June 1, 1970

To the Graduate Council:

I am submitting herewith a dissertation written by Vernon Glenn Chappell, Jr., entitled "The Identification and Evaluation of Factors Affecting Economic Growth in the Tennessee Valley Region, 1950-1960." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Agricultural Economics.

Joe A. Martin
Major Professor

We have read this dissertation
and recommend its acceptance:

Irving Hurbov

H. B. Badenhop

William Cable

Accepted for the Council:

Hilton A. Smith
Vice Chancellor for
Graduate Studies and Research

THE IDENTIFICATION AND EVALUATION OF FACTORS AFFECTING ECONOMIC GROWTH IN THE TENNESSEE VALLEY REGION, 1950-1960

A Dissertation

Presented to

the Graduate Council of

The University of Tennessee

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Vernon Glenn Chappell, Jr.

August 1970

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Dr. Irving Dubov, Dr. M. B. Badenhop, and Dr. William E. Cole for their comments and suggestions during the preparation of the manuscript.

Special recognition is due Dr. Joe A. Martin for his guidance throughout the study. Without his constructive criticisms and encouragement, this study would not have been completed.

Acknowledgment is also made of the assistance provided by the secretarial staff of the Department of Agricultural Economics and Rural Sociology for typing several drafts of this manuscript.

Appreciation is expressed to Dr. T. J. Whatley and the Department of Agricultural Economics and Rural Sociology for financial assistance during the graduate program.

Finally, I wish to express my deep gratitude to my wife, June, and my children, Christopher and Kimberly Anne, for the many sacrifices made and the encouragement provided throughout the course of my graduate study.

926805

ABSTRACT

Despite large investments in natural resource projects, economic growth in the Tennessee Valley region during the decade from 1950 to 1960 was not evenly spread across the region. In an attempt to analyze the effectiveness of natural resource investments in stimulating regional economic growth, this study was undertaken.

The overall objectives of this study were: (1) to ascertain the factors that determine the rate of economic growth in the counties making up the Tennessee Valley Region and (2) to assess the multistatistical approach as a method of analysis in regional research. The specific objectives of the study were: (1) to determine the relationship of various types of socio-economic changes with the level and pace of economic development and growth in the Tennessee Valley Region, (2) to estimate the relationships between various kinds and amounts of investments in natural resources and the economic growth of the Tennessee Valley region, and (3) to discover the causes of disparity in the rate of economic growth between counties in the Tennessee Valley region.

In the study, it was hypothesized that regional economic growth is a function of the socio-economic characteristics of the region in question. To test this hypothesis, observations on the various socio-economic variables thought to affect regional growth were collected from secondary sources for the 150 county study region. The data were subjected to various multivariate statistical procedures.

Principal component analysis was applied to the data that was designed to measure the magnitude of economic change which occurred in the region between 1950 and 1960. The results of this analysis provided an index of economic growth in which the most important variables were: (1) change in median income of all families, (2) rate of population change, and (3) change in total payroll per capita. On the basis of this index, the counties in the study area were divided into four groups representing different rates of economic growth. An examination of these groups indicated that economic growth occurred most rapidly in urban-industrial complexes located along rivers or waterways which had received natural resource investments in the form of dams or steam generator plants.

Multiple discriminant analysis was used to test the significance of the four groups of counties identified by the principal component analysis. The results of this analysis indicated that the four groups were significantly different from each other. The analysis also indicated that the most important variables in discriminating among the four groups were: (1) median school years completed, (2) average size of household, (3) non-worker to worker ratio, (4) location of TVA steam generator plants and location of TVA dams.

Factor analysis was used as an objective method for formulating hypotheses about regional economic growth and in selecting dependent variables which constituted the basic or functional unities of a complex system of interrelated variables. The factor analysis was applied to two sets of variables.

The first factor analysis was applied to 25 variables and extracted seven factors. These factors were identified as: (1) Extent of Poverty, (2) Sub-Marginal Agriculture, (3) Labor Force Maturity, (4) Inadequate Natural Resource Investments, (5) Extent of Metropolitan Work Opportunities, (6) Labor Force Productivity, and (7) Extractive. The relationship between the seven factors and regional economic growth was analyzed by regressing the index of economic growth derived by principal component analysis on the seven factors. This analysis resulted in a regression equation with an R^2 of 0.76 and four significant independent variables at the .01 level. The four independent variables having significant regression coefficients were: (1) Extent of Poverty, (2) Sub-Marginal Agriculture, (3) Labor Force Maturity, and (4) Labor Force Productivity.

The second factor analysis derived 10 factors from 38 variables: (1) Level-of-Living, (2) Natural Resource Investments, (3) Population Maturity, (4) Sub-Marginal Agriculture, (5) Agricultural Employment, (6) Ruralism, (7) Extensiveness, (8) Welfare Needs, (9) Local Government Ineptness, and (10) Immobility.

In a regression equation, these 10 factors explained 98 percent of the variation in the index of economic growth and had eight significant regression coefficients at the .01 level. The two variables with non-significant regression coefficients were: (1) Natural Resource Investments and (2) Immobility.

The findings of the study supported the following hypotheses: (1) regional economic growth is a function of the socio-economic characteristics of the region, and (2) the observed socio-economic characteristics

can be analyzed by various multivariate statistical procedures as a means for gaining an understanding of the process of regional economic growth.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	Introduction	1
	The Problem	2
	Objectives	7
	Procedures	7
II.	THEORY	10
	Introduction	10
	Theories of Regional Growth	10
	The sector theory	10
	The export-base theory	13
	Natural resources and economic growth	14
	The retardation hypothesis	16
	Balanced growth	18
	Unbalanced growth	19
	Short-Run Model of a Regional Economy	21
	Introduction	21
	The model	21
	Related studies	27
	A Simple Long-Run Model of a Regional Economy	29
	Introduction	29
	The model	29
	Conclusions	37

CHAPTER PAGE

III. METHODOLOGY 38

Introduction 38

The Data Base 39

The study region 39

Sources of data 39

Principal Component Analysis 42

Introduction 42

Description of principal component analysis 43

Mathematical model for principal component analysis . . . 46

General Factor Analysis 50

Introduction 50

The technique of factor analysis 51

The statistical model for factor analysis 55

Multiple Discriminant Analysis 57

Introduction 57

The linear discriminant function 58

The statistical model for linear discrimination 60

Summary 63

IV. PRINCIPAL COMPONENT ANALYSIS OF ECONOMIC GROWTH IN THE TENNESSEE VALLEY REGION 65

Introduction 65

Data Requirements 65

The Analysis of the 150 County Tennessee Valley Region . . 66

Computational procedure 66

Results of the analysis 67

CHAPTER PAGE

V. USE OF DISCRIMINANT ANALYSIS TO EVALUATE THE ECONOMIC GROWTH OF THE 150 COUNTY TENNESSEE VALLEY REGION 79
Introduction 79
Data Requirements 79
Computational Procedure 80
The Analysis of the 150 County Tennessee Valley Region . . 83
Summary 88
VI. FACTOR ANALYSIS OF ECONOMIC GROWTH IN THE TENNESSEE VALLEY REGION 97
Introduction 97
Data Requirements 98
Computational Procedure 98
Factor Analysis of the 25 Independent Variables 100
Identification of factors 103
Factor 1 104
Factor 2 104
Factor 3 104
Factor 4 108
Factor 5 108
Factor 6 111
Factor 7 111
Discussion of results 111
Relationship of the seven factors to the index of economic growth 114
The regression analysis 115

CHAPTER PAGE

Factor Analysis of the 38 Socio-Economic Variables 117

Identification of factors 118

Factor 1 121

Factor 2 121

Factor 3 121

Factor 4 125

Factor 5 125

Factor 6 125

Factor 7 129

Factor 8 129

Factor 9 129

Factor 10 133

Discussion of results 133

Relationship of the 10 factors to the index of economic growth 135

The regression analysis 135

Summary 137

VII. SUMMARY AND CONCLUSIONS 140

BIBLIOGRAPHY 150

APPENDIXES 159

Appendix A 160

Appendix B 163

Appendix C 170

VITA 175

LIST OF TABLES

TABLE		PAGE
I-1	Per Capita Income in the Tennessee Valley Region as a Percent of the National Average, 1929-1960	6
IV-1	Component Coefficients for the First Principal Components of the Measure Variables for the 150 County Tennessee Valley Region	68
IV-2	Component Correlation Coefficients for the First Principal Component of the Measure Variables for the 150 County Tennessee Valley Region	70
IV-3	The Rank Order of Each County in the Tennessee Valley Region Ordered by the Size of the First Principal Component of Economic Growth	72
V-1	Mean Scores of 25 Selected Socio-Economic Variables by Group for the 150 County Tennessee Valley Region	84
V-2	Classification Function Coefficients for 25 Selected Socio-Economic Variables by Group for the 150 County Tennessee Valley Region	87
V-3	Classification Matrix for the 150 County Tennessee Valley Region	89
V-4	Evaluation of the Discriminating Function for Group I for Each County in the 150 County Tennessee Valley Region . . .	90
V-5	Evaluation of the Discriminating Function for Group II for Each County in the 150 County Tennessee Valley Region . . .	91

TABLE PAGE

V-6 Evaluation of the Discriminating Function for Group III for Each County in the 150 County Tennessee Valley Region . . . 92

V-7 Evaluation of the Discriminating Function for Group IV for Each County in the 150 County Tennessee Valley Region . . . 93

VI-1 The Rotated Orthogonal Factor Loadings Matrix for the 25 Independent Socio-Economic Variables Used in the Analysis of the 150 County Tennessee Valley Region 101

VI-2 Distribution of Communality Coefficients (h^2) for the Factor Analysis of the 25 Independent Socio-Economic Variables Used as the Analysis of the 150 County Tennessee Valley Region 102

VI-3 Socio-Economic Variables Used to Identify Factor 1 and Factor Loadings Derived from the Factor Analysis of the 25 Independent Variables 105

VI-4 Socio-Economic Variables Used to Identify Factor 2 and Factor Loadings Derived from the Factor Analysis of the 25 Independent Variables 106

VI-5 Socio-Economic Variables Used to Identify Factor 3 and Factor Loadings Derived from the Factor Analysis of the 25 Independent Variables 107

VI-6 Socio-Economic Variables Used to Identify Factor 4 and Factor Loadings Derived from the Factor Analysis of the 25 Independent Variables 109

TABLE PAGE

VI-7 Socio-Economic Variables Used to Identify Factor 5 and Factor Loadings Derived from the Factor Analysis of the 25 Independent Variables 110

VI-8 Socio-Economic Variables Used to Identify Factor 6 and Factor Loadings Derived from the Factor Analysis of the 25 Independent Variables 112

VI-9 Socio-Economic Variables Used to Identify Factor 7 and Factor Loadings Derived from the Factor Analysis of the 25 Independent Variables 113

VI-10 The Rotated Orthogonal Factor Loadings Matrix for the 38 Socio-Economic Variables Used in the Analysis of the 150 County Tennessee Valley Region 119

VI-11 Distribution of Communality Coefficients (h^2) for the Factor Analysis of the 38 Socio-Economic Variables Used in the Analysis of the 150 County Tennessee Valley Region 120

VI-12 Socio-Economic Variables Used to Identify Factor 1 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 122

VI-13 Socio-Economic Variables Used to Identify Factor 2 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 123

VI-14 Socio-Economic Variables Used to Identify Factor 3 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 124

TABLE PAGE

VI-15 Socio-Economic Variables Used to Identify Factor 4 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 126

VI-16 Socio-Economic Variables Used to Identify Factor 5 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 127

VI-17 Socio-Economic Variables Used to Identify Factor 6 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 128

VI-18 Socio-Economic Variables Used to Identify Factor 7 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 130

VI-19 Socio-Economic Variables Used to Identify Factor 8 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 131

VI-20 Socio-Economic Variables Used to Identify Factor 9 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 132

