned.

The ultimate objectives of the President's "war on poverty" are clear but there are and will continue to be many unanswered questions. There are uncertainties about many things, including the way proposed programs should be designed and implemented and about the degree of success which alternative approaches will have in operation. There is a need also for periodic assessment of a program's achievement.

Our state of knowledge concerning the rural poor is so inadequate that we are not even very sure how best to approach research related to the

problem. This thesis is somewhat of a pilot study designed to provide some basic information about the levels and patterns of consumption among low income rural families. It examines samples of households drawn from among "open country" dwellers in three low income areas of Kentucky and Texas. Knowledge gained from the study is intended to help public program administrators to better understand consumer behavior of low income rural people and provide for a more realistic adaptation of antipoverty programs.

Qualifications for Public Assistance

The process of classifying individuals or households as to eligibility for particular benefits presents many problems. A concise, simple definition of need will not detect important distinctions in actual well-being. More precise or more realistic definitions, on the other hand, may defy statistical analysis with data currently available. Nevertheless, a boundary about the poor must somehow be described, even if it only grossly defines a level below which family resources must not fall if a minimum living standard is to be maintained. This boundary or minimum standard is not established simply on the basis of characteristics of the low income population, but by an ultimately arbitrary determination made in the political decision making process and involving a wide range of matters including administrative and political feasibility.

The process by which needy families are certified as eligible to participate in a particular program is typically a large scale undertaking. The qualifying test for participation has to be applied to many thousands of families; therefore, it is not feasible to apply a very complex definition of need to each individual family. Instead, the measures used to establish need should be objective and readily obtainable, requiring a minimum expenditure of time by State and local certifying agencies. Thus,

a limited number of characteristics associated with low income families and some specific set of needs (e.g., housing, food, medical care, etc.) are selected for use in identifying the population to be served by a specific program.

Perhaps the most common measuring rod used for gauging the living level, and thus the need for assistance, is the current money income. However, money income is not a good measure of family living status because nonmonetary income often comprises a large portion of rural family resources. Furthermore, money incomes fluctuate a great deal among low income families and thus the current income rate does not necessarily reflect average family well-being.

The asset level of the family is another factor that is commonly taken into account to help determine the family's need for public aid. Where there are sufficient liquid assets, the consumption level may be reduced very little following a sharp reduction in family income. On the other hand, a family with few or no assets to draw upon may have to reduce its spending severely when there is a large drop in income.

Family size is still another variable that is frequently associated with classifying families when their need for assistance is being determined. It seems reasonable to use higher income cutoffs for larger families and lower cutoffs for smaller families. This approach presents a more realistic approximation to the minimum needs of a family than would any single across-the-board income cutoff for all families. Differences in family size are also important in characterizing the low income population. The incidence of low incomes is particularly high among the aged and in rural and nonwhite families. 1/ The aged are generally associated with small families, while rural and nonwhite families tend to be larger than the national average.

1/ Source: Economic Report of the President, transmitted to Congress January 1964, p. 81, Tables 11, 12.

As not

program gua

monetary in

important d

which educ

level of li

well-being

annual exp

family pos

more susc

It is

level educ

of consumi

complex pr

freedoms,

relation

or experi

and the p

those mos

"level of

one or th

The

characteri

with exper

relation

Objectives of the Study

As noted in the previous section, when families are certified for program participation, among the measures most commonly obtained are their monetary income, family size, and the level of their liquid assets. An important objective of this study is to measure the degree of association which exists between characteristics such as these and the rural family's level of living. For the purpose of this study, the level of living or well-being of the rural family will be measured alternatively by their annual expenditures on family living and by the value of their total family consumption for one year. These two measures will be described more fully in Chapter II.

It is recognized that in the family experience, the actual living level encompasses more than either the level of expenditure or the level of consumption. As Davis points out, the level of living is made up of a complex combination of consumption, working conditions, possessions, freedoms, and "atmosphere" and the balance or harmony among them, in relation to needs and felt wants. 1/ In this study, however, consumption or expenditure are the only components which are measured and available, and the term "level of living" is used in the thesis when referring to those measures in a generic sense. The terms "level of consumption" and "level of expenditure" are reserved for use when referring specifically to one or the other of the two measures.

The poor in this study will also be classified by various additional characteristics to see which family traits have a degree of association with expenditure and consumption. The measures and traits that show a relation with the living level can then be combined in a multivariate

1/ Joseph S. Davis, "Standards and Content of Living," American Economic Review, Vol. XXXV (March 1945), p. 7.

regression

makers and

characteri

to the pro

quantifying

limitations --

It also is

for the mark

A furth

methods for

was consider

strategic in

with groups

Certain

factor which

example, in a

in food experi

an applied ex

correspondence

concluded that

particularly c

extreme lowfix

Another s

lowfrequency plan

case survey ac

regression model. The results should provide information of use to policy makers and administrative decision makers in identifying the salient characteristics of the rural poor. This kind of information is necessary to the process of identifying the poor, both within an area and in specifying the geographic distribution of the poor between political subdivisions -- and thus in designing a system for allocating program funds. It also is necessary in the process of designing or redesigning programs for the rural poor.

A further objective of the study is to gain a better insight into methods for studying low-income, low socioeconomic status families. It was considered likely that a different set of influences would be strategic in determining the behavior of low-income families, as compared with groups belonging to higher levels of income.

Previous Work Done

Certain evidence does exist that lower income households behave in a manner which is distinct from that of the main body of households. For example, in a study which examined factors associated with the variance in food expenditures among a national sample of households, Herrmann found an atypical expenditure behavior on the part of households from the lowest socioeconomic categories of that study. 1/ On the basis of his study, he concludes that neither social class nor education of the homemaker seem particularly significant except insofar as they separate out the "deviant" extreme low-income group. He suggests separate study of the deviant group.

Another study that appears to attribute a distinct mode of behavior to low-income families was reported by Rockwell. In his study based on the same survey as that examined by Herrmann -- namely, the 1955 Household Food

1/ Robert O. Herrmann, Household Socioeconomic and Demographic Characteristics as Determinants of Food Expenditure Behavior. Unpublished Ph. D. Thesis. Michigan State University, 1964, p. 82.

Consumption Survey -- Mr. Rockwell reports as follows: "The income elasticity of demand for fluid whole milk was found to be only moderately high, and then only in low income households." 1/ In a more general statement he says: "Low-income nonfarm families whose incomes increase significantly tend to expand their food purchases more, percentagewise, than do high-income families receiving proportionate income increases -- as one would expect on the basis of Engel's law. Low-income farm families, being less dependent on purchased food, tend to increase their food consumption very little in response to significant income increases."

In still another study where Martin David examined family composition and consumption, he describes his results that related to automobile consumption. 2/ David found significantly different regressions for high and low income groups -- the marginal propensity to consume out of the incremental income was found to be greater among the lower income brackets.

Nearly all research relating to low-income families is reported in a manner similar to that of the three examples cited above. In other words, low-income family behavior is usually examined within a study of broader scope, and the results pertaining to low-income families are thus reported as an incidental part of a larger study. The results are in general what one would expect from economic theory.

Besides the analyses that have been based on individual cross-sectional studies, there has also been considerable work done to compare the findings of studies of a similar nature conducted in two or more periods of time. The results of the 1955 Household Food Consumption Survey have been compared, for example, with a similar U. S. national food consumption survey which was

1/ George R. Rockwell, Jr., Income and Household Size--Their Effects on Food Consumption. U. S. Department of Agriculture, MRR-340, June 1959. Summary.

2/ M. H. David, Family Composition and Consumption, "Contributions to Economic Analysis" No. XXV; Amsterdam: North Holland Publishing Co., 1962, pp. 86-89.

made in 1
made betw
within the
comparison
farm hous
consumed a
between 1
significant
farmers tha
farmers (e
declined o
survey repo
to attribut
which may
Decli
consumption
of consumer
Statistics,
so that th
current. F
1952, corre
12,900 farm
Many
have been
studies of

made in 1942. Most of the comparisons between the two points in time are made between similar groups of families such as the same income class and within the same urbanization class. Marguerite Burk reports an interesting comparison between the 1942 and 1955 studies that concerned low-income farm households. She sought to explain a rise in the market value of food consumed at home which had occurred among low-income farm households between 1942 and 1955. 1/ After considerable research, Burk found the significant element to be the different sample composition among low-income farmers that existed in the two survey periods. In 1955 the ratio of cotton farmers (especially sharecroppers) to tobacco and general farmers, had declined substantially from what it had been in 1942. A study of separate survey reports for Southern farming areas provided the information needed to attribute a large part of the food consumption increase to the decrease which had occurred in the proportion of cotton farmers.

Besides the large scale studies that have been made relative to food consumption only, there have also been a number of national surveys made on consumer expenditures for all goods and services. The Bureau of Labor Statistics, U. S. Department of Labor, conducts such surveys periodically so that the weighting factors used for the consumer price index may be kept current. Perhaps the most comprehensive BLS study was made in 1950 and 1951, covering 1,500 items of budget information collected from each of 12,500 families and individuals in all income and occupational classes.

Many volumes of statistical data based on the 1950 and 1951 surveys have been published, 2/ and there has also been a series of research studies of the economic and social aspects of consumption and savings.

1/ Marguerite C. Burk, Influences of Economic and Social Factors on U. S. Food Consumption, Minneapolis: Burgess, 1961, p. 93.

2/ The data tabulations and cross-tabulations are available in Studies of Consumer Expenditures Incomes and Savings, Philadelphia: University of Pennsylvania, 1956, 18 volumes.

One of the important research papers based on the 1950 BLS cross-section data was one in which Crockett and Friend fitted multivariate regressions to ungrouped data. 1/ One of their chief aims was to measure the net effect of income on consumption, while holding constant a number of other family characteristics which were correlated with income and which therefore would bias the income effect if not taken explicitly into account. The main part of the two author's regression analysis was devoted to after tax incomes in the $1,000 to $9,999 range. It was considered likely that such incomes would be reported more accurately and be less affected by transitory or abnormal elements than the incomes at either extreme.

The two researchers did, however, also investigate the consumption behavior of the extreme low-income families, as well as those who reported incomes over $10,000. Their appraisal of the overall consumption is as follows: "A complete demand function may thus be presented by 3 straight lines, one corresponding to extremely low incomes, the second corresponding to incomes in the $1-10 thousand range, and the third corresponding to the upper income extreme. In general, both the extreme slopes are flatter than the central slope giving an S-curve effect." 2/

The authors could see the likelihood of certain confounding factors prevailing at the lower end of the income scale. They surmised that when they were dealing with the poorer families, assets and borrowing capacity would become much more important relative to income in the determination of consumption. They also hypothesized that the very low-income families would have larger values of "transitory" income and that the observational errors would be much larger.

1/ Jean Crockett and Irwin Friend, "A Complete Set of Consumer Demand Relationships," Proceedings of the Conference on Consumption and Saving, Vol. I, Irwin Friend and Robert Jones (Eds.) Philadelphia: University of Pennsylvania, 1960, pp. 1-92.

2/ Ibid, pp. 10-11.

The

However,

wide sum

wide sum

wider wid

information

In Ky

rate, wide

and the Ky

alienable

and 400" s

low-income

agricultural

interaction

study of a

capitalized

collected

under $5,000

In recen

multiactivity

study of cou

Three such s

of Kentucky

The 1950 and 1951 BLS data were restricted to U. S. urban areas only. However, in 1955 the U. S. Department of Agriculture conducted a nation-wide survey of 3,845 farm-operator families. 1/ The major objective of this survey was to obtain expenditure data for the development of price index weights; however, it has also served very readily as a source of information for studying rural family living expenditures. 2/

In 1960 and 1961 a new national survey of consumer expenditure was made, this time as a joint project between the Bureau of Labor Statistics and the Department of Agriculture. To date, the BLS has published a considerable number of general purpose tables in its "BLS Report Number 237 and 238" series. It is anticipated there will be sufficient numbers of low-income households within each urbanization class 3/ so that considerable analytical work with low-income families will be feasible for any single urbanization class. The sample is not adequate, however, to do a separate study of a geographic subdivision say as small as a standard metropolitan statistical area. In the Detroit SMSA, for example, a report based on data collected from one-half the full sample had only 19 families with incomes under $3,000. 4/

In recent years a number of surveys have been made which were conducted exclusively in economically depressed areas. They had as their goal the study of consumption behavior of needy families as a group unto themselves. Three such surveys were made in the period 1957-59 in depressed rural areas of Kentucky and Texas where low farm income was considered a major problem.

1/ For further information about this study see: U. S. Bureau of Census. U. S. Census of Agriculture, 1954. Vol. III, Special Reports Part II, Farmer's Expenditures in 1955 (Washington: USGPO, 1956).

2/ See, for instance, Laura Mae Webb's Report to the 38th Annual Outlook Conference, U. S. Department of Agriculture, Washington, D. C., November 16, 1960.

3/ The specified urbanization classes are: urban, rural nonfarm, total nonfarm, and rural farm.

4/ See Survey of Consumer Expenditures 1960-61, U. S. Department of Labor, BLS Report Number 237-1, November 1962, (Advance Report).

It was from these surveys that the primary data for the present study are drawn. These data are discussed in the next section.

Since 1961, a number of research agencies of the U. S. Department of Agriculture have done several studies centering around the Pilot Food Stamp Program. Research related to this program included surveys of sales in retail food stores and of attitudes and opinions of low-income families, members of the food industry, welfare workers, and others. There has also been a series of special studies designed to measure the economic and dietary effects of the program. 1/

The present study aims to contribute additional findings which can help public administrators and others who are concerned with programs devoted to the welfare of needy families.

The Primary Data

The surveys which form the primary basis for this study were not conducted specifically for the purpose to which they are employed here. Nevertheless, there is much to recommend these data for a preliminary study of low-income rural families. Our systematic knowledge of low-income rural dwellers is very meager at this stage, and we can profitably look at the information at hand as a pilot study before embarking on a more ambitious project.

The data employed for the present analysis were collected in surveys of household living expenditures and consumption conducted by the Consumer and Food Economics Research Division, Agricultural Research Service, U. S. Department of Agriculture. The surveys provide information for studies of family living in low income rural areas of South Central Kentucky, East Texas, and the Northern Blackland area of Texas. They are based on samples

1/Three reports evaluating the Pilot Food Stamp Program have been published by the Department of Agriculture: The Food Stamp Program, AMS-472, April 1962; Effect of the Pilot Food Stamp Program on Retail Food Store Sales, AER-8, April 1962; and Food Consumption and Dietary Levels Under the Pilot Food Stamp Program, AER-9, June 1962.

reasonably a

About 150

The a

since this

Pennsylvan

areas of t

to other r

Texas and t

from the o

this type o

There

determinat

boulder-si

large anch

belt of th

indicate o

are possib

Alleged ac

Army Corps

The cou

activate.

effective

Date fo

Reply Card

September

County Pers

Re Income

randomly chosen from a five county area in each of the three regions. About 350 open country inhabitants were interviewed in each region.

The areas sampled represent appropriate populations for our purpose, since they have a high incidence of poverty within their boundaries. Furthermore, they may be considered as somewhat typical of two depressed areas of the United States. South Central Kentucky may be fairly similar to other rural areas in the poverty stricken Appalachian region. East Texas and the Texas Blackland, on the other hand, are closely associated with the old cotton producing areas of the South and may fairly represent this type of depressed region.

There was a further assurance of high poverty concentration since the observations were drawn from the rural or "open country" portions of the counties that were surveyed. Rural areas contain a disproportionately large number of the Nation's poor -- 32 percent of the rural and 16 percent of the urban population had money incomes under $3,000 in 1959. The incidence of poverty is particularly high among those who are farmers or are connected with farming. In 1962, forty-five percent of farmers and farm managers earned money incomes of less than $3,000. Poverty was even greater among farm laborers and foremen, with 56 percent earning below $3,000. 1/

The study sample included only those families headed by both husband and wife. A further qualification was that the family had to have been in existence for at least one year at the time the interview was conducted.

Data from the three sample areas are not strictly comparable. The family living data collected in Kentucky were for the 12 month period, September 1956 through August 1957. However, income data for the same study were for calendar 1956 and thus predate the rest of the data somewhat. The income figures for Kentucky are given as net before taxes. In contrast,

1/ Source: Economic Report of the President, Transmitted to Congress January 1964, p. 81, Tables 11, 12.

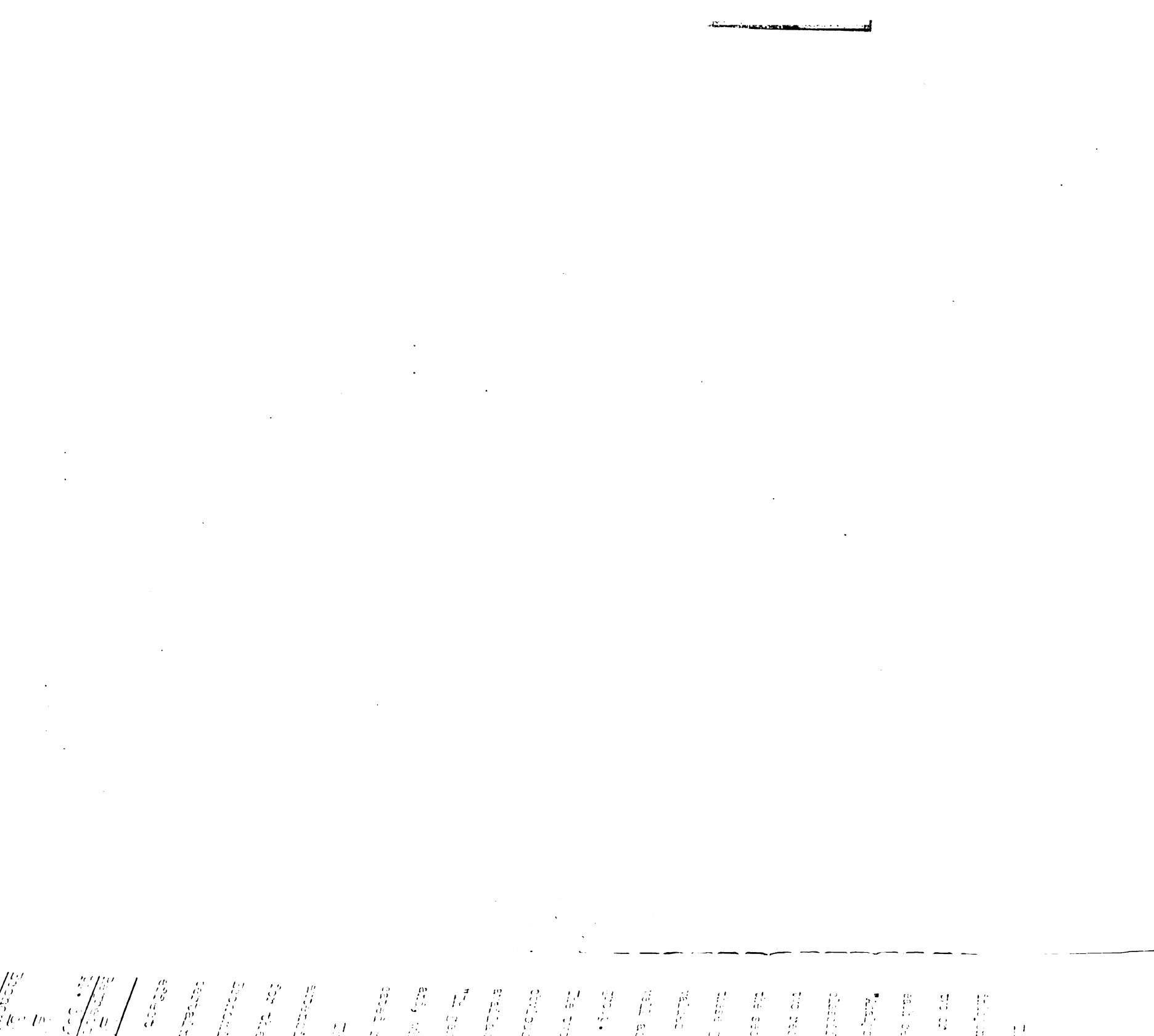
income figures for the two Texas studies are given as net after taxes. In further contrast to the Kentucky study, the income for the two Texas studies was for the same period as all other data in the same sample -- calendar 1958 for East Texas and calendar 1959 for the Blacklands.

Kentucky income figures differ from those of the Texas studies in yet another respect. The farm component in the Kentucky income figures consists only of net cash farm receipts. The farm component of the income figures of the Texas households, on the other hand, reflects changes in farm inventory as well as cash receipts.

The questionnaires used in the three surveys were designed to obtain all expenditures for family living for one year. An additional feature permitted the determination of a more comprehensive, more stable measure of the level of living. This measure, termed the value of consumption, includes, besides money expenditures on nondurables, the valuations placed on home produced food, fuel and clothing, as well as a year's depreciation on the family home, its furnishings and equipment, its stocks of clothing, and the family share of automobiles and trucks. These valuations were determined by professional home economists of the Family Economics Branch, Consumer and Food Economics Research Division, U. S. Department of Agriculture.

The relative merit of the annual consumption value over annual expenditures, as a measure of the level of living, should show up especially in the present study of low income families. 1/ Within rural life the consumption of home produced (rather than purchased) goods and services, is most frequently a major component of family consumption in the rural low income sector.

1/ Data from the Kentucky and East Texas studies have been used in comparing the two measures of the level of living in an article by Mary Jane Ellis, "Value of Consumption: An Improved Measure of the Level of Living." Family Economics Review, U. S. Department of Agriculture, June 1961, pp. 3-8.



Design and Methodology

The technique most commonly used to isolate the effects of given independent variables on a dependent variable, involves cross-classifying two or more independent variables and averaging the dependent variable for each cell defined. Ezekial and Fox describe this process as follows: "Analysis by averages where there are two independent variables, involves classifying the records first by one variable, then breaking each of the two resulting groups into several smaller groups according to the values of the second variable. If a third independent variable were to be considered, these groups would be broken up into still smaller groups, according to the values of the third variable. Then the values of the dependent variable as well as each of the independent variables would be averaged for each subgroup. This process is known as subclassification or cross-classification."1/ The two authors regard cross-classifying and averaging as an intuitively obvious method to employ for analysis when large numbers of observations are available. 2/ Regarding this same technique of analysis, Daniel Suits says: "In general, the function treated as a table is the easiest to understand, and, where adequate data are available, is the most complete and sophisticated representation possible." 3/

The method of cross-classification and averaging has been used extensively in analyzing the three surveys which contain the primary data for the present study. The Family Economics Branch, Consumer and Food Economics Research Division, USDA, has used this technique to cross-classify certain basic variables with other family characteristics to determine the effect of these factors on family expenditures and consumption. Since a

1/ Mordecai Ezekial and Karl A. Fox, Methods of Correlation and Regression Analysis--Third Edition. New York, Wiley and Sons, Inc. pp. 389-390.

2/ Ibid, p. 388.

3/ Daniel B. Suits, Statistics: An Introduction of Quantitative Economic Research, Chicago: Rand McNally and Co., 1963, p. 87.

fairly detailed breakdown of items of expenditure and consumption was employed, the data could be presented by only two or three cross-classifying variables. 1/

The procedure of holding, say, two variables constant while the effects of a third variable are being examined, is not an adequate analytical tool when it would be desirable to control the effects of 3 or more variables. With respect to consumer behavior there are usually several important factors that are closely correlated with the factor whose effects are being studied. Thus, if the studied factor is to be properly isolated, a method must be found that will hold the effects of the correlated factors constant. As we shall see in subsequent chapters, failure to control the effects of correlated characteristics may affect the estimates obtained to a marked extent.

The main analytical technique employed in the present study is a single equation multiple regression. It is obvious that economic phenomena are affected by many factors and the fluctuations in a dependent variable may be due to the interaction of many forces. Food expenditures, for example, may vary with numbers and ages of family members, the childhood or ethnic background of parents, the price of food in stores, the kinds and values of home produced foods, and many more factors. It is essential, therefore, that a multivariate function be employed so that a number of different independent variables can be explicitly specified as influencing the dependent variable.

But we can never list and measure all of the factors which influence the dependent variable. In addition to the factors mentioned above, there are literally hundreds of other variables which have a greater or lesser effect on food purchases. Of these, only a few can be specified in any

1/ In connection with these studies the Agricultural Research Service has published Home Economics Research Report No. 26: Rural Family Spending and Consumption in a Low-Income Area of Kentucky, by Jean L. Pennock. The same author is preparing a report based on the two Texas studies as well.

one function being considered. The dependent variable is, therefore, not an exact function of the designated independent variables, and is for that reason called a statistical or stochastic function. The specified variables comprise the systematic component of such a function and the non-specified variables make up the stochastic component.

The statistical function employed takes the general form:

$$Y = f(X_1, X_2, \ldots X_k, X_k + 1, \ldots X_n) + u$$

where Y is family expenditure or consumption, X_1 to X_k consist of continuous independent variables, $X_k + 1$ to X_n represent one or more sets of discrete independent variables, and u represents the myriad of nonspecified, non-measured variables. The continuous variables, X_1 to X_k, are also referred to as numerical, interval, or scaled variables, whereas the discrete sets $X_k + 1$ to X_n, are variously referred to as dummy or categorical variables.

The first problem in specifying the function lies in the decision of which dependent variables one should specify. The answer to this problem depends largely on what it is that we really want to explain or study, e.g., are we interested in only food expenditures, or do we want to look at the broader total of family consumption?

The next problem is in specifying the independent variables which on theoretical grounds or from previous empirical evidence are believed to affect the dependent variable. This process often ends up in a compromise between the measures that are thought to be pertinent and those which are actually measurable and available to be related to the dependent variable.

A third important specification problem revolves around the designation of the proper form that should be attributed to each variable, i.e., should data be presented in their original form or with a log, square root, or some other transformation; should the function be fitted linearly, linearly in logs, or curvilinearly; etc. The objective is to obtain an appropriate fit for each variable.

The r

procedure.

at the end

available.

and this mo

$Y = b_0$

$Y = b'$

$Y = b''$

.

.

.

The vari

in the "goo

the equation.

An im

(a) a variabl

enters the eq

regression eq

ment. The fi

ating an ad

included in th

Certain v

special consid

equation and i

inevitably, or

are added or

The regression analysis employed in this study is a stepwise procedure. 1/ Besides the complete multiple regression equation obtained at the end of a computer run, a number of intermediate equations are also available. These equations are obtained by adding one variable at a time and this results in the following intermediate equations:

$$Y = b_o + b_1X_1$$

$$Y = b'_o + b'_1X_1 + b'_2X_2$$

$$Y = b''_o + b''_1X_1 + b''_2X_2 + b''_3X_3$$

.

.

.

The variable added is that one which makes the greatest improvement in the "goodness of fit." The coefficients represent the best values when the equation is fitted by the specific variables included in the equation.

An important property of the stepwise procedure involves the facts that (a) a variable may be indicated to be significant in an early stage and thus enters the equation and (b) after several other variables are added to the regression equation, the initial variable may be indicated to be insignificant. The insignificant variable will be removed from the equation before adding an additional variable. Therefore, only significant variables are included in the final regression.

Certain variables such as income and family size are singled out for special consideration and study. These variables can be forced into the equation and held there even if their coefficients are insignificant initially, or if they should lose their significance as other variables are added or removed. Therefore, the final multiple-regression equation

1/ For a discussion of this procedure see M. A. Efroymson, "Multiple Regression Analysis" in Mathematical Methods for Digital Computers, Anthony Ralston and Herbert S. Wilf, (Eds.) John Wiley and Sons, Inc. 1960, pp. 191-203.

will contain all forced variables regardless of their significance, and any others of the specified variables which came into the equation and remained significant at a designated F test level.

For those variables not forced into the regression but brought in as the stepwise technique selects them, the criterion or decision rule for selection is as follows: If the variance reduction obtained by adding a variable to the regression is significant at the specified F level, this variable is entered into the regression.

CHAPTER II

SPECIFYING THE MODEL

The investigator of consumption behavior does not have a fully systematic or adequate body of knowledge to which he can turn for guidance. However, he may obtain guideposts for his own purposes if he examines the various theories and postualtes found not only in economics but also in the other behavioral sciences.

Economists tend to take as their starting point the enabling conditions of income, assets, and credit which are available to the consumer household. Their point of view received its beginnings in the economic conditions that existed over a century ago when nearly all income went into subsistence. The masses of people were so poor that demand for a good was thought to be limited only by the consumer's ability to buy, or otherwise gain command of the good.

More recently, economistst have given recognition to the fact that social and psychological variables play an important role in economic matters. Demand is now considered a function not only of the ability to buy but also of the willingness by the consumer to make the purchase. In other words, it is now granted that consumers have considerable latitude in their actions, and many of their demands are influenced by their perceptions, attitudes, and expectations.

Questions concerning consumer behavior are now being explored by an ever widening group of social scientists on the basis that consumer actions are motivated by many factors rather than one factor of overwhelming importance. They are attempting to determine the extent to which the consumer's actions are associated with rationality, and to what degree they may be responding merely as creatures of habit and inertia. They are also interested in the extent to which the consumer may be a product of his

sociocultural setting and of the group or groups to which he belongs. Various hypotheses are postulated concerning the role that should be assigned to such attributes as the physiological drive, the social mores, and to such facts that the consumer grew up in an affluent society. 1/

The validity of many hypotheses can tentatively be tested by the employment of mathematical models of behavior. A behavior model, however, becomes researchable only when each of the explanatory variables can be represented by some objective measure. Thus, variables such as income, which ideally come in quantified continuous form, may plug into this model in an essentially unmodified form. Some other variables may be given numerical content by various scaling procedures. Although the resultant scales are usually quite arbitrary, they often do represent a useful further step in the direction of quantification.

Variables can often be quantified and become part of an empirical model if the characteristic considered is presented as a set of mutually exclusive subclassifications. These subclassifications, which are termed categorical or dummy variables, are supplied with numerical values and can then be placed in a regression equation. The procedure for using the dummy variable technique is briefly but well outlined by Suits. 2/

There is a tendency to restrict the explanatory variables to readily measured demographic and financial variables. Statistical groupings based on characteristics like age and income are not always the most pertinent classifications we could select, but in many cases at present they are the only variables readily available for statistical study of consumption.

1/ Some notable cross disciplinary investigations have been made through work carried on at the Survey Research Center, University of Michigan. One product based partly on this work was a book by George Katona: The Powerful Consumer (New York: McGraw-Hill, 1960).

2/ Daniel B. Suits, "Use of Dummy Variables in Regression Equation." Journal of the American Statistical Association. Vol. 52, pp. 548-551. December 1957.

behavior. Many other family traits or characteristics that economic or social theory consider to be relevant to consumer behavior are so complex and often so subjective that they are difficult to observe or to measure. Even though considered important, some phenomena may not lend themselves to any adequate description.

The Role of Proxy Variables

Hopefully, though, the classifications we do use in our models can capture a number of subjective intervening variables that relate to consumption or expenditures. This is accomplished when certain of the specified variables act as a proxy for forces that are not directly specified and in many cases are not even observed. The proxy variable is expected to explain much the same portion of variance in consumption that would be explained by a direct measure of the unobserved variable.

In no case would the proxy variable be able to capture exactly the theoretical construct in which we are interested. It is possible though, that a given variable can be proxy, in an imprecise way, to several different constructs. An example of such a proxy variable is illustrated by age classifications of the family homemaker.

When homemakers are categorized into age groups, the main consumption effect we hope to capture usually relates to stages of the family life cycle. These same age classifications, however, may reflect consumption behavior that is not directly attributable to the life cycle into which they fall. Following are three examples of behavior that may be traced to other theoretical constructs, and yet they may be captured by life cycle classifications:

(1) The age group that grew to maturity in the 1930's could have attitudes toward family expenditures which were uniformly different from the group which reached maturity in the post World War II period. These

attitudinal differences are suggested by the fact that vastly different economic conditions prevailed in the two periods.

(2) An older group could reflect many effects and a younger group very few of various influences that relate to the ethnic background of the older group -- even though they were of the same ethnic origin.

(3) Different age groups may exhibit different degrees of adapability to a subsistence type of living, and this fact might be expected to reflect in their family spending behavior, e.g., an older generation may be able to live on less cash income, by virtue of the fact that they grew up in a situation that demanded a great degree of subsistence living and, therefore, they learned to get along with less cash.

Along a similar vein, if we classify the family decision makers by their home ownership or their racial origin, we may also be delineating groups with similar attitudes and aspirations. Renters and sharecroppers who have practically no prospects for ownership, could be expected to regard the future somewhat differently from landowners or homeowners. Furthermore, there is apt to be a functional relationship between an individual's achievement motivation and his racial status. Kurt Lewin and his co-workers suggest that levels of aspiration are reality oriented in that they are only slightly higher or slightly lower than actual levels of accomplishment. 1/ We know that, on an average, the accomplishment levels in our present society are quite different for whites and nonwhites. They also vary with tenure status, especially when we compare sharecroppers with owners.

1/ Kurt Lewin; Tamara Dembo; Leon Festinger; and P. S. Sears, "Level of Aspiration." In J. McV. Hunt (Ed.), Personality and Behavior Disorders. (New York: The Ronald Press Company, 1944).

Specifying the Dependent Variable

As pointed our in Chapter I, an important objective of this study is to examine the relationship between the family living level and the characteristics thought to be associated with the living level. To achieve this objective, it is necessary first of all to specify the dependent variable which is to represent the level of living.

The current income is sometimes used as a measure of the family living level. When it relates to medium or high income people with a steady income, money income may be fairly adequate as a level of living index. But it can prove quite unsatisfactory when used to measure the well-being of low-income families, or persons who earn their livelihood mainly from farming or other endeavors where year to year income can vary greatly. Money incomes fluctuate a great deal if the household wage earners are employed only intermittently, or if they frequently work at "odd jobs" rather than at a regular line of work. The amount of assistance that poor families receive from public or private sources also varies with fluctuations in the family's eligibility to receive help.

Farm incomes exhibit considerable instability since they vary with the weather, prices received and paid, the type of farm program currently in effect, as well as other matters. Low-income farmers are also subject to various degrees of underemployment, varying with the amount of capital the family can bring to bear on their farming operations, and with the opportunities for off-farm employment.

A better measure of the level of living is found in the amount of goods and services consumed over a given period of time, such as one year. The measure most commonly available for this purpose is the family's total outlay for the goods and services which it has purchased with private resources at its command. Public services received, such as education and

police protection, are difficult to evaluate and are generally not included in any level of living measure. There are of course other factors, such as proximity to good highways or schools, which are not easy to evaluate and therefore are not usually included in a scaled index of living.

The three surveys from which we draw our primary data have provided for two different numerical measures of the living level:

(1) The money expenditures made during the survey year for goods and services that are normally associated with family living, regardless of the amounts of such goods and services actually consumed by the family in the same year.

(2) A measure of the value of goods and services actually consumed in the survey year, regardless of the family outlay for such goods and services during the year.

The first of these measures will be referred to as "family expenditures;" the second as "value of family consumption."

Family expenditures include the total outlay during the survey year for food that was consumed both at home and away from home. Included also, are expenditures on clothing, clothing services, costs of operating the household, repairs and replacements to the home, and home furnishings and equipment. Rent paid on the home was also included. Improvements and additions to the home were regarded as a capital investment, and none of it was charged to current living expenses. Other items included medical care, recreation, reading, education, personal care, and tobacco. The family share of automobile and truck operating costs was also included along with other transportation expenses. If an automobile had been purchased during the year, ten percent of its purchase price was added to family expenditures.

The value of family consumption includes the valuations attached to consumption of durables as well as nondurable goods and services. Home produced food, and food received as pay or gift are assigned values then added to purchased food to arrive at a value of total food consumed. Home produced fuel is also given a value and this amount is added to cash outlays in order to determine the total cost of household operations.

The value of durables consumed was determined by assigning an annual "use-value" to each of the durable goods used by the family. To the extent that it appeared feasible, house furnishings and equipment were inventoried and a use-value assigned on the basis of original cost and age. The use-value of clothing was based on an estimate of the length of time garments are kept in inventory, and the average value of the garments. Use-value of the family dwelling was made up of two components: (1) depreciation on the current value of the dwelling and (2) an allowance for maintenance, insurance, taxes, and return on investment. Similarly, a depreciation rate was applied to cars and trucks and a proportion was determined as the family share of use-value of automobiles and trucks.

A detailed description of the derivation of various consumption values can be found in U. S. Department of Agriculture's Home Economics Research Report No. 26. 1/

Subclasses of the Dependent Variable

It seemed appropriate to subdivide the total expenditure and total consumption into additional dependent variables for examination and study of additional hypotheses. For instance, the amount of food that is

1/ Pennock, Jean L., Rural Family Spending and Consumption in a Low Income Area of Kentucky, Home Economics Research Report No. 20, 1964. Appendix B.

purchased or consumed by the family appears to be a pertinent dependent variable to consider. Food is the single most important item in the family budget, and it tends to become a larger proportion of the total budget as the total purchasing power of the family becomes smaller.