VI-21 Socio-Economic Variables Used to Identify Factor 10 and Factor Loadings Derived from the Factor Analysis of the 38 Variables Included in the Study 134

A-1 List of Variables Subjected to Multivariate Statistical Analysis 161

TABLE PAGE

A-2 Correlation Matrix for Variables 1-10 Used in the Analysis of Economic Growth in the 150 County Tennessee Valley Region 164

A-3 Correlation Matrix for Variables 11-20 Used in the Analysis of Economic Growth in the 150 County Tennessee Valley Region 165

A-4 Correlation Matrix for Variables 21-30 Used in the Analysis of Economic Growth in the 150 County Tennessee Valley Region 166

A-5 Correlation Matrix for Variables 31-40 Used in the Analysis of Economic Growth in the 150 County Tennessee Valley Region 167

A-6 Correlation Matrix for Variables 41-50 Used in the Analysis of Economic Growth in the 150 County Tennessee Valley Region 168

A-7 Correlation Matrix for Variables 51-60 Used in the Analysis of Economic Growth in the 150 County Tennessee Valley Region 169

A-8 Identification of the Counties Included in the Study Region . 171

LIST OF FIGURES

FIGURE		PAGE
I-1	The 201 County Tennessee Valley Region	3
II-1	The Relation Between Total Expenditures on Consumption, Investment, and Government and Total Income in a Regional Economy	25
II-2	The Aggregate Demand for and Supply of the Regional Economy's Production	26
II-3	The Production Possibilities Curve for a Regional Economy . .	31
III-1	The 150 County Study Region	40
IV-1	The Geographical Distribution of the Fast Growth, Medium Growth, Slow Growth, and No Growth Counties in the Tennessee Valley Region, 1950-1960	77

CHAPTER I

INTRODUCTION

I. INTRODUCTION

The development and conservation of the nation's natural resources historically has been the responsibility of the federal government. In the past, public investments in natural resources have been justified on the grounds that they improve economic efficiency. The provision of natural resource protection such as flood prevention or soil conservation, results in external economies, public goods, and indivisibilities. Without these three phenomena maximum economic efficiency cannot be achieved. However, since the provision of the benefits to a given individual cannot be made contingent upon payment of a fee to the person undertaking the natural resource investment, there is no incentive for the investment in a market system. Thus, the government is justified in making such investments on the grounds that economic efficiency will be improved.

During recent years there has been an increasing interest by federal, state, and local governments in the problems of the slow growing or chronically depressed regions. With this increased concern federal agencies administering natural resource programs have become interested in the impact that their activities had had on regional economic development. While much of the work of these agencies has centered around their attempt to conserve, develop, and manage the nation's supply of soil, water, timber, and mineral resources, the overall objective of these investments in natural resources has become

one of economic development. This fact is emphasized by Senate Document Number 97 which states:

> National economic development, and development of each region within the country, is essential to the maintenance of national strength and the achievement of satisfactory levels of living. Water and related land resources development and management are essentail to economic development and growth . . . [93, pp. 12-13].

Thus, the stimulation of regional economic development in the slow growing and chronically depressed regions of the United States has become a second justification for investments in natural resources by the federal government.

A prevalent philosophy of long standing is that the full development of natural resources eventually will generate economic activities within the region of this development that will fully exploit the potential of the augmented natural resource base. As a consequence of this exploitation, so the reasoning goes, economic development will be stimulated and the problems of low income, unemployment, and underemployment will be alleviated. To the extent that regional economic development is a national policy goal, it is important to evaluate the hypothesis that investments in natural resources serve as a stimulus to regional economic development.

II. THE PROBLEM

The Tennessee Valley Region is composed of 201 counties in the Tennessee River Watershed or served by distributors of Tennessee Valley Authority electric power (Figure I-1). The region has a population of

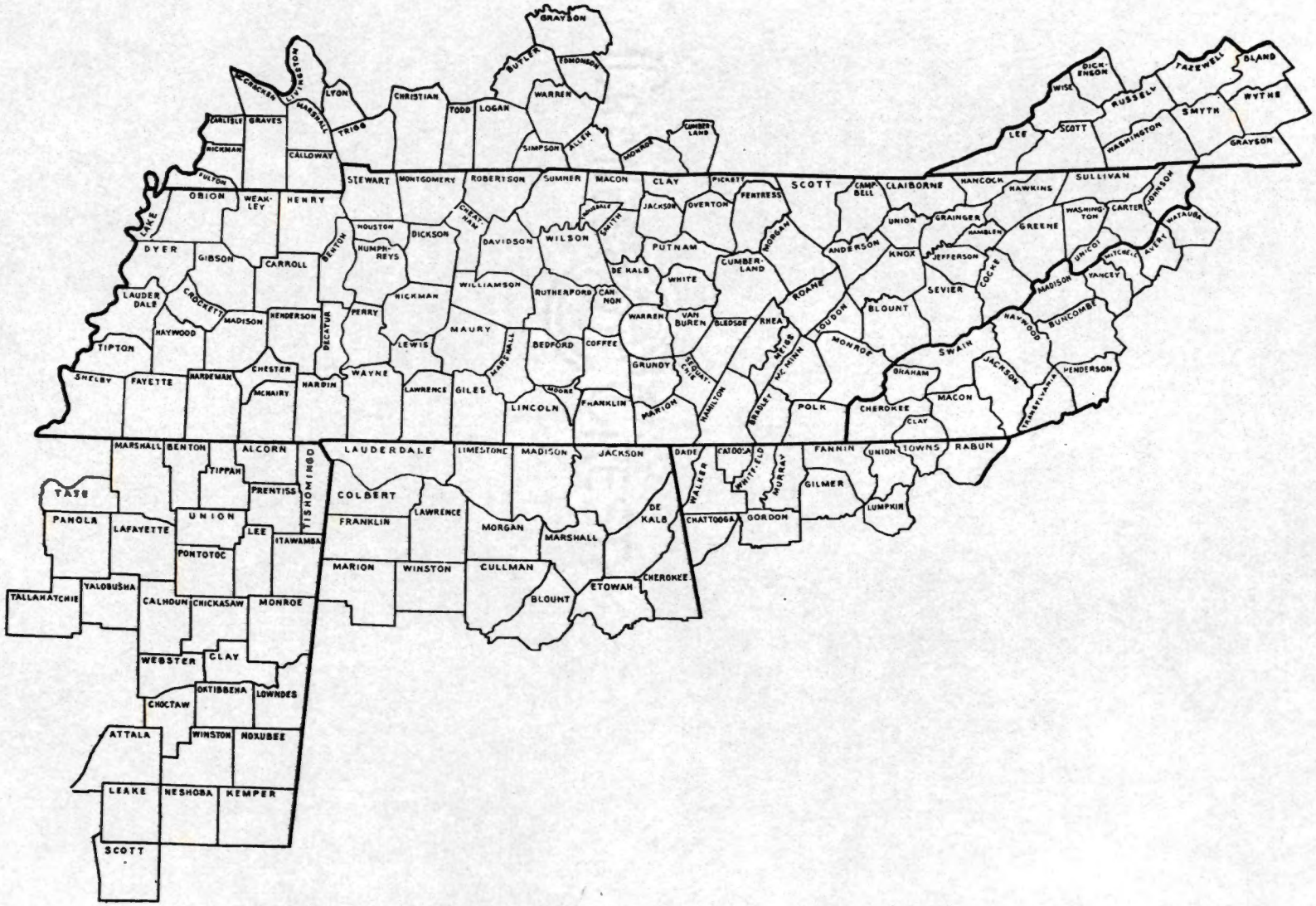


Figure I-1. The 201 county Tennessee Valley region.

almost seven million people, or about 3 percent of the nation's total [63, p. 3]. The region is homogeneous in the sense that it comprises one entire river basin (the Tennessee) and part of another (the Cumberland), but it is somewhat diverse in topography and economic activity.

In keeping with the National policy to stimulate regional economic development, the Tennessee Valley Authority was established by Congress in 1933 to organize and administer a regional resource basin of the Tennessee River. The objectives of TVA included flood control, navigation improvement on the Tennessee River, development and production of new fertilizers, electric power production, resource development, and improvements in the economic welfare of the people in the area [7, p. 475]. The efforts of TVA represent a sizable public investment in natural resources. For example, between 1938 and 1967 TVA invested over 700 million dollars in water resource developments as a part of their resource development programs.[1] While TVA has been the largest investor in natural resources in the Tennessee Valley other investments in natural resources have been made by the Soil Conservation Service and the Army Corps of Engineers.

Although large amounts of money have been invested in the Tennessee Valley Region for the purposes of economic and resource development, many areas of the Valley remain in poverty and can be classified as depressed areas. The underdeveloped nature of the Tennessee Valley is

---

[1] This figure was compiled from data provided by the Tennessee Valley Authority.

indicated by the fact that the region's labor force declined by 36,000 people from 1950 to 1960. Further, the number of workers in the young adult age group (18-34 years) declined by 188,000 while the number in the older age groups (35-64 years) increased by 152,000. Total population of the region increased during the 1950-1960 period but 135 of the 201 counties experienced population declines. The major metropolitan areas of the region accounted for about half of the total population increase [63, pp. 3-7].

Further evidence of the underdeveloped nature of the Tennessee Valley Region's economy is contained in its per capita income statistics (Table I-1). In 1929 per capita income in the Tennessee Valley Region was $317 or about 45 percent of the national average of $703. By 1945 the income gap had narrowed and per capita income in the region was $784 or about 64 percent of the national figure of $1,234. By 1966 the per capita income of the region was $2,075 while that of the nation was $2,963. Thus, since 1921 the income gap has been declining so that by 1966 the Tennessee Valley Region's per capita income reached 70 percent of the national average [63, p. 13]. However, of the 201 counties in the Tennessee Valley Region, only 34 had incomes as great as the regional average. Of the 167 counties below the regional average, many had per capita incomes amounting to 45 percent or less of the national average [63, p. 18].

Thus, it is apparent that while economic growth and development have taken place in the region, this growth has not been evenly spread across the region. Some areas appear to have shared more fully than others in

TABLE I-1

PER CAPITA INCOME IN THE TENNESSEE VALLEY REGION AS A PERCENT OF THE NATIONAL AVERAGE, 1929-1960

Year	Tennessee Valley Region	National Average	Region as a Percent of the National Average
	(dollars)	(dollars)	(percent)
1929	317	703	45
1933	168	375	45
1945	784	1,234	64
1950	890	1,496	59
1960	1,420	2,215	64
1966	2,075	2,963	70

Source: Tennessee Valley Authority, The Tennessee Valley Region, Highlights of Growth and Change (Knoxville: Tennessee Valley Authority, 1968), p. 13.

the region's prosperity. Therefore, because poverty, low income, unemployment, underemployment, and lagging economic development characterize several areas in the United States, examination of the Tennessee Valley Region's economic development should provide some insight into the problem of ascertaining how effective various kinds and amounts of natural resource investments are, or would be, in alleviating these conditions.

III. OBJECTIVES

The overall objectives of this study were (1) to ascertain the factors that determine the rate of economic growth in the counties making up the Tennessee Valley region and (2) to assess the multistatistical approach as a method of analysis in regional research. More specifically, the objectives of the study were:

1. to determine the relationship of various types of socio-economic changes with the level and pace of economic development and growth in the Tennessee Valley region;

2. to estimate the relationships between various kinds and amounts of investments in natural resources and the economic growth of the Tennessee Valley region; and

3. to discover the causes of disparity in the rate of economic growth between counties in the Tennessee Valley region.

IV. PROCEDURES

In order to accomplish the objectives of this study a multistatistical approach was employed. Data were collected from secondary sources

on 60 variables (Appendix A, Table A-1) that were thought to be dimensions of regional economic growth. The data were then subjected to a sequence of different statistical techniques in order to both provide and test hypotheses concerning regional economic growth.

As a preliminary step in the multivariate statistical analysis, correlation analysis was applied to the 60 variables in an attempt to eliminate those variates exhibiting similar relations.[2] The resulting reduced set of variables was employed in the remainder of the study.