A combination of the basic items, food, clothing, and shelter are another appropriate subclassification. Since we are concerned primarily with poor people, we should investigate the extent to which our sample of low-income families appears to approach a subsistence type of living. We regard a subsistence economy as one in which most people devote all their income to food, clothing and shelter and where consumer discretion is therefore largely absent. The designation "more basic" expenditures or consumption is used to identify this group of subsistence items.

It might be argued that the "more basic" designation ought to include some other items such as medical care. But we know that many poor people avail themselves of proper medical care only in the most dire of circumstances. In fact, it is readily apparent that many rural people with limited means strive harder for such modern day "necessities" as refrigerators, automobiles, and TV sets, than they do for adequate medical services. 1/

A final subclassification was designated "less basic" expenditures or consumption, since it consisted of the residual after the so-called more basic items of food, clothing, and shelter had been subtracted out of the total. The three most important items included in this less basic designation were transportation, medical care, and home furnishings and equipment.

1/ The author is not able to cite any research in support of this thesis, but a number of observations from his experience would tend to support the proposition. A rationalization for the priority given these "necessity" items is not hard to find: people who have lived in rural areas can attest to the importance of both car and refrigerator to the rural dweller. And as Katona points out, op cit, p. 26, watching TV was one of the least expensive of leisure time activities in the post World War II period.

Thus, there are eight separate dependent variables which will be examined in eight separate regression equations for every phase of the analysis that is undertaken. The variables are as follows:

Y_1 - Total family expenditure

Y_2 - Total family consumption

Y_3 - Total food expenditure

Y_4 - Total food consumption

Y_5 - More basic family expenditure

Y_6 - More basic family consumption

Y_7 - Less basic family expenditure

Y_8 - Less basic family consumption

The "more basic" designation includes all food, all clothing, and the costs of household use, operations and maintenance. The "less basic" designation makes up the remainder of all family expenditure or consumption for the survey year.

Specifying the Independent Variables

We have varying reasons for specifying the particular independent variables that are selected to go into a statistical model. A first requirement is that the variable be numerical, scaled, categorical, or that it be convertible into one of these quantified forms. It is equally apparent that the variable selected must have relevance in theory or an explanatory value in the judgment of the researcher.

Of course, in the case where secondary data are used the nature and limitations of the data as they were gathered often modify the form in which an independent variable is finally specified for the model. One has to contend with the limited number of observations that are available, both in total and within given subclassifications. Errors of observation

must be considered along with sampling variability when selecting the factors best equipped to reduce the residual error. The way in which the scaled and dummy variables are presented, is also considerably dependent on the nature of the survey data.

Variables for Preliminary Examination

Family income usually ranks among the most important of explanatory variables in the study of consumption behavior. It is immediately apparent though, that total income is comprised of a number of components. Money or cash income comes from the earnings of land, labor, and capital and the remainder derives from various transfer payments. Nonmoney income usually comes in several forms of "income in kind."

Information on total cash income was collected in the field surveys. The figure used for East Texas and Blackland Texas was money income after deduction of personal taxes. In the Kentucky study, information on taxes was not available and therefore the income figure used was "cash income before taxes." Among our designated "poor" people, however, the income before taxes figure should not run a great deal higher than income after taxes.

For the purpose of preliminary examination, money income was studied as a continuous variable and fitted as a linear function in the regression equation.

Among economically deprived people, especially those living in rural areas, cash income is apt to fluctuate a great deal. Crops and business returns can have severe ups and downs. The availability of jobs to family wage earners may be very intermittent. Outside aid coming from private and public sources may arrive only in sporadic fashion.

Two items that can add stability to the income picture of rural low-income families are nonmoney income and the assets which are held by a

family. In a rural consumption study, the more important components of nonmoney income are usually home produced food and fuel plus imputed rents received from the family dwelling. A lesser item, home sewing, can cut back substantially the amount of cash outlay needed for clothing.

Among the families interviewed in all three surveys, home produced food was the most significant item of nonmoney income. Furthermore, a large majority of families produced at least some of the food they consumed. Hence, the ratio "home produced food as a percentage of total food consumption" was utilized to indicate the effect of nonmoney income. Although home fuel production and home sewing were significant, only a small proportion of families engaged in these activities.

The home produced food variable was subdivided into 4 classes and these were run as dummy variables in the regression equation. It was hypothesized that a linear specification might have produced a biased regression coefficient due to an expected curvature in the relationship--hence the specification of dummy variables which could conform with any curvature that was present.

Assets can also help stabilize the welfare status of families who are subject to large fluctuations in the incomes of the family earners. Earnings from interest, dividends, royalties, and real estate rentals all represent a fairly steady income flow. Assets which do not earn direct returns can, furthermore, serve as the basis for obtaining credit when liquid assets do not suffice. Unfortunately, it was not possible to include an asset variable in the regression model since the surveys did not include much information on family assets.

The quality of homes in which low-income families resided, was considered a measure that might be closely related with the living level in general. Since quite detailed housing characteristic data had been

collected, the author undertook to build an index intended to reflect the adequacy of the housing in which each family lived. A scaling procedure was used which rated 17 different aspects of housing, placing each on the plus or minus side of an arbitrarily selected average. The net index calculated for each home was independent of the tenure status of the home's occupant. A complete account of the technique used to determine the so-called "housing index" is found in Appendix B.

It was expected that the housing index would fit into the regression model as a continuous linear function.

Source of income was chosen as another variable that might relate importantly to the living level. Families were classified according to the source from which they received the major portion of their money incomes. The classifications used were: (1) salaries and wages, (2) professions and business, (3) interest, rent, and transfer payments, and (4) income from farming operations.

Age of the homemaker was the characteristic selected to represent variations caused by stages in the family life cycle. Age could readily have been run as a continuous variable, but this was not considered the correct form in which to place this variable into the equation. Different bundles of goods and services apply to different stages of the life cycle because of events that conform to various age spans. Therefore, discrete classifications to conform with life cycle stages seemed more appropriate than a continuous mathematical function. The families were classified so as to represent three groups -- headed in turn, by a younger, a middle-aged, and an older homemaker. The age ranges for each of the categories were selected in part on the basis of the frequency distribution of the homemakers' ages in the surveys.

The life cycle stages were thought to be more closely associated with the wife's age than with the age of her husband. She marries at a considerably younger age than does her husband, especially in rural areas. And the standard deviation from the average age is smaller for the women than it is for men. In the Kentucky sample, which was the most rural of our surveys, the average age of the homemaker at her first marriage was 20.4 years. The standard deviation around this average was 5.4 years. The average husband was married at age 25.5 with a standard deviation of 7.9 years.

Since the age at which the wife marries is more nearly predictable than the husband's, it is also easier to delineate the period of household formation on the basis of the wife's age. Likewise, her age becomes a better guide to predicting the arrival and departure of children from the home. The shorter age range in which women can beget children, helps further to predict reliably the time when children are in the home, and their probable ages.

The education of the homemaker is also often considered a factor that has a bearing on the level at which a family spends or consumes. Increased education presumably provides the homemaker with better skills in managing the home and in making purchases for the family. The extent of her dependence on habit and inertia in management and buying, should also undergo certain modifications with greater education. Her tastes would change along with her attitudes toward such things as advertising and the introduction of new products. A better knowledge of nutrition could be expected to bring about some changes in food purchases and in home production of food.

There are at least two reasons for considering the education of the wife in preference to that of her husband. With regard to family purchases

and consumption, the wife is assumed to be connected more closely with the various decision making activities. Furthermore, in the samples of families interviewed, the wives were not so highly concentrated in the lower education levels as were their husbands. Thus, they could provide more observations to test the effects of the middle and upper levels of education.

Tenure status may influence the living level in a number of ways. Home owners can usually have more adequate housing, since the decision to make repairs or improvements to the home is their own and does not need to have the concurrence of a landlord. Owners of entire farms also have greater flexibility in the way they use their land or other farm resources. For instance, they may devote considerable resources to producing income in kind such as food or fuel. They do not have to answer to a landlord, should production of these products for family consumption cut significantly into commercial farm production. For similar reasons, farm owners are more free than renters to accept off-farm employment.

Race was regarded as an important characteristic for a model on spending or consumption behavior. It is recognized that this variable has considerable intercorrelation with a number of other classifications that were specified. Nonwhites, for example, tend to be concentrated among the lower income, lower education groups. They are typically engaged in non-farm pursuits or in very small, uneconomic size farming efforts. The quality of their home is vastly lower than that of whites, on an average.

But a number of differences between whites and nonwhites are, no doubt, attributable to different motives, attitudes, and expectations. The segregated conditions which have existed between the two races have probably produced a number of traits in each race that can be expected for the present to generate different consumption patterns.

Analytic Procedure

In studies that are dependent upon previously collected survey data, the variables specified are of necessity confined to factors which are measurable with the data collected in the survey. The variables we can afford, therefore, are often only partially effective in capturing the theoretical constructs we might wish to examine.

The nature and extent of the data set limits on the analysis, but, within these limits, the analyst must seek out every opportunity to approach the pertinent classifications and the proper functional forms. There is a further need to direct the analysis in such a way as to locate the important interactions among the variables. Then a new set of classifications, including combination variables, can be chosen to represent the constructs which seem to underlie consumption or spending behavior.

Steps in Building the Model

The steps used in building the model consist of four successive phases: Phases I, II, and III are exploratory, aimed at discovering the more decisive factors which determine the family level and pattern of living; Phase IV analyzes and discusses a final regression model.

Phase I considers three independent variables which, it was believed, should be analyzed separately from several other explanatory factors. Money income, family size, and housing adequacy were intended to serve in place of the three factors, income, family size, and assets, which are normally considered by case workers in establishing eligibility for family assistance.

In the surveys used for this study, measures of income and family size were collected but no measure was obtained which could properly account for net assets controlled by the family. Instead of studying the effect of assets, the influence of the housing index (which was discussed

earlier in this chapter) will be examined. It was hypothesized that housing adequacy may serve partially as a proxy for assets. But even if it should not capture the same effects as a direct measure of assets, the housing variable appeared to be an appropriate variable to consider in any case. Due to its importance in low income budgets, the adequacy of a family's housing may of itself, to a significant degree, reflect their need to participate in anti-poverty programs.

In the initial analysis, the factors of money income, housing, and family size are run as independent variables against the eight previously described dependent variables. Since the investigator can find no contrary evidence to begin with, we start with the simplifying assumption that the effects of the three variables are linear and that they can be fitted as continuous variables. The initial analytic process can, therefore, give an indication of the effectiveness of money income, housing index, and family size as predictors of the living level. In addition, we are afforded the opportunity to start out with a model that is simple and manageable.

After examining the eight regression equations of the first analysis, more freedom is permitted in the way the independent variables may operate. If there are indications, for example, that the effects of a continuous variable arenot linear, transformations are permitted to reflect possible curvilinear effects. Alternatively, the use of dummy variables can be employed to conform with any curvature that is present. To accomplish the latter, the conventionally measured variable is partitioned into intervals, a set of dummy variables is defined in terms of the intervals, and the dummy variables then serve in place of the continuous variable. 1/ Treating a numerical variable as a categorical variable is particularly desirable

1/ This method is suggested by Suits in his journal article; op cit, p. 551.

if the researcher has doubts concerning the validity of the numerical scale that goes with the measured variable.

In Phase II several new explanatory factors were added to the three already specified. This step recognized that additional variables should be taken out of the unspecified category and placed in the equation. The new variables were all specified as categorical variables because there was no a priori notion about the particular mathematical form which the true function would take.

The method of regression analysis employed was the stepwise procedure discussed in Chapter I. After forcing the three Phase I variables to come into the regression, the categorical variables were then permitted to enter only if they could contribute to explained variance at a pre-designated F test level. Thus, for instance, if the category of families with older homemakers entered the regression it would be an indication that the families of this cateogry consumed at a level significantly different from the combined group of families headed by middle aged and younger homemakers. If the regression coefficient associated with the older homemaker group is negative, it indicates the older families consumed less than the average of the remaining families; if the coefficient is positive, it indicates the opposite effect. The combined younger and middle aged group is the "base" group with which the older group compares.

Suppose now that after the older homemaker group entered the regression equation, there was still a significant difference (at the specified F level) in the consumption levels of the groups headed by middle aged and younger homemakers. Then in a subsequent sweep of the stepwise process we could expect either one of these two categories to enter into the regression. Let us say the middle aged group came in with a positive

coefficient; and let us say that the coefficient for the older group, although it would necessarily adjust, continued significant at the specified F level. The younger family group would now become the "base" category with which the other two categories compared. The negative coefficient associated with the older group indicates they consume at a lower level than the younger group; the positive coefficient with the middle aged group indicates they consume at a higher level than the younger group.

The entry of a single category from a set of dummy variables indicates that this particular variable significantly increases the sum of squares explained. We know this is so by definition of the way the stepwise regression technique operates. However, lack of significance in any single class of a set of dummy variables reveals nothing about the possible significance that the set of classes would exhibit if taken jointly or as a group. The procedure for testing the null hypothesis that a subset of the coefficients in a regression are all (jointly) equal to zero is straightforward, and may be found in many standard references. In the present investigation, time and other resources did not permit the many computer runs that would have been required to test the significance of all pertinent subsets of coefficients.

The objectives of the Phase II analysis are only exploratory, searching for the more decisive characteristics which influence consumption behavior. To accomplish this purpose, the stepwise technique was regarded as adequate to give an indication of the factors that should be considered in the final regression model to be tested in Phase IV of the analysis. The categorical variables that enter the Phase II regressions may be regarded as representing the more important effects which determine fluctuations in family living.

The F level specified in Phase II is the 5 percent level throughout, and a number of subclassifications from several dummy variable sets will enter the regression at this level. When all the classifications have entered which can significantly contribute to variance reduction, a final (complete) multiple regression equation is printed out. In each of the three analysis chapters the complete equation for each of the 8 dependent variables is reported in a table.

There are numerous intercorrelations among the independent variables specified. Hence, any test to see if a given factor contributes significantly to explained variance, is contingent upon what other factors have been considered, i.e., the apparent influence of any one independent variable depends on what other variables have been allowed for. Thus, for instance, the effect of money income on family spending is modified in the regression equation every time we add or subtract a variable that is correlated with money income. We may, for example, introduce a second independent variable "income in kind" and allow for its effects. This will cause the regression coefficient and standard error of the coefficient for money income to be altered, depending on the degree of correlation between the two types of income. The addition of a third correlated variable, say nonwhite consumers, would further mediate the influence of money income on spending. This process could be continued until all hidden correlated variables had been taken out of the unspecified category.

In Phase III the analysis becomes even more complex when we begin to search for hidden interactions. 1/ To expose interactions, we must not only specify the important interacting variables, but also discover how

1/ Hidden interactions and correlations are discussed by Daniel B. Suits in his book -- Statistics: An Introduction to Quantitative Economic Research. Rand McNally & Company, 1963, pp. 111-115.

the impact of one variable depends on the values of the others. Students of consumer behavior already recognize a number of interaction effects, e.g., (1) substantial reductions in money income receipts are accompanied by large or small declines in consumption, depending on the amount of assets and income in kind at the family's disposal; (2) consumption increases with the size of the family, more or less depending on whether the family is well to do or has severe income limitations. Other examples could be cited.

The stepwise regression technique is employed again in Phase III, and this time it selects from among several sets of dummy variables made up of interaction terms. They are parallel sets to those employed in Phase II because the interaction terms are combinations of the categorical and continuous variables which were specified in Phase II. For example, if the continuous variable family size were designated X and three mutually exclusive homemaker age groups were A_1, A_2, and A_3, then the interaction terms XA_1, XA_2, and XA_3, would be one set of dummy variables.

Suppose XA_3 were selected under the decision rules to be brought into the regression equation. It would mean that the consumption function for the A_3 group of families sloped significantly differently from the families in A_1 and A_2. Note that when dealing with interaction terms we are looking for differences in the slope of the consumption function 1/, whereas in Phase II we were looking for differences in the level of consumption. Hence, whereas in Phase II the purpose of the selection process was to identify family groups which consumed at different levels, in Phase III the purpose is to identify groups with a different pattern of consumption with respect to some (continuous) variable.

1/ See Suits journal article, op cit, p. 550.

A more complete explanation of the way the stepwise selection technique seeks out interaction effects is presented in the Phase III discussion of Chapter III.

The investigator needs to direct his analysis so as to discover hidden interactions and thereby achieve the greater insight that the interactions afford. We are assured by noted workers in this field that we can expect distorted results from analyses that assume away all interaction effects. 1/ Morgan and Sonquist observe that: "Those of us who have looked closely at the nature of survey data, have been increasingly impressed with the importance of interaction effects and the useful way in which allowing for interactions between measured factors gets us closer to the effects of more basic theoretical constructs." 2/

In Phase IV the final regression analysis attempts to incorporate the insights gained from the preliminary analyses. Evidence of existing interaction warns us of the need for something other than a simple additive regression equation. This problem is partially handled by the creation of complex variables which make it possible to build some of the interaction effects right into the new classifications. It is hoped that the new classes specified, will also represent a valid approximation to the underlying theoretical structure of consumption behavior among low-income rural families.

Chapter III follows the general procedure outlined in this section in analyzing the survey that was conducted in East Texas. Chapters IV and V are devoted to a similar analysis of the Texas Blackland and South Central Kentucky respectively. Chapter VI will make comparisons among the three areas and will attempt to arrive at those generalizations that appear likely on the basis of the analysis of these surveys. Chapter VII will draw conclusions.

1/ See Suits book, op cit, p. 110.

2/ James N. Morgan and John A. Sonquist, "Problems in the Analysis of Survey Data and a Proposal." Journal of the American Statistical Assn.,

Most of the discussions and analyses that follow deal with families reporting less than $2,500 money income. These families were selected to represent the so-called "low-income" portion of the total sample in each survey. Whereever the discussion deals with the total sample rather than the subsample of families reporting under $2,500 income, it will be so indicated.

CHAPTER III

EAST TEXAS

A General Description of the Area

The five counties surveyed for this study were Anderson, Cherokee, Nacogdoches, Rusk, and Smith. They are a set of contiguous counties located approximately in the East Central portion of the State of Texas. (See Chart I, Appendix A)

The U. S. Census of Population reveals that these five counties had a total population of 212,000 in 1960. Of those with native parentage, nonwhites made up 28 percent. Males 25 years and older had a median education of 9.2 years 1/; females 25 and older averaged 10.3 years. Nonwhites 25 years and older (male and feamle combined) averaged 7.7 years of schooling.

Manufacturing is the chief industry group, employing about 18 percent of all employed persons in the five counties. Retail trade employs about 16 percent and agriculture ranks third with 10 percent. But among nonwhites, the classification "personal services" was the largest industry group employing about 28 percent of all employed nonwhites, manufacturing employed 17 percent; agriculture, forestry, and fisheries employed 16 percent, and wholesale and retail trade employed 12 percent.

The median annual income for the five counties was $3,886, but 27 percent of the families made less than $2,000. 2/ Among nonwhites, nearly 52 percent made less than $2,000 per year.

About half the population was classified "rural" and 30 percent of the rural people were nonwhite. The median rural income was a little over $3,000 and 36 percent of all rural families had incomes under $2,000.

1/ The median for each county was multiplied by the number of males 25 and older in that county, and the 5 products were added to make one total. This total was then divided by total males 25 and older in all 5 counties to yield a weighted median.

2/ These multicounty income statistics are weighted medians constructed in the same manner as the education data above.

Families classified "rural farm" made up less than one-fifth of the total rural population. The median rural farm income was $2,570 and 42 percent had incomes below $2,000.

Description of the Survey Sample

The total sample consisted of 350 randomly selected families who lived in the open country portion of the five counties at the time the survey was conducted in the spring of 1959. Their net incomes after taxes, in 1958, ranged from a negative $4,300 to nearly $20,000. One hundred eighty-nine of these families had incomes below $2,500, and for the purpose of this study they are labeled "low-income" or "poor" families.

It should be recognized that an across-the-board cutoff point will overemphasize the number of small families and underestimate the number of large families that are classified as impoverished. But in order to avoid a complicated classification process, a concept of absolute poverty was adopted here. The cutoff level of $2,500 annual money income is the same level used by Robert Lampman for a family of four in his report to the Joint Economic Committee, Congress of the United States in 1959. 1/

When all families in the sample were classified according to the chief source of their income, wages and salaries accounted for 54 percent; professions and business 8 percent; interest, rent, and transfer payments made up 23 percent; returns from farming 15 percent.

Roughly 22 percent of the families did not produce any of the food they consumed, and an additional 16 percent produced less than one-tenth of their food supply. Nineteen percent of the families produced more than half the total food they consumed.

1/ Study Paper No. 12 -- The Low Income Population and Economic Growth. U. S. Government Printing Office, Washington, D. C., 1959.

Nonwhites comprised 22 percent of the sample families, and they all produced at least some of the food consumed in their households. The bulk of nonwhite households fell in the class "10 to 29.9 percent of home produced food."

In the subsample of low-income families, wages, salaries, professions, and business provided the chief income for about 40 percent. Another 40 percent of the families received their main income from transfer payments, interest, and rent. The remaining 20 percent obtained most of their income from farming.

Fourteen percent of the low-income families produced no food of their own, and an additional 11 percent produced less than one-tenth of the total amount of food they consumed. Twenty-six percent produced more than half their food. The proportion of nonwhites rose to nearly one-third of the 189 low-income families.

The Analysis

Phase I -- Money Income, Housing Index, and Family Size as Explanatory Factors

The first set of regressions were intended to test the effectiveness of money income, housing index, and family size as predictors of the family living level. We have a special interest in these three household traits since they could very well serve as qualifying characteristics when certifying needy families for public aid programs. The object is to find out how closely the three factors correlate with the spending or consumption levels of the low-income families in this rural sample.

Table 3.1 gives results of regressions for eight dependent variables on income, housing, and family size. Each independent variable was run as a linear continuous function and there was found to be little intercorrelation among the three variables. The simple correlation coefficient between money

income and housing index was -0.11; between money income and family size, 0.06; between housing index and family size, -0.26.

It is notable that the regression coefficient for money income was insignificant, at the 5 percent level, in all cases except equation 3 where food expenditure was the dependent variable. Normally one would expect the money income variable to show high significance in all equations, because income is such an important enabling condition for a family's purchases or consumption.

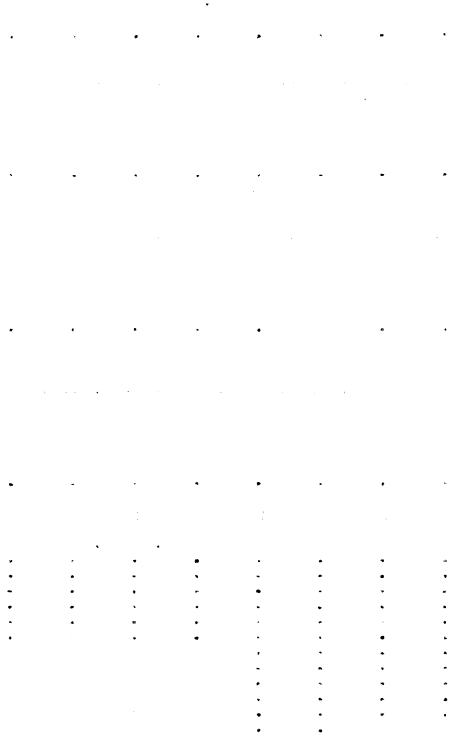
Failure of the income variable to show significance in the Table 3.1 regression, can be interpreted as an indication that money income has little effect on a low-income rural family's living level. Another possibility is that the money income function had so much curvature that it was not amenable to study by linear approximation.

Table 3.1--Computed t values in regressions of the living level on three household characteristics, low-income rural families, East Texas, 1958

Dependent variables	Independent variables			$\bar{R}^2$
	Money income	Housing index	Family size	
(1) Total expenditure	1.672	6.653	5.401	.235
(2) Total consumption	-0.514	10.433	7.356	.412
(3) Food expenditure	2.363	4.287	7.005	.237
(4) Food consumption	-0.567	5.559	7.289	.258
(5) More basic expenditure	1.555	6.222	6.940	.266
(6) More basic consumption	-0.945	11.242	9.750	.486
(7) Less basic expenditure	1.252	4.949	2.151	.109
(8) Less basic consumption	0.081	6.086	2.522	.159

Critical t value at 5 percent level = 1.973
Critical t value at 1 percent level = 2.600

In contrast to the insignificant effects of money income, the coefficients for the housing adequacy index were highly significant in



every equation. The computed "t" values ($t = b/s_b$), pertaining to the housing index, ranged from a low of 4.29 in equation 3 to 11.24 in equation 6. The critical t value at the 5 percent level was 1.973; at the 1 percent level it was 2.600.

The family size variable was also significant in all eight equations if the 5 percent test level is employed. But at the 1 percent level it failed to show significance in the last two equations. Apparently the effects of family size are not so significant when they relate to less basic expenditures or consumption.

The effects of the three independent variables were able to explain only about 12 percent of the variation in less basic expenditure. But in equation 6, the coefficient of multiple determination, $R^2 = 0.494$, indicates that nearly 50 percent of the total variation in more basic consumption depends upon the effects of the three variables. The R^2 is largest in equation 6; both housing index and family size have their greatest influence in this equation. 1/

An important reason for the housing index coming in so strongly in equation 6 is the high correlation which exists between the index on one hand and the shelter component of more basic consumption on the other. The explanation for the high correlation is found in the way the two factors

---

1/ Throughout this study, the $\bar{R}^2$ (adjusted coefficient of multiple determination) is presented in the tables rather than the unadjusted R^2. Thus, when the amount of variance explained by one regression is compared with the amount explained by another, the $\bar{R}^2$ compensates for any differences in degrees of freedom, i.e., the adjustment in R^2 compensates for the bias resulting from unequal degrees of freedom in the equations compared.

There is an exact functional relationship between R^2 and $\bar{R}^2$ since

$$\bar{R}^2 = 1 - \left[(1 - R^2) \frac{N - 1}{N - P - 1} \right]$$ where

N = number of observations and

P = number of independent variables.

were calculated -- essentially the same aspects were considered in determining both the value of the shelter component and also the housing index.

It is important that we not attach too much significance to the high correlation between housing index and shelter cost. The value of the shelter consumption, after all, comprises only about one-eighth the value of total family consumption. The value of shelter consumption, furthermore, is included in only two of the eight dependent variables -- equations 2 and 6.

Further Investigation of the Money Income Variable

In order to learn more about the money income variable, an analysis parallel to that of Table 3.1 was run on the total sample of 350 families. In this run, the money income had a highly significant effect (significant at the 1 percent level) on all the dependent variables except food consumption. Computed t values were as high as 9.47 in equation 1.

The comparative significance of money income in the total sample and its lack of significance in the subsample of poor families, as well as the signs on the coefficients, suggested that a best fitting income function might be roughly U-shaped rather than linear. This estimate of the situation was largely confirmed when still another set of regressions was run on the 123 families which had reported less than $1,500 income. The coefficients for money income were coming in negative in all the equations, and were in fact significant at the 1 percent level in the four equations predicting the consumption levels -- equations 2, 4, 6, and 8.

It was hard to find a logical explanation that could account for a negative slope of consumption on income. Therefore, further investigation was undertaken to discover the peculiarities which pertained to this particular situation.

A theory has been propounded for some time now that consumption is more nearly a function of long-term average income than of the current income. 1/ The author believes that this theory applies particularly well to the case of farmers. Many of them are plagued by large fluctuations in their year to year incomes due to vagaries of the weather and other causes.

A household by household investigation of the East Texas survey showed that many of the very low incomes could be traced to farming operations which had turned out poorly for the year that was reported, 1958. There was additional evidence that the larger of the negative incomes tended to be associated with some rather sizeable farming operations. It appeared, in fact, that a number of these farm families did not properly belong with a classification of "poor" families at all. Adverse conditions had put them in this class in 1958, but their long time average could, and probably would, place them with a different group entirely.

Here then, is a possible and plausible explanation for the comparatively high consumption levels reported by families with large deficits in their 1958 farming returns. Presumably, they were consuming according to a long-term income potential, and in several cases it was considerably above a so-called low-income level. Nearly all families with incomes approximately in the range of $1,000 to $1,500 were mainly older couples and nonwhite families whose long-term income potential was not very different from the incomes they did report in 1958. Their consumption levels, therefore, reflected more nearly an income potential that one would associate with the genuinely poor.

1/ Milton Friedman's permanent income hypothesis is among the better known of hypotheses along this vein. In his book, A Theory of the Consumption Function, Princeton University Press, 1957, pp. 21-22, he describes a permanent component of income as analogous to the "expected" value of a probability distribution. He also describes a transitory component to be interpreted as all "other" factors which the affected unit would treat somewhat as "accidental" or "chance" occurrence.

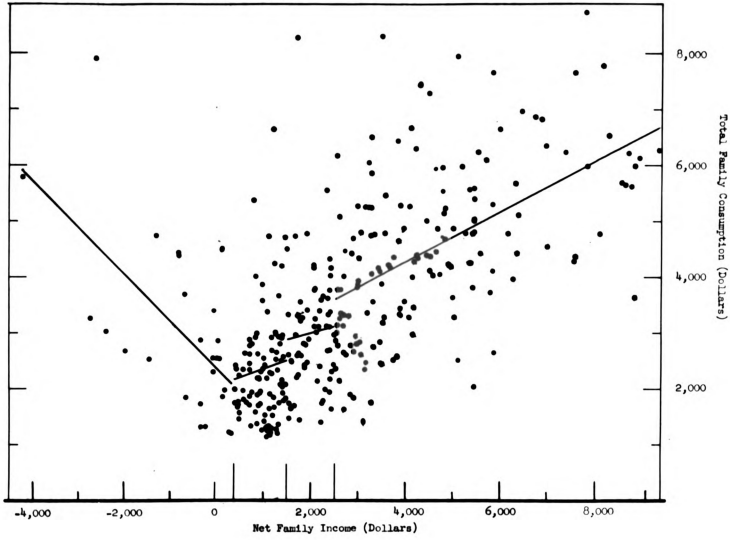
In order to get an overall picture of the way money income was related to consumption in East Texas, a scatter diagram (Figure 3.1) was drawn. Each dot in the chart depicts the total family consumption associated with a particular income level. The household which had nearly $20,000 income was omitted entirely, leaving an income range from -$4,300 to $9,240 among the remaining 349 observations.

In an effort to delineate groups with somewhat similar characteristics, four classifications of income groups were defined as follows: (1) incomes below $400, (2) $400 - $1,499, (3) $1,500 - $2,499, and (4) $2,500 and above. A simple least squares regression line was fitted to each group. An inspection of Figure 3.1 reveals a sharply negative slope in the regression line for the lowest income category, a slightly positive slope in the next two classes, and a fairly significant upward slope in the class $2,500 and above.

It is apparent from Figure 3.1 that if money income were to be treated as a continuous variable, a quadratic rather than a linear function could better capture the curvature that is present. A parabolic function, for instance, would represent a relatively good fit, but this is not the type of function we would expect to fit the more typical case. In a study dealing with rural families that are truly poor people, there may be little or no empirical support for a parabola type function. Certainly, a sample of strictly poor families would not contain any observations like the larger farmers that contributed so greatly to the negatively sloping portion of the regression in Figure 3.1.

Since we do not have a good logical basis for selecting a particular quadratic function, the use of dummy variables represents a very flexible alternative approach for analyzing our income-consumption relationship. This method tends to avoid the biases which result when an incorrect

Figure 3.1--Total family consumption as a function of net family income: Comparisons of functions for 4 income classes, rural families, East Texas, 1958



mathematical function is fitted to a set of statistical data. 1/ The regression coefficient associated with a dummy variable merely reflects the level at which the respective income class is consuming; it does not have reference to the regression slope of consumption on income for that particular class. The level of consumption can therefore be compared among the several income categories, without regard to the mathematical function that would fit best if continuous income, instead of the larger intervals, were used.

It is necessary when dealing with a set of dummy variables to omit one of the subclassifications from the regression equation. The omission of one class prevents the system from becoming overdetermined, and thereby ensures the proper conditions for a determinate estimate of each parameter. The omitted class becomes the base with which other "specified" classes in the same set are compared. Each estimated coefficient represents the difference in the living level between the omitted class, and the class with which the coefficient is associated.

The results from a new set of regressions are given in Table 3.2. In this analysis, money income was run with the employment of dummy variables in the manner described above. The variable selected for omission was the class with incomes below $400. The housing index was again run in linear continuous form as had been done previously. The family size, this time, was entered with a log transformation rather than as actual family size.

The log transformation produced a curvilinear regression which has consumption increasing at a decreasing rate as the household size goes up. Theoretically, this type of function should provide a reasonably good fit in an analysis of low-income families. When there are inadequate resources to devote to family living, as is often the case with poor families, the

1/ See Suits' journal article, Op cit, p. 557, for a more complete discussion in the use of dummy variables to capture curvilinear effects.

Table 3.2--Computed t values in regressions of the level of living on certain household characteristics, low-income rural families, East Texas, 1958

Dependent variable	Independent variables				$\bar{R}^2$
	Continuous variables		Dummy variables		
			Income class a/		
	Housing index	Log family size	$400 to $1,499	$1,500 to $2,499	
(1) Total expenditure.....	7.114	6.694	0.072	2.251	.332
(2) Total consumption	11.235	8.369	-1.433	0.748	.502
(3) Food expenditure	4.543	7.747	1.449	2.911	.298
(4) Food consumption	5.764	7.870	-1.622	-0.040	.329
(5) More basic expenditure ..	6.583	7.883	0.009	1.927	.349
(6) More basic consumption ..	12.093	10.929	-2.178	-0.453	.563
(7) Less basic expenditure ..	4.382	3.267	0.135	1.870	.220
(8) Less basic consumption ..	6.329	2.944	-0.222	1.580	.207

a/ The omitted class was "families with incomes below $400."

$t_{.05} = 1.973$ $t_{.01} = 2.600$

per capita consumption is forced to decrease with an increase in the size of household. Furthermore, it is reasonable to expect that some economies of scale would occur in household operations as new family members are added. The costs of shelter and household utilities ought to rise by less than a constant proportion with increased family size. Some scale economies should also appear in the process of preparing food for larger families. One researcher found the per person cost for food to be about 20 percent less in a 6 person family than in a 2 person family with a given standard of dietary adequacy, and where the effects of income and family composition had been allowed for. 1/

1/ Janet Murray, "Per Person Food Cost Differentials in Large and Small Families," U. S. Department of Agriculture, Family Economics Review, (Sept. 1960), pp. 3-4.

Thus, it appears reasonable to try fitting a cruve whose slope declines with increased size of household. The general form for such a function can be represented by a semilogarithmic equation $Y = a + b\log X$, where Y represents the living level and X represents family size.

The results indicated that log family size did give a better fit than was obtained with the original data. The new form contributed substantially to the larger coefficients of multiple determination found in Table 3.2, compared with respective R^2's of Table 3.1. The housing index also performed considerably better when run in conjunction with log family size and classifications of income (Table 3.2), than when it was combined with linear forms of income and family size as reported in Table 3.1.

The computed t values associated with the income classes showed that the lowest income families (omitted class) were relatively high consumers on an average. The reason for this perverse relationship has already been suggested: a number of the negative incomes were reported by fairly large scale farmers. The t values in several equations indicated that a number of the dummy variable consumption levels were not significantly different (at the 5 percent level) from the level of the omitted class. But in the present case where we appear to have a U-shaped step function, a test of greater interest may be the one based on the null hypothesis: there is no significant difference in the level of consumption between the \$400 - \$1,499 and \$1,500 - \$2,499 classes. 1/ If this null hypothesis can be stated as $B_4 - B_3 = 0$, then the statistic

$$t = \frac{b_4 - b_3}{S_e\sqrt{C_{33} - 2C_{34} \quad C_{44}}}$$

has the t distribution with N-5 degrees of freedom; where b_4, b_3 are estimates of B_4, B_3; C_{33}, C_{44}, and C_{34} are corresponding elements of the

1/ This test was suggested by Dr. R. L. Gustafson, Department of Agricultural Economics, Michigan State University.

inverse moment matrix (properly scaled, of course, so that $S_e \sqrt{C_{ii}}$ is the standard error of the b_i); and S_e is the regression "standard error of estimate."

Notice that with $b_4 > 0$ and $b_3 < 0$, $b_4 - b_3$ is larger than either of the individual coefficients. When, for instance, this significance test was applied in the case of total consumption (equation 2), the computed t value for the null hypothesis $B_4 - B_3 = 0$ was 2.096 and this is significant at the 5 percent level. Ordinary t tests of the separate null hypotheses $B_4 = 0$ and $B_3 = 0$ produced t values of 0.748 and -1.433 respectively, indicating that neither B_4 nor B_3 were significantly different from zero. This demonstrates that it is quite possible for the difference between two income classes to test out significantly even though the individual coefficients do not test out significantly different from zero.

Although the dummy variable specification for money income tests out somewhat better than the linear continuous form, the tentative conclusion is that little correlation exists between the money income variable and the living level for low income rural families. Only among families reporting more than $2,500 income per year, does money income demonstrate considerable predictive power.

The error of observation in reported money incomes also contributes to low predictive power. It is the author's impression from his own experience that family incomes are among the least accurately reported of the usual kinds of data that are collected in a typical sample survey. While there will be a few overstatements of income, the understatements are certainly more common. And, whereas some of the understatements are due to modesty or faulty memory, the majority can be attributed to conscious or inadvertent concealment on the part of the respondent. Another difficulty arises when

the homemaker rather than the family head is interviewed. The homemaker often does not know the correct information relative to family income. 1/

It seems a prudent course, therefore, to exercise caution in evaluating results of analyses relating to family incomes where these incomes have been reported through sample surveys.

A parallel analysis to that of Table 3.2 was also run on the total sample. In this run, the log family size variable did not show an improved fit over that which actual family size had demonstrated. It should be recalled that in the subsample with incomes under $2,500, the transformed variable had shown a substantial improvement over original data. The reason for this difference can probably be attributed to different levels of resources available for family living. The large families of the top income class, usually had sufficient means to maintain a more or less stable per capita living level. But the large families among the poor, were often forced by their limited buying power to follow a declining per capita living level.

Larger families in the top income class were especially better fed and better housed than those from the three lower income classes. In the class with incomes above $2,500 and with 6 or more members per household, the average value of food consumed was $1,880; the average value placed on their shelter consumption was $422. The 27 families in this class averaged 7.2 members.

1/ In making these observations concerning the accuracy of incomes reported in sample surveys, the author is influenced primarily by his own experiences as an interviewer and also as a supervisor of interviewers. However, other authors have referred to this problem especially as it pertains to the underreporting of incomes. See, for instance: Bureau of Labor Statistics, Family Income, Expenditures, and Savings in 1950, Bulletin No. 1097 (Revised) pp. 7-8; Office of Statistical Standards, Statistical Evaluation Report No. 5, December 1964, pp. 7-10; Department of Agriculture, Report to the 38th Annual Outlook Conference given by Laura Mae Webb, pp. 1-2.

In the class with incomes below $2,500 and with 5 or more members in the household, the average value of food consumed was $1,542, and the average value placed on their shelter consumption was only $177. The 18 families in this group averaged 8.1 members.

Phase II -- Additional Explanatory Factors Specified in the Regressions

This phase of the analysis permits several additional independent variables to be included in the multivariate regressions. Taking extra variables into explicit account removes them from the unspecified category, thereby reducing the residual variance and increasing the variance explained by the regression effects. This process also reduces the biases which are always present so long as any variables correlated with both the dependent variable and at least one of the independent variables, remain within the unspecified group.

In addition to the three independent variables considered in the Phase I analysis, the following characteristics and their subclassifications were specified for a new set of regressions:

1. Chief source of income
   - (a) Wages and salaries
   - (b) Professions or business
   - (c) Transfer payments, rent, and interest
   - (d) Farm sources
2. Age of homemaker
   - (a) Under 40
   - (b) 40 to 54
   - (c) 55 and older
3. Education of the homemaker
   - (a) Less than 8th grade
   - (b) Grades 8 through 11
   - (c) Grade 12 and higher

4. Tenure of home by farm status

   (a) Owner farm

   (b) Renter farm

   (c) Owner nonfarm

   (d) Renter nonfarm

5. Race

   (a) White

   (b) Nonwhite

The reason for selecting these particular characteristics and these particular subclasses were discussed in Chapter II.

Dummy variables were defined on each subclass named under the above household characteristics, and then placed in a regression along with money income, housing index, and family size. The latter three variables were again entered in linear continuous form for the first set of analyses under Phase II.

Table 3.3 gives computed t values for all the variables that were specified and brought into the equations. Since the three continuous variables were forced and held in the equation, there is a t reading on each of them regardless of the influence they had on the dependent variable. The categorical variables, however, were brought in (one at a time) by a stepwise regression procedure, and only those variables entered the equation which represented a significant improvement in the goodness of fit of the regression.

If a classification variable enters the regression, it means that consumption of families in this particular classification is significantly different from families in other classifications (of the same dummy variable set) which did not enter the regression equation.

Depen varial	
Total expend	
Total consum	
Food expend	
Food consum	
More t expend	
More t consum	
Less t expend	
Less t consum	

$t_{.05}$

Table 3.3--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, East Texas, 1958

Dependent variables	Independent variables: Continuous variables: Money income	Housing index	Family size	Dummy variables: Chief source of income: Wages or salaries	Professions or business	Transfer payments, rent, or interest	Farm	Age of homemaker: Under 40	40 to 54	55 and over	Education of homemaker: Less than 8th grade	Grades 8 thru 11	Grade 12 and above	Tenure of home by farm-nonfarm status: Owner farm	Renter farm	Owner nonfarm	Renter nonfarm	Race: White	Nonwhite	$\bar{R}^2$
Total expenditure	-0.176	6.701	4.289	4.271									4.161							.348
Total consumption	-0.363	10.253	6.097			-4.016							3.974							.493
Food expenditure	0.707	3.297	7.154	3.344								-2.198					2.366		-2.824	.324
Food consumption	0.379	3.606	7.213			-2.366										-3.480			-2.613	.345
More basic expenditure	1.753	5.721	5.735			-3.345							4.050							.355
More basic consumption	-0.139	10.907	8.686			-3.185							2.798			-3.059				.548
Less basic expenditure	-0.104	5.010	1.094	3.950									2.670							.196
Less basic consumption	-1.442	6.121	1.292	4.258									3.900							.275

$t_{.05} = 1.973$ $t_{.01} = 2.600$

The classifications for which there is a t reading given in Table 3.3 were therefore the only categorical variables which were a part of the final multivariate regression reported in the table. The 5 percent significance level was the critical level designated both for entering variables, and also for removing any variables which had dropped below significance as other variables were entered.

The entry of these particular categorical variables caused a downward adjustment in the t values for the three continuous variables (Table 3.3 cf. Table 3.1). The adjustments in the t values occurred because the intercorrelations among the continuous and new categorical variables caused adjustments in the regression coefficients and also in the standard errors of these coefficients. The new parameter estimates retained fewer biases which were there because of the (previously hidden) correlations that were brought out by the more inclusively specified regressions.

When all the adjustments had been reflected in the final equations, the money income variable showed less than a significant effect on any of the 8 dependent variables. The influence of housing index and family size continued highly significant, however. Only in its effects on the less basic items of living, was family size relatively insignificant as an explanatory factor.

In each of the 8 equations at least one of the classifications delineating the chief sources of family income entered the multivariate regression. Families whose chief income was from wages and salaries were identified with a higher living level than the remainder of low-income families. This was true in four of the equations. As a counterpart in the remaining four equations, the recipients of transfer payments, rent, and interest were identified with a lower level of living than was enjoyed by the rest of the families.

None of the "age of homemaker" categories entered any of the equations at the 5 percent significance level. But there was a relatively high correlation between families with an older homemaker, and families whose chief income was from transfer payments, rent, and interest. This "source of income" category, therefore, contributed to an attenuation of the effect of the "age of homemaker" category and vice versa. The result is that only one of the two classifications comes in as a significant variable.

When the same Phase II variables were run using the total sample, an age of homemaker class entered 5 of the 8 equations. In each of the 5 instances, a chief source of income classification had also entered and both remained as significant variables in the final equation. It is probable, therefore, that if chief source of income were omitted entirely as an explanatory factor, the category "homemaker 55 and older" would come into all the equations as a significant variable..

The homemaker's education was another factor that contributed substantially to explaining variability in the level of living. Families whose homemaker had a grade 12 or higher education spent and consumed at a substantially higher level than those headed by lesser educated homemakers. Only when predicting total food consumption did none of the education categories enter at the 5 percent probability level.

Grades 8 through 11 homemakers appear to have spent significantly less on food than their better and lesser educated counterparts. The author was not able to find a suitable explanation for this. Neither was there a satisfactory explanation for the "owners nonfarm" to be consuming food and other more basic items at a lower level than families in the other classifications. Chance may have been the main factor in producing these results, but the difficulty in explaining these particular outcomes suggested the need for specifying an additional variable--one that would

help account for more of the variations that relate to food purchases and consumption. The proposed new variable, a measure of home produced food, is discussed more fully in Phase III of this chapter where it is first specified in the regressions.

The tenure and farm status classification did not demonstrate an identifiable pattern of behavior when only Table 3.3 results were considered. But with the help of two other separate analyses (one with the total sample; another with families whose incomes were $400 to $1,499) a pattern did seem to emerge. The pattern that emerged was that farmers spend less on the more basic living items, and also less on total expenditures, but they tend to consume more food than do nonfarmers.

These results lend support for the practice employed above, which combines home tenure with farm-nonfarm status, when both these traits are specified in the regression. If instead of the joint variable we were to specify owners vs. renters only, we could expect distortions produced when an interaction effect caused the regression outcome to be dependent on the farm-nonfarm makeup of the sample. An increased proportion of farmers would lower the per family amount spent on several items of living and raise the amount of food consumed. Increasing the proportion of nonfarmers would have the opposite effect. The difference in the living levels between the owner class and the renter class could therefore be great or small, depending on the proportion of farmers to nonfarmers that comprise each class.

Hidden interactions are discussed and analyzed more fully in the next phase of this chapter. In that part of the analysis another interaction was discovered involving owners and renters, but related this time to the housing index.

Race is the last variable that was specified in the regressions reported in Table 3.3. Nonwhites differ significantly from whites only

with respect to the food variable, i.e., nonwhites purchased less food and they also consumed less food. But in a similar analysis of the total sample, nonwhites were shown to purchase and consume not only less food, but less of the more basic items generally. A more detailed investigation indicated, however, that it was only the food and shelter items (rather than clothing) of which the white families used substantially more. The clothing purchases by nonwhites, in fact, actually exceeded those of the whites.

The $400 - $1,499 Income Class Analyzed

When the 96 families with incomes $400 to $1,499 were analyzed as a separate group, even the food variable no longer showed a significant difference between the races. The reason appeared to be that the whites of this small subsample were mostly older couples who like the nonwhites had a relatively low living level. These two classifications actually make up the bulk of the sample -- thirty-three of the 96 families were nonwhite, and 55 out of 96 had homemakers in the top age class.

The results of the eight final regressions on families with incomes $400 to $1,499 are given in Table 3.4. One category of each of the characteristics, "chief source of income" and "education of homemaker" again entered the regression of this smaller subsample of low income families. Tenure of home by farm status continued to make little contribution to explained variance. The race variable showed significance only with respect to less basic expenditures; but this time, interestingly, the nonwhites spent at a higher level than the whites.

A detailed investigation showed that nonwhite families of this small subsample had outspent the white families in several less basic categories including house furnishings and equipment, transportation, personal care,

Table 3.4--Computed t values in regressions of expenditure or consumption on household characteristics, low-income rural families with income $400 to $1,500, East Texas, 1958

Dependent variables	Independent variables: Continuous variables: Money income	Housing index	Family size	Dummy variables: Chief source of income: Wages or salaries	Professions or business	Transfer payments, rent, or interest	Farm	Age of homemaker: Under 40	40 to 54	55 and over	Education of homemaker: Less than 8th grade	Grades 8 thru 11	Grade 12 and above	Tenure of home by farm-nonfarm status: Owner farm	Renter farm	Owner nonfarm	Renter nonfarm	Race: White	Nonwhite	$\bar{R}^2$
Total expenditure	2.934	4.996	2.456			-4.392							3.017							.430
Total consumption	2.185	8.573	3.657			-4.981							4.412							.615
Food expenditure	2.775	4.356	4.397			-4.203						-3.364								.397
Food consumption	1.923	3.101	3.978			-4.135								2.996						.367
More basic expenditure	3.179	5.648	4.416			-3.900						-2.883								.433
More basic consumption	2.405	9.115	5.668			-4.695							2.559			-3.048				.636
Less basic expenditure	1.837	4.004	-0.480			-3.389							2.754					-2.174		.268
Less basic consumption	1.022	4.903	0.197	3.197									4.328							.375

$t_{.05} = 1.973$ $t_{.01} = 2.600$

recreation, and education. It is understandable, however, that whites should have spent less on these items because the white family component consisted mostly of families with older homemakers. These older families were small spenders on both the basic and nonbasic items.

The analysis of the families in the $400 to $1,499 income class affords a good illustration of some of the added insights which a stepwise regression can provide. The final equation of this procedure is reported in Table 3.4 for the present sample. But the stepwise method also points out three or four intermediate equations in addition to the final equation results. One can learn certain information from the intermediate equations which is not apparent from a study of the final equation alone.

The results found in Table 3.4 would indicate that money income contributes substantially to explaining variance in the dependent variables. But this same money income variable is shown to be less significant as an explanatory factor in the first intermediate equation where only housing index and family size have been controlled. It is the entry of the transfer payments, rent, and interest category which gives a sizeable boost to the money income coefficient in all eight equations.

In the stepwise analysis that has been employed for studying the present problem, the "transfer payment, rent, and interest" class was the first of the categorical variables to enter after the continuous variables had been forced to enter the equation. The second intermediate equation, therefore, gives parameters that were estimated in a regression where money income, housing index, family size, and a source of income classification were the independent variables.

Table 3.5 gives computed t values pertaining to money income, before and after the "source of income" variable was part of the regression. The table also gives $\bar{R}^2$'s for the first and second intermediate equations.

Table 3.5--Computed t values pertaining to the money income variable and $\bar{R}^2$'s for explained regression before and after the inclusion of a source of income variable, rural East Texas families with incomes $400 to $1,499 in 1958

Dependent variable	Computed t value for money income and $\bar{R}^2$			
	First intermediate equation a/		Second intermediate equation b/	
	t value	$\bar{R}^2$	t value	$\bar{R}^2$
Total expenditure	1.721	.286	2.727	.380
Total consumption	0.854	.460	1.874	.536
Food expenditure	1.540	.241	2.466	.329
Food consumption	0.693	.211	1.644	.311
More basic expenditure	2.032	.316	2.917	.388
More basic consumption	0.738	.498	1.828	.578

a/ The specified independent variables were money income, housing index, and family size.

b/ The specified independent variables were money income, housing index, family size, and the classification of families whose chief source of income was from transfer payments, rent, and interest.

The evidence has been that families with chief income from transfer payments, rent, and interest live on a considerably lower plane than the average of other classes. Furthermore, the families with this particular source of income are found more frequently near the $1,499 level of income than down near the $400 level. Thus, after this classification of families has been allowed for and controlled, there is an increase in the slope of the money income function.

A directly opposite effect was obtained when the entire East Texas sample was analyzed, using the same variables. The entry of the same

"source of income classification" into the regression, caused the money income coefficient to be adjusted downward instead of upward as in the previous case. The explanation is found once again in the frequency distribution of the various source of income categories. Families with incomes chiefly from transfer payments, rent, and interest were now concentrated more heaving near the lower end of the overall income distribution. Consequently, when their classification entered the regression, the money income coefficient was adjusted downward.

The opposite adjustments to the money income coefficient, when "source of income" entered, would probably not have been discovered if the print out of the intermediate equations had not been available for each analysis. The availability of the intermediate analysis also makes it possible to determine the amount that each entering variable contributes toward total explained variance.

The analytic value of the stepwise regression technique is further illustrated in Phase III where it proves very helpful in locating the hidden interactions.

The employment of stepwise regression, however, is not without certain drawbacks. For example, the procedure almost guarantees that some biases due to high correlations among independent variables will not be discovered. Thus, in the equation

$$Y = B_o + B_1X_1 + \ldots + B_iX_i + \ldots + B_jX_j + \ldots + B_pX_p$$

(a) if $B_i = 0$, $B_j = 0$ and X_i, X_j are highly positively correlated, the stepwise procedure will typically pick up one, not the other, and attribute the entire influence of both to the one. The ordinary regression procedure, where both variables are forced to come into the regression, would reveal the true situation.

(b) if $B_i = 0$, $B_j = 0$, and X_i, X_j are highly positively correlated, the stepwise procedure would typically pick out neither one. Ordinary regression again would generally reveal the situation.

In the present study, an example of (a) above is found in the case where "age of homemaker" classes are closely correlated with "chief source of income" classes. Thus, in the Phase II analysis of this chapter, a "source of income" class entered each of the equations and thereby helped to prevent any "age of homemaker" classes from also coming in. In Phase II of the next chapter, it will be observed that the opposite phenomenon occurred: age of homemaker classes came into the regression equations and prevented the chief source of income classes from entering as well.

The stepwise procedure is especially inadequate as a test to see if an entire set of dummy variables makes a significant contribution to variance reduction. Since the process selects only individual classes from a set, the individual coefficient t-tests (H_o: $B_i = 0$) can indicate nonsignificance even if the set (jointly) were to add significantly to variance reduction. Hence, when it is desired to test if a family characteristic (represented by a set of dummy variables) has a significant influence, it is necessary to test the null hypothesis that the set of coefficients are all (jointly) equal to zero. The test is straightforward and may be found in many standard references.

Phase III -- Search for Interaction Effects

Additional insights into family living behavior are obtained by deliberately searching for interaction effects among the independent variables. The meaning of interaction between two variables can be better understood if we examine the relationship between owner-renter status and farm-nonfarm status, as the two factors relate to the level of food consumption.

Table 3.5 shows food consumption as a function of home tenure and farm status. The dollar values in each cell are the average 1958 food

Table 3.5--Value of food consumption as a function of home tenure and farm status, low-income rural families, East Texas, 1958 (Figures in parentheses are standard errors)

Farm status	Owner	Renter
Farm	$1,142 (55)	$1,084 (101)
Nonfarm	$ 798 (57)	$1,050 (112)

consumption for the sample of families falling into a particular cell. In order to see the effects of tenure, compare the consumption of families who have the same farm status but have a different tenure. Conversely, to see the effect of farm status, compare families belonging to the same tenure class but different farm classifications. The table shows that food consumption is correlated with farm status: nonfarm families consume less food than farm families. But this effect differs between owners and renters; whereas farmers consume approximately equal amounts of food whether they rent or own, the nonfarm owners consume substantially less food than the nonfarm renters.

When the effect of one independent variable depends on the value of another, the two independent variables are said to interact. 1/ Thus, home tenure and farm status interact in their effects on food consumption.

The effects of two variables which interact are said to be "nonadditive", i.e., the influence of a change in one cannot be added directly to the effect of a change in the other. To get at the total effect, a third component, interaction effect, has to be brought into the analysis.

1/ See Suits' discussion of interaction in his book, Op cit, pp. 83-85.

The presence of interaction between tenure and farm status seems evident from an inspection of the cell means of Table 3.5. However, this case could represent a chance result generated by random behavior of non-interacting variables. It is necessary, therefore, to use a significance test to calculate the risk that chance variation may have produced the results observed.

Search by Variance Analysis

One way of testing for interaction by statistical methods is to employ a variance analysis in which the effects of two factors are studied simultaneously. In the present case, the sums of squares for rows, columns, and interaction are computed by treating each mean in Table 3.5 as a single observation. The computing procedure yields the variance analysis of Table 3.6.

Table 3.6--Analysis of variance of the data from Table 3.5

Source of variation	Sum of squares	d.f.	Variance	F	F.05	F.01
Farm status	35,664	1	35,664	5.55	3.89	6.78
Tenure	8,497	1	8,489	1.32	3.89	6.78
Interaction	25,110	1	25,110	3.91	3.89	6.78
Error		185	6,423			

The analysis indicates a significantly different level of food consumption, at the 5 percent level, between families with a different farm status, but the effect of tenure is apparently insignificant. Interaction effect is found to be just barely significant at the 5 percent level.

It should be pointed out that the cell comprised of owner-nonfarm families, which had consumed substantially less food than the groups of

the other 3 cells, contained a larger than proportionate number of nonwhite families. Earlier analysis had shown nonwhites consumed significantly less food than did the white families. It is therefore considered meaningless to attribute any net food consumption effects to race, home ownership, or farm status since there are substantial intercorrelations and interactions among these three factors.

Morgan and Sonquist indicate that if intercorrelations exist among the explanatory variables, it is difficult to assess the relative importance of different factors because their intercorrelations get in the way. 1/ Hanan Selvin finds that the level of significance is frequently given an erroneous positive interpretation . . . "To cite the level of significance without first being certain that the observed difference could not have been created by correlated biases is to convey the impression that the results are highly precise when, in fact, they may not be." 2/

Another case where there was thought to be interaction, is between education and age categories of the family homemaker. It was hypothesized that age and education interact in their effect on expenditures for less basic living items like transportation expenses and purchases of household furniture and equipment. Using the age and education classifications defined earlier, averages for less basic expenditures were calculated and recorded in the 9 cells of Table 3.7. The standard errors of the means and the number of observations for each cell are also indicated in the table. The great inequality of cell frequencies found here, leads to a heterogeneity of variances and consequently the F test employed in the

1/ Morgan and Sonquist, Op cit, p. 416.

2/ Hanan C. Selvin, "A Critique of Tests of Significance in Survey Research," American Sociological Review, Vol. 22 (October 1957) p. 523.

analysis of these data must be regarded as nonconservative, i.e., the hypothesis may be rejected too readily at a given level of statistical significance.

Table 3.7--Expenditures on less basic items of family living as a function of age and education of homemakers, low-income rural families, East Texas, 1958 (Figures in parentheses are standard errors)

Education	Age		
	Less than 40	40 to 54	55 and above
Less than 8th grade	N = 9 $441 (96)	N = 19 $867 (135)	N = 46 $578 (45)
Grades 8 through 11	N = 19 $878 (115)	N = 33 $757 (100)	N = 22 $632 (67)
Grade 12 and above	N = 15 $991 (127)	N = 12 $1190 (254)	N = 14 $785 (97)

The means in Table 3.7 suggest an interaction between education and age in a number of ways. For example, the younger homemakers with grades 8 through 11, spent double as much as their lesser educated counterparts; yet among older homemakers, the higher education had produced little additional spending. Among the middle age group, spending actually declined for the grade 8-11 group. If the spending of homemakers with grade 12 or above is compared with the spending in the next lower education tier, the biggest difference by far, is among the homemakers that are between 40 and 54 years of age.

It should be noted, however, that the standard errors of the means in Table 3.7 are rather large, especially among the younger and middle aged groups. In view of these large standard errors, we need a more precise evaluation of whether an interaction exists between education and age.

Such an evaluation is obtained through the variance analysis that has been calculated in Table 3.8. This analysis produces strong evidence that both an age effect and an education effect actually exist, but any evidence of interaction between the two is rather weak, i.e., it shows insignificance at the 5 percent level.

Table 3.8--Analysis of variance of the data from Table 3.7

Source of variation	Sum of squares	d.f.	Variance	F	F.05	F.01
Education	200,298	2	100,147	7.24	3.04	4.71
Age	113,837	2	56,919	4.12	3.04	4.71
Interaction	93,263	4	23,316	1.69	2.41	3.41
Error		180	13,834			

Search by Multivariate Techniques

Still another way to discover hidden interactions among independent variables is to identify subclasses of families whose living level behavior differs from other family groups with respect to some factor. The method used to identify these classes, is to specify so-called interaction terms in regression runs. 1/ Each interaction term consists of a continuous variable multiplied by a categorical variable. These combination predictors are then run along with the simple variables which are the components of the interaction terms. If a particular interaction term does demonstrate a significant effect, it is an indication the family group identified with this term has a living pattern that differs from that of other family groups.

It helps to use an example for demonstrating the significance of an interaction term coming into the equation. Suppose Y were family consumption, X_1 were money income, and X_2, X_3, X_4 represented three categories of

1/ Suits describes this technique in his journal article op cit, p. 550.

families headed by homemakers under 40, 40 to 54, 55 and over, respectively. Then suppose that the interaction term X_1X_2 entered the multivariate regression as a significant variable. This would mean that the consumption function for families with homemakers under 40 had a significantly different slope from the consumption function of families with homemakers 40 and older. From this it would follow that a significant interaction existed between money income and the age of the homemaker, as the two factors influence family consumption.

Then suppose X_1X_3 entered the equation and that X_1X_2 continued as a significant variable. The meaning of this combined occurrence, is that the slopes of the functions of both family groups (one group headed by a younger, another by a middle aged homemaker) differed significantly from the slope of the consumption function for the third category of families, namely the one headed by an older homemaker.

The latter class of families would now perform the role of the single class which is omitted from a set, if the regression system is not to become overdetermined. The consequences of overdeterminacy were discussed in Phase I of this chapter.

As the class which is omitted, the families of the older homemakers would become the "base" class with which the two other "specified" classes are compared. The coefficient estimated for each specified class, represents the difference in the slopes of the consumption function between the specified class and the omitted class. If the estimated coefficient is significant and negative, it means that the slope of the consumption function for the specified class is significantly smaller than for the class that was omitted. If the coefficient is positive, the reverse situation holds true.

The basic variables proposed for the Phase III analysis are essentially the same as those which were run in the regression of Phase II. One new

set of dummy variables was added, however. The characteristic "percentage of food home produced" was determined for each sample family and the families were then classified into 4 categories: those producing (1) 0 to 9.9 percent; (2) 10 to 29.9 percent; (3) 30 to 49.9 percent; (4) 50 percent and over, of all the food they consumed.

All the interaction terms introduced, were a combination of a continuous variable (either housing index or log family size) and a dummy variable (a subclass of one of several family characteristics). The purpose here was to test if interaction existed between housing and family size on the one hand, and certain categorical family characteristics on the other. If more computer capacity had been available, it might have been worthwhile to test also for interaction between (continuous) money income and dummy classifications. Additionally, with more machine capacity, one could have tested for interactions among certain classification variables themselves.

The stepwise regression procedure was used, and interaction terms as well as any of the categorical variables were permitted into the equation only if their contribution to explained variance was significant at the 5 percent level. However, housing index, log family size, and two classifications of money income were again forced into the equation, regardless of whether their contribution to regression effects was great or small.

The final regression results for each of 8 dependent variables are given in Table 3.9. Note that only those interaction terms which turned out to be significant in one or more equations are listed in the independent variables column. However, all other combinations of continuous x dummy variables were specified, except the "log family size x income categories" and "housing scale x home produced food categories."

One characteristic which interacts importantly with log family size is "chief source of income." The t values in Table 3.9 indicate that with either "less basic living items" or with "total all items" as dependent variables, the "wages, salaries, professions or business" function had a relatively greater slope with respect to family size. When this class of families with its steeper function was controlled, i.e., when the interaction term X_2X_6 had entered the equation, the coefficient of the log family size variable was reduced sharply. It appears that the families whose chief income was from wages, salaries, professions or business, had more adequate means with which to provide their larger families. Thus, when this group had been allowed for in the regression, the living level function for the remaining families showed considerably less correlation with family size.

The interaction of family size with (a) chief source of family income and (b) percentage of food home produced, is illustrated by the successive adjustments of the log family size coefficient, as new variables were added to the regression. For instance, with total consumption as the dependent variable, (equation 2) the interaction term X_2X_{14} was the first of the nonforced variables to enter the regression. This resulted in control of family size effects on consumption, for those families which produced 30 to 49.9 percent of their food. This class of families tended to include owners of the more substantial farms and also of the better homes. As a result of the control of their effect upon consumption, the t value associated with log family size was reduced from 8.369 to 7.351. A further reduction in t to 6.930 occurred when the next classification "families with homemakers grade 12 or higher" entered the equation. Then followed the entry X_2X_6, causing a further reduction in t associated with log family size to 4.184.

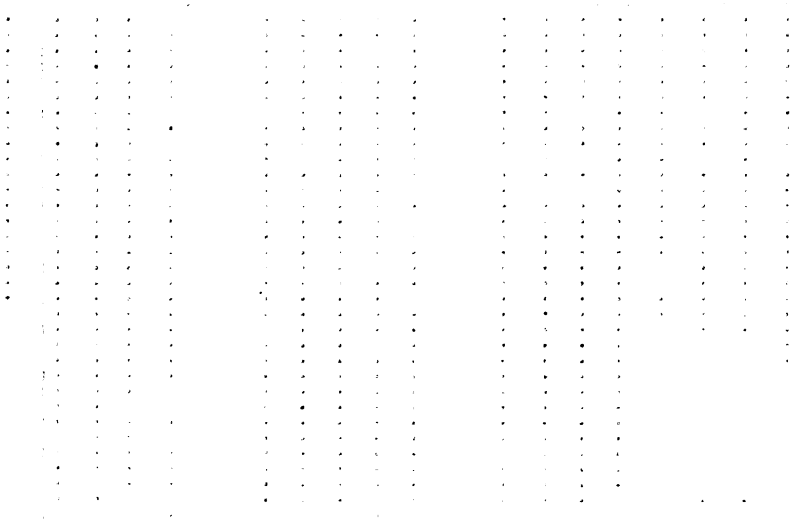
Table 3.9--Computed t values in regressions of family levels of living on household characteristics, low-income rural families East Texas, 1958

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_1 - Housing index	6.778	10.042	2.770	3.726	1.260	10.286	5.389	4.428
X_2 - Log family size	4.698	4.658	7.822	10.813	6.867	11.412	0.157	-0.693
Income Categories								
X_3 - Below $400	omitted variable							
X_4 - $400 to $1,499	-1.363	-1.899	1.356	-0.171	-0.253	-0.826	-0.631	-0.989
X_5 - $1,500 to $2,499	0.072	-0.354	2.082	1.234	1.245	0.763	0.577	0.004
X_1X_3								2.131
Chief Source of Income								
X_6 - Wages, salaries, professions, business								
X_7 - Transfer payments, rent, and interest			-2.052					
X_8 - Farm income								
X_1X_7	-2.103	-2.093						
X_2X_6	3.665	3.668					3.680	3.841
Education of Homemaker								
X_9 - Less than 8th grade								
X_{10} - Grades 8 through 11								
X_{11} - Grade 12 and higher	2.620	3.316						
X_2X_{11}				2.983	2.619			

Table 3.9--Computed t values in regressions of family levels of living on household characteristics, low-income rural familie East Texas, 1958 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumptio
Percent of Food Home Produced								
X_{12} - 0 to 9.9 percent				-4.173			2.289	
X_{13} - 10 to 29.9 percent								
X_{14} - 30 to 49.9 percent								
X_{15} - 50 percent and over								
X_2X_{12}			3.476			-4.079		
X_2X_{13}	-3.053	-2.216		-5.104	-2.106	-5.163		
X_2X_{14}		3.255					5.548	2.984
X_2X_{15}	-4.529		-3.844		-4.730			
Home Tenure								
X_{16} - Owner								
X_{17} - Renter								
X_1X_{16}					2.287			
X_1X_{17}		-1.973				-3.128		
X_2X_{16}				-2.499				
Race								
X_{18} - White								
X_{19} - Nonwhite								
X_2X_{18}					2.139	2.390		
X_2X_{19}			-3.426	-3.416				
$\bar{R}^2$	.454	.577	.437	.449	.467	.641	.255	.317

$t_{.05} = 1.973$ $t_{.01} = 2.600$



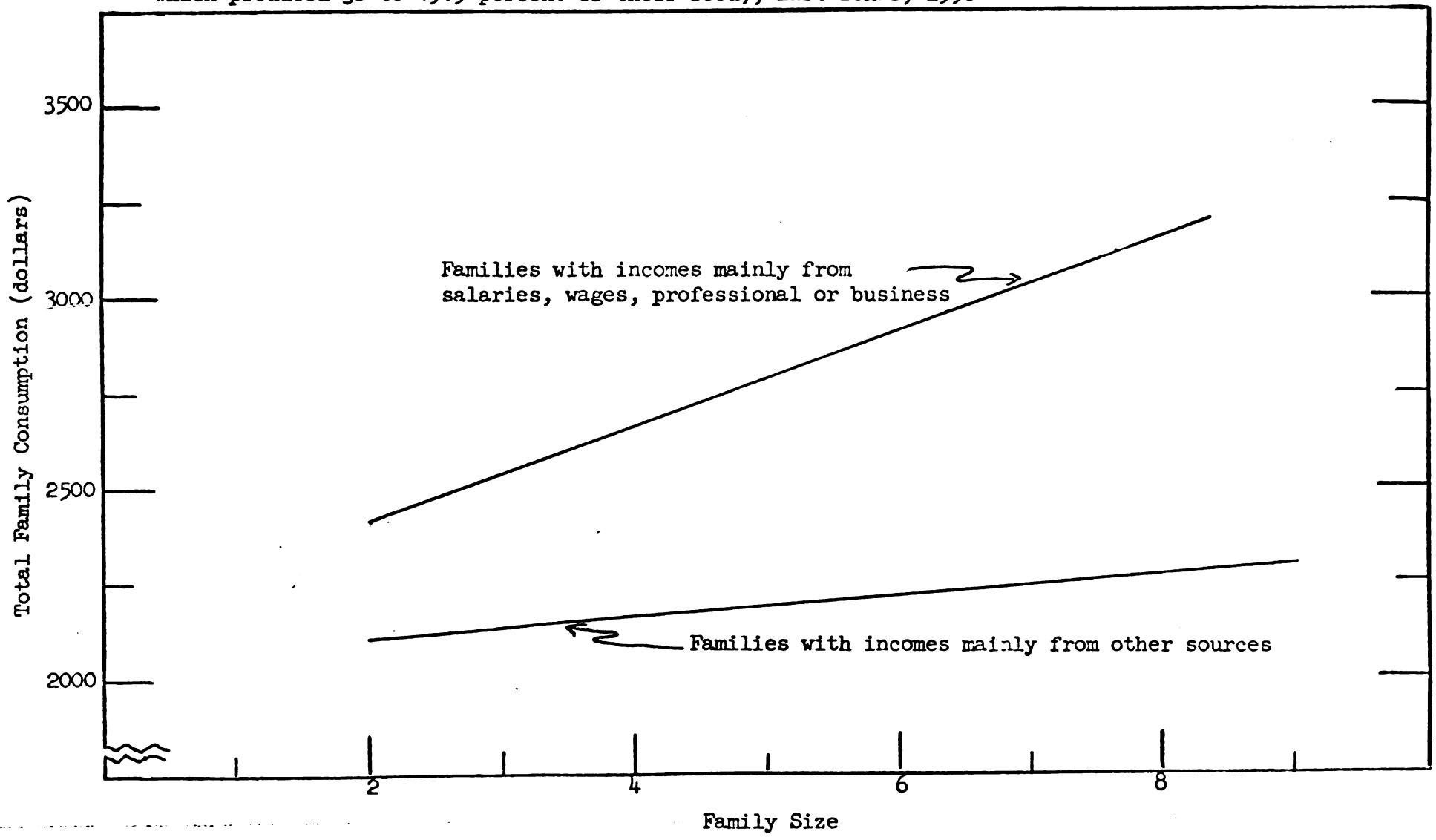
The last adjustment in t was a rather large one, and it suggests the existence of a significantly closer correlation between total consumption and log family size if the families being analyzed received their incomes mainly from wages, salaries, professions or business, rather than from some other source. Figure 3.2 illustrates in a diagram that consumption of the two groups differ substantially with respect to family size.

Families with homemakers with grade 12 education and above, and families which produced 30 to 49.9 percent of their food, were removed from the sample of 189 low-income families. The remainder were then classified into two groups: (1) those with chief source of income from wages, salaries, professions or business (52 families), and (2) those with chief income from other sources (67 families). One family with 17 members was omitted entirely. It was substantially larger than any other family, so it was omitted to make the Figure 3.2 diagram more compact.

The two functions appearing in Figure 3.2 represent simple regression lines for two classes of families. The divergence in the slopes of the two functions indicates an interaction between family size and source of family income. It is apparent that one single analysis which combined the two groups, would necessarily produce a weighted compromise effect of the family size influence. The slope of the function for the two groups combined would fit a resultant function falling somewhere between the two regression lines shown in Figure 3.2.