The next step in the multistatistical approach utilized in this study was to subject the reduced set of variables to principal component analysis. This analysis served two purposes. First it facilitated the removal of other variables exhibiting small variances. This permitted the study of only those variates which had large variances. Secondly, principal component analysis provided an objective procedure for deriving an index of economic growth. The results of the principal component analysis were utilized in a regression analysis in an attempt to provide a quantitative explanation of economic growth in the Tennessee Valley region.

In order to gain some insights into the interaction of various types of socio-economic variables with the level and pace of economic development, the technique of factor analysis was utilized. This analysis extracted from the set of variables employed in this study those factors which formed the basic underlying structure of the region's economy. The

---

[2]The correlation matrix for the 60 variables is contained in Appendix B, Tables A-2 through A-7.

results of this analysis were used in conjunction with the results of the principal component analysis in a regression equation designed to explain regional growth. The results of these analyses were compared to the results obtained using per capita income as the dependent variable.

Finally, discriminant analysis was applied to the variables to explore the forces affecting the region's prospects for development. More specifically this analysis was employed to investigate the differences in the value of the growth index derived by principal component analysis and thereby discriminate between four groups of counties representing different rates of economic growth.

CHAPTER II

THEORY

I. INTRODUCTION

Regional economic growth and development have become a growing concern of public policy. However, the social, political, and economic processes involved in the regional growth process are so complex and interrelated that theorizing about their nature is difficult. As a result, economists and others concerned with the problems of regional growth have developed their theories and analyses along different lines. Therefore, there is today no one theory of regional development and growth that is universally accepted [23, p. 209].

The remainder of this chapter will consider the major theoretical concepts which have been advanced to explain regional economic growth and development. This will be accomplished by reviewing some of the major theoretical explanations of regional growth and by constructing simple short-run and long-run models of a regional economy.

II. THEORIES OF REGIONAL GROWTH

The Sector Theory

The sector theory of regional growth was developed primarily by Clark [13] and Fisher [20]. They observed that with a rise in per capita incomes in different regions or at different times, the proportion of the region's population employed in the agricultural sector generally

declined. Also observed to accompany this rise in per capita income was an increase in the proportion of the population employed in secondary activities and in the service industry [50, pp. 58-59]. Clark and Fisher hypothesized that this shift was due to (1) the existence of different income elasticities of demand for the various sectors' products, and (2) the differential rates of change in the productivity of labor in each sector. Thus, in the sector theory of regional economic growth attention is focused on internal development where economic growth is seen to be the result of specialization of labor in the production of each sector's product. Therefore, the sector theory hypothesizes that the underlying cause of regional growth are shifts in the various sectors of the region's economy [50, p. 59].

Hoover and Fisher [33, pp. 180-184] have combined the sector theory with the theory of location to derive a hybrid theory of regional development stages. They conceive the development of a region's economy as occurring in a sequence of five stages. The first stage in the development process is the self-sufficient subsistence economy in which there is little investment or trade. In this stage the majority of the population is employed in agriculture and is located according to the distribution of natural resources. As transportation improves the regional economy moves into the second stage in which production is increased through specialization in agriculture and interregional trade. With the increasing interregional trade and production, a third stage is reached in which the region's agriculture shifts from extensively cultivated crops to enterprises requiring intensive cultivation. In this stage

increasing population pressure coupled with diminishing returns in agriculture induces the introduction of secondary industries. As secondary industry continues to grow the emphasis in the region's economy shifts from secondary processing industries to heavy industry and the fourth stage in the development of the region's economy is entered [33, p. 182]. The fifth and final stage is reached when the region's economy attains an advanced level of economic development in which it exports services, personnel, and capital to other less developed regions [50, p. 59].

In summary, the sector theory places emphasis on internal rather than external development. Regional economic growth is conceptualized "as primarily an internal evolution of specialization and division of labor . . . " [50, p. 59]. The theory of developmental stages, which is an extension of the sector theory, implies that the policies and activities of government and private groups can determine the rate at which the regional economy moves from one stage to another [50, p. 60]. However, North [46] has objected to the developmental stages theory on two grounds. First, the actual development of regions in America bears little resemblance to the theoretical stages of the theory. Secondly, the sequence of stages does not provide a normative model of how regions should grow and, therefore, offers no means of analyzing the factors responsible for a region's development [46, p. 245]. Thus, it would appear that the developmental stages theory is useful primarily in the sense that it may provide an impressionistic view of the regional growth process.

Therefore, the main value of the sector theory of regional development is that it emphasizes internal development and that it may provide

a convenient frame of reference for the study of regional growth. However, while the sector theory may serve to point out some of the more important economic growth relationships, it is unsatisfactory as a framework for a broad-scale analysis of regional development and growth [50, p. 62].

The Export-Base Theory

North [46] has hypothesized that the factor which initiates growth in a region is the export-base. According to this theory, the growth of a region is initiated as a result of the response of the industries within the region to an increase in demand for their output which arises from sources outside the region. The increased demand induces the industries within the region to expand their operations and it also induces additional economic activity in the trade and service industries. As the demand for the region's exportable commodities increases, capital from outside the region will be invested in the export industries within the region in an effort to develop the production processes of the export industries and increase specialization of services to the export commodities. These investments result in increases in income which, in turn, lead to an increased demand for secondary products. As this demand increases, investment is induced in other industries [50, p. 57]. Therefore, North maintains that "the export base plays a vital role in determining the level of absolute and per capita income of a region" [46, p. 250].

The export-base theory has also been used extensively to explain the forces behind urban economic growth [6]. In this context the economic

base refers to those industries or activities of the urban community which export their output to areas or individuals outside the economic boundaries of the urban community [6, p. 161].

While the export-base theory is useful in emphasizing the fact that regional development and growth are directly affected by developments within the national economy, it "does not provide sufficient scope for the internal growth sequences" of the regional economy [50, p. 60]. Tiebout [65] has argued that the export-base theory has limited applicability because the quantitative importance of exports as an explanatory variable in regional growth depends on the nature and size of the region under study [65, p. 164]. He has also pointed out that the export base of a region is not the sole source of regional economic growth and may not even be the most important factor in regional growth [65, p. 161]. Thus, while the export-base theory is useful for its emphasis on the tie between regional and national growth, it is too aggregative for analysis in depth because it tends to neglect the functional relationship between regions [50, p. 58].

Natural Resources and Economic Growth

Closely related to the export-base theory is the role played by natural resource endowment in economic growth. This role has been analyzed from many viewpoints. At one extreme such as in the classical dynamics [8, pp. 51-69; 9, pp. 13-21] natural resources are accorded a dominating role in economic growth. At the other extreme as exemplified by Herrod's models [27] natural resources are given no role whatsoever in

the process of economic growth. However, as Schultz [58, pp. 125-145] has shown neither of these contrasting views has any general validity because they were both applied to a particular economy at a given period in history.

In discussing the role of natural resources in economic growth, Schultz [57, pp. 1-9] has emphasized the following points. First, as a country or region progresses in terms of per capita income and wealth, the relative importance of natural resources declines. Second, because newer forms of capital, such as investment in education, embody technology superior to older forms, their capacity to increase output when combined with natural resources, tends to rise rather than fall so long as the economy is technically quite progressive [57, pp. 4-5]. Third, in a dynamic economy, the supply price of the services of natural resources may fall despite continuing population growth [58, p. 52]. Finally, the value of natural resources at factor cost has been declining relative to the aggregate value of all resources [58, p. 52]. These observations lead Schultz to the conclusion that

> . . . although natural resources are mainly an integral part of the so-called backward sectors of the economy of most poor countries, it does not follow that the production possibilities of these countries are such that their natural resources act as a burden on their economic growth [58, pp. 52-53].

In a recent article Cotner [15] has reformulated the classical theory on resource scarcity so that it includes the faith in science doctrine about resource availabilities. He demonstrates that the assumptions behind the static theory of production lead to the conclusion that "limited resources would have a negative or dampening effect on economic

expansion and growth" [15, p. 93]. However, when the faith in science doctrine is included in the classical theory on resource scarcity resource availability ceases to be a limitation on economic growth. Since science and technology can be used to create resources or to make more productive use of existing resources, economic growth becomes a function of adequate resources to absorb investments. Therefore, Cotner concludes that in order to achieve sustained economic growth adequate institutional arrangements must be provided to insure that resource creation and improved use of existing resources are not blocked [15, p. 97].

In summary, the literature on the role of natural resources in economic growth indicates the following: (1) natural resources have exerted their influence on economic growth by subtly shaping the composition of growth, rather than its rate, (2) the influence of natural resources on growth is part of a complex of many influences, and (3) the relative importance of natural resources in economic development declines as an economy progresses.

The Retardation Hypothesis

Another concept of regional development and growth is the retardation hypothesis of T. W. Schultz [56]. Schultz maintains that economic development occurs in a specific locational matrix. Furthermore, he concludes that there can exist more than one locational matrix in the economy from which growth occurs and that this growth will occur at different rates in alternative matrices [56, pp. 204-205]. Schultz hypothesizes that these nodes of growth are industrial-urban in composition

and, therefore, that the center of regional economic growth is not in rural areas but in urban areas [56, p. 206]. This leads Schultz to the conclusion that the economic organization of a region works best at the center of these urban-industrial growth nodes. The implication of Schultz's hypothesis is that a region's lack of economic development is due to inefficiencies in the economic organization of the urban-industrial centers of the region.

Schultz's retardation hypothesis is closely related to the study of the role of cities in regional economic growth [22, 41, 62]. These studies conclude that in general the concentrated growth of cities results in external economies which can induce further expansion and growth of economic activity. Also, the social environment of the city is hypothesized to favor creativity and innovation to the extent that the city is in the forefront of modernization. Thus, the concentration of political, social and economic power in the city leads to further concentration and, therefore, further economic development and growth [50, p. 211].

While some empirical studies have tended to support the hypothesis that urbanization causes more rapid progress other studies have not. Tang [61] has concluded that input markets in agriculture operate more efficiently near urban centers than at locations farther away. On the other hand, Berry [10] has shown that a highly integrated system of cities is not a necessary condition for economic development. However, he did find a strong association between urbanization and economic development.

Balanced Growth

Still another theoretical concept which has been used to explain regional economic development is the balanced growth theory. This concept as formulated by Rosenstein-Rodan [54] and Nurske [48] is based on the idea of external economies which are defined by Higgins [30, p. 327] as

> . . . benefits which accrue to the society as a whole, or to some members of it, in a fashion that does not bring a direct return to the investor concerned.

Balanced growth as conceived by Rosenstein-Rodan [54] emphasized the limitations imposed by the size of the market. Because of indivisibilities in the supply of social overhead capital, complementarity of demand, and indivisibility in the supply of savings, Rosenstein-Rodan concluded that large-scale, planned industrialization is necessary to stimulate economic development [30, p. 328].

Carrying this argument further Nurske [48] has concluded that the size of market in an undeveloped area limits the inducement to invest [48, p. 6]. Thus, Nurske reasons that the small size of the market discourages individual investment in any single line of production considered separately. He concludes that this problem can be overcome through "a more or less synchronized application of capital to a wide range of different industries" [48, p. 11]. Nurske reasons that

> People working with more and better tools in a number of complementary projects become each others' customers. Most industries catering for mass consumption are complementary in the sense that they provide a market for, and thus support, each other . . . . The case for "balanced growth" rests on the need for a "balanced diet" [48, p. 11].

The balanced growth doctrine has been criticized for its failure to recognize that undeveloped regions may have limited resources. In regions where there is a shortage of resources the relationship between industries is competitive rather than complementary as is assumed in the balanced growth doctrine [21, p. 279]. This is due to the fact that the limitation of factor supplies coupled with the introduction of large-scale production units may give rise to diseconomies rather than economies in other industries competing for the same market [21, p. 281].