In a similar way, a compromise slope with respect to the family size variable would be obtained if white families and nonwhite families were considered together and no allowance were made for interaction effects. To illustrate this fact, we use food expenditures as the dependent variable (equation 3) and again examine the adjustment taking place in the coefficient of log family size as new variables are added. After the forced variables

Figure 3.2--Total family consumption as a function of family size: Comparisons of functions for two family classes, low income rural families with homemakers having less than grade 12 education (except families which produced 30 to 49.9 percent of their food), East Texas, 1958

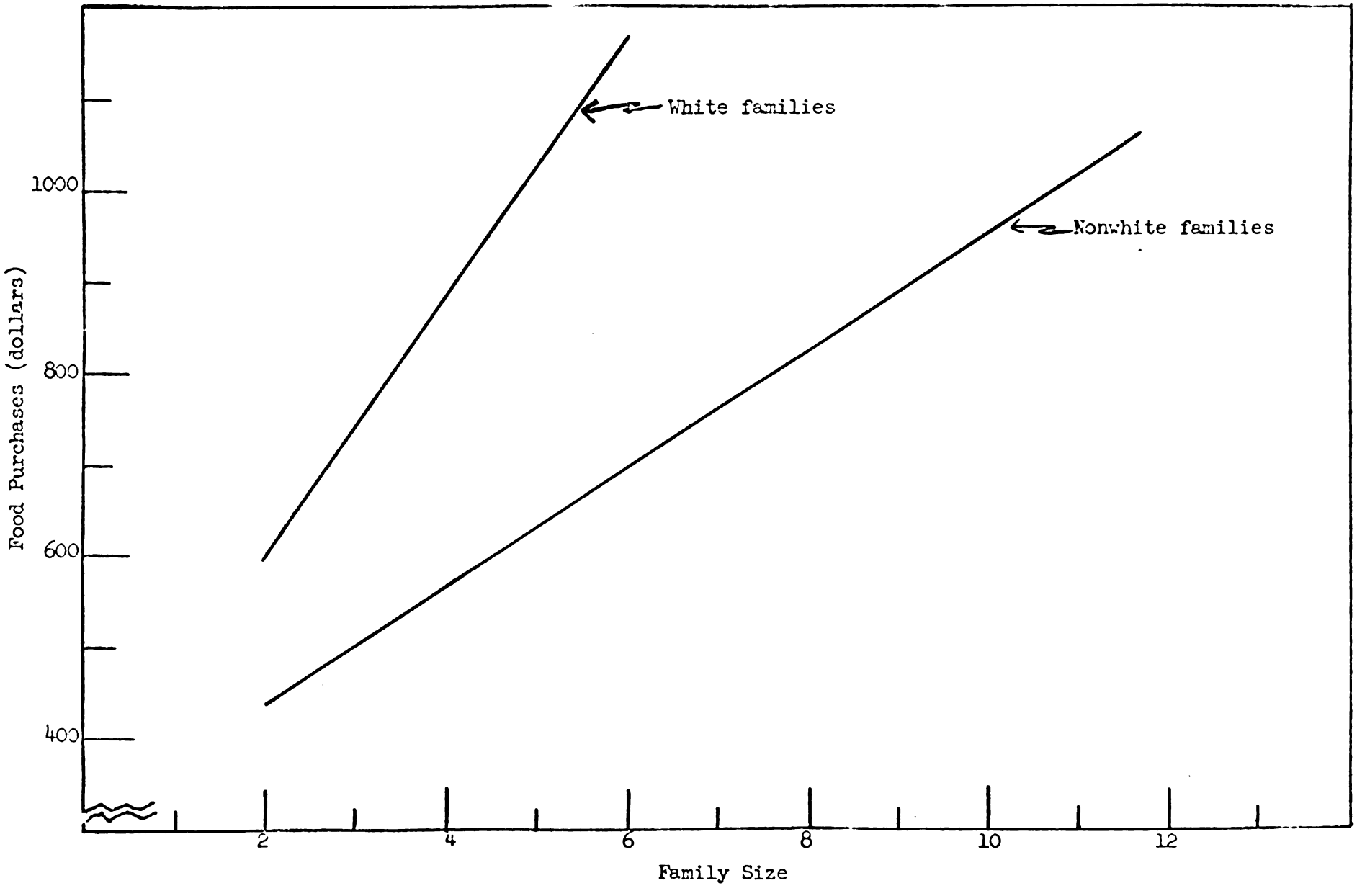


had been allowed for in the equation, there followed the two interaction terms X_2X_{12} and X_2X_{15}. This amounted to a control of the family size effects for two classes of families: (a) the class producing less than 10 percent of their food, and (b) the class producing 50 percent or more. The next variable to enter was X_2X_{19} (log family size x nonwhite) and the effect of this interaction term was to increase the coefficient of log family size by one-third. This means that when family size effects of nonwhites had been controlled, the log family size variable, now pertaining to white families only, assumed greater significance. In other words, when those families are considered which produce between 10 to 50 percent of their food, the white families demonstrate a significantly greater correlation between their food purchases and their family size.

Figure 3.3 shows two simple regression lines for food purchases on family size -- one representing whites (57 families); the other nonwhites (36 families). The difference in slopes of the two functions suggest that family size has considerably more value as a predictor of food purchases for white families than it has for nonwhite families.

The foregoing analysis has provided evidence that family size interacts significantly with race, chief source of income, and the proportion of food home produced. But there is considerable intercorrelation among the latter three characteristics. For example, the class of families that home produced 10 to 29.9 percent of their food, had considerably more than a proportionate number of nonwhites among them; the 30 to 49.9 percent class had a greater than proportionate number of farmers; the 50 percent and over class, had a greater than proportionate number of smaller farmers and families who had received most of their income from transfer payments, rent, and interest.

Figure 3.3--Food purchases as a function of family size: Comparisons of two family classes, low income rural families which produced 10 to 49.9 percent of their food, East Texas, 1958



It is difficult, therefore, to assess the relative importance of each of the 3 sets of classifications as they relate to family size and the level of living. However, it is unlikely that the decision to produce more or less food at home, had much to do with the size of their families. When "percentage of home produced food" and "family size" were ranked and then related in a correlation analysis, it was found there was not a significant correlation between them. A Spearman's rank correlation coefficient that was computed came to 0.026. More likely, the farm and ownership status of the family helped to determine the amount of food home produced. These factors, in turn, are correlated with race and chief source of income.

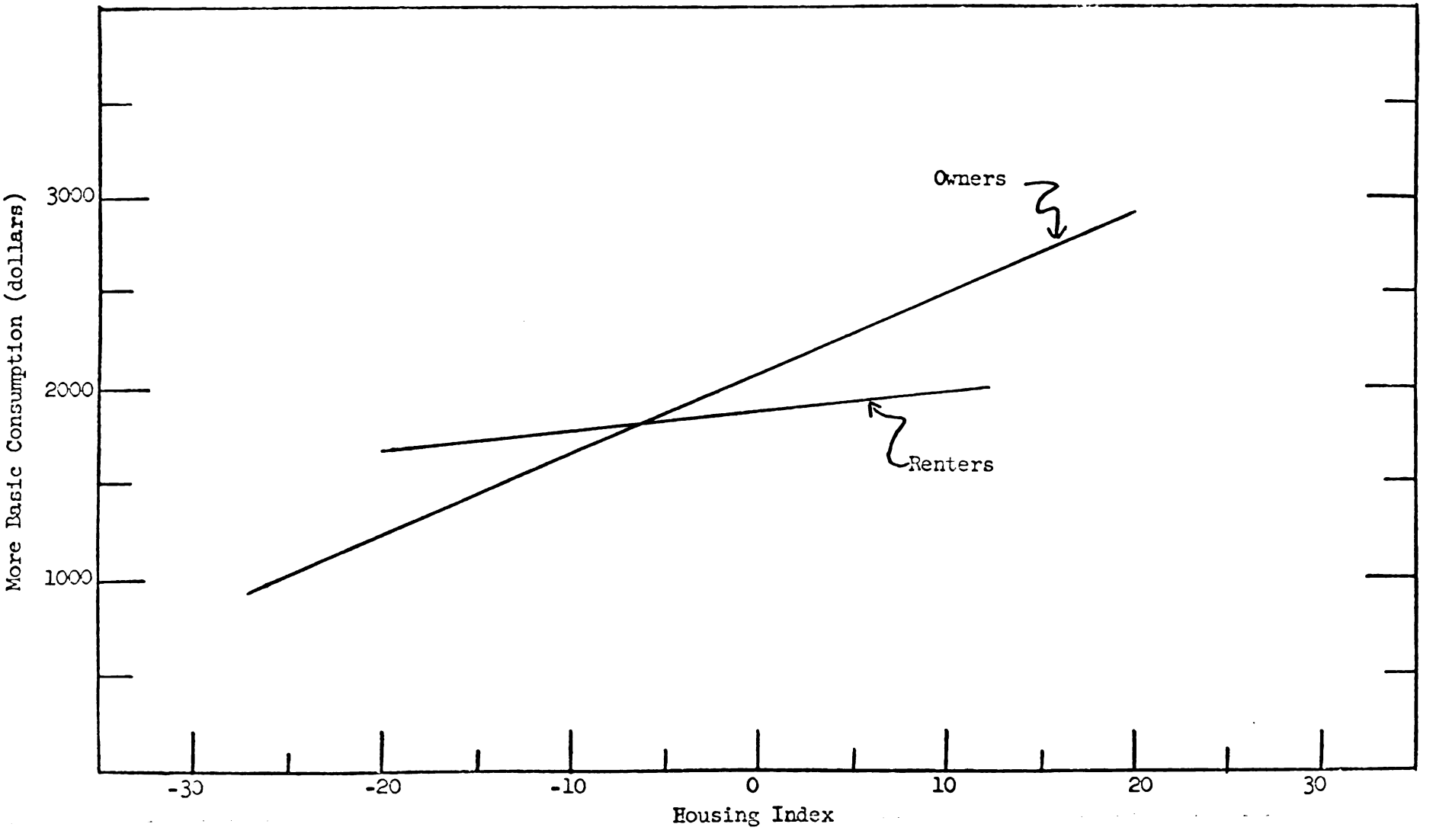
Another interesting case of interaction brought out by the analysis of Table 3.9 is between housing index (a continuous variable) and home tenure classifications. The X_1X_{17} interaction term came in with a significantly negative coefficient, when either "more basic consumption" or "total consumption" was the dependent variable. This is interpreted to mean that among families who rent their homes, the housing index is a less useful predictor of consumption, at least for the more basic items of consumption.

Figure 3.4 illustrates a difference in consumption behavior between owners and renters with respect to their housing index. Families producing less than 30 percent of their food were not included in generating the two regression lines in the diagram, since their effect had been controlled by the previous entry of X_2X_{12} and X_2X_{13} into the equation. It was at this point in the stepwise procedure that X_1X_{17} entered the equation and indicated an interaction between housing index and home tenure.

The differences in slope of the owner and renter functions are the indicator of interaction between home tenure and housing index. However, since the two functions cross each other, there apparently is not any appreciable correlation between these two factors. 1/ This marks a

1/ See Morgan and Sonquist, op cit, p. 419 for a diagramatic representation of their interpretation of multicolinearity and interaction phenomena.

Figure 3.4--More basic consumption as a function of housing index: Comparisons of two family classes, low income rural families which produced 30 percent or more of the food they consumed, East Texas, 1958



contrast with Figures 3.3 and 3.2 where the two functions, although having different slopes, did not intersect within the relevant area. Thus, there were indications of both interaction and intercorrelation between race and family size in Figure 3.3, and between chief source of income and family size in Figure 3.2.

Interaction between tenure and housing index is also indicated when more basic expenditure is the dependent variable. But in this case, a succession of several interaction terms entering the equation, reduced the t value associated with housing index from 6.583 right down to a relatively insignificant 1.260. The interaction terms entering, and the sequence in which they entered the regression were: (a) X_2X_{15}, (b) X_2X_{18}, (c) X_2X_{11}, (d) X_1X_{16}, and (e) X_2X_{13}. The accompanying adjustments in t associated with housing index, went from 6.583 to (a) 6.719, to (b) 4.665, to (c) 4.457, to (d) 1.293, to (e) 1.260. The biggest adjustment happened in step (d) when the interaction term, "housing index x owners," entered. This entry brought with it a control in the effects of housing index as it applied to all homeowner families. Hence, only renters remained, and the housing index coefficient now reflected the slope of the spending function for renters only. For this residual (renter) group, the housing index variable apparently did not contribute significantly to predicting changes in the level of more basic expenditures.

Phase IV -- The Final Regression Model

The first three phases of this chapter were a preliminary study of apparent relationships, before proceeding to a final multiple regression model. The results have clearly shown that in a study of the level of living of low-income rural families, the hidden intercorrelations and interaction effects are so numerous and so important that they must be explicitly allowed for in any adequate treatment of the subject. It was

brought out, for instance, that money income and family size would be very inadequate if they were to serve as the sole predictors of the family living level. The two factors are extensively correlated with other variables, as is evidenced by the substantial adjustments that occur in their coefficients when additional variables enter the multiple regression.

Failure to take account of the more important correlated variables biases any regression analysis and may produce misleading results. For example, it was demonstrated that food purchased by a family of given size and money income depended a great deal on race, chief source of income, and the amount of food that is produced at home by the family. Omitting one or more of these factors would materially alter estimates of the parameters in the regression equation. There are, of course, a number of other correlated variables that remain in the unspecified category, but for the present study they either are not available for explicit specification or this investigator was unable to define properly the relevant variables that should be specified.

In the final model an effort was made to include the several independent variables which had shown significant correlation with one or more of the 8 dependent variables. The forms in which the independent variables are specified were determined by the preliminary findings, and also by the goal of keeping the number of variables to a manageable level. One method employed to keep down the number of explanatory variables, was to combine several factors into complex constructs which would hopefully explain more than other possible sets.

Still further reduction in the number of variables is possible if continuous, instead of discrete, variables are employed. From among the three variables run in continuous form originally, only two are still presented in this form. In the preliminary runs, housing index seemed to

show a good fit as a linear continuous variable, and is therefore presented in that form here. Family size showed an improvement in fit by a curved regression over a linear one. It is therefore entered with a log transformation, but still in continuous form.

Money income of the family was initially run in two different forms--as a linear continuous variable, and in subclasses as dummy variables. Neither of these forms made a very significant contribution to explained variance in the living level of low-income families. Both forms became highly significant, however, when all the families in the sample, including those with incomes over $2,500 were analyzed. Thus, in the final model, the only money income classifications employed were: (1) below $2,500, (2) $2,500 to $3,999, and (3) $4,000 and above. These income classes were not separately specified, but were studied in conjunction with other factors in compound variables.

A series of joint (compound) variables were drawn up which combined several interrelated factors that had also shown significant correlation with the dependent variable. Each combination of values of the independent variables thus specified, defines the dimensions of a newly created cell designed for a particular specification of families. The chief aim behind the creation of these composite variables, is to provide classifications within which families would display less heterogeneous behavior than might exist under alternative classifications. A second goal is to reduce the amount of intercorrelation and interaction among the independent variables.

Regression Model for the Total Sample

The preliminary studies demonstrated the need for something other than a simple additive regression equation; yet, it is difficult to allow for many interaction effects. The usual way to handle this problem has been to build an equation of the joint function type which has the important

interaction effects built right into the various classifications. Our sample is too small, however, to permit all of the joint variables we suspect are needed to capture the many interaction effects.

It is necessary that there be sufficient observations in any particular subgroup, so that we can be sure the effect of that subgroup actually matters. Therefore, certain cross-classifications cannot be specified because there are not sufficient cases to fit into the defined cell. With this limitation in mind, the author undertook to combine a number of characteristics and subclasses into family groups that were presumed to display individual modes of behavior.

The major split in families suggested by the earlier analyses is one between whites and nonwhites. Nonwhites consumed at a considerably lower level than whites, especially the important items of food and shelter.

Another group that seemed to have uniformly lower levels of consumption were the older families. But the classification "families with homemakers 55 and older" showed less significance as an explanatory variable, than did the closely correlated class "families who received their chief income from transfer payments, rent, and interest." This result suggested that a split at age 65 (the age at which transfer payments in the form of old age benefits become available) was more appropriate than a split at age 55. Therefore, all white families were classified into categories with family heads "under 65" and "over 65."

Chief source of income was another factor that had shown considerable effect on consumption or expenditure behavior. It was anticipated, however, that a slightly different classification procedure would delineate groups that were more nearly homogeneous in their behavior. The procedure used was to subdivide all white families with heads below 65 years of age into (1) commercial farmers, (2) noncommercial farmers, and (3) nonfarmers.

Gross farm sales plus change in farm inventory were used to define the 3 classes: if this figure was $2,500 or more, the families were grouped with commercial farmers; if it was between $250 and $2,500, they were classed with noncommercial farmers; if it was below $250, they were classed with nonfarmers.

One last factor, money income, was incorporated into the several family groups that were finally delineated. Table 3.10 shows the eleven subgroups into which the 349 families were classified, and it also gives the number of families belonging to each group. Each subgroup was judged to have adequate observations so that a valid comparison could be made between it and the omitted group.

Table 3.10--Subgroups and the number of observations per group, rural families, East Texas, 1958

Group	Number of observations
Nonwhite, income below $2,500	60
Nonwhite, income above $2,500	17
White (over 65) income below $2,500	47
White (over 65) income above $2,500	13
White (under 65) commercial farmers	31
White (under 65) noncommercial farmers, income below $2,500	24
White (under 65) noncommercial farmers, incomes $2,500 to $3,999	11
White (under 65) noncommercial farmers, income $4,000 and above	15
White (under 65) nonfarmers, incomes below $2,500	41
White (under 65) nonfarmers, incomes $2,500 to $3,999	35
White (under 65) nonfarmers, incomes $4,000 and above	35

The subgroups defined in Table 3.10 are quite adequate in reflecting the effects of race, age, source of income and the level of income. But there are a number of other factors that should be specified as well. Two categories of families (1) with homemakers less than the 12th grade, and (2) with homemakers grade 12 and above, were specified to test the effects of education. Student t-tests had shown that a further subdivision of grades 0 through 7 and grades 8 through 11 were not warranted on the basis of statistical significance. The same conclusion could also be drawn from earlier multivariate analyses.

With regard to "percentage of home produced food" the 4 categories delineated earlier were specified in the final model. Their effect on several of the 8 dependent variables had been indicated as quite substantial. A further reason for specifying this characteristic is the relationship the amount of home produced food ostensibly has with nonmonetary income.

Housing index and family size were specified as continuous variables, even though the analysis in Phase III had suggested they ought to be specified jointly with the categorical variables. The joint variables, thus created, would have a number of important interactions built right into them. Specification of all such interaction terms, however, would nearly triple the number of independent variables, and the complexity of the model would be greatly increased.

It was, therefore, considered expedient to assume away the interactions and to specify the housing index and family size as "additive" continuous variables. It is recognized, of course, that this specification is not as effective in reducing the predictive error as the alternative one which would include interaction terms. Furthermore, the coefficients of the two continuous variables must be interpreted with caution since they represent a weighted compromise effect among the several different effects that really exist.

The results of the final regression model for all families are given in Table 3.11. The class that was omitted from the joint variables classifications was "white (under 65) commercial farmers"; the class omitted from the two education categories was "families with homemaker less than 12th grade"; the class omitted from the percentage of home produced food groups was "families producing 10 to 29.9 percent."

The two continuous variables were significant, at the 5 percent level, in all but two equations: (1) housing index did not significantly affect food purchases, and (2) log family size did not significantly affect less basic consumption.

Among families that had reported less than $2,500 income, nonwhites and also whites (over 65) spent and consumed at a level substantially lower than white (under 65) commercial farmers. There was one exception: Nonwhite purchases or consumption of the less basic items were not significantly below the commercial farm level. This result supports a similar Phase II finding -- that whites and nonwhites consume at levels not very different from each other where only the less basic items are considered.

Two other classifications with incomes under $2,500, noncommercial farmers and nonfarmers, had a lower living level than commercial farmers. However, in about half the equations estimated, the differences were not significant at the 5 percent level.

The living levels of families with incomes $2,500 to $3,999 were about the same as the levels of the commercial farmers, but families with incomes $4,000 and above lived considerably better. Only in connection with food and shelter did the commercial farmers approach the same level of consumption as that enjoyed by families with income $4,000 and above.

The most apparent difference between noncommercial farmers and nonfarmers (each reporting $4,000 or more income) shows up in connection with their

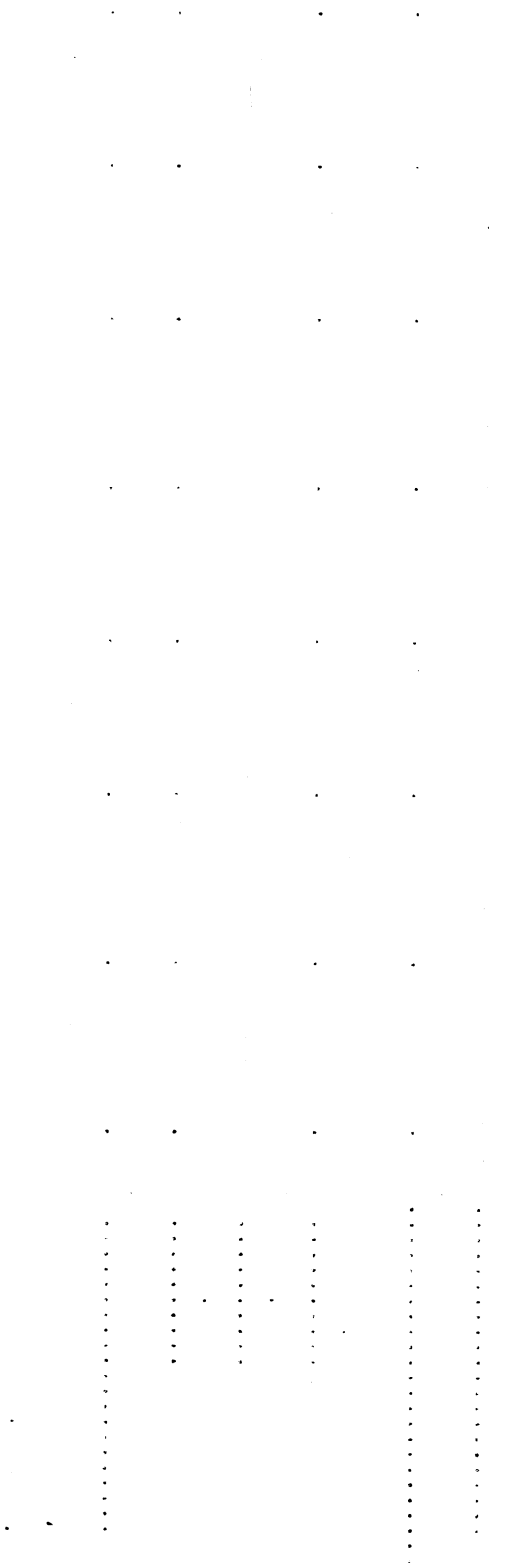
Table 3.11--Computed t values in regressions of family levels of living on household characteristics, rural families, East Texas, 1958

Independent variables	Dependent variables: Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_1 - Housing index	5.072	9.434	1.612	2.371	4.535	10.279	3.700	5.342
X_2 - Log family size	6.524	8.308	8.543	9.687	8.526	12.462	1.939	1.633
X_3 - Nonwhite, income below $2,500	-2.457	-2.334	-3.124	-3.873	-2.734	-3.814	-1.252	-1.723
X_4 - Nonwhite, income above $2,500	0.145	-0.541	-1.623	-2.226	-0.906	-1,300	1.200	0.734
X_5 - White (over 65) income below $2,500	-3.132	-4.230	-2.058	-2.439	-2.443	-3.355	-2.693	-3.504
X_6 - White (over 65) income above $2,500	-0.276	-1.225	-0.762	-1.444	0.533	-0.953	-1.032	-1.031
X_7 - White (under 65) commercial farmers	omitted variable							
X_8 - White (under 65) noncommercial farmers, income below $2,500	-2.023	-2.330	-1.566	-2.275	-1.514	-2.677	-1.303	-1.655
X_9 - White (under 65) noncommercial farmers, income between $2,500 and $3,999	0.603	1.088	0.192	0.021	0.660	0.406	0.319	1.306
X_{10} - White (under 65) noncommercial farmers, income $4,000 and above	3.143	3.204	1.568	1.237	1.636	1.317	3.511	3.737
X_{11} - White (under 65) nonfarmers, income below $2,500	-2.070	-2.961	-1.150	-1.576	-1.745	-2.755	-1.643	-2.092
X_{12} - White (under 65) nonfarmers, income between $2,500 and $3,999	0.133	-0.784	-0.229	-0.714	-0.207	-1.323	0.442	-0.029
X_{13} - White (under 65) nonfarmers, income $4,000 and above	5.444	4.103	3.851	1.948	4.669	1.852	4.237	4.645

Table 3.11--Computed t values in regressions of family levels of living on household characteristics, rural families, East , Texas, 1958 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_{14} - Families with homemaker less than 12th grade	omitted variable							
X_{15} - Families with homemaker grade 12 or above	1.079	2.223	-1.036	-0.453	1.363	1.101	0.372	2.427
X_{16} - Families producing 0 to 9.9 percent of their food	2.861	1.591	2.770	-0.509	3.043	0.350	1.600	2.404
X_{17} - Families producing 10 to 29.9 percent of their food	omitted variable							
X_{18} - Families producing 30 to 49.9 percent of their food	0.906	2.314	-0.774	3.672	-0.106	3.440	1.637	1.262
X_{19} - Families producing 50 percent of their food	-2.986	0.591	-3.998	3.732	-3.452	2.203	-1.384	-1.049

$t_{.05} = 1.971$ $t_{.01} = 2.598$



more basic expenditures: whereas noncommercial farmers spent only an insignificantly greater amount than commercial farmers, the nonfarmer expenditures were greater by a highly significant amount.

Any effect on the living level produced by education appeared to be considerably smaller than had been produced by the education categories in earlier analyses. The chief reason for the smaller effect in the present analysis seems to be that education classifications are so intercorrelated with the joint variables classes, that the regression effects of education are attenuated in substantial degree. Thus, it is not correct to describe education as an unimportant determinant of the living level. This is simply a situation where the total regression influence is shared by education classes and the joint classes which combined race, age, income source, and income level. Due to the intercorrelations, a net effect cannot be assigned to the individual factors.

The "percentage of home produced food" categories appeared to add substantially to regression effects in the present model. One or more specified subclasses came in significantly different from the omitted subclass in 7 out of the 8 equations.

The individual coefficient t-tests are not really adequate evidence to show that either the "education effect" or the "proportion of home produced food effect" adds significantly to "explanation" of the dependent variable. The coefficients of the dummy variables are estimates of their respective parameters, and it is possible for these to test out individually as non-significant even though the complete set of dummy variables might add significantly to the regression effects. Hence, to establish definite statistical significance of the influence of either education or the proportion of food home produced, it would be necessary to test the hypothesis

that the entire set of dummy variables under each of these two characteristics is not significant.

Computer and time resources were not available in the present investigation to carry out these tests.

Exactly the same independent variables, X_1 through X_{19} of Table 3.11, were run in a second set of regressions -- this time specifying the log of each of the 8 dependent variables. The log fit produced an increase in explained variance in all 8 equations, ranging from about one percent when estimating less basic expenditure, to 12 percent in the equation predicting total food purchases. Table 3.12 compares the coefficients of multiple determination for the two sets of regression equations. It indicates that predictive error is reduced in every case when a log transformation of the dependent variable is employed.

Table 3.12--A comparison of coefficients of multiple determination: Regressions of natural Y and log Y on the independent variables listed in Table 3.11

Dependent variable	R^2 pertaining to	
	Natural Y	Log Y
Y_1 - Total expenditure	.612	.633
Y_2 - Total consumption	.680	.702
Y_3 - Food expenditure	.518	.586
Y_4 - Food consumption	.435	.466
Y_5 - More basic expenditure	.604	.640
Y_6 - More basic consumption	.668	.688
Y_7 - Less basic expenditure	.410	.414
Y_8 - Less basic consumption	.526	.530

The log transformation runs produced some fairly consistent changes in computed t values relating to independent variables. Nearly without exception, the t values increased (in absolute value) for classes representing the lower living levels, and decreased for those classes of families which consumed and spent at a higher level. Table 3.13 gives computed t values associated with the 19 independent variables for number 1 of the eight

Table 3.13--A comparison of computed t values: Regressions of total family expenditure (natural form and log transformation) on the independent variables described in Table 3.11

Independent variables	t values when Y is in form of	
	Natural Y	Log Y
X_1 - Housing index	5.072	5.182
X_2 - Log family size	6.524	6.681
X_3 - Nonwhite, income below $2,500	-2.457	-3.728
X_4 - Nonwhite, income $2,500 and above	0.145	0.541
X_5 - White (over 65) income below $2,500	-3.132	-4.023
X_6 - White (over 65) income $2,500 and above	-0.276	-0.420
X_7 - White (under 65) commercial farmers		
X_8 - White (under 65) noncommercial farmers, income below $2,500	-2.023	-3.140
X_9 - White (under 65) noncommercial farmers, income between $2,500 and $3,999	0.603	0.642
X_{10} - White (under 65) noncommercial farmers, income $4,000 and above	3.143	2.204
X_{11} - White (under 65) nonfarmers, income below $2,500	-2.070	-2.429
X_{12} - White (under 65) nonfarmers, income between $2,500 and $3,999	0.133	0.013
X_{13} - White (under 65) nonfarmers, income $4,000 and above	5.444	3.741
X_{14} - Families with homemaker less than 12th grade		
X_{15} - Families with homemaker grade 12 and above	1.079	1.827
X_{16} - Families producing 0 to 9.9 percent of their food	2.851	2.296
X_{17} - Families producing 10 to 29.9 percent of their food		
X_{18} - Families producing 30 to 49.9 percent of their food	0.906	0.702
X_{19} - Families producing 50 percent or more of their food		
	R^2= 0.630	R^2= 0.650

equations -- run first in natural, then in log form. The changes produced by the log form were very similar in the remaining 7 equations, both in the direction and in the extent of changes to the t values.

Regression Model for the Subsample

When only those families were considered which had reported less than $2,500 income in 1958, the definitions of the subgroups making up the new joint variables were slightly modified. The level of money income was not considered in the new cells, because earlier analysis had shown little correlation between income and the living level when dealing with the "below $2,500" families. The chief source of income, however, was extensively used as one of the dimensions for creating more homogeneous family groups.

The major split was again the one which separated whites from nonwhites. The nonwhite families were divided further into a group receiving their main income from salaries, wages, professions or business, and a second group that received their chief income from other sources. Whites were again divided between groups with family heads "over 65" and "under 65." The over 65 group was further classified according to whether their major income came from transfer payments, rent, and interest, or whether it was from other sources.

The white (under 65) classification was subdivided into groups according to their farm status and their major source of money income. The same definitions as for the previous (total sample) model were used to define commercial farmers, noncommercial farmers, and nonfarmers. The new composite variables and the number of families belonging to each are given in Table 3.14.

The subsample model was very similar to the one that was specified for the total sample. The nine groups defined in Table 3.14 were specified to replace the eleven joint variables of the earlier model. The

classes omitted from each set of dummy variables to prevent overdetermination in the analysis were also the same ones in both models.

Table 3.14--Subgroups and the number of observations per group, low-income rural families, East Texas, 1958

Group	Number of observations
Nonwhite, major income from salaries, wages, professions or business	31
Nonwhite, major income from other sources	29
White (over 65) major income from transfer payments, rent, and interest	30
White (over 65) major income from other sources	17
White (under 65) commercial farmers	17
White (under 65) noncommercial farmers, major income from salaries, professions or business	13
White (under 65) noncommercial farmers, major income from other sources	11
White (under 65) nonfarmers, major income from salaries, wages, professions or business	26
White (under 65) nonfarmers, major income from other sources	16

The estimated t values for the subsample model are given in Table 3.15. There is considerable similarity between the results presented here and those produced by the total sample (see Table 3.11). Once again, the only equation in which housing index did not demonstrate a significant effect was the one predicting food purchases. And log family size, though highly significant in predicting six dependent variables, showed no explanatory capacity when estimating purchases or consumption of less basic items (equations 7 and 8).

Table 3.15--Computed t values in regressions of family levels of living on household characteristics, low-income families, East Texas, 1958

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_1 - Housing index	3.759	7.009	1.403	2.120	2.589	7.303	3.624	4.313
X_2 - Log family size	3.871	5.481	5.515	6.659	5.643	8.573	0.931	0.363
X_3 - Nonwhite, major income from salaries, wages, profession or business	-1.111	-2.023	-3.096	-4.305	-2.725	-3.360	0.811	0.415
X_4 - Nonwhite, major income from other sources	-2.736	-3.869	-4.145	-5.413	-4.059	-5.045	-0.672	-0.355
X_5 - White (over 65) major income from transfer payments, rent, and interest	-3.298	-4.645	-4.216	-5.371	-3.332	-5.353	-1.695	-2.091
X_6 - White (over 65) major income from other sources	-1.673	-2.129	-1.282	-1.334	-1.736	-2.208	-1.095	-1.320
X_7 - White (under 65) commercial farmers	---------- omitted variable ----------							
X_8 - White (under 65) noncommercial farmers, major income from salaries, wages, professions or business	-1.385	-1.845	-2.365	-3.522	-1.850	-3.364	-0.498	0.229
X_9 - White (under 65) noncommercial farmers, major income from other sources	-2.189	-2.893	-2.581	-3.529	-2.623	-3.379	-1.031	-0.961
X_{10} - White (under 65) nonfarmers, major income from salaries, wages, professions or business	-1.065	-1.297	-2.069	-3.055	-2.133	-3.579	-0.314	-0.302
X_{11} - White (under 65) nonfarmers, major income from other sources	-2.201	-3.052	-2.003	-3.274	-2.345	-3.783	-1.371	-1.303

Table 3.15--Computed t values in regressions of family levels of living on household characteristics, low-income families, East Texas, 1958 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_{12} - Families with homemaker less than 12th grade	omitted variable							
X_{13} - Families with homemaker grade 12 and above	1.623	2.303	-0.024	0.407	1.733	1.436	1.008	2.329
X_{14} - Families producing 0 to 9.9 percent of their food	1.915	1.023	2.419	0.128	1.978	0.128	1.252	1.518
X_{15} - Families producing 10 to 29.9 percent of their food	omitted variable							
X_{16} - Families producing 30 to 49.9 percent of their food	0.892	2.978	-0.164	2.864	-0.020	3.180	1.501	1.753
X_{17} - Families producing 50 percent or more of their food	-2.313	1.337	-3.455	4.045	-3.058	3.183	-0.862	-0.871

$t_{.05} = 1.973$ $t_{.01} = 2.600$

White (under 65) commercial farmers enjoyed a considerably higher level of living than any of the other eight joint classifications. It is true that the difference was nonsignificant for less basic items, but with respect to food, clothing, and shelter the commercial farmers consumed significantly more than any one other group. Since the three "more basic" items made up 60 percent of low-income total purchases, and about two-thirds of their total consumption, it is conclusively demonstrated that commercial farmers were better off than all others reporting under $2,500 income.

At this point, it is important to remind the reader that several of the commercial farmers and others are not classified properly when identified with low-income families even though their 1958 income was reported at less than $2,500. Their long time average income may be expected to exceed $2,500 by a considerable margin; and if classified by a different criterion, several of these households would not appear with the "poor."

"White (over 65) with incomes chiefly from transfer payments, rent, and interest" were the group which spent and consumed at a level below all others. At only a slightly higher level, were the classification "non-whites with income chiefly from sources other than wages, salaries, professions or business." In general, the families which came closest to the living level of commercial farmers, were the ones which received their main income from wages, salaries, professions or business.

Little seemed to be gained in explanatory effect by classifying noncommercial farmers separately from nonfarmers. This was in contrast to the total sample model where the same distinction appeared to be one of considerable consequence.

As was true for the total sample, education categories appeared to add comparatively little explanatory effect when the joint variables had been allowed for. Indications were, however, that home produced food

categories made a significant further contribution to variance reduction even though these categories were intercorrelated with some of the 9 joint variables. Only when less basic expenditure or consumption were the dependent variables, were none of the three specified classes significantly different from the fourth (omitted) variable.

The actual estimating equations in regressions of family living levels on household characteristics of low-income rural families are given in Table 3.15a.

When a log transformation of the 8 dependent variables was run in place of their natural form, the results ranged from essentially no change in $\bar{R}^2$ in the case of less basic purchases or consumption, to a 16 percent increase in the explained variance of food purchases. The adjusted coefficients of multiple determination for each of the 8 equations are compared in Table 3.16.