Unbalanced Growth

Another argument which has been advanced to explain regional development is Hirschman's strategy of economic development [31]. According to Hirschman the scarce resource in economic development is decision-making. He maintains that savings and other resources are latently available in any given underdeveloped region [31, p. 5]. Therefore, he reasons that what is needed in dealing with problems of development is a strategy that will induce and conserve decision-making.

Hirschman maintains that development is an unbalanced process in which growth is "communicated from the leading sectors of the economy to the followers, from one industry to another, from one firm to another" [31, pp. 62-63]. Thus, in order to keep the growth process moving ahead, investment sequences must be initiated which maintain "tensions, disproportions, and disequilibria" [31, p. 66]. Hirschman reasons that such investment sequences enable an industry within a region to enjoy the benefits of external economies "created by previous expansion, and, at

the same time creates new external economies to be exploited by other operations" [31, p. 67].

Thus, Hirschman's scheme as applied to regional development is concerned with identifying the leading sectors in the region's economy which may be capable of inducing additional investments in the region. These leading sectors create investment opportunities which are easy to exploit and which depend on the external economies created by the leading sector. Hirschman calls these induced industries satellites and states that they may be created through forward and backward linkage effects [31, p. 102]. Where the satellite industry uses the products of leading industry as an input, development and growth occur as a result of forward linkage. If, however, the leading industry is an important user of inputs produced within the region, it may induce additional investment in the supplying industries of the region and economic growth occurs as a result of backward linkage. Therefore, Hirschman maintains that regional growth will be maximized if investment sequences can be induced which maximize the backward and forward linkage effects of the leading sector.

Using Hirschman's framework of analysis Christian and Vaughn [12] have argued that both the rate and kind of industrial growth in the Tennessee Valley reflect the sizable investment made in the electric power industry of the region. They hypothesized that the electric power industry acted as a leading sector in the region's economy by inducing additional investments in other industries. These investments were generated in industries which thrived in the pecuniary external economies passed on to them by the leading industry. Therefore, Christian and Vaughn

reason that these investment sequences have enhanced the economic development and growth of the Tennessee Valley region.

III. SHORT-RUN MODEL OF A REGIONAL ECONOMY[1]

Introduction

Metzler [43] and Airov [4] have presented a short-run model of a regional economy which reveals much about the central relationships between different economic magnitudes. The assumptions behind the model are that (1) the time period considered is of such a length that only changes in the location of economic activity, changes in investment, or changes in export demand will result in changes in total regional income and employment, and (2) there will be no changes in technology, wages, prices, and the distribution of income in the economy.

The Model

The logical starting point of this analysis is the relationship between regional consumption and regional income. The net income or product of a region is equal to the sum of consumption expenditures, investment, government expenditures, and the value of exports of the region, less the value of imported raw materials, intermediate goods, and goods for final use. This relationship can be expressed mathematically [47, pp. 160-161] as

$$Y = E + X - M \qquad (1)$$

---

[1] This section draws heavily upon [4], [43], and [47, Chapter 7].

where

Y = net regional product or income

E = expenditures on net investment, consumption, and government

X = exports

M = imports.

Since this is a short-run model prices, tastes, and preferences are assumed constant. This means that total consumption, investment, and import expenditures are a function of the total level of regional income. Mathematically these relationships can be represented as

$$E = a + bY \tag{2}$$

$$M = c + eY \tag{3}$$

$$X = \overline{X} \tag{4}$$

where

a = level of expenditures when Y is zero

b = marginal propensity to spend

c = level of imports when Y is zero

e = marginal propensity to import

$\overline{X}$ = given level of exports.

Thus, the equilibrium level of regional income can be obtained by substituting equations (2), (3), and (4) into (1). The result of this substitution is

$$Y = a + bY + \overline{X} - c - eY \tag{5}$$

which can be simplified to

$$Y = \frac{a - c + \overline{X}}{1 - (b - e)} \quad . \tag{6}$$

With the expenditure relations a linear function of the regional income level and with a given level of exports, a single determinate value of regional income can be obtained from equation (6). For some other value of the expenditure relations, a different value of income would result. Therefore, to the extent that the expenditure relations behave as indicated in equations (2), (3), and (4) and that the other simplifications employed thus far are inconsequential, fluctuations in regional income in the short-run may be said to be the result of fluctuations in total expenditures on consumption, investment, government, and imports. The analysis can be carried further in that the numerical magnitudes of the fluctuations in regional income due to changes in $\bar{X}$, a, or c can be obtained. The change in the equilibrium level of regional income with respect to a change in $\bar{X}$, a, or c is called the multiplier. The mathematical expression for the multiplier is obtained by taking the first derivative of equation (6) with respect to $\bar{X}$, a, or c. Thus, when the level of exports, $\bar{X}$, changes, the change in equilibrium level of regional income is equal to

$$\frac{\partial Y}{\partial \bar{X}} = \frac{1}{1 - (b - e)} \qquad (7)$$

This expression says that with the given assumptions an increase in the given level of exports would cause the equilibrium level of regional income to rise $\frac{1}{1 - (b - e)}$ times the initial increase in the given level of exports.

Graphically, the short-run model of a regional economy is depicted in Figures II-1 and II-2. In Figure II-1, the relationship between total expenditures on consumption, investment, and government and total income is represented by line EE. The relationship between imports and regional income is shown by line MM. Since exports depend upon the income in other regions, the level of exports is taken as given and is represented by line XX.

Figure II-2 represents the relationship between aggregate demand and aggregate supply in the regional economy. Since the aggregate demand for a region's production is equal to the sum of the expenditures and exports minus imports, line E'E' in Figure II-2 represents this summation. In other words, line E'E' shows the relationship between the aggregate demand for a region's products and its net income in the short-run.

The aggregate supply curve of the regional economy is depicted as line OY in Figure II-2. Remembering that prices and wages are assumed to be constant in this model, the horizontal axis can also be visualized as measuring the use of resources. Thus, as income increases more use is made of the region's resources. Therefore, the line OY depicts the level of factor income which is generated through the use of the region's resources.

In sum the line E'E' in Figure II-2 indicates the level of desired expenditures in the regional economy, while line OY shows the actual level of expenditures or the level of income that is generated from production in the region. When the level of regional income falls below

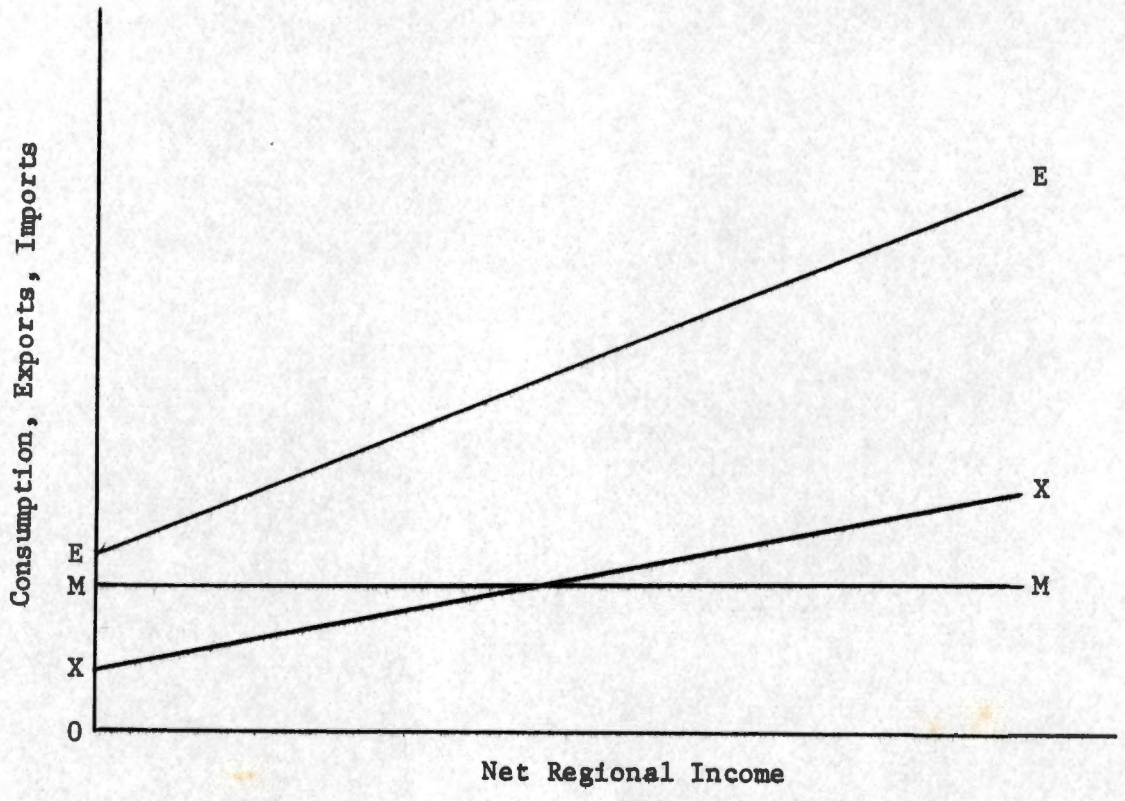


Figure II-1. The relation between total expenditures on consumption, investment, and government and total income in a regional economy.

Source: Hugh O. Nourse, Regional Economics: A Study in the Economic Structure, Stability, and Growth of Regions (New York: McGraw-Hill Book Company, 1968), p. 156.

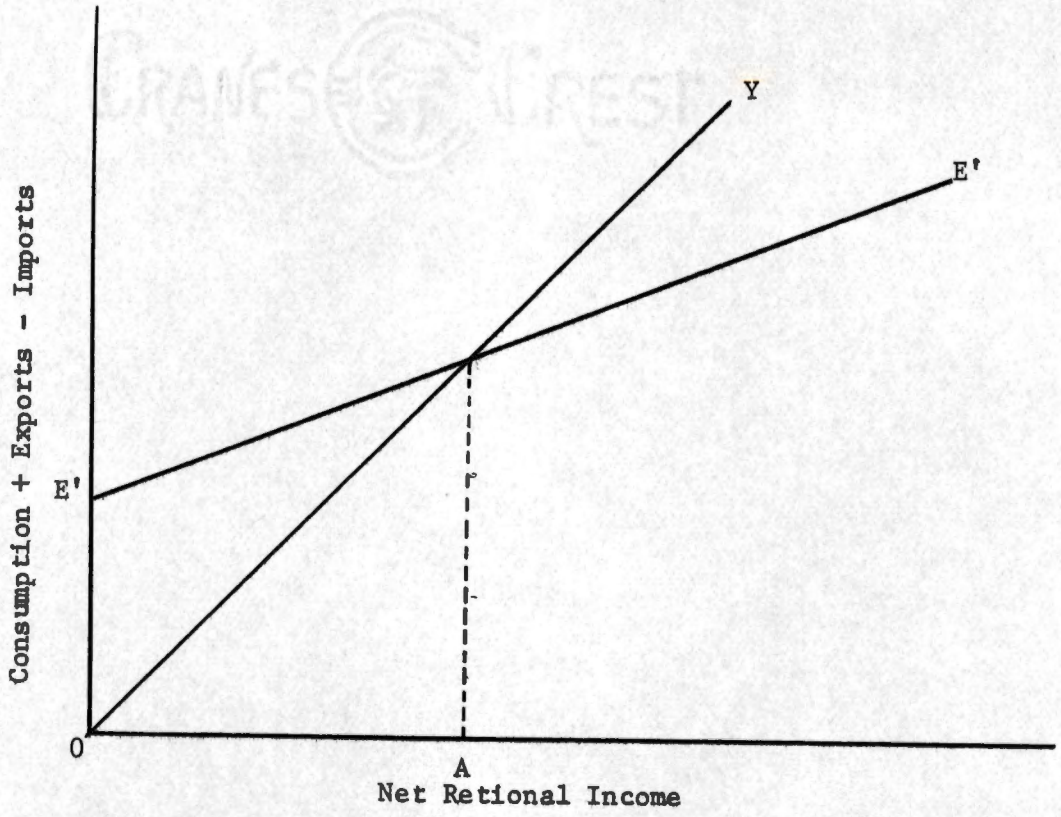


Figure II-2. The aggregate demand for and supply of the regional economy's production.