The amount of variance in the dependent variable that was explained by the final model (subsample) was considerably reduced from the amount that could be explained by the final model (total sample) [Table 3.16 cf. Table 3.12]. The greatest reduction occurred in the areas of less basic expenditure and less basic consumption. It is understandable, since the total sample includes families with higher incomes, that they should display more variation in the area of the less basic items. There is also a fairly obvious reason why the subgroups of the total sample should be able to explain more of the total variation. The subgroups specified for the total sample model are able to utilize the considerable predictive power provided by the money income classifications (1) below $2,500, (2) $2,500 to $3,999, and (3) $4,000 and above. Previous analysis had indicated a further subdivision of the group with incomes under $2,500 was not warranted

Table 3.15a--Actual estimating equations in regressions of family levels of living on household characteristics, low-income rural families, East Texas, 1958

$Y_1 = 503.037 + 22.600\ X_1 + 1295.809\ X_2 - 286.839\ X_3 - 702.608\ X_4 - 783.260\ X_5$
$- 415.449\ X_6 - 374.197\ X_8 - 643.014\ X_9 - 253.333\ X_{10} - 598.365\ X_{11}$
$+ 218.694\ X_{13} + 278.927\ X_{14} + 140.097\ X_{16} - 342.302\ X_{17}$

$Y_2 = 611.108 + 42.923\ X_1 + 1869.475\ X_2 - 532.314\ X_3 - 994.240\ X_4 - 1124.293\ X_5$
$- 536.950\ X_6 - 507.899\ X_8 - 865.605\ X_9 - 467.098\ X_{10} - 845.311\ X_{11}$
$+ 316.081\ X_{13} + 151.832\ X_{14} + 476.205\ X_{16} + 201.650\ X_{17}$

$Y_3 = 132.423 + 3.503\ X_1 + 767.087\ X_2 - 332.268\ X_3 - 434.439\ X_4 - 416.079\ X_5$
$- 131.832\ X_6 - 265.566\ X_8 - 314.959\ X_9 - 204.462\ X_{10} - 226.227\ X_{11}$
$- 1.351\ X_{13} + 146.430\ X_{14} - 10.679\ X_{16} - 212.491\ X_{17}$

$Y_4 = -281.792 + 6.647\ X_1 + 1162.051\ X_2 - 579.956\ X_3 - 712.873\ X_4 - 665.423\ X_5$
$- 243.299\ X_6 - 496.344\ X_8 - 540.679\ X_9 - 379.079\ X_{10} - 464.304\ X_{11}$
$+ 28.634\ X_{13} + 9.741\ X_{14} + 234.501\ X_{16} + 312.282\ X_{17}$

$Y_5 = -64.998 + 9.480\ X_1 + 1108.061\ X_2 - 412.965\ X_3 - 600.556\ X_4 - 540.836\ X_5$
$- 252.152\ X_6 - 293.158\ X_8 - 451.836\ X_9 - 298.343\ X_{10} - 373.978\ X_{11}$
$+ 136.921\ X_{13} + 169.042\ X_{14} - 1.853\ X_{16} - 265.466\ X_{17}$

$Y_6 = -188.417 + 26.433\ X_1 + 1728.320\ X_2 - 600.399\ X_3 - 857.398\ X_4 - 808.469\ X_5$
$- 329.157\ X_6 - 547.336\ X_8 - 686.125\ X_9 - 512.763\ X_{10} - 619.231\ X_{11}$
$+ 116.535\ X_{13} + 11.188\ X_{14} + 300.630\ X_{16} + 283.677\ X_{17}$

$Y_7 = 568.163 + 13.117\ X_1 + 187.665\ X_2 + 126.130\ X_3 - 102.065\ X_4 - 242.442\ X_5$
$- 163.307\ X_6 - 81.051\ X_8 - 191.194\ X_9 + 45.005\ X_{10} - 224.401\ X_{11}$
$+ 81.777\ X_{13} + 109.879\ X_{14} + 141.949\ X_{16} - 76.841\ X_{17}$

$Y_8 = 799.667 + 16.490\ X_1 + 141.063\ X_2 + 68.088\ X_3 - 136.857\ X_4 - 315.844\ X_5$
$- 207.803\ X_6 + 39.424\ X_8 - 179.497\ X_9 + 45.660\ X_{10} - 226.094\ X_{11}$
$+ 199.550\ X_{13} + 140.637\ X_{14} + 175.575\ X_{16} - 82.032\ X_{17}$

Y_1 through Y_8 and X_1 through X_{17} are the same variables listed in Table 3.15.

because such a division did not appear to add greatly to explained variance of the living level.

Table 3.16--A comparison of adjusted coefficients of multiple determination: Regressions of natural Y and log Y on the independent variables listed in Table 3.15

Dependent variable	$\bar{R}^2$ pertaining to	
	Natural Y	Log Y
Y_1 - Total expenditure	.361	.387
Y_2 - Total consumption	.528	.563
Y_3 - Food expenditure	.404	.479
Y_4 - Food consumption	.484	.524
Y_5 - More basic expenditure	.426	.449
Y_6 - More basic consumption	.632	.653
Y_7 - Less basic expenditure	.159	.157
Y_8 - Less basic consumption	.236	.239

Hence, money income is a dimension which adds little to explain the variation in consumption behavior of low-income East Texas families. But it appears to have considerable capacity for predicting the living level of families from the same area if their incomes exceed $2,500.

Allowance for Interaction Effects

It is apparent that the more important interaction effects brought out by the Phase III analysis, were not adequately allowed for in the two final models presented in Phase IV. The Phase III regression runs demonstrated that the slopes of the functions for the housing index and family size variables, could vary considerably for different values of other family characteristics. In other words, it was shown that housing

index and log family size could not really operate additively with race, age, chief source of income, and percentage of home produced food. Consequently, it is indicated that the explained sum of squares can be maximized only if additional predictors are specified which have the important interactions built into them.

Crockett and Friend have suggested that if the effects of important variables are considerably different for different population groups, it is desirable either to compute separate regressions for these groups or to introduce nonlinear terms specifying the nature of the interaction. 1/ These two authors felt they could not attempt the second alternative in their particular analysis because they had an insufficient basis from either a priori considerations or previous empirical work, for assigning a particular mathematical form to the interactions.

For the present study, however, it has been demonstrated that much hidden interaction can be captured in the regression effects by specifying interaction terms which combine continuous with categorical variables. By this technique, we can define a separate slope of the continuous variable for each designated family group.

One final regression was computed using the 189 low-income observations and specifying the same categorical variables X_3 through X_{17} found in Table 3.15; but the two continuous variables, X_1 and X_2 of that same table, were specified entirely differently. Interaction terms were introduced in which the categorical variables were each teamed first with housing index, then with log family size. The resulting equation could capture the various slopes of housing index and log family size functions peculiar to each designated categorical variable.

1/ Crockett and Friend, op cit, p. 11.

Thus, the final independent variable specifications consisted of 12 categorical variables and 24 interaction terms. Three categorical variables and six interaction terms were omitted to prevent overdetermination of the system. The eight dependent variables were all specified in natural form.

The model with the interaction terms achieved a sizeable reduction in predictive error over the model which used Table 3.15 predictors. Table 3.17 gives $\bar{R}^2$ values pertaining to the two sets of regressions. Since the second set of $\bar{R}^2$ values was considerably larger than the first, it is indicated that the gain in explained variance more than compensated for the

Table 3.17--A comparison of adjusted coefficients of multiple determination: Regressions of Y on (1) X_1 through X_{17} of Table 3.15 and (2) X_3 through X_{17} of Table 3.15 plus interaction terms

Dependent variables	$\bar{R}^2$ pertaining to	
	(1) a/	(2) b/
Y_1 - Total expenditure	.359	.399
Y_2 - Total consumption	.528	.596
Y_3 - Food expenditure	.403	.494
Y_4 - Food consumption	.484	.536
Y_5 - More basic expenditure	.426	.455
Y_6 - More basic consumption	.632	.666
Y_7 - Less basic expenditure	.159	.207
Y_8 - Less basic consumption	.235	.250

a/ Specified independent variables were X_1 through X_{17} of Table 3.15.
b/ Specified independent variables were X_3 through X_{17} of Table 3.15, plus interaction terms X_1X_3 through X_1X_{17} and X_2X_3 through X_2X_{17}.

loss in degrees of freedom. On the whole, though, the 24 interaction terms add considerable complexity to the Phase IV model, since it would be difficult to take cognizance of so many different family groups when studying the

effects of housing index and family size. It may therefore be desirable to compute separate regressions for larger subgroups of families that are identified as different with respect to certain important variables. An analysis like that employed in Phase III can be helpful in identifying such groups.

CHAPTER IV

BLACKLAND TEXAS

A General Description of the Area

Six counties were surveyed in the Northern Blackland area of Texas. Collin, Delta, Fannin, Grayson, Hunt, and Lamar counties lie in an area extending north and east of Dallas to the Oklahoma line. (See Chart I, Appendix A).

According to the U. S. Census of Population, the six counties had 218,000 people in 1960. Nonwhites made up 13 percent of those with native parentage. Males 25 and older had a median education of 9.4 years; 1/ females 25 and older had a median of 10.0 years. Nonwhites (male and female) 25 and older averaged 7.5 years of schooling.

Manufacturing was the chief industry, employing about one-fifth of all employed persons. Wholesale and retail trade was next with 19 percent of the employed persons, and agriculture was third with 15 percent.

Among nonwhites, "personal services" was the chief classification, employing 30 percent of all nonwhite employed persons; agriculture, forestry, and fisheries employed 18 percent; the wholesale and retail trade employed 13 percent; manufacturing employed 10 percent.

The median income for the six counties was $3,769, with 27 percent of the families earning less than $2,000. 2/ Of the nonwhite families, 54 percent made less than $2,000 per year.

About 46 percent of the population was classified "rural" and of these, one-third were classified "rural farm." The median rural income was $3,063 and for the rural farms the median was $2,643. More than one-third of the rural families had incomes below $2,000.

1/ The median for each county was multiplied by the number of males 25 and older in that county, and the 5 products were added to make one total. This total was then divided by total males 25 and older in all 5 counties to yield a weighted median.

2/ These multicounty income statistics are weighted medians constructed in the same manner as the education data above.

Description of the Survey Sample

The total sample was made up of 348 families living in the open country of the six counties. Their net incomes after taxes in 1959 ranged from -$3,700 to $25,000. Two hundred three of these families had incomes below $2,500.

Wages, salaries, professions or business comprised the chief source of income for 48 percent of all the families; transfer payments, rent, and interest were the chief income for 19 percent; returns from farming were the chief income for the remaining 33 percent.

About 8 percent of the families produced none of the food they consumed; another 20 percent produced less than a tenth of their food; 15 percent of the families produced more than half of their total food supply. Among nonwhite families, 20 percent produced less than a tenth, and 22 percent produced more than half their food supply.

In the subsample of low-income families, the chief source of income for about 40 percent of the families was derived from wages, salaries, professions or business. Another 25 percent received their main income from transfer payments, rent, and interest. The remaining 35 percent obtained most of their incomes from farming.

In the subsample of 203 low income families the proportion of nonwhites rose to 14 percent. The relative proportions of food produced at home by nonwhites continued essentially unchanged from those in the total sample. For the entire subsample, 5 percent produced no food of their own; 14 percent produced less than a tenth; 19 percent produced more than half the food they consumed.

The Analysis

Phase I -- Money Income, Housing Index, and Family Size as Explanatory Factors

The first set of regressions employed money income, housing index, and family size as predictors of consumption and expenditure behavior. The three variables were specified in linear continuous form, and the resulting regressions produced t values and $\bar{R}^2$ values as indicated in Table 4.1.

Table 4.1--Computed t values in regressions of the level of living on three household characteristics, low-income rural families, Blackland Texas, 1959

Dependent variables	Independent variables			$\bar{R}^2$
	Money income	Housing index	Family size	
(1) Total expenditure	2.319	7.116	5.396	.238
(2) Total consumption	2.309	11.088	8.128	.421
(3) Food expenditure	2.237	3.323	7.765	.244
(4) Food consumption	1.916	4.812	9.794	.336
(5) More basic expenditure	1.964	7.212	6.944	.275
(6) More basic consumption	2.207	12.480	11.712	.526
(7) Less basic expenditure	2.060	5.035	2.211	.113
(8) Less basic consumption	1.623	5.065	1.262	.132

$t_{.05} = 1.973$ $t_{.01} = 2.600$

There was considerable similarity between this study and East Texas, both in the manner and the extent to which the three variables helped to "explain" low-income rural family living levels. A comparison with Table 3.1 of the previous chapter showed that only in one regression (equation 4) was there a sizeable difference in $\bar{R}^2$ for comparable equations

of the two studies. The greater influence of the family size variable in equation 4 appeared to contribute most to a larger $\bar{R}^2$ for the Blackland.

Money income in the Blackland sample demonstrated slightly greater influence in the four consumption equations. Not only was it more effective in explaining food consumption (equation 4), but also in explaining other items of consumption (equations 2, 6, and 8). In the equation that estimated family expenditures, however, (i.e., equations 1, 3, 5, and 7), the t values associated with money income were about the same for both studies.

As was true in East Texas, little intercorrelation was found to exist among the independent variables -- the simple correlation coefficient between money income and housing index was -0.11; between money income and family size, 0.11; between housing index and family size, -0.33.

For the same reasons discussed in Chapter III, a new set of specifications was drawn up to allow for the possibility that family size and money income were not properly specified when presented in linear form. Thus, a log transformation was applied to family size, and income was subdivided into three classifications of families. Dummy variables were assigned to the three income classes: (1) below $400; (2) $400 to $1,499; and (3) $1,500 to $2,499. Housing index was again specified as a linear continuous variable. Results of the regressions with the new specifications are found in Table 4.2. The class with incomes below $400 was the one selected for omission from the dummy variable set.

A comparison of adjusted coefficients of multiple determination indicated that Table 4.2 specifications produced a significant improvement in regression goodness of fit over Table 4.1 specifications -- explaining from 5 to 41 percent more variance in the dependent variable. This improvement in fit was more modest, however, than had been obtained in East Texas where from

Table 4.2--Computed t values in regressions of the level of living on certain household characteristics, low-income rural families, Blackland Texas, 1959

Dependent variables	Independent variables				$\bar{R}^2$
	Continuous variables		Dummy variables		
			Income class a/		
	Housing index	Log family size	$400 to $1,499	$1,500 to $2,499	
(1) Total expenditure	6.652	5.756	-1.242	1.611	.282
(2) Total consumption	10.818	8.811	-1.519	1.446	.471
(3) Food expenditure	2.948	7.917	-0.229	1.329	.256
(4) Food consumption	4.503	10.748	-1.565	0.386	.389
(5) More basic expenditure ..	6.824	7.261	-0.759	1.179	.300
(6) More basic consumption ..	12.369	12.559	-1.040	1.469	.566
(7) Less basic expenditure ..	4.514	2.413	-1.411	1.614	.161
(8) Less basic consumption ..	5.181	1.495	-1.492	0.895	.161

a/ The omitted class was "families with incomes below $400."

$t_{.05} = 1.973$ $t_{.01} = 2.600$

16 to 100 percent more variability could be explained by the second set of specifications. In East Texas, each one of the three independent variables showed a better fit under the second specification set. In the Blackland, however, housing index fit less well when combined with log family size and categorical income, than it did when each of the three variables had been specified in linear continuous form.

Computed t values under the dummy variables of Table 4.2 illustrate that in the Blackland, even more so than in East Texas, money income follows the behavior of a U-shaped step function, i.e., families of the middle income class ($400 to $1,499) spend and consume less on the average, than either the group which earns less than $400 or the group earning $1,500 to $2,499.

In Table 4.2 the lowest income class was omitted, and t values indicate that consumption or spending in the other two classes was not significantly different from this omitted class. A different result is obtained, however, if we test for a difference between the two specified classes. Since the coefficients associated with $400 to $1,499 class are negative and the coefficients associated with the $1,500 to $2,499 class are positive, it appeared likely there would be a "true" difference between these two classes. A test of the null hypothesis that there was not a significant difference, was rejected at the 5 percent level in every case except in the equation that predicts food expenditure. This demonstrates again, that the difference between coefficients of two dummy variables can test out significant even though the individual coefficients are not significantly different from zero.

A plausible reason for the somewhat U-shaped regression of the living level on money income was suggested in Chapter III as it pertained to the East Texas situation. The same reason appears to apply to the Blackland situation. The lowest income class had a high proportion of farmers, and many of them reported net incomes considerably lower than their expenditures for the survey year. Their current purchases were therefore financed from past savings, or from credit based on future earnings. There is also the possibility that a number of the incomes were underreported. In any event, it was apparent that several families with sizeable farming operations were not properly classified when they were grouped with extremely poor families.

Phase II -- Additional Explanatory Factors Specified in the Regressions

This phase specified a number of family characteristics as independent variables in addition to the three explanatory factors considered in Phase

I. The additional factors of age, education, source of income, home tenure, race, and the proportion of food home produced were each divided into subgroups and dummy variables were defined on them. Table 4.3 gives computed t values for all the variables that were specified and entered into the equation, but the categorical variables entered only if they contributed to the regression at the 5 percent probability level. Those variables for which there is a t entry in Table 4.3 are, therefore, the only ones that were part of the final eight regressions.

When the several dummy variables were added to the regressions, adjustments occurred in the coefficients of the continuous variables. (Table 4.3 cf. Table 4.1). These adjustments did not follow a consistent pattern, but on an average the t values associated with housing index and family size adjusted downward; those associated with money income adjusted upward. After all adjustments had been allowed for, money income had become a significant variable in 7 out of the 8 equations; in East Texas, money income had ended up as a significant variable in only one equation. The housing index and family size variables performed similarly in both studies.

In contrast with East Texas, the "chief source of income" classes were nonsignificant in the present sample. Instead, the closely correlated "age of homemaker" categories came in significant. Computed t values indicated that in the first 6 equations, families with older homemakers consumed at a lower level than younger and middle-aged families. Conversely, in equations 7 and 8, the younger families were shown to consume and spend at a higher rate than families with older homemakers.

The education classes demonstrated a small predictive effect in the first equation only. But when the same regressions were run using the total sample, the category "homemakers with education less than 8th grade" was

Table 4.3--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, Blackland Texas, 1959

Dependent variables	Independent variables: Continuous variables: Money income	Housing index	Family size	Dummy variables: Chief source of income: Salaries profes-sions or business	Transfer payments rent or interest	Farm	Age of homemaker: Under 40	40 to 54	55 and over	Education of homemaker: Less than 8th grade	Grade 8 thru 11	Grade 12 and over	Tenure of home by farm status: Owner farm	Renter farm	Owner non-farm	Renter non-farm	Percent of food home produced: 0 to 9.9 per-cent	10 - 29.9 per-cent	30 to 49.9 per-cent	50 per-cent and over	Race: White	Non-white	R^2
Total expenditure	2.489	6.349	4.230						-3.222	-2.560							2.176			-2.706	2.189		.376
Total consumption	2.945	9.544	6.919						-3.360						-2.494	-2.447					2.419		.493
Food expenditure	2.152	3.408	8.157						-2.389								3.393			-5.291	3.477		.458
Food consumption	3.005	2.633	9.027						-2.619							2.188			2.169	4.091	3.722		.466
More basic expenditure	1.645	7.623	5.869	2.221					-3.152								2.227			-3.600		-3.152	.444
More basic consumption	2.952	10.663	10.943						-2.725						-2.619	-3.706					2.646		.596
Less basic expenditure	2.272	5.308	0.778				3.738																.168
Less basic consumption	2.185	4.895	-0.819				4.345	2.430					2.036										.204

$t_{.05} = 1.973$ $t_{.01} = 2.600$

significant (with a negative coefficient) in five additional equations. Only when estimating food expenditure or consumption was the education effect insignificant in both the total sample and the subsample.

Among the "tenure x farm status" categories, the nonfarm classes tended to consume at levels significantly below the farm classes in equations 2, 4, 6, and 8. These equations estimated family consumption. But neither farm status nor tenure appeared to have a significant effect in equations 1, 3, 5, and 7 which estimated expenditures.

"Percent of home produced food" classes contributed importantly to predicting purchases of more basic items. In the expenditure equations 1, 3, and 5, the "0 to 9.9 percent" class spent less while the "50 percent and over" class spent more than families in the middle "10 to 49.9 percent group." In equation 4, the families producing 30 percent or more of their food consumed at a significantly higher level than families which produced less than 30 percent of their food.

White families purchased and consumed the "more basic" items at a higher level than did the nonwhite families. However, any difference between the two races that might have existed with respect to the "less basic" items was found to be insignificant at the 5 percent probability level.

Phase III -- Search for Interaction Effects

The purpose of Phase III is to discover the more important interaction effects which exist among the independent variables. One hypothesis that is tested again in this chapter is that home ownership and farm status are independent in their effect on food consumption. Average dollar values of food consumption for each of four classifications is presented in Table 4.4. These statistics are generally not unlike the comparable ones of Table 3.5

which pertained to East Texas. However, the standard errors of the mean for the two renter classes are substantially smaller than they were in East Texas.

Table 4.4--Value of food consumption as a function of home tenure and farm status, low-income rural families, Blackland Texas, 1959 (Figures in parentheses are standard errors)

Farm Status	Owner	Renter
Farm	$1,137 (57)	$1,144 (63)
Nonfarm	$ 724 (63)	$1,016 (87)

Table 4.4 suggests a definite correlation between food consumption and farm status. Correlation between food consumption and tenure, however, was confined to nonfarm families with owners consuming substantially less than renters. "Owners nonfarm" consisted almost entirely of older 2-person families, a class which in earlier analyses was shown to be made up predominantely of low level consumers.

A variance analysis conducted on the cell means of Table 4.4 is reported in Table 4.5. The analysis indicates a very significant correlation between food consumption and farm status, but the tenure and interaction effects were not significant at the 5 percent level.

Table 4.5--Analysis of variance of data from Table 4.4

Source of variation	Sum of squares	d.f.	Variance	F	$F_{.05}$	$F_{.01}$
Farm status	73,143	1	73,143	7.66	3.89	6.78
Tenure	22,425	1	22,425	2.35	3.89	6.78
Interaction	32,435	1	32,435	3.40	3.89	6.78
Error		199				

Another variance analysis tested the null hypothesis that age and education did not interact in their effect on less basic expenditures. In order to test this hypothesis, education was reclassified into two classes: (a) families with homemaker grade 8 or less, and (b) families with homemaker above the 8th grade. The age classes employed were the same as defined earlier.

Table 4.6 provides the cell means of less basic expenditures for a 3 x 2 factorial experiment in which age and education are studied simultaneously. The table further includes an analysis of variance to test for

Table 4.6--Expenditures on less basic items as a function of age and education of homemakers, low-income rural families, Blackland Texas, 1959 (Figures in parentheses are standard errors)

Education	Age		
	Less than 40	40 to 54	55 and above
8th grade or less	N = 12 $385 (140)	N = 42 $479 (56)	N = 41 $735 (76)
Above 8th grade	N = 25 $1,131 (117)	N = 43 $1,050 (79)	N = 40 $679 (56)

Source of variation	Analysis of variance					
	Sum of squares	d.f.	Variance	F	$F_{.05}$	$F_{.01}$
Age	92,803	2	46,401	4.34	3.04	4.71
Education	72,930	1	72,930	6.81	3.89	6.76
Interaction	69,801	2	34,900	3.26	3.04	4.71
Error		197	10,703			

age, education, and interaction effects. The evidence from the analysis was that each effect is significant at the 5 percent level; and education, in fact, showed significance at the 1 percent level.

Search Using Multivariate Techniques

A further search for interaction was made with the help of a multivariate regression model. The model specified essentially the same set of variables as were tested in Phase II. Age and education, however, were specified as joint rather than additive variables because of the significant interaction they displayed in predicting less basic expenditures (Table 4.6). The regressions also included a number of interaction terms, each of which combined a continuous and a dummy variable. These interaction terms entered and remained in the final regressions, only if they contributed to regression effects at the 5 percent level. But housing index, log family size, and the money income categories were forced into the model regardless of the extent of their regression effects.

The final regression results in estimating each of the 8 dependent variables are given in Table 4.7. Note that only those interaction terms which turned out to be significant in one or more of the 8 equations are listed in the independent variables column. All other combinations of continuous x dummy variables were specified except: X_1X_3 through X_1X_5; X_2X_3 through X_2X_5; X_1X_6 through X_1X_{11}; X_1X_{12} through X_1X_{14}; and X_1X_{19} through X_1X_{22}.

Several important interactions were brought out by the analysis. For instance, log family size interacted with "chief source of income" in the first six equations. When predicting expenditures ("more basic" or "total") as a function of log family size, the slope was steeper for families with incomes chiefly from salaries, wages, professions or business. When predicting consumption ("food" or "total"), the least slope was generated by families receiving their incomes mainly from farming.

In the first six equations, log family size also interacted with the "percentage of home produced food" calsses. In predicting expenditures

Table 4.7--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, Blackland Texas, 1959

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_1 - Housing index	6.000	9.922	2.921	4.179	7.970	10.469	3.776	4.440
X_2 - Log family size	-1.141	1.785	4.329	10.314	-0.682	8.361	0.823	-0.328
Income Categories								
X_3 - Below $400	omitted variable							
X_4 - $400 to $1,499	-1.390	-0.871	-0.691	-1.449	-1.246	-0.872	-1.018	-1.043
X_5 - $1,500 to $2,499	1.359	2.039	0.750	0.791	0.966	1.892	1.677	0.929
Age and Education								
X_6 - Under 40, 8th grade or less								
X_7 - Under 40, above 8th grade	2.869	2.735					3.767	4.506
X_8 - 40 to 54, 8th grade or less								
X_9 - 40 to 54, above 8th grade	4.187	4.052			3.847		3.777	3.796
X_{10} - 55 and over, 8th grade or less								
X_{11} - 55 and over, above 8th grade								
X_2X_7				-2.407				
Chief Source of Income								
X_{12} - Salaries, wages, professions or business								
X_{13} - Transfer payments, interest, rent								
X_{14} - Farming								
X_2X_{12}	3.114		3.257		4.695	2.511		
X_2X_{14}		-2.375		-2.698				

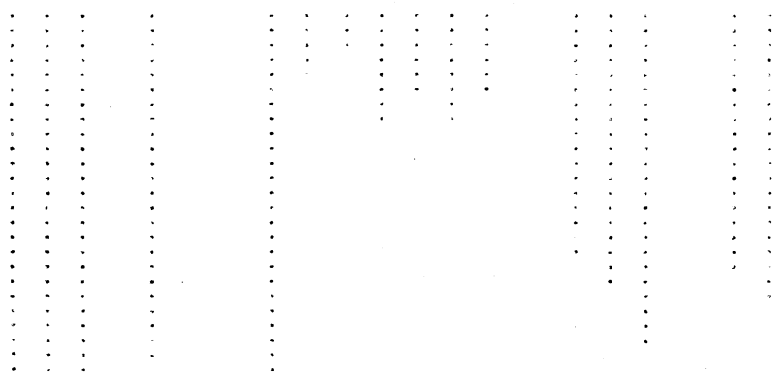


Table 4.7--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, Blackland Texas, 1959 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
Tenure and Farm Status								
X_{15} - Owner farm								
X_{16} - Renter farm								
X_{17} - Owner nonfarm								
X_{18} - Renter nonfarm						-2.367		
X_1X_{16}		-3.601		-3.022	-3.366	-2.921		
X_1X_{18}	2.270		3.075					
X_2X_{15}		2.089			2.242			
X_2X_{18}				-3.114				
Percentage of Home Produced Food								
X_{19} - 0 to 9.9 percent								
X_{20} - 10 to 29.9 percent								
X_{21} - 30 to 49.9 percent								
X_{22} - 50 percent and above	-3.839		-6.322		-4.960			
X_2X_{19}	2.880		5.283		4.538			
X_2X_{20}				-3.050		-2.399		
Race								
X_{23} - White	-2.766				-2.524			
X_{24} - Nonwhite		2.345						
X_2X_{23}	3.904	3.478	4.107	4.069	4.095	3.469		
$\bar{R}^2$	.491	.510	.528	.531	.547	.633	.235	.259

$t_{.05} = 1.973$ $t_{.01} = 2.600$

("more basic" or "total"), the group with the greatest sloped function was the "0 to 9.9 percent" class. But when more basic consumption (especially food) was estimated, the class producing "10 to 29.9 percent" had a smaller slope than any other class. Many of the larger families in the latter class had "homemakers under 40, 8th grade or less." Since this class tended to be low level consumers per family member, this fact helps to explain the small relative slope in the function for the 10 to 29.9 percent family group.

Interaction between log family size and race is found both in expenditures and consumption ("more basic" and "total"). The level of living function slopes more steeply for white families than it does for nonwhite families, except when predicting living levels pertaining only to less basic items where there is no significant difference between whites and nonwhites.

Interactions between housing index and categorical variables showed significance only when dealing with the "housing tenure x farm status" categories. The negative t values in four equations indicate that housing index has least explanatory power among "renter farm" families. This result supports the general conclusion reached in the last chapter: that housing index is not as useful a predictor of the living level among renters as among owners. The general proposition was not supported, however, by the regression results which estimated food purchases among "renters nonfarm" i.e., the t values associated with X_1X_{18} in equations 1 and 3, have an opposite sign from what was expected.

A peculiar set of circumstances appear to explain this unexpected result. Nearly half the subsample nonwhites were "nonfarm renters." Since their housing was considerably poorer than that of whites, they tended to concentrate near the left end of a scatter diagram for food purchases on

housing index. Furthermore, since nonwhite food purchases tend to be relatively low, it follows that whites (concentrated on the right side of the scatter distribution) had purchases placing them relatively high in the diagram. Hence, a best fitting straight line of the function "food purchases on housing index" shows a much steeper slope in this case than would be expected under less peculiar circumstances.

Phase IV -- The Final Regression Model

The specifications employed in the final regression models of this chapter are very similar to those used in the last phase of Chapter III. By following this procedure, ready use can be made of comparisons between the two surveys. Furthermore, the analytical processes that are conducted in Phase IV serve as a logical followup to the analysis carried out in the first three phases of this chapter.

Regression Model for the Total Sample

A set of joint variables specified for the final model were designed to capture the combined effects produced by race, age, source of income, and level of income. The object was to create family classifications that displayed less heterogeneity within classes than would result from alternative classifications. The compound variables are defined by the 12 subgroups of families presented in Table 4.8. The number of cases belonging to each classification were considered adequate for making valid intergroup comparisons.

The first important split made in the sample was one which placed nonwhites in a separate classification. The white families were then further divided into families with heads "over 65" and heads "under 65."

Table 4.8--Subgroups and the number of observations per group, rural families, Blackland Texas, 1959

Group	Number of observations
X_3 - Nonwhite families	36
X_4 - White (over 65) income under $2,500	47
X_5 - White (over 65) income over $2,500	19
X_6 - White (under 65) large commercial farmers	25
X_7 - White (under 65) small commercial farmers, over $2,500, major income -- farming	21
X_8 - White (under 65) small commercial farmers, under $2,500, major income -- farming	39
X_9 - White (under 65) noncommercial and small commercial farmers, over $2,500, major income -- nonfarm	37
X_{10} - White (under 65) noncommercial and small commercial farmers, under $2,500, major income -- salaries, wages, professions or business	32
X_{11} - White (under 65) noncommercial farmers, under $2,500, major income -- other than salaries, wages, professions or business	29
X_{12} - White (under 65) nonfarmers, under $2,500	21
X_{13} - White (under 65) nonfarmers, incomes between $2,500 and $3,999	20
X_{14} - White (under 65) nonfarmers, incomes $4,000 and above	22

The income levels were designated "under $2,500" and "over $2,500," but the latter group was in some instances further divided into (1) incomes between $2,500 and $3,999 and (2) incomes of $4,000 and above. The "source of income" designations related primarily to the extent to which incomes were derived from farm sources. Gross farm sales plus change in inventory were used to delineate the classes: where this figure exceeded $10,000 the families were classed as large commercial farmers; between $2,500 and

$10,000 as small commercial farmers; between $250 and $2,500 as noncommercial farmers; below $250 as nonfarmers.

Age and education were specified jointly rather than additively because of the evidence that a degree of interaction existed between the two factors (see Table 4.6). "Percentage of home produced food" was specified in four categorical variables using the same classes defined earlier. Housing index and log family size were each specified as a continuous variable.

Computed t values for the final regressions of the total sample are given in Table 4.9. Comparisons with the East Texas study can be made by referring to Table 3.11 in Chapter III. The variables that were omitted from each of the three sets of dummy variables are indicated in the table.

The negative t values associated with X_1 through X_{24} indicate that large commercial farmers in the Blackland had a higher living level than any other classification. The only group that consumed at a level close to that of large commercial farmers was "white (under 65) nonfarmers with incomes $4,000 and above." All other groups consumed and spent significantly less under the "total" and "less basic" designations. With regard to the "more basic" designation, all groups were on a lower plane than large commercial farmers although in some instances they were not on a significantly lower plane.

The age x education classifications did not demonstrate a significant influence in predicting (a) food purchases, (b) food consumption, and (c) more basic consumption. In the other 5 equations, the class "homemaker 40 to 54, above 8th grade" spent or consumed at a significantly higher level than the omitted class.

Among "percent of home produced food" classes, at least one of the specified categories was significantly different from the omitted category in 7 out of the 8 equations. Only in estimating total family consumption

Table 4.9--Computed t values in regressions of family levels of living on household characteristics, rural families, Blackland Texas, 1959

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_1 - Housing index	5.700	9.518	3.083	2.892	6.676	11.884	3.201	3.518
X_2 - Log family size	6.138	7.665	11.489	12.603	8.535	11.728	2.284	0.694
X_3 - Nonwhite families	-6.958	-7.347	-4.594	-5.731	-6.091	-6.062	-5.687	-5.780
X_4 - White (over 65) income under $2,500	-6.815	-7.159	-3.494	-3.559	-5.438	-5.301	-6.027	-6.228
X_5 - White (over 65) income over $2,500	-3.637	-3.641	-1.425	-1.551	-2.248	-2.740	-3.782	-3.125
X_6 - White (under 65) large commercial farmers	---------- omitted variable ----------							
X_7 - White (under 65) small commercial farmers, over $2,500, major income from farming	-4.796	-4.815	-1.932	-2.252	-4.178	-3.709	-3.938	-4.048
X_8 - White (under 65) small commercial farmers, under $2,500, major income from farming	-6.454	-6.194	-2.698	-2.738	-5.327	-4.125	-5.554	-5.843
X_9 - White (under 65) noncommercial and small commerical farmers, over $2,500, major income--nonfarm	-2.149	-1.976	0.317	-0.093	-1.332	-0.655	-2.232	-2.515
X_{10} - White (under 65) noncommercial and small commercial farmers, under $2,500, major income from salaries and wages	-3.080	-2.892	0.799	-0.554	-1.163	-0.710	-3.843	-3.925
X_{11} - White (under 65) noncommercial farmers, under $2,500, major income other than salaries and wages	-7.370	-7.608	-3.353	-3.745	-5.537	-5.647	-6.815	-6.606

Table 4.9--Computed t values in regressions of family levels of living on household characteristics, rural families, Blackland Texas, 1959 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_{12} - White (under 65) nonfarmers, under $2,500	-6.495	-6.859	-2.970	-3.994	-5.015	-5.563	-5.889	-5.490
X_{13} - White (under 65) nonfarmers, incomes between $2,500 and $3,999	-4.159	-4.315	-1.732	-1.790	-2.659	-3.150	-4.250	-3.798
X_{14} - White (under 65) nonfarmers, incomes $4,000 and above	-0.523	-1.209	0.794	-0.281	0.339	-1.238	-1.117	-0.714
X_{15} - Homemaker under 40, 8th grade or less	1.364	-0.038	0.267	-0.195	0.445	-0.912	1.763	0.837
X_{16} - Homemaker under 40, above 8th grade	1.046	0.304	-1.870	-1.521	0.062	-0.928	1.593	1.399
X_{17} - Homemaker 40-54, 8th grade or less	omitted variable							
X_{18} - Homemaker 40-54, above 8th grade	3.684	2.487	0.296	-0.792	3.065	0.727	3.180	3.261
X_{19} - Homemaker 55+, 8th grade or less	1.000	0.168	0.789	-0.010	-0.084	-0.676	1.647	0.934
X_{20} - Homemaker 55+, above 8th grade	0.755	0.266	0.714	-0.430	0.381	-0.333	0.859	0.754
X_{21} - 0 to 9.9 percent home produced food	3.438q	1.756	4.748	0.107	3.366	0.560	2.502	2.257
X_{22} - 10 to 29.9 percent home produced food	omitted variable							
X_{23} - 30 to 49.9 percent home produced food	-0.581	1.075	-1.798	3.546	-1.390	1.375	0.287	0.366
X_{24} - 50 percent plus, home produced food	-3.276	1.055	-6.617	5.169	-4.538	3.109	-1.235	-1.373

were none of the three designated classes significantly different from the fourth (omitted) class.

The same independent variables, X_1 through X_{24} of Table 4.9 were run in a new set of regressions -- this time specifying the log of each of the 8 dependent variables. Table 4.10 gives a comparison of coefficients of determination between this set of runs and the previous set when the natural form of the dependent variables was employed. The log form produced increases in $\bar{R}^2$ in seven of the equations. The increase could be attributed to a slightly better fit for housing index and log family size; a larger t value was also associated with the class "families producing 50 percent or more of their food."

Table 4.10--A comparison of adjusted coefficients of multiple determination: Regressions of natural Y and log Y on the independent variables listed in Table 4.9

Dependent variable	$\bar{R}^2$ pertaining to	
	Natural Y	Log Y
Y_1 - Total expenditure	.584	.615
Y_2 - Total consumption	.617	.634
Y_3 - Food expenditure	.555	.608
Y_4 - Food consumption	.513	.523
Y_5 - More basic expenditure	.606	.666
Y_6 - More basic consumption	.641	.571
Y_7 - Less basic expenditure	.396	.415
Y_8 - Less basic consumption	.402	.407

But in equation 6 which estimated more basic consumption, the log transformation brought about a substantial reduction in $\bar{R}^2$. In this equation, the t values associated with all independent variables were reduced when the log form of the dependent variable was specified.

Regression Model for the Subsample

When only the 203 families with incomes below $2,500 were considered, a new set of joint variables were defined to incorporate the effects of race, age, and source of income. Nonwhites were again classified separately and whites with family head over 65 were split from whites with head under 65. Chief source of income was the third factor used in defining the complex variables. Definitions of these variables and the number of observations falling under each classification are given in Table 4.11.

Table 4.11--Subgroups and the number of observations per group, low-income rural families, Blackland Texas, 1959

Group	Number of observations
X_3 - Nonwhite families	29
X_4 - White (over 65) major income from transfer payments, rent, or interest	29
X_5 - White (over 65) major income from other sources	18
X_6 - White (under 65) commercial farmers, major income from farm with gross sales plus inventory change below $6,000	28
X_7 - White (under 65) commercial farmers, major income from farm with gross sales plus inventory change above $6,000	17
X_8 - White (under 65) noncommercial and commercial farmers with major income from salaries, wages, professions or business	32
X_9 - White (under 65) noncommercial farmers with major income from sources other than salaries, wages, professions or business	29
X_{10} - White (under 65) nonfarmers	21

The family groups defined in Table 4.11 were specified as a set of dummy variables in the final regression model. Effects of education and home food production also were allowed for by specifying two additional sets of dummy variables. Finally, housing index and log family size each were specified as continuous variables.

The estimated t values for the subsample regressions are given in Table 4.12. The two continuous variables contributed toward these regressions, much as they did in East Texas: housing index was significant throughout, but log family size was significant in the first six equations only. Education effects and "percent of home produced food" effects were also similar to those observed in the East Texas subsample: the influence of education was indicated to be small, and the amount of food home produced was a significant factor in the first six equations only.

The joint variables, X_3 through X_{10}, behaved somewhat differently in the two studies. In East Texas, the living level of commercial farmers ranked above all the others. But, as pointed out in Chapter III, this classification contained a number of households which consumed at substantially higher levels than are normally associated with "the poor." Such "misclassified families", though present in the Blackland, were not as important as in East Texas. Hence, Blackland commercial farmers with gross sales either above or below $6,000, ranked lower than the classification "white (head under 65) noncommercial and commercial farmers, major income from salaries, wages, professions or business." It is notable that in East Texas the classification ranking next to commercial farmers was also "families having salaries, wages, professions or business as chief income source."

The actual estimating equations for the final subsample model are given in Table 4.12a.

Table 4.12--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, Blackland Texas, 1959

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_1 - Housing index	4.741	7.833	2.250	2.076	5.641	9.850	2.359	2.638
X_2 - Log family size	4.948	7.659	8.840	10.520	7.212	12.123	1.220	0.134
X_3 - Nonwhite families	-1.639	-2.292	-2.610	-3.787	-1.698	-3.240	-1.051	-0.420
X_4 - White (over 65) major income from transfer payments, rent, or interest	-2.657	-2.753	-1.867	-1.772	-2.116	-2.816	-2.301	-1.561
X_5 - White (over 65) major income from other sources	0.127	0.306	0.230	0.421	-0.054	-0.023	0.256	0.503
X_6 - White (under 65) commercial farmers, major income from farm with gross sales below $6,000	omitted variable							
X_7 - White (under 65) commercial farmers, major income from farm with gross sales above $6,000	2.690	4.628	1.207	2.543	2.307	3.972	2.175	3.371
X_8 - White (under 65) noncommercial and commercial farmers with major income from salaries, wages, professions or business	4.588	5.020	3.703	3.817	5.193	4.859	2.533	3.116
X_9 - White (under 65) noncommercial farmers with major income from sources other than salaries, wages, professions or business	-2.123	-2.156	-1.280	-1.246	-1.301	-2.332	-2.204	-1.908
X_{10} - White (under 65) nonfarmers	-1.709	-2.059	-1.086	-1.906	-0.961	-2.759	-1.854	-0.526

Table 4.12--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, Blackland Texas, 1959 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_{11} - Homemaker 8th grade or less	omitted variable							
X_{12} - Homemaker above 8th grade	1.407	0.181	-0.646	1.238	1.455	-1.178	0.905	1.440
X_{13} - Families producing 0 to 9.9 percent of their food	3.111	1.248	3.624	0.435	3.298	0.949	1.927	1.028
X_{14} - Families producing 10 to 29.9 percent of their food	omitted variable							
X_{15} - Families producing 30 to 49.9 percent of their food	0.390	2.008	-1.973	2.773	-0.620	2.093	-0.047	0.629
X_{16} - Families producing 50 percent or more of their food	-3.574	0.438	-3.524	3.590	-4.751	2.065	-1.311	-1.338

Table 4.12--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, Blackland Texas, 1959 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_{11} - Homemaker 8th grade or less	omitted variable							
X_{12} - Homemaker above 8th grade	1.407	0.181	-0.646	1.238	1.455	-1.178	0.905	1.440
X_{13} - Families producing 0 to 9.9 percent of their food	3.111	1.248	3.624	0.435	3.298	0.949	1.927	1.028
X_{14} - Families producing 10 to 29.9 percent of their food	omitted variable							
X_{15} - Families producing 30 to 49.9 percent of their food	0.390	2.008	-1.973	2.773	-0.620	2.093	-0.047	0.629
X_{16} - Families producing 50 percent or more of their food	-3.574	0.438	-5.524	3.590	-4.751	2.065	-1.311	-1.338

Table 4.12a--Actual estimating equations in regressions of family levels of living on household characteristics, low-income rural families, Blackland Texas, 1959

$$Y_1 = 1565.122 + 267.666\ X_1 + 1471.992\ X_2 - 330.893\ X_3 - 508.086\ X_4 + 26.808\ X_5 + 580.465\ X_7 + 834.859\ X_8 - 402.166\ X_9 - 362.186\ X_{10} + 149.350\ X_{12} + 432.695\ X_{13} - 49.882\ X_{15} - 497.661\ X_{16}$$

$$Y_2 = 2095.647 + 457.423\ X_1 + 2356.688\ X_2 - 478.626\ X_3 - 544.459\ X_4 + 66.942\ X_5 + 1033.105\ X_7 + 944.605\ X_8 - 422.416\ X_9 - 451.389\ X_{10} + 19.874\ X_{12} + 179.457\ X_{13} + 226.057\ X_{15} + 63.026\ X_{16}$$

$$Y_3 = 372.739 + 45.080\ X_1 + 933.371\ X_2 - 187.055\ X_3 - 126.722\ X_4 + 17.209\ X_5 + 92.448\ X_7 + 239.089\ X_8 - 86.059\ X_9 - 81.683\ X_{10} - 24.345\ X_{12} + 178.887\ X_{13} - 89.587\ X_{15} - 322.420\ X_{16}$$

$$Y_4 = 399.532 + 57.171\ X_1 + 1526.856\ X_2 - 373.041\ X_3 - 165.321\ X_4 + 43.354\ X_5 + 267.713\ X_7 + 338.826\ X_8 - 115.163\ X_9 - 197.093\ X_{10} - 115.894\ X_{12} + 29.537\ X_{13} + 173.118\ X_{15} + 243.861\ X_{16}$$

$$Y_5 = 805.169 + 185.080\ X_1 + 1246.889\ X_2 - 199.228\ X_3 - 235.177\ X_4 - 6.683\ X_5 + 289.383\ X_7 + 549.061\ X_8 - 143.228\ X_9 - 118.428\ X_{10} + 89.768\ X_{12} + 266.507\ X_{13} - 46.135\ X_{15} - 384.432\ X_{16}$$

$$Y_6 = 1218.946 + 359.305\ X_1 + 2330.366\ X_2 - 422.766\ X_3 - 347.908\ X_4 - 3.089\ X_5 + 553.927\ X_7 + 571.238\ X_8 - 285.370\ X_9 - 377.893\ X_{10} - 80.776\ X_{12} + 85.243\ X_{13} + 173.068\ X_{15} + 185.762\ X_{16}$$

$$Y_7 = 759.941 + 82.587\ X_1 + 225.134\ X_2 - 131.668\ X_3 - 272.905\ X_4 + 33.493\ X_5 + 291.082\ X_7 + 285.795\ X_8 - 258.935\ X_9 - 243.758\ X_{10} + 59.581\ X_{12} + 166.186\ X_{13} - 3.747\ X_{15} - 113.232\ X_{16}$$

$$Y_8 = 876.699 + 98.1180\ X_1 + 26.327\ X_2 - 55.861\ X_3 - 196.550\ X_4 + 70.031\ X_5 + 479.178\ X_7 + 373.366\ X_8 - 137.045\ X_9 - 73.495\ X_{10} + 100.650\ X_{12} + 94.213\ X_{13} + 52.990\ X_{15} - 122.737\ X_{16}$$

Y_1 through Y_8 and X_1 through X_{16} are the same variables listed in Table 4.12.

When a log transformation was applied to each dependent variable and these were specified in place of the natural form, the change in $\bar{R}^2$ were not all in the same direction. Table 4.13 shows an increase in $\bar{R}^2$ as large as .037 when estimating log Y_3, but a decrease of .102 when estimating log Y_6. When log Y_6 was specified in place of natural Y_6, the t values associated with X_1, X_2, X_7, and X_8 of Table 4.12 were reduced substantially. On the other hand, there was a slight increase (in absolute values) in t values associated with X_3, X_4, X_9, and X_{10}.

Table 4.13--A comparison of adjusted coefficients of multiple determination: Regressions of natural Y and log Y on the independent variables listed in Table 4.12

Dependent variables	$\bar{R}^2$ pertaining to	
	Natural Y	Log Y
Y_1 - Total expenditure	.512	.508
Y_2 - Total consumption	.622	.603
Y_3 - Food expenditure	.544	.581
Y_4 - Food consumption	.543	.545
Y_5 - More basic expenditure	.568	.599
Y_6 - More basic consumption	.702	.600
Y_7 - Less basic expenditure	.258	.271
Y_8 - Less basic consumption	.264	.277

In general, the changes observed here when specifying the log form of dependent variable are similar to the changes observed in East Texas analysis, i.e., the t values tended to increase (in absolute value) for the classes of families representing the lower levels of living, and they decreased for those classes whose members spent and consumed at higher levels.

Allowance for Interaction Effects

The Phase III analysis showed that significant interaction existed between either of the two continuous variables and certain of the categorical variables. In order to allow for these interaction effects, a new set of regression runs were computed which specified joint effects of continuous and categorical variables, as well as the individual effects of all categorical variables. The new runs specified X_3 through X_{16} of Table 4.12, and added two sets of interaction terms: (1) X_1X_3 through X_1X_{16}, and (2) X_2X_3 through X_2X_{16}. The resulting 8 equations could take account of differences in the slope of functions on either housing index or log family size as these are fitted to separate family groups. The separate family groups are defined by the dummy variables X_3 through X_{16} of Table 4.12.

The new model with 11 categorical variables and 22 interaction terms achieved an increase in the explained sums of squares over that produced by the model using Table 4.12 predictors. A comparison of adjusted coefficients of multiple determination produced by the two models given in Table 4.14. The values of $\bar{R}^2$ show that in six of the equations the addition to explained variance more than compensated for loss in degrees of freedom. However, in equations 2 and 6 the gains were just offset by the losses.

As a practical matter, any analysis that specifies many interaction terms becomes unwieldy and difficult to interpret. The sign which accompanies the regression coefficient, in some instances, turns out contrary to logic when certain specified variables are closely intercorrelated. Hence, it may prove more profitable to do separate analyses with groups segregated according to their race, chief source of income, and possibly other criteria.

Table 4.14--A comparison of adjusted coefficients of multiple determination: Regressions of Y on (1) X_1 through X_{16} of Table 4.12 and (2) X_3 through X_{16} of Table 4.12 plus interaction terms

Dependent variables	$\bar{R}^2$ pertaining to	
	(1) a/	(2) b/
Y_1 - Total expenditure	.511	.522
Y_2 - Total consumption	.621	.622
Y_3 - Food expenditure	.543	.589
Y_4 - Food consumption	.543	.582
Y_5 - More basic expenditure	.568	.583
Y_6 - More basic consumption	.701	.704
Y_7 - Less basic expenditure	.257	.293
Y_8 - Less basic consumption	.264	.284

a/ Specified independent variables were X_1 through X_{16} of Table 4.12.

b/ Specified independent variables were X_3 through X_{16} of Table 4.12, plus interaction terms X_1X_3 through X_1X_{16} and X_2X_3 through X_2X_{16}.

CHAPTER V

SOUTH CENTRAL KENTUCKY

A General Description of the Area

The survey counties -- Barren, Hart, Cumberland, Metcalfe, and Monroe--lie in an area extending from the Tennessee line north to approximately the center of the state (see Chart II, Appendix A). The five counties are in a depressed agricultural area on the western fringe of an economically depressed, mountainous coal mining region.

Data from the 1960 U. S. Census put the total population of the five counties at 70,423. Males 25 years and older had a median education of 7.4 years 1/; females 25 and older averaged 8.2 years. Nonwhites 25 and older (male and female) averaged 6.5 years of schooling. Of persons with native parentage, nonwhites comprised 6.8 percent.

Agriculture was the chief industry, employing about 42 percent of all employed persons. Wholesale and retail trade employed 15 percent and manufacturing was third with 11 percent.

If nonwhites only are considered, "personal services" employed nearly one-third of all employed persons. Agriculture employed a like number. Wholesale and retail trade employed 12 percent, and manufacturing was fourth with only 4 percent.

The median annual income of all families was $2,342 and 47 percent had incomes below $2,000. 2/ Among nonwhites, 55 percent made less than $2,000.

Six-sevenths of the total population was classified "rural" and 63 percent of these rural people were classified "rural farm." The median income

---

1/ The median for each county was multiplied by the number of males 25 and older in that county, and the 5 products were added to make one total. This total was then divided by total males 25 and older in all 5 counties to yield a weighted median.

2/ These multicounty income statistics are weighted medians constructed in the same manner as the education data above.

for rural families was $2,138 and for rural farm families it was $1,896. The median income for nonwhite families was $1,843.

Description of the Survey Sample

The total sample consisted of 346 randomly selected open country dwellers. Their family incomes before taxes in 1956 ranged from a slightly negative amount to about $16,000. Two hundred forty-one of the families had incomes below $2,500 and thus were included with so-called low-income families.

No information was available concerning the origin and breakdown of income from nonfarm sources. Therefore, a classification was adopted based on the proportion of income coming from farm sources only. A definition of the classes and number of observations in each class are as follows:

100 percent of income from farm sources -- 92 families

50 to 99 percent of income from farm sources -- 97 families

1 to 49 percent of income from farm sources -- 82 families

0 percent of income from farm sources -- 75 families

Only 9 percent of the families did not produce any of the food they consumed, while an additional 5 percent produced less than one-tenth. But nearly 55 percent produced one-half or more of their total food supply.

Besides home produced food, another important form of "income in kind" was home produced fuel. These two variables were closely correlated, having a simple correlation coefficient between them of 0.93. Thus, it is presumed the regression influence of home produced food also captures much of the regression influence that would be attributable to home produced fuel.

Nonwhites comprised only 3 percent of the total sample. Therefore, in carrying out the analysis of the sample, the nonwhites were not separately classified.

In the subsample of low-income families the same classification for source of income was used as in the total sample. The number of observations falling into each class were as follows:

100 percent of income from farm sources -- 74 families

50 to 99 percent of income from farm sources -- 65 families

1 to 49 percent of income from farm sources -- 47 families

0 percent of income from farm sources -- 55 families

Of the 241 families, 19 produced none of their food; 12 families produced up to one-tenth; 137 families produced 50 percent or more of their total food supply.

The Analysis

Phase I -- Money Income, Housing Index, and Family Size as Explanatory Factors

The initial analysis ran money income, housing index, and family size as independent variables against the same 8 dependent variables described in the previous chapters. The first set of runs, which are reported in Table 5.1, specified the independent variables in a linear continuous form. The simple correlation coefficient between money income and housing index was 0.14; between money income and family size, 0.08; between housing index and family size, -0.21.

The combined regression effects of the three independent variables were similar to those reported in the two Texas surveys. In other words, the three studies had R^2 values close to the same size in seven out of the eight equations. The one exception was the food consumption equation, where the three independent variables were able to explain twice as much of the total variance in the Blackland as they did in Kentucky. Apparently, it was the better fit of housing index and family size which made the prediction of food consumption in the Blackland considerably more effective.

Table 5.1--Computed t values in regressions of the level of living on three household characteristics, low-income rural families, South Central Kentucky, 1957

Dependent variables	Independent variables			R^2
	Money income	Housing index	Family size	
(1) Total expenditure	2.887	5.343	7.075	.248
(2) Total consumption	3.836	8.936	9.583	.415
(3) Food expenditure	1.864	2.989	8.077	.229
(4) Food consumption	2.365	1.331	6.648	.166
(5) More basic expenditure	2.778	5.839	7.795	.278
(6) More basic consumption	4.013	9.364	10.745	.454
(7) Less basic expenditure	1.983	2.964	3.862	.090
(8) Less basic consumption	2.165	5.021	4.529	.156

$t_{.05} = 1.973$ $t_{.01} = 2.600$

Housing index was, in general, not as good a predictor of the living level in Kentucky as in the Texas areas; money income, on the other hand, was a somewhat better predictor in Kentucky. It should be recalled at this point that in the Kentucky survey, incomes were defined differently from the Texas surveys. In the Kentucky study, incomes (1) predated all other data by several months, (2) did not take farm inventory change into account, and (3) were quoted as "income before taxes," while in the two Texas studies (1) incomes and all other data were for the same period (2) change in farm inventory was taken into account, and (3) incomes were quoted as "incomes after taxes."

Only two Kentucky families had reported negative net incomes, but there were 17 others whose money incomes were below $400. The same income

classifications as used earlier in the Texas areas were specified and were then run as dummy variables along with housing index and log family size. The calculated t values from the resulting regressions are reported in Table 5.2.

Table 5.2--Computed t values in regressions of the level of living on certain household characteristics, low-income rural families, South Central Kentucky, 1957

Dependent variables	Independent variables				$\bar{R}^2$
	Housing index	Log family size	Income class a/		
			$400 to $1,499	$1,500 to $2,499	
(1) Total expenditure	5.270	7.831	-0.336	0.798	.274
(2) Total consumption	8.270	10.525	0.515	2.000	.451
(3) Food expenditure	2.797	8.124	-0.098	0.646	.235
(4) Food consumption	1.355	6.633	1.069	1.492	.171
(5) More basic expenditure ..	5.683	8.234	-0.237	0.797	.293
(6) More basic consumption ..	9.397	11.634	0.933	2.102	.483
(7) Less basic expenditure ..	2.944	4.587	-0.315	0.509	.111
(8) Less basic consumption ..	4.747	5.068	-0.186	1.093	.186

a/ The omitted class was "families with incomes below $400."

$t_{.05} = 1.973$ $t_{.01} = 2.600$

In general, the living levels in the two lower income classes seemed to average out about the same, with expenditures slightly higher and consumption slightly lower in the "below $400" class. The $1,500 to $2,499 class consumed at a higher level than both the others, although, for the most part, not a significantly higher level.

There was still the question whether the preferred form in which to specify money income was as a linear continuous variable or as a set of dummy

variables. To find the answer, a new set of regressions was run in which the only change from Table 5.2 specifications was to specify money income as a continuous variable. The resulting $\bar{R}^2$ values were approximately the same as those in Table 5.2. Hence, when run in conjunction with housing index and log family size, money income proved about equally effective whether it was specified as a continuous variable or in large class intervals. The linear continuous form was superior, however, when the total Kentucky sample was analyzed.

Phase II -- Additional Explanatory Factors Specified in the Regression

Other family characteristics, including age and education of the homemaker, source of income, home tenure, and amount of home produced food were now specified as additional independent variables defined by the column headings of Table 5.3. These additional variables were thought to be relevant in their own right, but equally important, the allowance of their effect would remove some of the biases from Phase I regression coefficients which were there because the added variables had not been taken into account.

Notice that the class limits under "percent of home produced food" are different from those in the Texas studies. More extensive home food production in Kentucky made it necessary to redefine the classes so that observations would distribute fairly equally into four specified groups.

A stepwise regression procedure was used in the analysis and only significant categorical variables are included in the final regression equations reported in Table 5.3. Three continuous variables are necessarily included in every regression since provision was made to force in the continuous variables in all final equations.

The three continuous variables were significant at the 5 percent level in all cases but one -- housing index proved insignificant when predicting

ble 5.3--Computed t values in regre

Dependent variables	Continuous variables		
	Money income	Housing index	Family size
tal penditure ..	3.136	4.704	7.389
tal nsumption ..	3.691	8.653	10.229
tal food penditure ..	2.376	2.373	9.251
tal food nsumption ..	2.028	1.569	6.967
re basic penditure ..	3.344	6.131	8.779
re basic nsumption ..	3.854	9.232	11.455
ss basic penditure ..	2.048	2.768	2.696
ss basic nsumption ..	2.078	5.343	4.421

$t_{.05} = 1.973$ $t_{.01} = 2.600$

Table 5.3--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, South Central Kentucky, 1957

Dependent variables	Independent variables: Continuous variables: Money income	Housing index	Family size	Dummy variables: Source of income: 100% from farm	51 to 99% farm	1 to 50% farm	all non-farm	Age of homemaker: Under 40	40 to 54	55 and over	Education of homemaker: Less than 8th grade	Grades 8 thru 11	Grade 12 and over	Tenure of home by farm status: Owner farm	Renter farm	Owner non-farm	Renter non-farm	Percent of food home produced: 0 to 29.9 percent	30 to 49.9 percent	50 to 64.9 percent	65 percent and over	$\bar{R}^2$
Total expenditure	3.136	4.704	7.389								-2.301									-3.231	-5.221	.344
Total consumption	3.691	8.653	10.229				-4.483				-2.531											.468
Total food expenditure	2.376	2.373	9.251									2.422								-6.046	-9.791	.472
Total food consumption	2.028	1.569	6.967				-2.865											-2.083				.237
Core basic expenditure	3.344	6.131	8.779								-2.858						2.221			-4.703	-8.453	.475
Core basic consumption	3.854	9.232	11.455				-4.980				-2.210											.540
Non basic expenditure	2.048	2.768	2.696					2.377			-1.986											.118
Non basic consumption	2.078	5.343	4.421								-2.051					-2.238						.210

$t_{.05} = 1.973$ $t_{.01} = 2.600$

food consumption. Education of the homemaker also indicated a significant effect in all equations except the one predicting food consumption. An examination of the results concerning other variables tested, revealed only one other factor which had important regression effects -- the amount of "income in kind" received by the family. This factor was represented by "percent of home produced food" in equations 1, 3, and 5 and by "source of income" in equations 2, 4, and 6. The source of income class "all nonfarm income" is highly correlated with the class "0 to 29.9 percent home produced food." Hence, it is likely that the latter class would have come in significant in equations 2, 4, and 6 had "source of income" not been specified. Amount of home produced food can, therefore, be regarded as an important predictor of both expenditure and consumption of the more basic living items.

It is understandable that nonmoney income in the form of income in kind should show up as an important explanatory factor. A closer examination of low-income Kentucky families showed that their total food consumption was more than double the amount of their food purchases. Home produced fuel was also indicated to be substantial, since the value of household operations (which included fuel) exceeded the cash operating costs in the home by 50 percent. Since food and household operations made up one-half of all family consumption, it is apparent that income in kind constituted an important component of the total resources of South Central Kentucky families.

Phase III -- Search for Interaction Effects

The same tests for interaction made in the last two chapters are repeated here. The first one, tests the hypothesis that home tenure and farm status each has an independent effect on food consumption. The results are given in Table 5.4, with the variance analysis indicating no apparent interaction

between home tenure and farm status. Tenure effect by itself is also shown to be relatively insignificant, but a high correlation is indicated between farm status and food consumption. The relatively small standard errors of the mean for "owners farm" and "renters farm," indicate that substantially less variation existed in the amounts of food consumed by farm families than by nonfarm families.

Table 5.4--Food consumption as a function of home tenure and farm status, low-income rural families, South Central Kentucky, 1957 (Figures in parentheses are standard errors)

Farm Status	Owner	Renter
Farm	$1,014 (40)	$1,081 (48)
Nonfarm	$ 661 (82)	$ 798 (80)

Source of variation	Analysis of variance					
	Sum of squares	d.f.	Variance	F	$F_{.05}$	$F_{.01}$
Farm status	101,124	1	101,124	20.30	3.89	6.78
Tenure	10,302	1	10,302	2.07	3.89	6.78
Interaction	1,232	1	1,232	0.25	3.89	6.78
Error		237	4,982			

The lack of significance in tenure effect can be attributed partly to distortion caused by a further hidden interaction. The ownership classes (especially farmers) command substantially more assets than do the renters, but the average size of their families is smaller -- owners averaged 3.08 members; renters averaged 4.03 members. Since family consumption tends to vary directly with both housing index and family size, the two forces offset each other in the present case. This gives an impression of less tenure effect than actually exists.

Another variance analysis tested the hypothesis that age and education do not interact in their effect upon less basic expenditures. The results are given in Table 5.5.

Table 5.5--Expenditures on less basic items as a function of age and education of homemakers, low-income rural families, South Central Kentucky, 1957 (Figures in parentheses are standard errors)

Education	Age		
	Less than 40	40 to 54	55 and above
Grades 0 to 7	N = 38 $691 (66)	N = 34 $603 (50)	N = 46 $493 (59)
Grade 8 or above	N = 46 $831 (62)	N = 46 $711 (60)	N = 31 $517 (65)

Source of variation	Analysis of variance					
	Sum of squares	d.f.	Variance	F	$F_{.05}$	$F_{.01}$
Age	66,089	2	33,044	8.58	3.04	4.71
Education	12,322	1	12,322	3.20	3.89	6.76
Interaction	3,677	2	1,839	0.48	3.04	4.71
Error		235	3,854			

In order to get adequate observations for each cell, the two education groups "grades 8 - 11" and "grade 12 plus" were combined into one class. The resulting categories: (1) grades 0 - 7 and (2) grade 8 and above were set up in rows, and the age categories were set up in columns.

The nonsignificant interaction effect indicates that education differences have not been found to vary from one age group to the next. Education effect itself, shows insignificance at the 5 percent level, while age effect was highly significant.

The biggest differences among age groups occurred with respect to the items "home furnishings and equipment" and "transportation." During the survey year, the group of families with homemakers under 40 spent $186 on furnishings and equipment and $201 on transportation. The comparable figures for families with homemaker 55 and over were $92 and $81, respectively. With respect to cost of medical care, however, the older group averaged more than the younger families -- the "55 and over" group spent $233 while the "under 40" group spent only $181.

Search Using Multivariate Techniques

A further search was undertaken to discover any significant interactions that might exist between the continuous and dummy variables employed in Table 5.3. Interaction terms which combined each continuous with each dummy variable were added to the specifications of Table 5.3 and the new set of regressions are reported in Table 5.6. A stepwise regression procedure was used, and the 5 percent level was stipulated for entering any interaction terms which differ significantly from other terms of their respective set. Only the significant interaction terms are listed in the independent variables column of Table 5.6, but all the other terms were specified except: X_2X_4 through X_2X_7; X_3X_4 through X_3X_7; X_2X_8 through X_2X_{10}; X_2X_{11} through X_2X_{13}; and X_2X_{16} through X_2X_{19}.

Housing index showed no significant interaction with any of the dummy variables. Money income also had no significant interaction except for one isolated instance: in estimating less basic consumption, money income showed greater explanatory effect in the class producing 30 to 49.9 percent of its food than in the three other classes. Since there was only this one isolated entry, we can assume there is little or no interaction between money income and the categorical variables at the 5 percent significance level.

Table 5.6--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, South Central Kentucky, 1957

Indpendent variables	Dependent variables							
	Total expend-iture	Total con-sumption	Food expend-iture	Food con-sumption	More basic expend-iture	More basic con-sumption	Less basic expend-iture	Less basic con-sumption
X_1 - Money inoome	2.991	3.572	2.319	1.948	3.162	3.737	1.878	1.559
X_2 - Housing index	4.739	8.912	2.500	1.509	5.605	9.495	2.511	4.229
X_3 - Log family size	8.876	11.701	12.226	7.154	12.270	12.577	5.205	6.356
Source of Income								
X_4 - 100 percent farm								
X_5 - 51 to 99 percent farm								
X_6 - 1 to 50 percent farm								2.263
X_7 - 0 percent farm		-2.424		-2.595		-2.585		
Age of Homemaker								
X_8 - Under 40								
X_9 - 40 to 54								
X_{10} - 55 and older								
Education of Homemaker								
X_{11} - Less than 8th grade								
X_{12} - Grades 8 through 11								

Table 5.6--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, South Central Kentucky, 1957 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
Education of Homemaker (Con!t.)								
X_{13} - Grade 12 and higher								
X_3X_{11}	-3.349	-3.892			-3.491	-3.545	-2.340	-2.915
X_3X_{12}			2.600					
Tenure of Home								
X_{14} - Owner								
X_{15} - Renter								
Percentage of Home Produced Food								
X_{16} - 0 to 29.9 percent								
X_{17} - 30 to 49.9 percent								
X_{18} - 50 to 64.9 percent	-3.367							
X_{19} - 65 percent and above	-5.622							
X_1X_{17}								2.868
X_3X_{16}		-2.313		-2.201		-2.656		
X_3X_{18}			-6.357		-4.952			
X_3X_{19}			-10.343		-9.162			
$\bar{R}^2$	.392	.524	.497	.244	.504	.561	.128	.233



The third continuous variable, log family size, did show significant interaction with "education of the homemaker" and also with "percentage of home produced food." The slope of the function, level of living on log family size, was significantly flatter when it applied to families headed by a homemaker with an education level below the 8th grade. Only when predicting total food consumption did log family size not interact significantly with education. Apparently there was a similar pattern of food consumption in both the higher and lower education tiers.

Family size also interacted with the amount of food raised at home. When the proportion of home produced food was large (50 percent or more), the log family size variable explained relatively less of the variance in more basic expenditure. But an opposite effect was observed with respect to more basic consumption. Here, log family size demonstrated relatively less regression effect with consumption when it was associated with the group producing the smallest proportion of food (less than 30 percent).

The interaction terms that came in significant in Table 5.6 actually serve as displacements for related categorical variables that were significant earlier (in Table 5.3). For example, where "less than 8th grade" had been the significant category in Table 5.3, the closely correlated "less than 8th grade x log family size" became the significant variable in Table 5.6. Similarly the interaction terms "percentage of home produced food x log family size" in the latter table, had closely correlated counterparts among the categorical variables of the earlier Table 5.3.

The $\bar{R}^2$ values in Table 5.6 were only moderately greater than comparable $\bar{R}^2$ values in Table 5.3 -- the increase ranged from less than 5 percent in equation 4 to 20 percent in equation 8. About one-third of this increase in explained variance could be attributed to a better fit produced by log family size in Table 5.6 (cf., natural family size in Table 5.3). This left

two-thirds of the increase attributable to added regression effects produced by the several interaction terms.

The analyses of Phase III suggest that interaction effects in Kentucky are of comparatively little importance. Therefore, the influence of changes in independent variables can, for the most part, be regarded as "additive" influences. This fact sets Kentucky off from the two Texas studies where the effects of interaction were of considerable consequence.

Families in the open country of South Central Kentucky were considerably more homogeneous with respect to most characteristics than the families in the open country of the two Texas areas. A much higher proportion of Kentucky families devoted most of their resources to farming and to the production of income in kind. Opportunities for off-farm earnings were very poor in Kentucky, whereas they were relatively favorable in the Texas study areas.

Phase IV -- The Final Regression Model

This phase of analysis aims to build a final multiple regression model which maximizes predictive value while holding the number of predictors down to a manageable level. An important portion of the predictor variables consists of joint classifications which attempt to combine the effects of several explanatory factors. The joint variables represent subgroups of families displaying uniformity of behavior within the group, but with sufficiently large differences among the groups to warrant separate classifications.

The set of joint variables try to combine the behavioral effects attributed to variations in age, source of income, and level of income. Two additional sets of dummy variables are also specified: one reflects the effects produced by different education levels; a second reflects the effects produced by different levels of home produced food. Two continuous variables, housing index and log family size, are also specified in the final model.

Regression Model for the Total Sample

A split between white and nonwhite families was not implemented in the Kentucky study because the number of nonwhites in the sample was considered inadequate for special consideration. The first major split used, therefore, was one which separated families with head over 65 years of age from those with head under 65. Table 5.7 gives a description of the subclassifications where age, source of income, and level of income were the classifying criteria. The table also gives the total number of cases falling under each subgroup.

Table 5.7--Subgroups and number of observations per group, rural families, South Central Kentucky, 1957

Group	Number of observations
Over 65, income below $1,000	28
Over 65, income $1,000 and above	33
Under 65 (entire income from farm) income below $1,500	35
Under 65 (entire income from farm) income $1,500 to $2,499	35
Under 65 (entire income from farm) income over $2,500	15
Under 65 (50 to 99 percent of income from farm) income below $1,500	29
Under 65 (50 to 99 percent of income from farm) income $1,500 to $2,499	25
Under 65 (50 to 99 percent of income from farm) income $2,500 and above	28
Under 65 (1 to 49 percent of income from farm) income below $2,500	32
Under 65 (1 to 49 percent of income from farm) income $2,500 and above	30
Under 65 (no income from farm) income below $2,500	38
Under 65 (no income from farm) income $2,500 and above	18

In contrast to the Texas studies, the Kentucky survey had not collected any data on the breakdown of income coming from nonfarm sources. Neither was there any information on gross sales of farm products or changes in farm inventories. Hence, the only "source of income" criterion that was available for the present classifications was the proportion of net income derived from farm versus nonfarm sources. In addition, designators of the "level of income" were provided by class intervals based on money income before taxes.

Twelve designated joint variables became X_3 through X_{14} of Table 5.8. Six other categorical variables, X_{15} through X_{20}, accounted for education effects and for the effects of nonmoney income. The two continuous variables X_1 and X_2 made allowance for the family's size and the adequacy of their housing.

The group, "family head under 65 (50 to 99 percent of income from farm) income $1,500 to $2,499," was arbitrarily selected from among the joint classifications as the variable to be omitted. It, therefore, became the base of comparison with eleven other variables. Computed t values indicated that most other classifications of families lived on a lower level than the families of the omitted variable. Two classes in particular (1) over 65, income below $1,000; and (2) under 65 (no income from farm) income below $2,500, consumed and spent at a significantly lower level than omitted families. But when estimating "less basic" items, only the first of these classes (over 65, income below $1,000) was significantly below the omitted variable.