Source: Hugh O. Nourse, Regional Economics: A Study in the Economic Structure, Stability, and Growth of Regions (New York: McGraw-Hill Book Company, 1968), p. 157.

OA, a situation is encountered where desired expenditures are greater than incomes. In this situation firms in the region would increase production to offset the fact that their inventories were falling below desired levels. When the level of regional income exceeds OA, firms would encounter a situation in which their inventories are higher than the desired levels and, therefore, they would reduce output in order to unload their inventories. In this situation income and employment would fall in the short-run. Thus, OA represents the equilibrium level of regional income. However, the assumptions of this model do not insure that this level will be one at which full employment of the region's resources occurs.

Related Studies

Modifications of the above model have been used in a variety of regional studies. In regions where there is insufficient data to work out the above model and calculate the multiplier, economic base studies have been developed [37, 52, 66]. These studies stress "the interrelations of sectors within a regional economy and the spread of impulses originating in any one sector to all other sectors either directly or indirectly" [37, p. 189]. The usefulness of these types of studies is that they show how growth in one sector induces growth in another.

Another modification of the above short-run analysis combines the economic base study with an input-output framework to derive multipliers that will indicate the impact on a regional economy of a change in demand for a particular activity [47, p. 163]. Several of these studies have

attempted to analyze the structure of a particular state's economy [25, 22]. Kalter has used interindustry analysis to estimate the economic effects of water-based recreation investments on local political subdivisions [38]. Wilson has employed a Leontief input-output model to establish the quantitative interrelationships among agriculture and agricultural related industries and non-agricultural industries within the state of Tennessee [95].

In sum the short-run theory of regional income determination is difficult to implement because of the scarcity of data. However, it may be modified and become the base for several techniques of regional analysis all of which have their shortcomings. The use of the short-run framework in the economic base studies results in the error of using a short-run model to analyze a long-run problem. In this context the economic base study presents only a gross projection of the region's growth and development [47, p. 180].

The use of the short-run theory as a base for impact or intersectoral flow analysis is also useful. While the multipliers estimated from this type of analysis are more refined than those derived from economic base studies, they are still subject to the problems that result from deriving the estimates from averages rather than marginal propensities [47, p. 180].

IV. A SIMPLE LONG-RUN MODEL OF A REGIONAL ECONOMY[2]

Introduction

As was illustrated in the previous section, the short-run analysis of regional income determination was concerned with the utilization of the regional economy's full capacity during a time period in which capacity was not allowed to change. However, in long-run analysis the region is assumed to reach its full-employment equilibrium level of income. Thus, the concern of long-run analysis becomes one of explaining how the full-employment level of regional income can vary over time. Therefore, in contrast to the short-run analysis, the long-run analysis assumed that population, wages, prices, resources, technology, and the distribution of income will vary. Thus the changes that occur in these variables will hold the key to understanding changes in the potential level of regional income [47, p. 181]. To gain some insight into these relationships a simple long-run model of the regional economy will be set up. This model will utilize the production possibilities curve as a framework for analysis.

The Model

The assumptions of this model are that (1) the region produces two commodities: a product for export and a local service, (2) the export and local service market are the final market sectors of the regional economy, and (3) the sum of the production in these two final sectors less imports is equal to net regional product [47, p. 182].

---

[2]This section draws extensively from Chapter 8, [47].

The many combinations of amounts of exports and local services that could be provided with the given skills, resources, and economic organization of the regional economy can be represented by the production possibilities curve in Figure II-3. Any point on AB shows the maximum feasible quantities of exports and local services which can be produced with the resources available. If all resources in the regional economy were devoted to export production, OA exports could be produced. On the other hand, if all resources were put into local services, an output of OB units would be possible. If resources are divided between the two products, the regional economy can produce various combinations of exports and local services, represented by points on line AB.

By shifting resources from the production of exports to the production of local services (or vice versa) output is shifted from one to the other. AB is concave upward due to the specialization of some factors of production [47, p. 182]. This means that as one moves up the curve from B toward A, one must sacrifice larger and larger quantities of local services for each additional unit of exports produced. This will happen because the regional economy's resources are not all equally well adapted to producing the two goods.

Thus, AB separates feasible output combinations from combinations which are not possible. It represents the maximum boundary of production possibilities for the regional economy. The regional economy can never operate outside its production frontier and it will not want to operate inside the frontier. Combinations of production which fall within the curve AB are possible but they leave some of the economy's resources

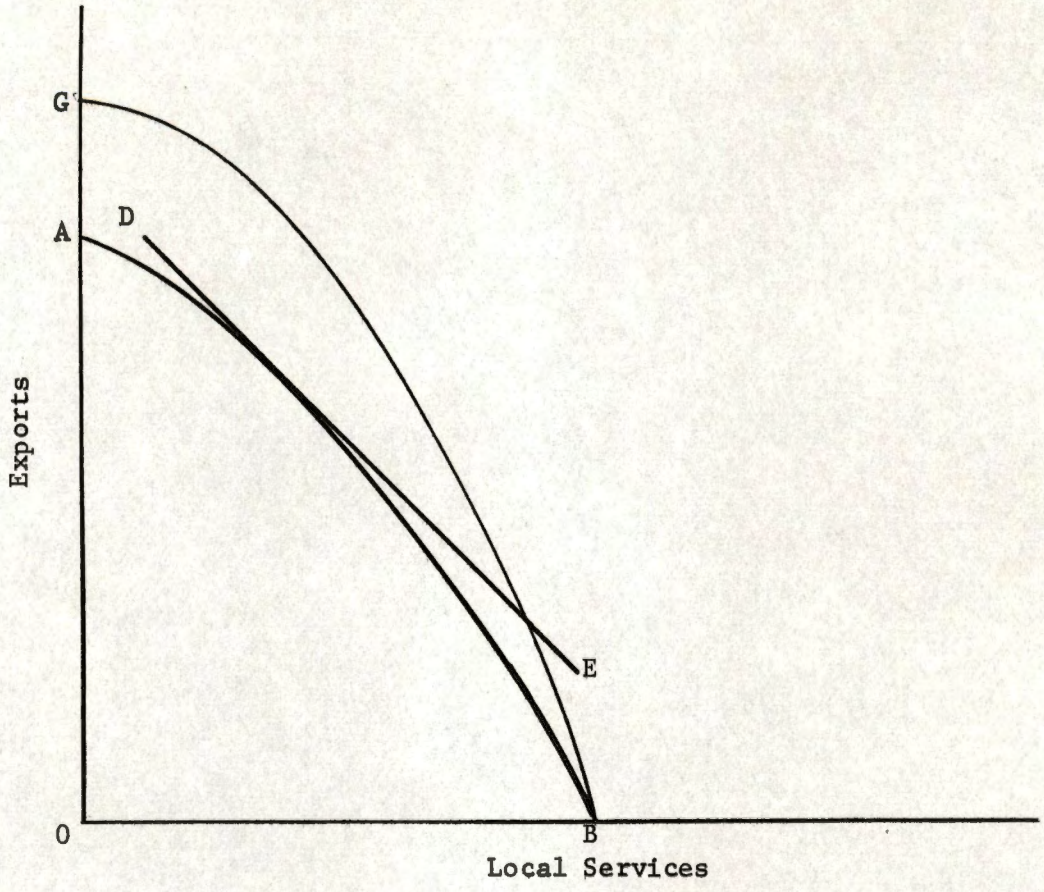


Figure II-3. The production possibilities curve for a regional economy.

Source: Hugh O. Nourse, Regional Economics: A Study in the Economic Structure, Stability, and Growth of Regions (New York: McGraw-Hill Book Company, 1968), p. 183.

underemployed and represent less than full employment in the regional economy [47, p. 183]. Therefore, in order to achieve full employment of its resources a regional economy must operate exactly on the boundary of its production possibilities curve.

In a regional economy which is operating at full employment the mix of production is determined by relative prices. In the present model that combination of exports and local services will be produced at which the ratio of prices of the products equals the ratio of their marginal costs [47, p. 183].

Following Henderson and Quandt [29, pp. 68-70] this relationship can be expressed mathematically. Assuming that the production possibilities curve, AB, in Figure II-3 can be represented as

$$x^{o} = h\ (q_1, q_2) \tag{8}$$

where

x^{o} = the locus of output combinations that can be obtained from the given input of the region's resources X

q_1 = output of local services

q_2 = output of exports.

Further assume that the slope of the tangent to a point on the production possibilities curve AB is represented by line DE. Line DE represents the rate at which exports (q_2) must be sacrificed to obtain more local services (q_1) (or q_1 sacrificed to obtain more q_2) without varying the amount of the region's resources, X, that are used. The negative

of the slope is defined as the rate of product transformation (RPT) and can be expressed as

$$RPT = \frac{dq_2}{dq_1} \tag{9}$$

Taking the total differential of (9),

$$dx = h_1\, dq_1 + h_2\, dq_2 \tag{10}$$

Since $dx = 0$ for movements along the production possibilities curve,

$$RPT = \frac{-dq_2}{dq_1} = \frac{h_1}{h_2} \tag{11}$$

Equation (11) says that the rate of product transformation at a point on a production possibilities curve equals the ratio of the marginal cost of local services (q_1) in terms of X to the marginal cost of exports (q_2) in terms of X at that point.

Employing the inverse-function rule, the rate of product transformation (RPT) can be expressed in terms of the marginal products or,

$$\frac{\partial q_1}{\partial x} = \frac{1}{h_1} \qquad \frac{\partial q_1}{\partial x} = \frac{1}{h_2} \tag{12}$$

Substituting (12) into (11),

$$RPT = \frac{-dq_2}{dq_1} = \frac{\partial q_2/\partial x}{\partial q_1/\partial x} \tag{13}$$

Equation (13) states that the rate of product transformation equals the ratio of the marginal product of X in the production of exports (q_2) to the marginal product of X in the production of local services (q_1).

Assuming that the outputs of the regional economy are sold at fixed prices, the revenue of the economy is given by the linear equation

$$R = p_1 q_1 + p_2 q_2 \tag{14}$$

where

p_1 = the price of local services (q_1)

p_2 = the price of exports (q_2).

Thus, equation (14) represents an isorevenue line which is defined as "the locus of output combinations that will earn a specified revenue" [29, p. 69]. In Figure II-3, page 31, line DE represents an isorevenue line and has a slope equal to the negative of the ratio of the output prices or $-p_1/p_2$. Thus, the slope of the isorevenue line is equal to the slope of a tangent to the production possibilities curve. The matter of determining the point on the production possibilities curve at which the economy will ultimately settle becomes a constrained-maximization problem which may be solved by forming the function

$$w = p_1 q_1 + p_2 q_2 + \mu [x^o - h (q_1, q_2)] \tag{15}$$

where μ is an undetermined Lagrange multiplier. Next, the function's partial derivatives are set equal to zero:

$$\frac{\partial w}{\partial q_1} = p_1 - \mu h_1 = 0$$

$$\frac{\partial W}{\partial q_2} = p_2 - \mu h_2 = 0$$

$$\frac{\partial W}{\partial \mu} = x^{\circ} - h(q_1, q_2) = 0$$

Solving these equations simultaneously,

$$RPT = \frac{p_1}{p_2} = \frac{h_1}{h_2} \tag{16}$$

or substituting from (12),

$$\frac{p_1}{p_2} = \frac{\partial q_2/\partial x}{\partial q_1/\partial x} = RPT \tag{17}$$

Equation (17) says that the rate of product transformation must be equated with the fixed price ratio. Graphically this means that the specified production possibilities curve AB (Figure II-3, page 31) must be tangent to an isorevenue line, DE, in order for the regional economy to maximize its income.

Thus, the point on the production possibilities curve at which the economy will ultimately settle can be defined as

$$\mu = \frac{p_1}{h_1} = \frac{p_2}{h_2} \tag{18}$$

or substituting from (12)

$$\mu = p_1 \frac{\partial q_1}{\partial x} - p_2 \frac{\partial q_2}{\partial x} \tag{19}$$

where μ is the total derivative of R with respect to x. Therefore, the equilibrium point on the production possibilities curve occurs where the slope of the tangent to the curve AB, such as DE in Figure II-3, page 31, is equal to the inverse ratio of the prices of the products [47, p. 185].

Utilizing the above framework it is possible to look at the means by which a region may grow in income and employment. Any change which results in the production possibilities curve shifting away from the origin increases the capacity of the economy to produce. These changes may include improvements in technology and the productivity of the labor force, increases in the quantity of inputs, new discoveries of natural resources, improved resource accessibility, and improved economic organization [47, pp. 185-186]. The impact of these changes may affect one product of the regional economy and not the other. For example, if an improvement in the technology of export production is developed, the production possibilities curve in Figure II-3, page 31, would shift from AB to GB. This shift would represent an increase in the economy's capacity to produce and, therefore, economic growth would result.

In sum, the simple long-run growth model outlined above indicated that regional economic growth will result from either an increase in available inputs or an increase in the productivity of inputs. While the simple long-run model of regional growth is useful for relating the factors affecting the growth of a regional economy, it is not suitable as a framework for a broad analysis of a regional economy.

V. CONCLUSIONS

The theoretical concepts which have been reviewed in this chapter emphasize the fact that there is today no one generally accepted theory of regional development and growth. Because the processes involved in regional growth are extremely complex, only partial theories have been advanced to explain regional growth and development.

However, one may conclude from the review of existing theories of regional economic development and growth that if the researcher has (1) acquired sufficient knowledge of the problem in question, (2) familiarized himself with the data, and (3) developed meaningful hypotheses, logical inferences may follow which will lead to predictions of importance. Thus, the conclusion of this chapter is that consideration must be given to the means of investigating the relationships involved in regional growth. If meaningful hypotheses can be developed, then logical inferences about the nature of these relationships can be drawn and tested. The next chapter will develop the methodology employed in this study to analyze the relationships involved in regional economic growth.

CHAPTER III

METHODOLOGY

I. INTRODUCTION

Since one of the major objectives of this study was to assess the usefulness of the multivariate statistical approach in the analysis of regional economic growth, it is appropriate to describe in detail the methodological procedures used. The multistatistical approach may be regarded as a family of statistical techniques which are applied to sets of dependent variates in an attempt to analyze the problem under investigation [39, p. 6].

As it was pointed out in the previous chapter, in regional analysis the complexity of the problem may prevent the development and testing of logical hypotheses. In this situation some consideration must be given to statistical techniques which may be used to reduce error and thereby provide hypotheses which are meaningful and testable. Multivariate statistical analysis[1] is one such family of statistical techniques.

The remainder of this chapter will present a detailed discussion of the statistical models used in this study. While the methodology used is not new there are few instances where it has been used in an analysis of regional economic growth.

---

[1]The use of multivariate statistical analysis in the study of water resource developments has been suggested by Cox in [16].

II. THE DATA BASE

In this study it was hypothesized that regional economic growth is a function of the socio-economic characteristics of the region under consideration. Therefore, as a test of this hypothesis a set of socio-economic variables were observed and analyzed as a basis for drawing inferences and hypotheses about the methods of analysis used and hopefully, some conclusions about the development and growth of the Tennessee Valley region.

The Study Region

One hundred fifty counties[2] (Figure III-1, page 40) of the 201 county Tennessee Valley power service region were selected for analysis. It was not possible to use all of the 201 counties in the analysis due to the limitations imposed by the computer programs utilized in the study. The 150 county region is shown in Figure III-1. In sum, the study region included a majority of the counties which were located in a seven state region composed of the counties in the Tennessee Valley watershed or served by distributors of Tennessee Valley Authority electric power.

Sources of Data

Data on the 150 counties located in the study region were obtained from secondary sources [69 - 92]. The following criteria were met in

---

[2]For a list of the counties included in the study area see Appendix C, Table A-8.

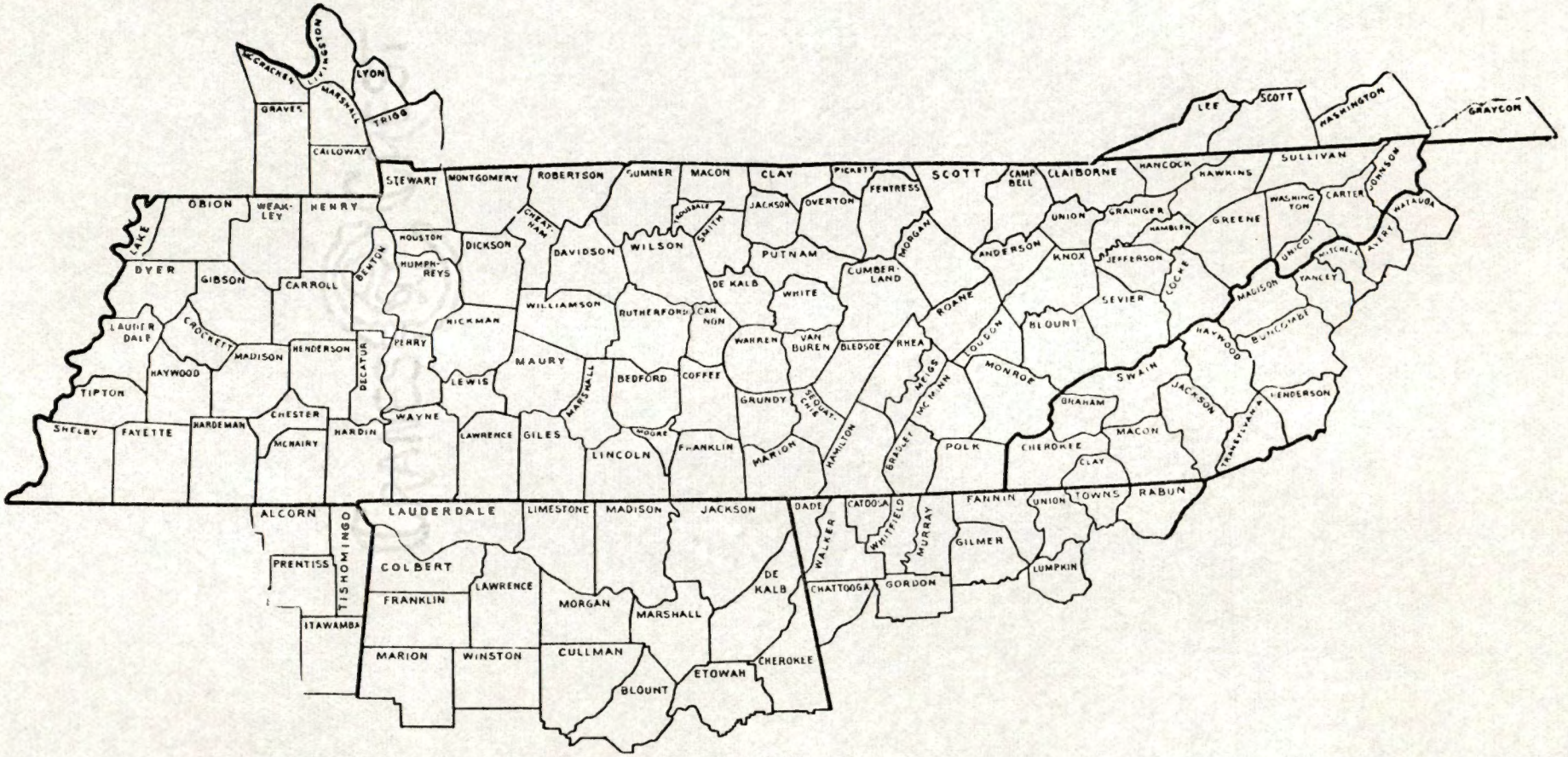


Figure III-1. The study region.

the selection of the data: (1) the data had to be readily available from secondary sources, (2) the data had to be available for the same time period, and (3) the data had to be available by county for all counties in the study region. Additional advantages of the use of secondary county data are the possible extension of similar studies to other areas within the study region, or as was done in this study, the application of other techniques to the same study region.

Sixty socio-economic variables were initially selected to represent the socio-economic characteristics of the study region. It was hypothesized that the socio-economic data on these variables were reflections of man's reactions to his environment. If this is true observations can be made on data of county variables in such a manner so as to determine the economic performance of the various counties within the study region as it relates to regional economic growth.

The set of 60 socio-economic variables were made up of two subsets. The first subset included variables 1 through 23 and were designed to measure the magnitude of economic change which occurred in the study region from 1950 to 1960. These variables subsequently became the dependent variables in the models used to analyze the data. The second subset of variables included variables 24 through 60 and were entered into the data analysis as independent variables. Two of these variables, the location of TVA dams and the location of TVA steam plants were entered into the analysis as dummy (zero-one) variables. This procedure was followed in an attempt to determine what effect the location of a dam or steam plant within the study region had on regional economic growth.

III. PRINCIPAL COMPONENT ANALYSIS

Introduction

In the exploratory phases of a study with a large number of variables under consideration the question arises as to how these variables can be organized into a more fundamental set of independent variables. One method that has been developed for this purpose is principal component analysis. This method was originally developed in the field of psychology by Pearson [49] as a means of fitting planes by orthogonal least squares. Hotelling [35, 36] adopted this methodology and applied it to the analysis of correlation structures. Girshick [24] showed that Hotelling's methodology could be applied to the solution of other problems outside the field of psychology. In the field of economics, applications of principal component analysis are limited. Stone [60] took data for the United States comprising, for each of the years 1928 - 1938, 17 series regarded as constituent elements of national income. He did a principal component analysis on the observations taken about their mean and extracted three principal components, thus reducing the effective dimensions of variation from 17 to three. The three dimensions were interpreted as total income, rate of change in total income, and a trend term which expressed expansion or contraction of the economy. Thus, Stone was able to gain considerable economy in the number of variables required to explain the variations of national income. This illustrates the basic objective of principal component analysis which is to economize in the number of variables [39, p. 10].

Component analyses may also be useful in the field of general index numbers. Many questions concerned with regional economic growth require a knowledge of the broad economic relationships that exist in the regional economy. It is obviously impossible to include all of the necessary variables in any one analysis. Therefore, it becomes necessary to rely on indexes as substitutes for certain groups of variables. The method of principal component analysis can be utilized to derive indexes that are statistically valid [2, pp. 109-110]. In this study the method of principal component analysis was used to derive an index of economic growth from the measure (dependent) variables. Principal component analysis was also applied to the independent variables to derive an index of functional relations.

Description of Principal Component Analysis

Principal component analysis is a statistical technique which provides an objective procedure for discarding variables which exhibit small variances and using only those with large variances. For example, suppose there are p variables $x_1 \ldots x_p$ each observed on n individuals. The jth observation on the ith variable can be denoted as xij and arrayed in a matrix:

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} \qquad (1)$$

Given that the objective is to reduce the number of variables, this objective is accomplished by deriving a set of principal components from the data matrix through a linear transformation of the variables [39, p. 10]. The resulting principal components are thus defined as linear combinations of statistical variables which have special properties in terms of variances. These special properties come about through the transformation of the original vector variable to the vector of principal components through a rotation of the original response coordinate system to a new coordinate system. This rotation partitions the total variance orthogonally into successively smaller portions and, when these portions are distinct, only one set of coefficient vectors evolve [44, p. 232]. Thus, the method of component analysis involves the rotation of coordinate axes to a new frame of reference in the total variable space. This rotation is an orthogonal transformation in which each set of the p original variables is describable in terms of the p new components [26, p. 136].