It is interesting to note in equation 4 that with (log) family size controlled the older families were not consuming significantly less food than families in the omitted class. The only class which did consume significantly less food was the "under 65 (no income from farm) below $2,500"

Table 5.8--Computed t values in regressions of family levels of living on household characteristics, rural families, South Central Kentucky

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X1 - Housing index	6.312	10.876	2.725	2.418	6.188	10.760	4.349	6.868
X2 - Log family size	7.596	11.059	9.483	9.345	8.241	12.175	4.406	5.367
X3 - Over 65, income below $1,000	-3.654	-3.496	-2.269	-0.832	=2.589	-2.928	-3.553	-2.901
X4 - Over 65, income $1,000 and above	-1.436	-1.164	-1.041	-0.852	-0.645	-0.924	-1.785	-1.033
X5 - Under 65 (entire income from farm) income below $1,500	-1.013	-1.697	-0.985	-1.591	-0.765	-1.663	-0.936	-1.093
X6 - Under 65 (entire income from farm) income $1,500 to $2,499	-1.347	-1.586	-0.866	-0.995	-0.612	-0.984	-1.666	-1.766
X7 - Under 65 (entire income from farm) income over $2,500	0.637	1.066	-0.068	0.266	1.074	1.001	-0.030	0.743
X8 - Under 65 (50 to 99 percent of income from farm) income below $1,500	-1.762	-2.327	-1.407	-1.948	-1.482	-2.139	-1.470	-1.682
X9 - Under 65 (50 to 99 percent of income from farm) income $1,500 to $2,499	---------- omitted variable ----------							
X10 - Under 65 (50 to 99 percent of income from farm) income $2,500 and above	1.936	2.478	1.017	1.171	2.566	2.918	0.638	0.955
X11 - Under 65 (1 to 49 percent of income from farm) income below $2,500	-1.402	-1.685	-1.307	-1.828	-1.466	-2.089	-0.871	-0.511
X12 - Under 65 (1 to 49 percent of income from farm) income $2,500 and above	1.601	1.368	0.523	0.122	1.589	0.686	1.083	1.737
X13 - Under 65 (no income from farm) income below $2,500	-2.347	-3.343	-2.486	-2.617	-2.059	-3.409	-1.868	-1.977
X14 - Under 65 (no income from farm) income $2,500 and above	0.950	-0.665	0.639	-0.920	1.992	-0.732	-0.450	-0.324

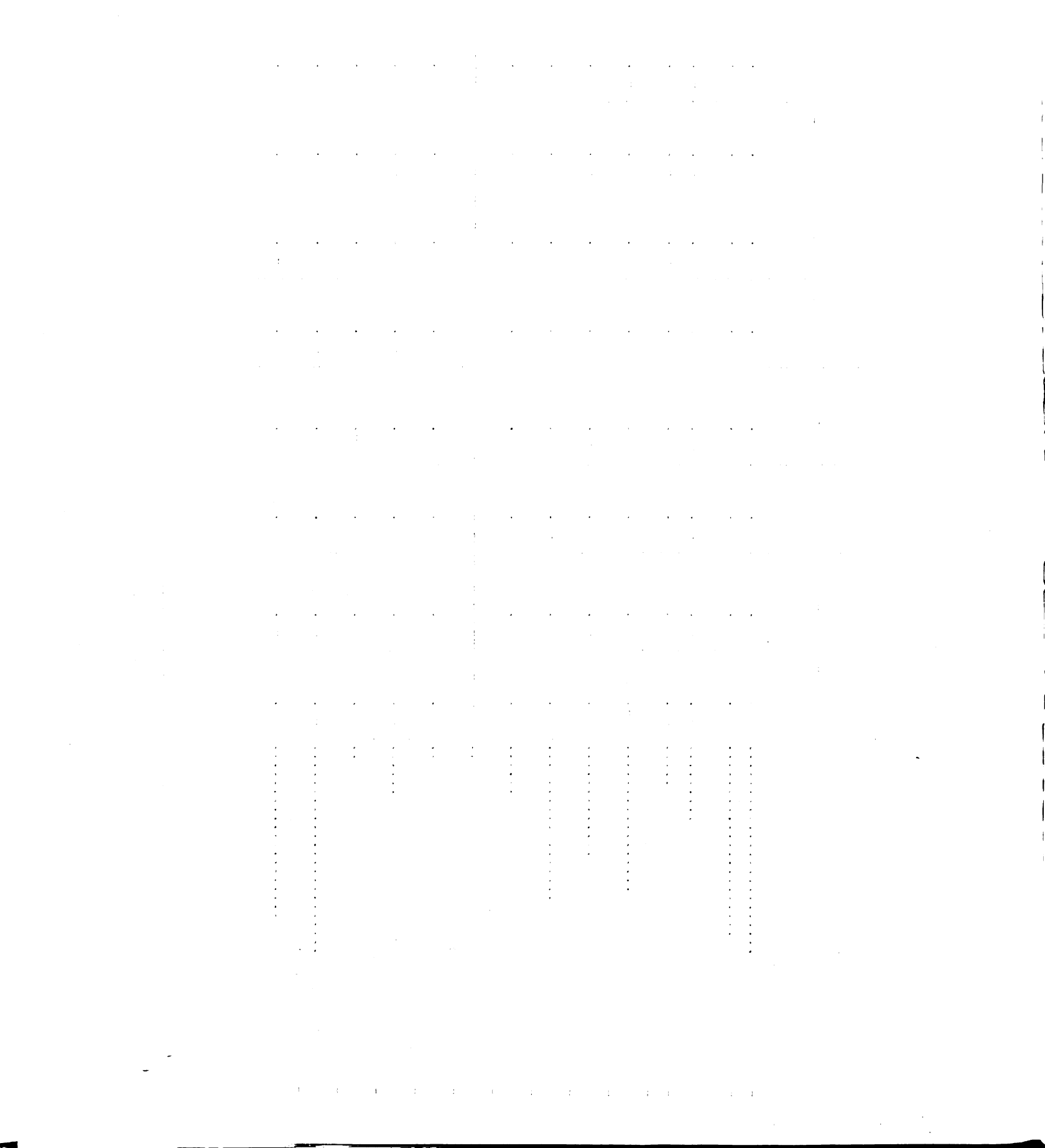


Table 5.8--Computed t values in regressions of family levels of living on household characteristics, rural families, South Central Kentucky (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_{15} - Homemaker less than 8th grade	omitted variable							
X_{16} - Homemaker 8th grade and above	0.946	1.451	0.624	0.241	1.057	1.224	0.517	1.194
X_{17} - 0 to 29.9 percent home produced food	2.167	-0.673	3.469	-1.632	3.040	-0.920	0.540	-0.093
X_{18} - 30 to 49.9 percent home produced food	omitted variable							
X_{19} - 50 to 64.9 percent home produced food	-2.096	-0.070	-3.890	0.573	-2.942	0.156	-0.520	-0.340
X_{20} - 65 percent plus, home produced food	-4.604	-0.383	-7.688	1.906	-6.202	-0.356	-1.412	-0.272

class. This group had averaged considerably less home produced food than any of the other classifications.

Education effects did not appear to be significant in any of the 8 equations. The "percent of home produced food" effects were significant only in equations 1, 3, and 5 where more basic expenditures were being estimated. Housing index and log family size were significant variables throughout.

A new set of regressions ran the same independent variables, X_1 through X_{20} of Table 5.8, against a log transformation of each dependent variable. Except when predicting Y_8, the logarithmic specification of the dependent variable resulted in substantially higher adjusted coefficients of multiple determination, particularly when estimating food purchases and consumption. Table 5.9 gives a comparison of $\bar{R}^2$ values.

Table 5.9--A comparison of adjusted coefficients of multiple determination: Regressions of natural Y and log Y on the independent variables listed in Table 5.8

Dependent variables	$\bar{R}^2$ pertaining to	
	Natural Y	Log Y
Y_1 - Total expenditure	.474	.514
Y_2 - Total consumption	.571	.602
Y_3 - Food expenditure	.507	.593
Y_4 - Food consumption	.310	.432
Y_5 - More basic expenditure	.527	.597
Y_6 - More basic consumption	.577	.613
Y_7 - Less basic expenditure	.234	.266
Y_8 - Less basic consumption	.329	.318

In the set of equations where log Y had been specified, the regression coefficient of log family size increased substantially in the first six equations, but decreased in equation 8. The log Y form also produced larger t values associated with family groups that spent or consumed at comparatively low levels, and produced smaller t values associated with groups living on a higher plane. This same phenomenon had been observed in connection with the two Texas studies as well.

Regression Model for the Subsample

A new set of joint variables combining the effects of age, source of income and level of income were defined for the subsample of low-income families. Definitions for these classifications and the number of observations falling into each class are given in Table 5.10. The major split employed was the one dividing the families into groups with "head over 65" and "head under 65."

Table 5.10--Subgroups and number of observations per group, low-income rural families, South Central Kentucky, 1957

Group	Number of observations
Over 65, income below $1,000	28
Over 65, income $1,000 to $2,499	19
Under 65 (entire income from farm) income below $1,500	35
Under 65 (entire income from farm) income $1,500 to $2,499	35
Under 65 (50 to 99 percent of income from farm) income below $1,500	29
Under 65 (50 to 99 percent of income from farm) income $1,500 to $2,499	25
Under 65 (1 to 49 percent of income from farm) income below $1,500	15
Under 65 (1 to 49 percent of income from farm) income $1,500 to $2,499	17
Under 65 (no income from farm) income below $1,500	19
Under 65 (no income from farm) income $1,500 to $2,499	19

The computed t values for the 8 equations of the subsample regression model are given in Table 5.11. The ten joint classifications are specified as a set of dummy variables, X_3 through X_{12}, in the final equations. The remainder of the specified independent variables were the same as in the model for the total sample. Likewise, the classes omitted from each set of dummy variables were the same ones in both models.

The regression performance of the independent variables was similar in the subsample and the total sample. Perhaps the greatest difference between them was that housing index demonstrated substantially less predictive power in the subsample -- probably because income in kind was assuming greater importance. In fact, for predicting food purchases or food consumption, housing index became insignificant at the 5 percent level. On the other hand, the effects of education and proportion of home produced food showed somewhat more explanatory effect in the subsample.

There is also strong evidence from the Table 5.11 analysis that a number of part-time farmers were misclassified when grouped with extremely low-income families. We find the X_{10} group "under 65 (1 to 49 percent income from farm) income $1,500 to $2,499" living at a lower level than the X_9 group "under 65 (1 to 49 percent income from farm) income below $1,500." In the latter group there were a number of families whose net income from the farm component was very small in 1956, and this contributed to their placement with a class whose net income from all sources was less than $1,500. Indications were that normal farm income for these particular X_9 families would be considerably higher than it was in the survey year. And normally they would, therefore, belong in a higher income classification.

The actual estimating equations for the regressions pertaining to the subsample of low-income families are listed in Table 5.11a.

Table 5.11--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, South Central Kentucky, 1957

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_1 - Housing index	4.140	8.099	1.848	1.343	4.921	8.833	2.013	3.301
X_2 - Log family size	6.871	10.550	7.908	6.340	7.954	11.778	3.534	4.722
X_3 - Over 65, income below $1,000	-4.116	-4.117	-2.532	-0.947	-3.065	-3.594	-3.534	-2.993
X_4 Over 65, income $1,000 to $2,499	-2.196	-1.426	-1.358	-1.159	-0.921	-1.060	-2.530	-1.249
X_5 - Under 65 (entire income from farm) income below $1,500	-1.092	-1.992	-0.992	-1.545	-0.810	-1.935	-0.954	-1.165
X_6 - Under 65 (entire income from farm) income $1,500 to $2,499	-1.486	-1.798	-0.891	-0.914	-0.693	-1.139	-1.650	-1.802
X_7 - Under 65 (50 to 99 percent of income from farm) income below $1,500	-2.000	-2.750	-1.536	-1.955	-1.767	-2.660	-1.502	-1.703
X_8 - Under 65 (50 to 99 percent of income from farm) income $1,500 to $2,499	omitted variable							
X_9 - Under 65 (1 to 49 percent of income from farm) income below $1,500	-0.376	-1.062	-1.004	-1.753	-0.815	-1.786	0.135	0.214
X_{10} - Under 65 (1 to 49 percent of income from farm) income $1,500 to $2,499	-2.193	-2.123	-1.570	-1.285	-2.111	-2.431	-1.496	-0.880
X_{11} - Under 65 (no income from farm) income below $1,500	-3.320	-4.226	-3.498	-2.579	-3.223	-4.161	-2.243	-2.531
X_{12} - Under 65 (no income from farm) income $1,500 to $2,499	0.000	-1.548	-0.375	-1.624	0.251	-2.292	-0.216	-0.046

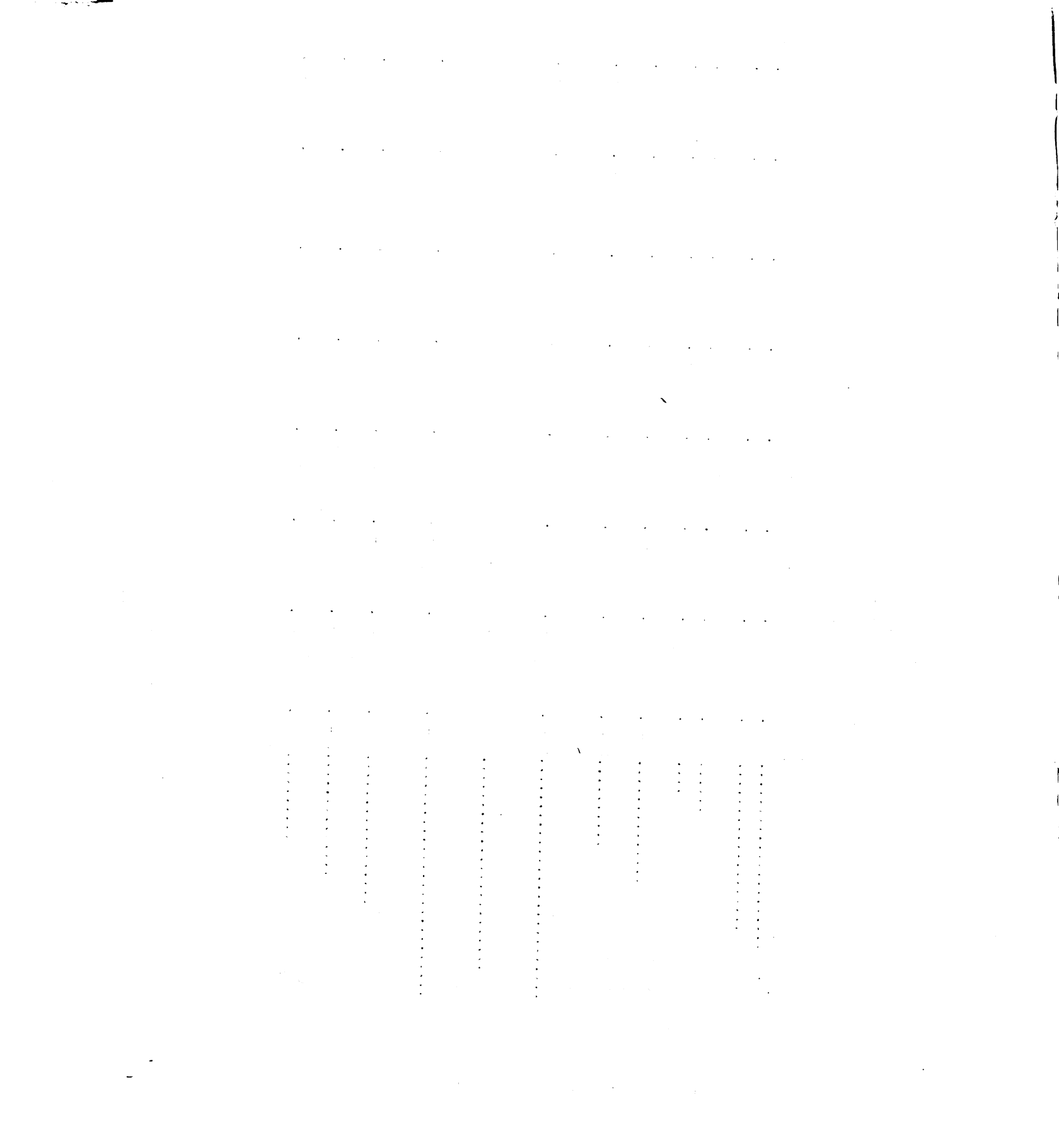


Table 5.11--Computed t values in regressions of family levels of living on household characteristics, low-income rural families, South Central Kentucky, 1957 (Continued)

Independent variables	Dependent variables							
	Total expenditure	Total consumption	Food expenditure	Food consumption	More basic expenditure	More basic consumption	Less basic expenditure	Less basic consumption
X_{13} - Homemaker less than 8th grade	omitted variable							
X_{14} - Homemaker 8th grade and above	1.687	2.669	1.043	0.785	1.802	2.593	0.993	1.634
X_{15} - 0 to 29.9 percent home produced food	0.345	2.386	2.203	-1.675	1.377	-1.953	-0.665	-1.883
X_{16} - 30 to 49.9 percent home produced food	omitted variable							
X_{17} - 50 to 64.5 percent home produced food	-2.317	-0.797	-3.674	0.346	-3.144	-0.004	-0.793	-1.374
X_{18} - 65 percent plus, home produced food	-4.906	-1.169	-7.492	1.449	-6.935	-0.791	-1.438	-1.114

Table 5.11a--Actual estimating equations in regressions of family levels of living on household characteristics, low-income rural families, South Central Kentucky, 1957

$Y_1 = 1308.438 + 17.653\ X_1 + 1499.268\ X_2 - 660.911\ X_3 - 358.262\ X_4 - 158.776\ X_5$
$- 215.898\ X_6 - 305.897\ X_7 - 69.541\ X_9 - 383.565\ X_{10} - 601.396\ X_{11}$
$+ 0.071\ X_{12} + 130.455\ X_{14} + 41.852\ X_{15} - 239.644\ X_{17} - 532.287\ X_{18}$

$Y_2 = 1843.830 + 39.870\ X_1 + 2657.735\ X_2 - 763.312\ X_3 - 268.645\ X_4 - 334.275\ X_5$
$- 301.704\ X_6 - 485.744\ X_7 - 226.962\ X_9 - 428.876\ X_{10} - 883.891\ X_{11}$
$- 327.995\ X_{12} + 238.353\ X_{14} - 333.335\ X_{15} - 95.205\ X_{17} - 146.494\ X_{18}$

$Y_3 = 337.376 + 2.881\ X_1 + 630.866\ X_2 - 148.664\ X_3 - 80.968\ X_4 - 52.714\ X_5$
$- 47.351\ X_6 - 85.891\ X_7 - 67.918\ X_9 - 100.438\ X_{10} - 231.649\ X_{11}$
$- 25.176\ X_{12} + 29.476\ X_{14} + 97.456\ X_{15} - 138.931\ X_{17} - 297.231\ X_{18}$

$Y_4 = 550.024 + 4.993\ X_1 + 1301.470\ X_2 - 132.665\ X_3 - 164.938\ X_4 - 195.818\ X_5$
$- 115.864\ X_6 - 260.779\ X_7 - 282.949\ X_9 - 196.008\ X_{10} - 407.364\ X_{11}$
$- 259.863\ X_{12} + 52.958\ X_{14} - 176.757\ X_{15} + 31.223\ X_{17} + 137.152\ X_{18}$

$Y_5 = 681.768 + 11.969\ X_1 + 989.911\ X_2 - 280.720\ X_3 - 85.684\ X_4 - 67.203\ X_5$
$- 57.468\ X_6 - 154.151\ X_7 - 86.069\ X_9 - 210.629\ X_{10} - 332.981\ X_{11}$
$+ 26.283\ X_{12} + 79.477\ X_{14} + 95.021\ X_{15} - 185.464\ X_{17} - 429.242\ X_{18}$

$Y_6 = 1145.122 + 29.047\ X_1 + 1969.823\ X_2 - 442.400\ X_3 - 132.568\ X_4 - 221.176\ X_5$
$- 126.879\ X_6 - 311.823\ X_7 - 253.408\ X_9 - 326.056\ X_{10} - 577.705\ X_{11}$
$- 322.405\ X_{12} + 153.985\ X_{14} - 181.168\ X_{15} - 0.325\ X_{17} - 65.817\ X_{18}$

$Y_7 = 626.642 + 5.684\ X_1 + 509.413\ X_2 - 380.181\ X_3 - 272.570\ X_4 - 91.574\ X_5$
$- 158.433\ X_6 - 151.745\ X_7 + 16.543\ X_9 - 172.940\ X_{10} - 268.411\ X_{11}$
$+ 26.215\ X_{12} + 50.981\ X_{14} - 53.173\ X_{15} - 54.182\ X_{17} - 103.047\ X_{18}$

$Y_8 = 698.712 + 10.823\ X_1 + 687.906\ X_2 - 320.913\ X_3 - 136.078\ X_4 - 113.098\ X_5$
$- 174.824\ X_6 - 173.921\ X_7 + 26.445\ X_9 - 102.819\ X_{10} - 306.186\ X_{11}$
$- 5.589\ X_{12} + 84.367\ X_{14} - 152.167\ X_{15} - 94.880\ X_{17} - 80.677\ X_{18}$

Y_1 through Y_8 and X_1 through X_{18} are the same variables listed in Table 5.11.

When a log transformation of the 8 dependent variables was run in place of their natural forms, the resulting increase in explained variance ranged from 6 percent in equation six to 84 percent in equation 4. A comparison of $\bar{R}^2$ values is given in Table 5.12.

Table 5.12--A comparison of coefficients of multiple determination: Regressions of natural Y and log Y on the independent variables listed in Table 4.11

Dependent variables	$\bar{R}^2$ pertaining to	
	Natural Y	Log Y
Y_1 - Total expenditure	.419	.461
Y_2 - Total consumption	.520	.563
Y_3 - Food expenditure	.495	.573
Y_4 - Food consumption	.214	.393
Y_5 - More basic expenditure	.509	.553
Y_6 - More basic consumption	.549	.583
Y_7 - Less basic expenditure	.161	.258
Y_8 - Less basic consumption	.223	.263

The log Y form produced sizeable increases in the absolute t values associated with family groups whose heads were over 65. Similarly, as pertains to the X_{11} group, there were substantial increases in the size of t, but only in the consumption equations 2, 4, 6 and 8. And in the X_{18} group, which produced 65 percent or more of its food, the log Y form produced larger t values in the expenditure equations 1, 3, and 5.

In the remainder of the analysis, there was a little evidence of systematic changes in the computed t values when going from a natural Y to a log Y specification.

Allowance for Interaction Effects

The apparent interaction effects in the Kentucky study were rather minor compared with those discovered in the Texas studies. It may be recalled that in the Texas studies there were two main dichotomies which interacted significantly with family size: (a) white vs. nonwhite families, and (b) families which received their chief income from wages, salaries, professions or business vs. those whose chief income was from other sources. It was not possible in the Kentucky study to specify these two-sided characters of the "race" and "source of income" traits.

Even if these dichotomous characteristics had been specified in the Kentucky analysis, it is unlikely that the same degree of interactions would have been apparent. An examination of the three populations (Appendix A), distinguishes the properties of the Kentucky counties as substantially different from the Texas areas. Furthermore, the samples drawn from these populations also suggest that separate race and source of income specifications would have had a smaller impact in South Central Kentucky. In the latter area the opportunities for off-farm employment, for instance, were not nearly so favorable as in the Texas areas.

On the whole, the Kentucky data did not contain sufficient evidence of interaction effects to warrant additional specifications of interaction terms to the independent variables specified in Tables 5.8 or 5.11. The Phase III analysis did indicate, however, that the error sum of squares could be appreciably reduced if the two education categories and also the four "percent of home produced food" categories were multiplied by log family size, and the combinations specified as interaction terms in the final model.

CHAPTER VI

THE STUDIES COMPARED

Population Comparisons

People in the two Texas survey areas were generally much better off than the ones in South Central Kentucky. For the survey counties as a whole, the median income was nearly 64 percent higher in Texas than in Kentucky, but the comparison shrank to 43 percent for the rural people and to 36 percent in the rural farm portion. Actual welfare discrepancy in the rural farm area is probably below this percentage, since income in kind was more significant in Kentucky than in Texas. The smallest discrepancy was among nonwhites -- the median income of Texas nonwhite families was only about 6 percent higher than in Kentucky.

Eighty-six percent of the Kentucky people were classified "rural" and three-fifths of these were "rural farm." Less than half the Texas people were classified "rural" and of these "rural farm" comprised only one-third in the Blackland; less than one-fifth in East Texas.

The median level of education was approximately two years higher in Texas, but this discrepancy with Kentucky shrank to one year among rural and nonwhite persons over 25. Of all employed persons, nonwhites made up roughly 6 percent in Kentucky, 11 percent in the Blackland, and 25 percent in East Texas.

Sample Comparisons

An examination of the samples drawn from the three populations indicates farming was the chief pursuit of open country dwellers in Kentucky, while nonfarm pursuits predominated in East Texas. Farming was the major source of income for 55 percent of the families in Kentucky, 33 percent in the Blackland, and 15 percent in East Texas. But 79 percent of

Kentucky, 74 percent of Blackland, and 51 percent of East Texas family heads were engaged in farming at least part-time.

Even a larger proportion of families raised some of their food -- 91 percent in Kentucky, 92 percent in the Blackland, and 79 percent in East Texas. In the Kentucky sample, 55 percent of the families produced more than half the food they consumed, but in the Texas samples this figure was only 17 percent.

East Texas seemed to have the best opportunities for off-farm earnings. The Blackland had a more viable agriculture, however, and this appeared to put open country dwellers in the two Texas areas more or less on an economic par. In Kentucky, the opportunities seemed about equally poor on or off the farm.

Analytic Comparisons

Phase I -- Money Income, Housing Index, and Family Size as Explanatory Factors

One objective of this study was to discover criteria that are readily obtainable and can serve as possible indicators of the family's need for public assistance. Initially three factors were selected as likely determinants of the family's living level, and hence as a guide to their probable need for assistance.

Money income, although regarded as an important enabling condition, proved to be quite ineffective as a predictor of the living level of low-income families. When specified as a continuous linear variable, its influence ranged from statistically insignificant in East Texas to significant at the 1 percent level in four out of eight equations in Kentucky. When specified alternatively as categorical variables, it produced considerable improvement over the linear fit in East Texas, but only a small improvement in the Blackland and no improvement in Kentucky.

The adequacy of the family home was also regarded as a factor that would be correlated with the general level of living. The variable used to measure the home's adequacy was the so-called housing index, and it turned out to be a good predictor of all 8 dependent variables. The area of dependent variables in which housing index was least effective as a predictor was connected with food purchases or consumption. Even here, though, there was only one case (in the equation predicting food consumption for Kentucky) where the housing index regression coefficient was not significant at the 1 percent level.

Family size also proved to be a good predictor of the dependent variables. In Kentucky it was a highly significant variable throughout, but in Texas studies it came in strongly only in the first six equations. In equations 7 and 8 which estimated "less basic" living, it was insignificant at the 1 percent level.

In each study, the log form of family size showed a better regression fit than the natural form in all estimating equations generated by low-income families.

The three independent variables, specified simultaneously, explained substantially more variance in the 4 consumption equations than in the 4 expenditure equations. But there was one case in Kentucky where the reverse was true: the coefficient of multiple determination was larger for the equation predicting food expenditure than for the one predicting food consumption.

Phase II -- Additional Explanatory Factors Specified in the Regressions

When several additional household traits were specified as independent variables, the gap between "explained expenditures" and "explained consumption" was narrowed considerably. The "percentage of home produced food"

variable did more than any other to bring up the explained variance in equations 1, 3, and 5 which estimated more basic expenditures. Home produced food could serve as a measure of nonmoney income, and when allowed for in the expenditure equations, did much to boost the explained sum of squares of more basic expenditures.

Age of the homemaker was a significant variable in both Texas studies. In the Blackland, families with homemaker 55 or older spent and consumed less than younger families. In East Texas, the same role was played by families which received their chief income from transfer payments, rent, and interest, i.e., mostly families receiving old age assistance. In Kentucky, the degree to which older families may have lived on a lower level was insignificant at the 5 percent level when several other factors had been controlled.

The homemaker's education was also significant. In East Texas, families whose homemakers had attained grade 12 or above, lived on a higher level than the lesser educated. In Kentucky the families with homemaker below the 8th grade lived on a lower level than the better educated. In Blackland, education effect was not so evident and it was necessary to run the total sample in order to get the class with least education to enter the stepwise regression at the 5 percent level. In none of the 3 studies was there a significant correlation between education and the level of food consumption.

In the Texas areas, nonwhites consumed and purchased considerably less food and shelter than whites. In Kentucky, there were too few nonwhites to permit separate analysis.

Phase III -- Search for Interaction Effects

A number of interesting and insightful interactions were discovered among the independent variables. One technique used to find the interactions was the "factorial experiment" in which two factors were studied simultaneously. The basis for this technique was to postulate hypotheses concerning interactions, then test them directly by an analysis of variance.

One hypothesis so tested was that farm status and home tenure were independent in their effect on the amounts of food consumed. At the 5 percent level, this hypothesis was rejected in East Texas and at the 10 percent level it was rejected in the Blackland as well. In Kentucky, however, no evidence was produced which would reject the null hypothesis. The conclusion was that tenure and farm status tended to interact in their effects on food consumption in the Texas studies, but in Kentucky these two effects could be regarded as being additive.

Another postulate was that age and education interacted in determining the less basic expenditures. The hypothesis tested was that the effects of age and education were independent. The F value calculated for the Blackland was significant at the 5 percent level and for East Texas it was significant at the 10 percent level. But in Kentucky one could only conclude that education differences did not vary from one age group to the next.

A further test for interactions employed the specification of interaction terms in multivariate regressions. An interaction term can be defined as the product of a continuous variable multiplied by a dummy variable. When this term comes in "significant," it indicates the existence of interaction between the continuous variable (which is one component of the term) and the set of categorical variables represented by the second component of the term.

The null hypothesis postulated in each case was that the slope of the dependent variable on a continuous independent variable was invariant for a particular set of mutually exclusive subclassifications of families. The technique used to test this hypothesis is the equivalent of an analysis of covariance which tests for homogeneity of regression coefficients. The hypothesis was rejected in a number of cases when the 5 percent level of significance was employed.

The most important interactions uncovered were between family size and several sets of dummy variables. One of these sets consisted of "white" and "nonwhite" families. In both Texas studies, log family size was a significantly better predictor of "more basic" living, when white rather than nonwhite families were considered in the regression.

Family size also interacted significantly with the "major source of income" categories. Log family size had more regression effect when it pertained to families receiving most of their income from wages, salaries, professions or business. In the Blackland, this relationship held in cases predicting the "more basic" items, whereas in East Texas, it held true when predicting "less basic" items.

The "percentage of home produced food" categories represented a third set of dummy variables which interacted with family size. Although there was considerable variation in the indicated relationships, the following generalization is nearly correct: in equations 1, 3, and 5 purchases rose more steeply with additional family members, for those families producing a small proportion of their food; alternatively, the purchase function was flatter for families producing a high proportion of their food. In general, an opposite effect could be observed in the consumption functions 2, 4, and 6. Logically this is the pattern one would expect among families with limited economic means.

In East Texas, intercorrelations were the cause of a couple of notable deviations from the generalization. There was a disproportionate number of nonwhites in the category producing 10 to 29.9 percent of their food. Consequently, this group tended to behave like the "nonwhite x family size" group, rather than as a group which produced a relatively small percent of its food. The category producing 30 to 49.9 percent of their food contained, in many respects, the most viable families of the 4 groups, and in fact behaved very differently from the group which produced 50 percent or more food. Thus, in equations 7 and 8, the 30 to 49.9 percent group had a steeper function than the other three groups.

In summary, it can be said that certain intercorrelations present in the East Texas study, caused Groups I and III (producing 0 to 9.9 percent and 30 to 49.9 percent respectively) to behave approximately as the generalization would attribute to the low level food producing groups. On the other hand, Groups II and IV behaved about as the high producers might be expected to act under the general postulate.

A fourth set of dummy variables that interacted significantly with family size were the ones defined under "education of the homemaker." But it was only in Kentucky where the interaction term "less than 8th grade x log family size" entered nearly all the equations. A parallel interaction term did not consistently enter either of the Texas analyses, and this was probably another case where intercorrelations got in the way. In Texas, interaction terms which multiplied log family size by "race" or "major source of income," came in significant and may have attenuated the interaction effects of the correlated "education" variable. In Kentucky it will be recalled, race and major source of income were not specified at all and were therefore unable to attenuate the education effect.

Another interesting set of interactions were discovered between housing index and home tenure. The general conclusion that could be drawn from the two Texas surveys was that housing index was a better predictor of "more basic" living when families owned rather than rented their homes. There was one exception to this rule in the Blackland, but certain circumstances indicated it may have been an apparent rather than a real exception.

Phase IV -- The Final Regression Model

In the final phase an attempt was made to combine some relevant characteristics and their subclasses into complex theoretical constructs. It was difficult to carry this task to its full logical conclusion because of the need for a minimum number of observations in each joint variable grouping. But the process did afford an opportunity to build the more apparent interactions and intercorrelations right into the combination predictors.

Preliminary analyses had suggested two dichotomous divisions which might properly be applied to the samples. One was a division by race into "whites" and "nonwhites"; a second was a division by age into families with heads "over 65" and "under 65." Subsequent analyses of the final regression models further strengthened the belief that these two were the more significant initial splits that should be adopted.

In order to achieve greater homogeneity within groups, some further subdivision appeared to be advisable. "White families with head under 65" were obviously still a very heterogeneous group and additional subclassifications were called for. In the Texas studies, "major source of income" became a classifier which suggested a division between families with incomes mainly from wages, salaries, professions or business, and those who received their incomes mainly from any other source.

Classification by net income created considerable distortion, especially when some large farmers were classified with very low-income families. To remove at least some of this distortion, a new classifier was added in the Texas studies. The "white under 65" families were subdivided into commercial farmers, noncommercial farmers, and nonfarmers by grouping according to their gross farm sales plus change in inventory. It was not possible to remove all the bias in this way, however, because "net incomes below $2,500" continued to be the designation for so-called low income families. Hence, some families that were not really poor continued to be included in this group.

In the Kentucky survey no information had been collected either on major sources of income or on gross farm sales. Thus, families with heads under 65 were classified by the amount of their net incomes which was derived from farming. Money income categories were also employed in the Kentucky study, mainly because in this survey there were few other factors available for subdividing the larger groups.

Limitations of the sample size did not permit further subdivisions in the set of joint variables specified for each final model. But it was decided that two other sets of mutually exclusive family classifications ought to be specified as dummy variables. One set reflected the level of the homemaker's education; a second set reflected the level of home food production. Two more factors, farm status and home tenure, were not specified explicitly because the analysis suggested that their effects were sufficiently reflected within the larger set of joint variables.

Housing index and log family size were specified as continuous variables, although it was recognized they could not operate additively with some categorical variables in the same equations. In other words, the regression coefficients of the continuous variables were known to represent

a compromise of several different effects attributed to interactions. The compromise coefficient, nevertheless, turned out to be significant in nearly all the equations.

Regression results showed that "whites with family head over 65" were very low consumers of family living items, particularly when the bulk of their income was from transfer payments, rent, and interest. Nonwhite families also ranked very low particularly with respect to food and shelter. With respect to other major living items, however, differences between low-income white and nonwhite families tended to be statistically insignificant.

A third group that represented very low level consumers were nonfarmers in the Kentucky sample. In the Texas areas, however, nonfarmers had a relatively good standing because of their superior off-farm earning opportunities. In fact, when the total samples (including all the higher incomes) were analyzed, the highest income nonfarmers in the Blackland equalled the living level of the large commercial farmers. In East Texas, the highest income nonfarmers far exceeded the level of commercial farmers, and also the level of high income noncommercial farmers.

Regression results in the final models further indicated that education was not a significant determinant of the dependent variables. But we know the specified complex variables had captured much of the education effect. For example, the divisions created by the "race" and "age" splits did much to create cells which were more nearly homogeneous with respect to the homemaker's education.

The proportion of food home produced turned out to be a significant predictor of more basic living in the final model. Here again there was some intercorrelation with the complex variables, but the degree of this relation was not sufficient to cause the "percent of home produced food" effects to appear insignificant.

When a log transformation was applied to the dependent variables, a small substantial improvement in regression fit was observed in nearly all cases. A notable exception occurred in the Blackland, however, where the natural form of "more basic" consumption had a better fit than the log form. It is difficult to account for this exception, but the explanation may be related to the erratic regression behavior in the variables when used to explain food consumption (food being an important part of the "more basic" consumption).

In Kentucky, where domestic food production was extensive, only two of the variables specified in the final model were indicated as significant predictors in estimating food consumption. When a log transformation was applied to food consumption, several other categorical variables also became significant and helped to increase explained variance by 64 percent. On the other hand, in the Blackland where home produced food was a much smaller proportion of total food, several variables contributed significantly to the fluctuations in food consumed. (See Table 4.12). When "log food consumption" was specified in this analysis, no change was effected in the percentage of variance that could be explained.

In general, the only predictor variables that were consistently significant in explaining food consumption were race, family size, and percent of home produced food. In regressions where these three factors were not controlled, farm status came in as a very decisive factor in determining the amount of food consumed.

Housing index as an explanatory factor, showed consistently less power to predict food expenditure or consumption than to predict several other major items of family living. If the housing variable is in fact a proxy for family assets, this suggests that as families acquire more net worth, they do not strive as hard for more and better food as they strive for many

other things. This is a pattern of behavior that one might expect among families with limited economic means who live in the open country. Up to a point they already possess the means to expand or improve their diets, so other items are given priority when new means become available to improve the family level of living.

It is probably that a high correlation exists between the quality of family housing and the level of family assets. Under normal accretion of goods, one would expect a family to devote new resources proportionally to the improvement of the home and to the acquisition of household durables and other items such as automobiles. The additions to home improvement would tend to be more lumpy than the acquisition of other durables, but on the average they could be expected to move together.

CHAPTER VII

CONCLUSIONS

This study has demonstrated that certain social, economic, and demographic characteristics contribute significantly to predicting the living level of low-income rural families. A useful byproduct of the study has been a few new insights into the interrelationship of various family consumption and expenditure characteristics in depressed rural areas.

The findings of the study should be of some help to the architects of the different antipoverty programs in their efforts to design or adapt programs for characteristics of rural poverty and the conditions of the rural areas. For example, certain results of the study have implications for the extent of participation that can be expected in the Food Stamp Program. People who are eligible to participate in the Food Stamp Program purchase food coupons approximating the value of "normal" food expenditures and then receive additional free or bonous coupons which can be used in purchasing additional food at retail stores. The proportion of bonus coupons that would need to be offered to induce participation can be expected to vary with such factors as "proportion of home produced foods" and "source of the family's major income." In other words, a so-called "normal" level for food expenditure actually varies a great deal among rural family groups. Hence, the guidelines for issuing coupons should ideally be different for the various groups. Administrative feasibility would, of course, set a limit to the extent that such action might be implemented.