The import t characteristic of the new components is that they account, in turn for a maximum amount of the variance of the variables. More specifically, the first principal component is that linear combination of the original variables which contributes a maximum to their total variance. Mathematically [44, p. 224], the first principal component of the observations (xij's) in the data matrix in equation (1) is that linear combination

$$C_1 = v_{i1}X_1 + \ldots + v_{p1}X_p \qquad (2)$$

where

c_1 = the first principal component

v_{i1} = the elements of the characteristic vector associated with the greatest characteristic root of the correlation matrix of the observations

x_p = the variables under consideration.

The second principal component accounts for as much of the residual variance as possible and can be represented mathematically as:

$$C = v_{i2}X_1 + \ldots + v_{p2}x_p \tag{3}$$

where

C_2 = the second principal component

v_{i2} = the elements of the characteristic vector associated with the second greatest characteristic root of the correlation matrix of the observations

x_p = the variables under consideration.

Additional components are derived until all of the variance has been analyzed [26, p. 136]. Thus, the jth principal component of the multivariate observations represented in the data matrix, x, of equation (1) is the linear combination

$$C_j = v_{ij}x_1 + \ldots + v_{pj}x_p \tag{4}$$

where

C_j = the jth principal component

v_{ij} = the coefficients of the jth component which are the elements of the characteristic vector associated with the jth largest characteristic root of the correlation matrix

x_p = the variables under consideration.

In the extreme case the first principal component would explain all the variation in the multivariate system. In the more usual case, the importance and usefulness of the component would be measured by the proportion of the total variance attributed to it. For example, if a significant proportion of the variation in a multivariate system of responses could be accounted for by a simple weighted average of the response values, it would appear that the variation could be expressed along a single continuum rather than in multidimensional space. Not only is this more efficient, but the relative importance of each original variable in the newly derived component would be indicated by the coefficients of the responses [44, pp. 224-225]. In calculating subsequent principal components, which are constructed so as to be linear compounds and orthogonal, the sum of the variances of all p principal components is equal to the sum of the variances of the original variables [26, p. 136]. A better understanding of principal component analyses can be facilitated by a more mathematically oriented specification of the model for component analysis.

Mathematical Model for Principal Component Analysis

As was mentioned previously the basic task of principal component analysis is to determine, first, an axis in the p-dimensional

sample space along which the variance is maximum; then a second axis, orthogonal to the first, which accounts for as much of the remaining variance as possible; and so on until all of the variance has been analyzed. Each new orthogonal axis accounts for a smaller proportion of the original variance.

Following Kendall [39, pp. 13-15] this basic task is accomplished by solving a set of p-homogeneous equations which result from partial differentiation of the function to be maximized, subject to a restriction that is introduced by use of LaGrange multipliers. Mathematically this can be represented in the following manner:

$$
\begin{aligned}
&v_1(1-\lambda) + v_2\gamma_{12} + v_3\gamma_{13} + \ldots + v_p\gamma_{1p} = 0\\
&v_1\gamma_{21} + v_2(1-\lambda) + v_3\gamma_{23} + \ldots + v_p\gamma_{2p} = 0\\
&v_1\gamma_{31} + v_2\gamma_{32} + v_3(1-\lambda) + \ldots + v_p\gamma_{3p} = 0\\
&\quad\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots\\
&v_1\gamma_{p1} + v_2\gamma_{p2} + v_3\gamma_{p3} + \ldots + v_p(1-\lambda) = 0
\end{aligned}
\tag{2}
$$

Equations (1) can be written as matrices as follows:

$$
\begin{bmatrix}
(1-\lambda) & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1m}\\
\gamma_{21} & (1-\lambda_i) & \gamma_{23} & \cdots & \gamma_{2m}\\
\gamma_{31} & \gamma_{32} & (1-\lambda_i) & \cdots & \gamma_{3m}\\
\vdots & & & & \\
\gamma_{pi} & \gamma_{p2} & \gamma_{p3} & \cdots & (1-\lambda_i)
\end{bmatrix}
\cdot
\begin{bmatrix} v_{1i}\\ v_{2i}\\ v_{3i}\\ \vdots\\ v_{pi} \end{bmatrix}
=
\begin{bmatrix} 0\\ 0\\ 0\\ \vdots\\ 0 \end{bmatrix}
\tag{3}
$$

Again, in matrix notation,

$$(R - \lambda iI)\ Vi = 0,\ i = 1, 2, \ldots p_j \qquad (4)$$

where

I = an identity matrix

0 = a null vector

R = the correlation matrix

λ_i = the ith eigenvalue.

There are p nontrivial solutions, each V_i representing the coefficients for converting the p variables to one of the new, uncorrelated components.

The trivial solution would be to set V = 0. For the nontrivial solutions, the determinant of the coefficient of V is zero.

$$(R - \lambda I) = 0 \qquad (5)$$

Equation (5) is the so-called characteristic equation of the correlation matrix, for which there are p possible values of λ. For a given root, λ_i, the corresponding vector V_i can be obtained by substituting the λ_i in the set of equations (2) and solving for V_i.

When the vectors are made normal, that is, when

$$\sum_{i=1}^{p} V_i^2 = 1 \qquad (6)$$

the variance of each set of components is λ_i. The vector V_1 produces the component of maximum variance, the variance being the value of the

largest characteristic root λ_1. Thus, the first principal component may be denoted as

$$C_1 = \sum_{j=1}^{p} v_{ij} x_j \tag{7}$$

where

C_1 = the first principal component

v_{ij} = the elements of the characteristic vector associated with the greatest characteristic root λ_1.

x_j = variables in the study.

It can be shown [39, p. 15] that the variance of C_1 is equal to λ_1.

If this process is continued until all the variance has been analyzed, the original variables $x_1, x_2 \ldots x_p$ will have been transformed into a new set of variables, $C_1, C_2 \ldots C_p$ which are uncorrelated and have variances $\lambda_1, \lambda_2, \ldots \lambda_p$ in decreasing order. Thus, the sum of the characteristic roots is equal to the total variance to be accounted for or

$$\sum_{i=1}^{p} \lambda_p = \sum_{j=1}^{p} r_{jj} \tag{8}$$

In conclusion, the value of principal component analysis is derived from its use as an objective means for summarizing the variation in a multivariate system in fewer variables [44, p. 228]. However, it must be remembered that while principal component analysis provides a means for determining the total variance and how it is partitioned, it does not provide a measure of a fundamental model for covariance structure.

In short, principal component analysis is merely a transformation of the original variables into new variables which may reduce the dimensions of variation of the original observations [39, p. 37]. In the present study this technique was utilized to derive an index of economic growth and is applied in Chapter IV.

IV. GENERAL FACTOR ANALYSIS

Introduction

In the study of regional economic growth the multiplicity of variables to be considered creates a complexity of relationships that requires simplification if the essential factors responsible for growth are to be understood. The effectiveness of researchers seeking this simplification is hampered by the lack of precision of conceptualization and by the lack of measurement.

In the study of regional growth some objective method is needed to provide insight into the chaos of multitudinous variables. The researcher cannot resort only to heuristic arguments based upon intuition and analogy in order to formulate testable hypotheses. Factor analysis is one statistical technique that is particularly useful in the early stages of a study for formulating hypotheses.

Factor analysis was originally developed by psychologists in connection with the determination of mental factors from scores on various psychological tests [26, p. 4]. With the advent of computer science factor analysis has been utilized in other social science fields. Haygood [28] used factor analysis to delineate subregions of the United

States. Schnore [55] performed a factor analysis to obtain factors describing world industrialization and urbanization. Anderson [5] and Kendall [40] have demonstrated the use of factor analysis in the analysis of economic time series. In a study of socio-economic and political change in underdeveloped countries Adelman and Morris [3] employed factor analysis to summarize the social, economic, and political characteristics of 74 underdeveloped countries. The results of this analysis were used in an econometric model in an attempt to obtain a statistical explanation of the economic and noneconomic forces which determine a country's capacity for economic growth. In a study of watershed development projects in Oklahoma, Cox and Badger [17] used factor analysis to reveal the underlying relationships important to watershed development. The derived factors were then utilized in a regression analysis in an attempt to predict the rate of change of upstream watershed development projects. In a study of the managerial behavior of farmers Morrow [45] used factor analysis to isolate, from observable and relevant management behavior of farm operators, basic mental processes which explained such behavior.

The Technique of Factor Analysis

The primary purpose of factor analyses is to reduce the original number of explanatory variables to a smaller number of independent factors in terms of which the whole set of variables can be understood [34, p. 17]. It accomplishes this purpose by identifying and summarizing the structural relationships of a system so as to combine highly interrelated variables into clusters called *factors*. Thus, factor analysis provides an objective

means for expressing the observed relationships presented in a correlation matrix in a smaller, but suitable linear space or dimension [34, p. 21].

Perhaps a better understanding of factor analysis can be achieved by comparing it with regression analysis. Factor analysis is similar to regression analysis in that it breaks down the original variance of a variable into variance components associated with the variation of a set of other quantities. This variable in regression analysis and the variables that account for its variation are known as the independent variables. However, in factor analysis, no distinction is made between independent and dependent variables. They all are dependent and independent in turn. Thus, while regression analysis is a study of dependence, factor analysis is a study of mutual interdependence [3, p. 131].

Factor analysis also differs from regression analysis in that the final explanatory variables of factor analysis are hypothetical constructs instead of observable magnitudes as is the case in regression analysis [26, p. 16]. These hypothetical constructs represent groupings of the original variables into clusters known as factors. Each cluster or factor is a result of a linear combination of the original variables included in the study [26, p. 14]. More specifically, each factor is an eigenvector of the correlation matrix among the original variables [26, pp. 137-143]. The mathematical procedure for constructing each factor from the observable variables involves the following: (1) highly intercorrelated variables are combined to form a single factor; (2) variables are allocated to a given factor so that they are most nearly independent

of the variables allocated to the other factors; (3) factors are derived so as to maximize the percentage of the total variance attributable to each successive factor; and (4) the factors are constructed so as to be uncorrelated with each other [3, p. 132].

As mentioned previously, factor analysis attempts to represent each variable as a linear combination of several underlying factors. As Harman [26, p. 15] has pointed out, two types of factors may be distinguished. Common factors are required to explain the intercorrelations among the variables while each unique factor accounts for the remaining variance of that variable. Since each variable can be expressed as a linear combination of the common factors and a unique factor, factor analysis can be interpreted as a regression of the observed variables on the unobserved common factors and on a specific factor [26, p. 16].

The coefficients of the factors are called factor loadings and serve to relate the observed variables to the common factors [26, p. 15]. These coefficients or factor loadings are interpreted in the same manner as regression coefficients in correlation analysis. The factor loadings are analyzed with respect to size and sign. The significant factor loadings determine the number of variables included in any one factor. Since the squared factor loadings represent the relative contribution of each factor to the total standardized variance of a variable, the summation of the squared row factor loadings indicates the extent to which the common factors account for the total unit variance of the variable [26, pp. 17-18]. These summations are called communalities and play the same role in factor analysis as that played by the coefficient of multiple

determination in regression analysis. Thus, the communalities are a measure of the variance of each variable explained by the extracted factors.

In summary, factor analysis is a statistical technique that can be used to simplify statistical data which arise from complex relationships encountered in a scientific endeavor. In the usual scientific study this simplification involves two steps. First, a theoretical model is formulated. Secondly, the data are analyzed and checked against the model. However, in the study of regional economic growth there is no adequate theory from which a model can be formulated and tested empirically. As Adelman and Morris [3, p. 133] have demonstrated, under these circumstances some alternative is required for formulating hypotheses. Factor analysis offers one such alternative. The principal usefulness of factor analysis in this regard has been pointed out by Thurston [64, p. 56]:

> Factor analysis is useful, especially in those domains where basic and fruitful concepts are essentially lacking and where crucial experiments have been difficult to conceive. The new methods have a humble role. They enable us to make only the crudest first map of a new domain. But if we have scientific intuition and sufficient ingenuity, the rough factorial map of a new domain will enable us to proceed beyond the exploratory factorial stage to the more direct forms of . . . experimentation . . . .