This study should also be of help in suggesting some of the characteristics (and an analytic procedure for determining other possible characteristics) that can be used in identifying the rural poor. Knowledge of the distribution of the relevant population groups is necessary in designing equitable

formulas or systems for allocating program funds between states and other smaller political jurisdictions.

Administrators of public programs can use methodology similar to or building upon that employed in this study for designing analysis to evaluate the programs and see if they are achieving their objectives. A base period measure of consumption by categories could be compared to later measures in order to obtain a notion of the changes effected by the programs. A similar comparison between two periods could also be made regarding the performance of the economy in the region involved.

In any case, it should be kept in mind that this analysis is basically a pilot study. Our present lack of fundamental knowledge on the consumption characteristics and behavior of the rural poor is such that there are still many open questions, even about the best way to design research to improve our knowledge. Thus, one of the more important results of the present study lies in what it suggests for the design of further research. More research is needed before our knowledge in this area of human behavior is really adequate to answer some of the more urgent questions that arise in the design and implementation of programs to aid the rural poor.

Empirical and Theoretical Conclusions -- Dependent Variables

Either family expenditures or family consumption were used as measures of the living level and both these criteria were set up as the dependent variable in a series of regression equations. Any given set of predictor variables could explain a greater percentage of variance in the consumption equations than in the expenditure equations. It was necessary during the course of the survey, however, to collect more data for determining the consumption measure, than would have been required if only current expenses were to be obtained.

There is a theoretical basis for preferring the consumption measure. When calculated as described in Chapter II, the value of consumption is practically unaffected by "dearth" and "windfall" elements in income that are so common in low-income families and among farmers. Therefore, "consumption value" takes on a character which reflects an average long-term value of resources available to the family. Family expenditures, on the other hand, are highly affected by "transitory" components in the income, and are consequently a poor guide to the true level of living.

The division of family living into "more basic" and "less basic" items had considerable analytic value. Particularly in the study of low-income families, it seemed appropriate to take a separate look at food, clothing, and shelter as a group, because these are so fundamental to a study of the family's need for assistance.

But in the course of the analysis, it was observed that this dependent variable ought to be subdivided even farther. Distortions creep into regression results if there are opposing forces working within a specified division of the studied variable. In East Texas, for example, nonwhites consumed significantly less food and shelter but more clothing than the whites. Hence, when the three items were combined into "more basic" consumption, the difference in level between the two races was indicated to be significant at the 5 percent level. Similarly, when "less basic" consumption was studied there were also some counteracting effects: families with older homemakers paid more for medical care but less for transportation, furnishings, and equipment than did the younger families.

There was further evidence that several major categories of consumption had to be studied individually if an adequate analysis was to be achieved, for example, (1) family size proved to be an important predictor of food or clothing but it bore no significant relation to the consumption of shelter;

(2) the index of housing adequacy was a highly significant variable when explaining "more basic" items in general, but it tended to approach insignificance when explaining the food component by itself.

There is an additional reason, external to this analysis, for exploring separately the major different categories of consumption. If the present global income definitions of poverty are ever to be related directly to the various antipoverty programs (in any useful manner at the operational level), the definition of poverty will have to be recast in terms of the primary categories of human need around which the programs are constructed. In short, an expenditure or consumption definition of poverty will be necessary in which the various categories of human need (food, shelter, medical care, etc.) can be identified and within which minimum acceptable standards can be agreed upon by the political decision makers -- in place of a simple overall income measure. This would provide a program level framework within which various programs could be related to each other and which could be aggregated to an overall income framework as well.

The indications are that at least the following consumption categories should be separately analyzed as dependent variables:

1. Food
2. Clothing
3. Shelter, including household operations
4. Medical care
5. Transportation
6. Furnishings and equipment

Empirical and Theoretical Conclusions -- Independent Variables

Since the study deals with low-income people, it was anticipated that family income would be a significant and important determinant of consumer behavior and the level of living. Results relating to the monetary component

of income, however, did not confirm this hypothesis. Several reasons for the comparatively small influence of money income were suggested and discussed at length in Chapters III-V. Briefly the major reasons are as follows: (1) A number of farmers with rather high consumption levels had reported substantially negative net money incomes; thus, they appeared at the extreme lower end of the income distribution. They were not properly classified when grouped with the genuinely poor families who also were at the low end of the income scale; (2) The observation error in reported family incomes was probably quite large. Prais and Houthakker have observed that errors in the determining variables may lead to an over statement of their variances and hence, bias the regression coefficient downward. 1/ Such a bias is believed to exist in the present study; (3) The transitory component in the incomes of many farmers and of the poor in general, tends to be large. Hence, it has the effect of obscuring the permanent component of income which is purportedly the one to which family consumption responds fairly closely.

The notion of a permanent income component is fairly well encompassed by a concept which Margaret Reid calls "normal" income. Normal income may be regarded as the long term average income or, more correctly, what the family considers their long term or expected income. Miss Reid would put explanatory variables, especially as they pertain to total expenditure, into 4 categories: (1) those that are chiefly proxies for normal income, e.g., value or rent of dwelling, occupation, education, race, and city size; (2) those that are in large measure proxies for normal income but seem likely to have some independent relation to total expenditures, e.g., tenure, age of head and family size, and cash assets; (3) those that might conceivably serve

1/ S. J. Prais and H. S. Houthakker, The Analysis of Family Budgets "University of Cambridge": Cambridge University Press, 1955, p. 62.

as indexes of the mixture of the income, e.g., income change and income expectation; and (4) those whose effects there is no basis of predicting, e.g., debt status, beyond its probable relation to the purchase of durable goods. 1/

If the variables in Dr. Reid's first two categories come as close to representing normal income as she implies, then nearly all the predictors examined in the present study can be regarded as proxy variables to normal income. The housing index in the present study is akin to Dr. Reid's value or rent of dwelling; "chief source of income" is akin to her occupation classification.

The concept of normal income might very well serve as a general barometer to register family need for assistance. And if the family characteristics tested in Phase II of Chapters III-V can act as proxies for normal income, these variables might well become the basis for establishing eligibility to participate in antipoverty programs.

More cross sections of data need to be examined to confirm or reject a number of hypotheses suggested by the present study. A few generalizations, however, seem to be warranted, based on only the three surveys examined here:

(1) The quality of the family's housing has an important relationship to its level of living. This relationship, however, is more apparent among those who own their homes rather than among those who rent them;

(2) Large families economize in their consumption, particularly food and housing. Some of the economies are realized through purchasing larger quantities at a discount or by economies of scale in food preparation; others are obtained through the use of poorer quality

1/ Margaret Reid: Comments on A Complete Set of Consumer Demand Relationships in "Proceedings of the Conference on Consumption and Saving" Vol. 1, Irwin Friend and Robert Jones, (Ed's.), Wharton School of Finance and Commerce, University of Pennsylvania, 1960, p. 146.

goods and services. The economies that are accomplished, however, are more apparent among: (a) Nonwhite families, (b) families which receive their chief income from sources other than wages, salaries, professions or business, (c) families which produce a high proportion of their food, and (d) families in the lower education tier;

(3) When family size has been allowed for, there appear to be only two "significant" life cycle stages: the stage before and the stage after old age assistance becomes available to the family. It should be recalled, however, that only husband-wife families were interviewed and therefore marital status was not a variable;

(4) It seems appropriate to specify only two levels of education but the dividing point between the levels should be held flexible. When the general education level is relatively high and off-farm employment opportunities are good (as in East Texas) the split might come at the completion of the 11th grade. When both on-farm and off-farm opportunities are fairly good (as in the Blackland) a split at the completion of the 8th grade is suggested. When the education level is low and employment opportunities poor (as in South Central Kentucky) a split somewhere between the 5th and 7th grade may be more appropriate.

Several explanatory factors were dropped from the final model, chiefly because they were intercorrelated with certain of the factors that were retained. The factors retained for the final regression were those considered most useful in reducing the error of estimating the living level. Some of these factors were combined to create composite variables representing theoretically meaningful family groupings.

The Phase IV analysis in Chapters III-V indicated that in building combination predictors the race, age, and major source of income factors

were most salient for identifying and segregating the "best" set of family subgroups. When gross farm sales are available, it seems useful to employ these data to segregate commercial from noncommercial farmers. In any case, nonfarmers should be separately classified from farmers.

It was concluded that "proportion of home produced food" categories should be specified as a separate set of dummy variables. Although there was considerable interdependence between these categories and farm status classifications they were, nevertheless, sufficiently independent that each would contribute significantly to explaining the dependent variable.

Housing index and log family size were important predictors in the final model but the slopes of the consumption function for these two variables were known to vary for different categories of other family characteristics. The true effect of housing index and family size can best be learned by running separate regressions for each group indicated. In the present study, separate computations were indicated for: (a) owners and renters, (b) whites and nonwhites, and (c) families earning most of their incomes from "wages, salaries, professions or business," and all other families as a group.

It would no doubt be useful to make a separate analysis for some other population groups, e.g., families with head over 65. Not all the difference in behavior between young and old persons need be attributed to age. Most of the older families learned in their youth to do without automobiles and certain other durables, and they made relatively small purchases of food, fuel, and clothing. Thus, their attitudes and preferences may well be different from that of today's youth when they reach the same age.

The variables specified in the final model accounted for about two-thirds of the variance in "more basic" consumption, but they could account for less than one-third of the variance in "less basic" consumption. At

first sight the latter proportion of explained variance may appear small, but if we reflect on some of the factors that were omitted it is surprising the specified variables could account for as much as they did.

An obvious omission from among the predictors is a variable that reflects the nature and extent of household durable stocks. It is probable that housing index can to some extent serve as a proxy for capital stock, but an actual measure of inventory would likely be more satisfactory. Likewise, a measure of liquid assets available to the family might serve as a useful explanatory factor.

It would be profitable, furthermore, if one could adapt the analysis to a dynamic setting and measure the effects of such diverse factors as income expectations, habit, tradition, inertia, and the nature of joint decision processes carried on by the household. Friedman describes the psychological variables which we wish to measure as states of mind or attitudes toward a specific social and economic structure. He concludes that we may be able to observe some manifestation of these attitudes but not the attitudes themselves. 1/

A reasonable assumption is that some manifestations of consumer attitudes are inherent in the social, economic, and demographic factors employed in the present study. In other words, certain variables like those of race, age, and education are proxy to a number of variables which are not directly observable. However, the degree of attitudes, expectations, hopes, and aspirations, which are manifest in socio-demographic variables is open to conjecture.

1/ Milton Friedman: Comment on Research on Size Distribution of Income in "Studies in Income and Wealth", Vol. 13, 1951. National Bureau of Economic Research, Inc., New York, p. 57.

Methodological Conclusions

The present state of our knowledge of low-income rural consumer behavior is sketchy; therefore, it is difficult to specify the correct form of relationship between consumption and its determinants. Since there are many forms which the relationship might take, it is expedient that the initial research technique be one which uses a general approach. The results of this study have demonstrated that a dummy variable regression technique is an appropriate one for this purpose.

It is clear that the coefficients of wrongly specified regressions will have little meaning and may be misleading. Suppose we were to specify a variable as "continuous linear" when, in fact, its influence turned out to be "continuous U-shaped." If we started with no "available information" we might well infer from the inital regression run that the variable in question is not a significant determinant and ought to be omitted from subsequent runs. However, if this explanatory variable were subdivided into, say, three classes and specified as dummy variables, an approximation to the true effect would likely be discovered during the first test run.

Again, if he had no prior knowledge of potential behavior, the researcher would probably specify additivity over the whole range of a variable. Such a specification could not detect possible interaction with other variables; furthermore, where interaction does exist the estimated coefficient would be a weighted compromise of several effects and could be misleading or meaningless. A way to discover the important interaction effects is to specify interaction terms in dummy variable regressions as described in Phase III of the analysis chapters. This method provides a fairly adequate test for the hypothesis of null interaction effects.

The stepwise regression procedure, so extensively employed in this study, proved to be an effective tool in finding some of the hidden interactions that

appear to be so common among the explanatory factors collected by sample surveys. The stepwise technique also represents a useful tool for a preliminary search of the more apparent influences that affect the dependent variable. A careful study of the intermediate equations will often provide additional insights as an aid to a better understanding of consumer behavior.

But the stepwise technique is not a suitable tool for a more methodical, more definite search of the factors that significantly determine the dependent variable. As a test to see if a given factor contributes significantly to regression effects, it is necessary to compute two separate regression equations. One equation includes the factor in question as one of its independent variables; the other does not. In other respects the two equations are alike. It is then possible, using a particular F test to see if the variable in question makes a significant contribution to explained variance at a given probability level. If the variable represents a characteristic that has been entered into the regression as a set of dummy variables, it is appropriate to test the null hypothesis that the subset of coefficients of this dummy variable set are all (jointly) equal to zero.

However, in testing the significance of an explanatory factor (regardless of the functional form in which it has been entered), the outcome of the test will be greatly influenced by what other variables have been allowed for in the regression. Usually there is substantial intercorrelation among the variables measured in survey data, and as a consequence the apparent influence of any one independent variable depends on what other variables have been controlled. If therefore, let us say, only the principal independent variable has been controlled and we introduce "race" as a factor to explain more basic consumption -- the white vs. nonwhite impact may be indicated as highly significant. But after several other factors which are

intercorrelated with race have been allowed for in the regression, the added influence of the race variable may conceivably fall below the level of significance. As Morgan and Sonquist put it in their article (op cit, p. 416) "It is difficult to assess the relative importance of different factors, since their intercorrelations get in the way."

APPENDIX A

Characteristics of the Three Survey Areas

The three areas analyzed in this study had one problem in common: low farm incomes were considered a major problem in each area. However, there were substantial differences in their social, economic, occupational, and racial makeups. The following table uses a selection of characteristics to compare the populations of the three areas. Data are from the 1960 U. S. Census of Population. 1/

Characteristics	Survey Areas		
	East Texas a/	Blackland Texas b/	South Central Kentucky c/
Total Population	212,221	217,663	70,423
Population with native parentage	208,724	213,490	70,089
White	150,252	186,317	65,309
Nonwhite	58,472	27,173	4,780
Years of school completed			
Male, 25 years old and over	57,055	61,145	19,452
0 through 7 years	21,305	21,382	10,834
8 through 11 years	17,653	21,169	5,963
12 years plus	18,097	18,594	2,655
Median school years completed	9.2	9.4	7.4
Female, 25 years old and over	62,974	68,056	19,966
0 through 7 years	18,410	18,643	8,594
8 through 11 years	21,559	25,597	7,843
12 years plus	24,005	23,816	3,529
Median school years completed	10.3	10.0	8.2
Selected industry groups			
Employed persons (both sexes)	74,020	76,759	23,926
Agriculture	7,829	11,223	9,289
Manufacturing	13,100	15,407	2,681
Wholesale and retail trade	14,484	14,671	3,691

See footnotes on page 180.

Characteristics	Survey Areas		
	East Texas a/	Blackland Texas b/	South Central Kentucky c/
Family income in 1959			
All families	55,677	59,764	18,657
Under $1,000	6,423	6,401	3,988
$1,000 to $1,999	8,554	9,585	4,348
$2,000 to $2,999	7,109	7,960	3,097
$3,000 to $3,999	6,716	7,691	2,189
$4,000 to $4,999	6,104	7,535	1,543
$5,000 to $5,999	5,613	6,242	1,072
$6,000 to $6,999	4,202	4,267	711
$7,000 to $7,999	3,089	2,953	479
$8,000 to $8,999	2,151	2,149	314
$9,000 to $9,999	1,443	1,385	199
$10,000 and over	4,273	3,596	717
Median income all families	$3,886	$3,769	$2,342
Total Nonwhite Population	58,492	26,415	2/3,630
Years of school completed			
Persons, 25 years old and over	28,641	12,876	1,876
0 through 7 years	15,377	7,112	1,233
8 through 11 years	7,970	3,693	513
12 years plus	5,294	2,071	130
Median school years completed	7.7	7.5	6.5
Selected industry groups			
Employed persons (both sexes)	18,563	8,085	1,359
Agriculture, forestry, fisheries	2,996	1,458	430
Manufacturing	3,119	793	51
Wholesale and retail trade	2,329	1,065	168
Personal services	5,248	2,422	422
Family income in 1959			
All families	12,784	5,706	825
Under $1,000	3,067	1,394	182
$1,000 to $1,999	3,550	1,665	274
$2,000 to $2,999	2,492	1,092	170
$3,000 to $3,999	1,671	687	89
$4,000 to $4,999	913	413	53
$5,000 to $5,999	477	193	28
$6,000 to $6,999	204	136	4
$7,000 to $7,999	171	62	8
$8,000 to $8,999	97	34	9
$9,000 to $9,999	66	5	0
$10,000 and over	76	25	8
Median income all families	$1,984	$1,940	$1,843

See footnotes on page 188.

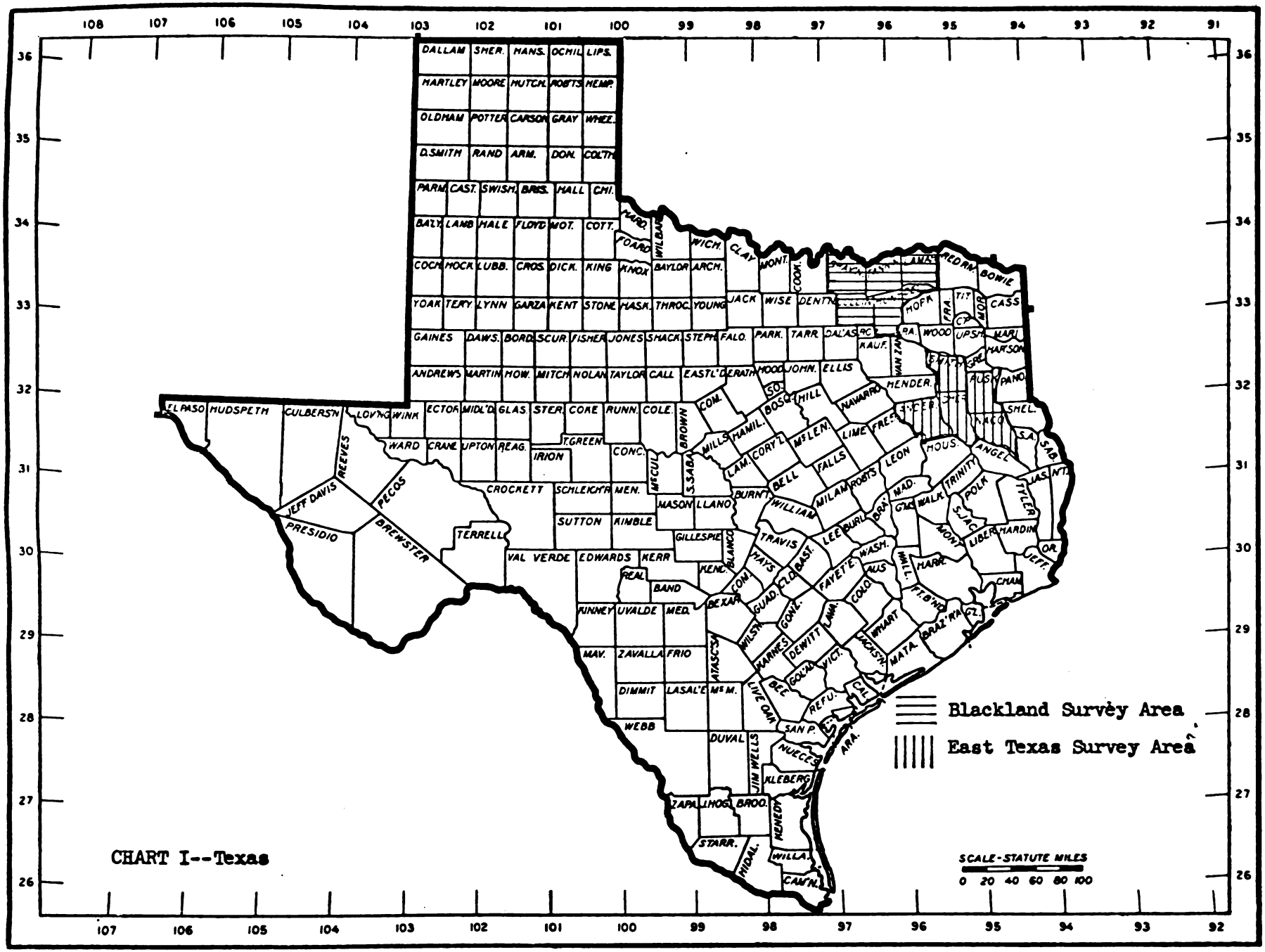
Characteristics	Survey Areas		
	East Texas a/	Blackland Texas b/	South Central Kentucky c/
Total Rural Population	107,444	99,198	60,354
Nonwhite rural population	31,823	10,036	3,798
Years of school completed			
Persons, 25 years old and over	61,378	59,747	33,642
0 through 7 years	25,261	20,484	17,127
8 through 11 years	21,685	22,795	11,820
12 years plus	14,432	16,468	4,695
Median school years completed	8.6	9.0	7.8
Family income in 1959			
All families	28,324	27,426	16,008
Under $1,000	4,479	4,182	3,706
$1,000 to $1,999	5,630	5,484	3,986
$2,000 to $2,999	4,034	3,890	2,741
$3,000 to $3,999	3,593	3,485	1,749
$4,000 to $4,999	3,011	3,264	1,258
$5,000 to $5,999	2,603	2,366	810
$6,000 to $6,999	1,692	1,589	499
$7,000 to $7,999	1,182	995	331
$8,000 to $8,999	687	638	227
$9,000 to $9,999	439	464	148
$10,000 and over	974	1,069	508
Median income all families	$3,035	$3,062	$2,138
Total Rural Farm Population	20,412	33,676	37,689
Median school years completed	8.6	9.2	7.7
Median family income in 1959	2,510	2,642	1,896

1/ All medians given in the table were weighted by a process described in footnotes 1/ and 2/ on page 41.

2/ Includes Barren and Hart counties only. Data on nonwhites were not given for the other 3 counties, because each of these counties had fewer than 1,000 nonwhites.

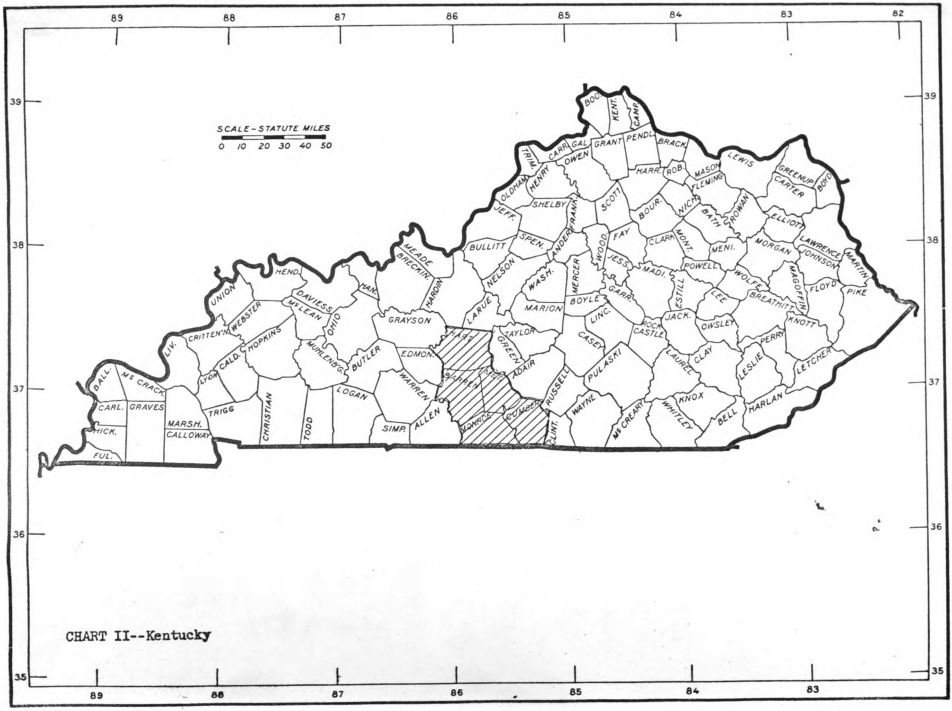
a/ Includes Anderson, Cherokee, Nacogdoches, Rusk, and Smith counties.
b/ Includes Collin, Delta, Fannin, Grayson, Hunt, and Lamar counties.
c/ Includes Barren, Cumberland, Hart, Metcalfe, and Monroe counties.

TEXAS



AD - 400.41

KENTUCKY



AD - 400.15

APPENDIX B

The Housing Index Variable

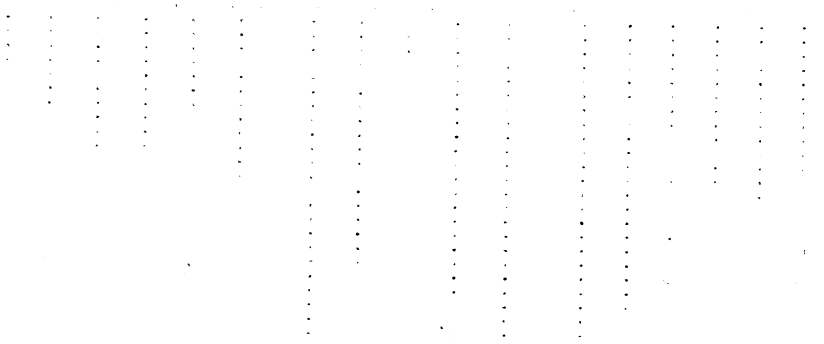
A measure of housing adequacy was fitted as an independent variable into regressions explaining fluctuations in the level of living. In order to achieve an operational variable, a scaling procedure was adopted which could provide a numeric measure reflecting the comparative qualities of homes occupied by the survey families. The method employed was to delimit a number of gradations that could be associated with different quality levels for each housing characteristic collected by the surveys. An arbitrary selection was then made, designating one of the gradations as presumably "average" or "normal." A numerical rating of zero was attached to this so-called average gradation for each specified housing trait. With the aid of this scheme each home could now be rated for any designated housing aspect receiving a zero, a plus, or a minus rating, depending on whether the home placed in the same classification as the (a) selected average, (b) above this average, or (c) below this average.

Seventeen different aspects of housing were rated according to the above procedure. The actual rating assigned to each specified characteristic ranged from "minus two" when the aspect in question was considered highly inadequate, to "plus two" when the same aspect was judged to be very adequate.

Table 1-B illustrates the procedure that was followed in rating the home of each family interviewed. The ratings across the top of the table (-2 through +2) are the values assigned to the various gradations in the body of the table, as these apply to the 17 characteristics appearing in the left hand column. A net numerical value was determined for each dwelling in the survey sample. In East Texas, these net values ranged from -27 to +27. The average rating for all homes of the total sample was -1.36; the average rating for the subsample (incomes below $2,500) was -5.99.

Table 1-B--Gradations used in the rating of dwelling characteristics and the scale used to measure housing adequacy of rural families in East Texas 1958, and Blackland Texas 1959

Dwelling characteristics	Rating				
	-2	-1	0	+1	+2
Age of house (years)	40+	30-39	20-29	10-19	0-9
Number of closets	zero	1 or 2	3 or 4	5 or 6	6+
Rooms per occupant	less than 1	1 to 1.5	1.6 to 2	2.1 to 3	3+
Living space (sq. ft.)	less than 500	500 to 800	801 to 1200	1201 to 1600	1600+
Foundation	Pier only		Pier and beam		solid masonry
Basement	none		under less than half the house	under more than half the house	
Heating	none				central
Electricity	none				wired
Electrical outlets per room	zero	less than 1	1 to 2.9	3 to 5	5+
Running water	none		cold		hot and cold
Plumbing	none	sink w/drain	flush toilet	bathtub or shower	complete bathroom
Number of bathrooms	none		½	1	1+
Water supply (adequacy)	never enough			usually enough	always enough
Grade of construction	very inferior	inferior		standard	superior
Physical depreciation	severe	moderate		minor	none
Current value (dollars)	less than 1000	1000 to 2999	3000 to 5999	6000 to 7999	8000+
Replacement value (dollars)	less than 3000	3000 to 5999	6000 to 8999	9000 to 11999	12000+



In the Texas Blackland sample, the net values for housing adequacy ranged from -23 to +22. The average rating for the total sample was -2.38; the average rating for the subsample was -4.76.

In the Kentucky study, the scaling employed had to be slightly modified because information on water supply adequacy and on the area of living space had not been obtained by the survey enumerators. A correction factor was applied, however, so that the Kentucky index was comparable to those of the two Texas studies. The net rating for all Kentucky sample dwellings ranged from -26 to +28. The average rating for the total sample was -5.34; the average rating for the subsample was -7.39.

In trying to arrive at a satisfactory weighting system for combining the various factors into a single index, several variations based on relative factor importance were tested. The net rating for each dwelling was found not to change a great deal as several plausible weighting systems were tried.

It was decided that the simple and straight forward method of equal weighting of the factors might be as good as any. The large number of factors which go to make up the index should obviate any substantial bias that might result from a failure to use correct weights for any one factor. Certain facilities, like indoor plumbing for example, probably should receive more weight than say the presence or absence of a basement. But on closer inspection, it is found that 4 of the 17 factors tested in Table 1-B do reflect directly or indirectly on the indoor plumbing. Similarly, several factors relate to the condition of the dwelling.

BIBLIOGRAPHY

Public Documents: Reports

U. S. Bureau of the Budget. Family Income Distribution Statistics Published by Federal Agencies, Statistical Evaluation Report No. 5. (Washington: Office of Statistical Standards, 1964).

U. S. Bureau of Census. Characteristics of Housing Units Classified by 1959 Income of Occupants, Series H-121, No. 11. (Washington: USGPO, 1964).

U. S. Bureau of Census. Methodology and Scores of Socioeconomic Status, Working Paper No. 15. (Washington: USGPO, 1963).

U. S. Bureau of Labor Statistics. Changing Consumption Patterns, BLS Report No. 238-11. (Washington: U. S. Department of Labor, 1965).

U. S. Bureau of Labor Statistics. Expenditure Patterns of Low-Consumption Families, BLS Report No. 238-10. (Washington: U. S. Department of Labor, 1965).

U. S. Bureau of Labor Statistics. Family Income Expenditures and Savings in 1950, Bulletin No. 1097. (Revised) (Washington: USGPO, 1953).

Public Documents: Articles

Ellis, Mary Jane. "The Outlook for Family Consumption Among Low-Income Families," U. S. Department of Agriculture, 38th Annual Outlook Conference (November 1960).

Ellis, Mary Jane. "Value of Consumption: An Improved Measure of the Level of Living," U. S. Department of Agriculture, Family Economics Review (June 1961), pp. 3-8.

Glick, Paul C. "The Life Cycle of the Family," U. S. Department of Commerce, Marriage and Family Living (February 1955), pp. 3-9.

Murray, Janet. "Per Person Food Cost Differentials in Large and Small Families," U. S. Department of Agriculture, Family Economics Review (Spetember 1960), pp. 3-4.

Pennock, Jean L. "Effects of Changing Incomes on Family Spending in a Low-Income Area," U. S. Department of Agriculture, 38th Annual Outlook Conference (November 1960).

Webb, Laura Mae. "The Outlook for Family Living Expenditures Among Low-Income Rural Families," U. S. Department of Agriculture, 38th Annual Outlook Conference (November 1960).

Books

Bowman, Mary Jean. Resources and People in East Kentucky (Baltimore: Johns Hopkins Press, 1963).

Burk, Marguerite C. Influences of Economic and Social Factors on U. S. Food Consumption (Minneapolis: Burgess, 1961).

Cochran, Willard G. and Carolyn Shaw Bell. The Economics of Consumption (New York: McGraw-Hill, 1956).

David, Martin H. Family Composition and Consumption ("Contributions to Economic Analysis," No. 25; Amsterdam: North-Holland Publishing Co., 1962).

Ezekial, Mordecai and Karl Fox. Methods of Correlation and Regression Analysis, Third Edition (New York: Wiley, 1959).

Foote, Nelson Northrup (ed.). Household Decision Making (New York: New York University Press, 1961).

Friedman, Milton. A Theory of the Consumption Function (Princeton: Princeton University Press, 1957).

Friend, Irwin and Robert Jones (eds.). Proceedings of the Conference on Consumption and Saving, Vol. I. (Philadelphia: University of Pennsylvania, 1960).

Katona, George. The Powerful Consumer; Psychological Studies of the American Economy (New York: McGraw-Hill, 1960).

Li, Jerome C. R. Introduction to Statistical Inference (Ann Arbor: Edwards Brothers, 1957).

Mills, F. C. Statistical Methods, Third Edition (New York: Holt, 1955).

Murdock, D. C. Linear Algebra for Undergraduates (New York: Wiley, 1957).

National Bureau of Economic Research. Studies in Income and Wealth Series, Vols. 13 and 15 (New York: 1951 and 1952).

Prais, S. J. and H. S. Houthakker. The Analysis of Family Budgets ("University of Cambridge, Department of Applied Economics Monographs," No. 4; Cambridge; Cambridge University Press, 1955).

Ralston, Anthony and Herbert S. Wilf (eds.). Mathematical Methods for Digital Computers (New York: Wiley, 1960)

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences (New York: McGraw-Hill, 1956).

Suits, Daniel B. Statistics: An Introduction to Quantitative Economic Research (Chicago: Rand McNally, 1963).

Walker, Helen M. and Joseph Lev. Statistical Inference (New York: Holt, 1953).

Weatherford, W. D. and Earl D. C. Brewer. Life and Religion in Southern Appalachia (New York: Friendship Press, 1962).

Articles and Journals

Crockett, Jean. "Demand Relationship for Food," Proceedings of the Conference on Consumption and Saving, Vol. I, Irwin Friend and Robert Jones (eds.), (Philadelphia: University of Pennsylvania, 1960), pp. 293-310.

Crockett, Jean and Irwin Friend. "A Complete Set of Consumer Demand Relationships," Proceedings of the Conference on Consumption and Saving, Vol. I, Irwin Friend and Robert Jones (eds.), (Philadelphia: University of Pennsylvania, 1960), pp. 1-54.

Davis, Joseph S. "Standards and Content of Living," American Economic Review, Vol. XXXV, (March 1945), pp. 1-15.

Friedman, Milton. "Comment on Research on Size Distribution of Income," Studies in Income and Wealth, Vol. 13, (New York: National Bureau of Economic Research, 1951), pp. 55-60.

Kahl, Joseph A. and James A. Davis. "A Comparison of Indexes of Socio-economic Status," American Sociological Review, Vol. 20, (June 1955), pp. 317-325.

King, Gordon A. "New Developments in Food Consumption Analysis," Journal of Farm Economics, Vol. XLV, No. 5, (December 1963), pp. 1408-1417.

Morgan, James N. and John A. Sonquist. "Problems in the Analysis of Survey Data and a Proposal," Journal of the American Statistical Association, Vol. 58, (June 1963), pp. 415-434.

Reid, Margaret. "Comments on a Complete Set of Consumer Demand Relationships," A paper by Crockett and Friend in Proceedings of the Conference on Consumption and Saving, Vol. I, Irwin Friend and Robert Jones (eds.), (Philadelphia: University of Pennsylvania, 1960), pp. 143-155.

Selvin, Hanan C. "A Critique of Tests of Significance in Survey Research," American Sociological Review, Vol. 22, (October 1957), pp. 519-527.

Suits, Daniel B. "Use of Dummy Variables in Regression Equations," Journal of the American Statistical Association, Vol. 52, (December 1957), pp. 548-551.

Bulletins

Adkins, William G. Incomes of Rural Families on the Blackland Prairies (College Station: Texas Agricultural Experiment Station in cooperation with U. S. Department of Agriculture, 1963).

Lampman, Robert J. The Low Income Population and Economic Growth, Study Paper No. 12, Prepared for the use of the Joint Economic Committee, 86th Congress (Washington: USGPO, 1959).

Montgomery, James E., Sara Smith Sutker, and Maie Nygren. Rural Housing in Garfield County, Oklahoma (A publication of Department of Housing and Interior Design; Stillwater: Oklahoma State University, 1959).

Pavlick, Anthony L. and Robert I. Coltrane. Quality of Rural and Urban Housing in the Appalachian Region, Agricultural Economics Report No. 52, (Washington: USDA, 1964).

Pennock, Jean L. Rural Family Spending and Consumption in a Low-Income Area in Kentucky, Home Economics Research Report No. 26, (Washington: USGPO, 1964).

Rockwell, George R. Income and Household Size -- Their Effects on Food Consumption, U. S. Department of Agriculture, MRR No. 340, (Washington: USGPO, 1959).

Unpublished Material

Herrmann, Robert O. "Household Socio-Economic and Demographic Characteristics as Determinants of Food Expenditure Behavior," (Unpublished Ph. D. thesis, Department of Agricultural Economics, Michigan State University, 1964).

Manderscheid, Lester V. "An Introduction to Statistical Hypothesis Testing," Department of Agricultural Economics, Michigan State University, East Lansing. (Mimeographed).