To facilitate a better understanding of factor analysis a mathematical specification of the statistical model is required. The next section is a technical specification of the classical factor analysis model.

The Statistical Model for Factor Analysis

To state the model more specifically, let z_j be the jth variable and z_{ji} be the value assumed by the jth variable on the ith observation. In the present study, the variables were the socio-economic indicators; the observations were the various counties included in the study region. Therefore, a particular z_{ji} represents the score assigned to the ith county on the jth socio-economic characteristic.

Recalling that the purpose of factor analysis is to represent each variable z_j (j-1, . . . n) as a linear combination of several underlying factors, the classical factor analysis model can be specified as [26, pp. 15-16; 45, p. 59]:

$$z_j = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jm}F_m + d_jU_j \quad (j = 1, 2, \ldots, n) \quad (9)$$

where

z_j = variable j

$a_{j1} \ldots a_{jm}$ = common factor coefficients

$F_1 \ldots F_m$ = common factors

U_j = unique factor

d_j = unique factor coefficient.

Assuming: (1) the common factors and unique factors have zero means and unit variances

(2) the unique factors are uncorrelated with each other and

(3) the total variance in the data matrix is composed of that due to common factors (communality) and

that due to a specific factor in each item and to error in measurements (uniqueness).

Following Morrow [45, p. 59], assumptions (1) and (3) can be expressed mathematically as:

$$s_j^2 = 1 = h^2_j + d_j^2$$

$$= a^2_{j1} + a^2_{j2} + \cdots + a_{jm} + d_j^2 \qquad (10)$$

where

s_j^2 = total variance of variable

h^2_j = communality or variance of variable; attributable to the common factors

d_j^2 = unique variance or variance of variable j attributable to unique factors.

Assumption (2) can be expressed mathematically as:

$$a_{j1}a_{k1} + a_{j2}a_{k2} + \cdots + a_{jm}a_{km} = r_{jk} \qquad (11)$$

where

$a_{j1} \cdots a_{jm}$ = common factor coefficients for variable j

$a_{k1} \cdots a_{km}$ = common factor coefficients for variable k

r_{jk} = correlation coefficient for variable j and k.

Equation (11) states that the correlations among the variables is accounted for by m common factors [26, p. 15]. Thus, the basic computational problem of factor analysis is the estimation of the mm factor loadings

of the common factors [26, p. 16]. This is accomplished by determining a matrix of common factor coefficients which will produce a reproduced correlation matrix that best approximates the observed correlations in the data matrix [45, p. 60].

In sum, the primary purpose of factor analysis is to reduce the original number of explanatory variables to a smaller number of independent factors in terms of which the whole set of variables can be understood. It accomplishes this objective by combining highly interrelated variables into factors which provide the researcher with a simpler, more compact explanation of the regularities apparent in the empirical results. Factor analysis was used in this study to simplify the complexity of interrelationships which evolved from a multiplicity of variables affecting regional economic growth. This technique is applied in Chapter V.

V. MULTIPLE DISCRIMINANT ANALYSIS

Introduction

In this study it was hypothesized that socio-economic data were reflections of man's reaction to his environment. If this is true, observations on data of county variables can be utilized to classify counties according to their performance in terms of economic growth. In order to make this classification some objective means was needed for detecting the differences among counties. One such statistical technique is discriminant analysis.

Multiple discriminant analysis is a statistical method which may be used in a multivariate comparison of several groups. The comparison

is made through an examination of the group membership of individuals on the basis of a set of attributes of those individuals [39, p. 144]. More specifically, multiple discriminant analysis determines statistically by an analysis of variance, those linear combinations of variables which best discriminate among various groups of individuals [53, p. 237].

While discriminant analysis is a fairly new tool in the economist's tool kit, it has been applied to a variety of problems. Durand [19] applied discriminant analysis to a set of financial data in an attempt to discriminate between good and bad loans. In a study of the differences between prices of producers' goods and prices of consumers' goods, Tintner [67] employed discriminant analysis as a means of distinguishing between the two groups of prices on the basis of certain measurements connected with their behavior during the business cycle. In a study of agricultural land prices in Oklahoma, Abdel-Badie and Parcher [1] used discriminate analysis to show whether it is possible to discriminate between groups of land sales which had a wide range in price. Adelman and Morris [2] used discriminate analysis to devise relatively objective criteria for selecting underdeveloped countries with immediate development potential. Cox and Badger [17] applied discriminant analysis to a set of socio-economic variables in order to discriminate between four groups of upstream watershed representing different rates of development.

The Linear Discriminant Function

As mentioned previously discriminant analysis is a statistical method that can be used to discriminate between i groups of observations

on the basis of several observed variables. The discriminant function accomplishes this by (1) separating the means of the i groups, and (2) concentrating the variables in each group about their respective means in order to make the probability of correct classification as large as possible. Thus, the discriminant function combines a set of variables in such a manner that the difference between the means of the i groups per unit of dispersion about those means is maximum [51, p. 408]. In this sense the discriminant function can be said to be an optimum discriminator [53, p. 288]. Furthermore, given certain assumptions, the multivariate populations of the groups which serve as inputs for determining the discriminant function can be used in a test of the effectiveness of the discriminatory power of the function [51, p. 408].

Some insight into the usefulness of the linear discriminant function can be obtained by briefly comparing it to linear regression analysis. The discriminant function differs from the linear regression function in its treatment of the dependent variable. In a linear regression function the values of the dependent variable are used to determine a linear function that will predict the values of the dependent variable. However, in the linear discriminant function a classification of the data among the various groups is used to determine the linear function. Therefore discriminant analysis overcomes the problems usually associated with quantitatively specifying a dependent variable [32, p. 180].

A better understanding of discriminant analysis is facilitated through a mathematical specification of the statistical model for linear discrimination. While the model presented in the next section is for

discriminating between two groups of individuals, it is equally applicable to a problem involving several groups.

The Statistical Model for Linear Discrimination

The statistical model for linear discrimination is based on several assumptions. First, it is assumed that the group assignment procedure is derived from a model of a multivariate normal distribution of observations within groups. It is further assumed that individuals from the groups appear in intermingled random order for assignment to one of the groups. Before assignment it is not possible to tell to which group an individual belongs. Observations of k variables, $x_1, x_2, \ldots x_k$, are made for each group and are multivariate normal in the k variables.

Let $\sigma^2_{ii.1}$ be the variance of the ith variable in the first group and $\sigma_{ij.1}$ be the covariance of the ith and jth variables in the first group. The counterparts for the second group are $\sigma^2_{ii.2}$ and $\sigma_{ij.2}$. The model assumes that the variance-covariance matrix of the first group equals, component by component, the variance-covariance matrix of the second group or

$$(\sigma_{ij.1}) = (\sigma_{ij.a}) \tag{12}$$

for all pairs of 1, j formed from the numbers 1, 2, . . . k.

Since the variance-covariance matrices are equal no distinction will be made between the variances and covariances of the g groups in the following discussion. In the remainder of the discussion σ^2_i will

represent the variance of the ith variable in any of the groups and σ_{ij} will represent the covariance of the ith and jth variables.

A second assumption of the model for linear discrimination is that the means of the two groups differ [32, p. 180]. Mathematically this means that the point denoted by $(\mu_{11}, \ldots, \mu_{k1})$ does not coincide with the point $(\mu_{12}, \ldots, \mu_{k2})$ where μ_{jj} represents the mean of the ith variable (i = 1, . . . , k) in the jth group (j = 1, 2).

Given the above assumptions it is possible to form two univariate normal groups of individuals. This is accomplished by taking a linear combination of the set of k variables by means of which it is desired to discriminate among the two groups of individuals or

$$Y = \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k \tag{13}$$

where

Y = the value of the linear combination of k variables

$\lambda_1 \ldots \lambda_k$ = the coefficients of the k variables in the linear combination

$x_1 \ldots x_k$ = the k variables.

When each k variable observation in the two groups of individuals is subjected to equation (13) the resulting groups will have identical variances or

$$\sigma^2_{y.1} = \sigma^2_{y.2} = \sigma^2_y \tag{14}$$

but their means μ_{y1} and μ_{y2} will differ. Thus the problem in discriminant analysis is to determine the λ's that will enable y to serve as an index for differentiating among members of the two groups [32, p. 180]. This is accomplished by solving a set of simultaneous linear equations derived from equations (13) and (14) of the form.

$$\begin{aligned}
&\lambda_1\sigma^2_1 + \lambda_2\sigma_{12} + \ldots + \lambda_k\sigma_{1k} = \mu_{11} - \mu_{12},\\
&\quad\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots\\
&\lambda_1\sigma_{k1} + \lambda_2\sigma_{k2} + \ldots + \lambda_k\sigma^2_k = \mu_{k1} - \mu_{k2}.
\end{aligned} \tag{15}$$

Hoel [32, pp. 181-183], and Kendall [39, p. 147] have shown that the values of σ_i determined by solving equations (15) will maximize the expression

$$\sigma = \frac{(\mu_{y1} - \mu_{y2})^2}{\sigma^2_y} \tag{16}$$

Therefore, the determination of the λ's will maximize the square of the difference between the means of the linearly transformed observations per unit of their variance. In order for the difference between the means to be a maximum the square of the difference has to be a maximum [51, p. 410]. This property is illustrated by equation (16). In sum, when the lambda coefficients of equation (13) are determined by the above procedure the resulting equation becomes a linear discriminant function.

In conclusion discriminant analysis can be used as a unified approach in solving a problem involving multivariate comparisons of several groups. In this study the technique was used to evaluate several different environmental variables which may affect group performance and to establish significant group differences. The discriminate analysis is performed in Chapter VI.

VI. SUMMARY

In this study it was hypothesized that man's reactions are a function of the socio-economic characteristics of his environment or community. It was also hypothesized that observations on these characteristics can be analyzed so as to gain insight into the mechanics of regional economic growth.

In order to accomplish this analysis three multivariate statistical models were employed along with stepwise regression analysis. Principal component analysis was proposed as a technique for transforming the original variables into new variables which may reduce the dimensions of variation of the original observations and serve as an index of economic growth.

Factor analysis was proposed as a technique for explaining the possible interactions or groupings of the socio-economic variables so as to provide a simpler, more compact explanation of the regularities apparent in the data. The established patterns or regularities can then be used as a basis for an analysis of the forces affecting regional economic growth.

Multiple discriminant analysis was proposed as the statistical method to use in a multivariate comparison of several groups of counties. The results of this analysis can be utilized to classify counties on the basis of their performance as it relates to economic growth.

The examination of the three statistical models suggested that they were highly complementary and provided a way to comprehensively and objectively analyze the relationship between man's socio-economic environment and regional economic growth. These analyses included not only the delineation of important variables and factors contributing to regional economic growth, but also the capability to predict future economic growth of the region.

CHAPTER IV

PRINCIPAL COMPONENT ANALYSIS OF ECONOMIC GROWTH IN THE TENNESSEE VALLEY REGION

I. INTRODUCTION

The problem of measuring the rate of regional economic growth is difficult. The usual approach uses some proxy variable such as the level or change in per capita income as an overall indicator of economic growth. However, while such an approach is parsimonious it does not always suffice as the sole criterion of the level of regional economic growth [30, p. 10]. The problem of deriving an index of economic growth may be approached by the use of principal component analysis. The purpose of this chapter is to illustrate how the technique was applied to the study of economic growth in the Tennessee Valley region.

II. DATA REQUIREMENTS

Principal component analysis was applied to 13 of the 23 measure variables in an attempt to construct an index of economic growth for the 150 county Tennessee Valley region. Using correlation analysis as a screening technique on the initial 60 variables, the following 13 measure variables were selected for the principal component analysis: (1) rate of population change, (2) change in median income of all families, (3) change in median income of rural families, (4) change in median income of rural farm families, (5) change in total property tax, (6)

change in median value of owner occupied housing units, (7) change in the percent of dwellings structurally sound with all plumbing facilities, (8) change in the percent employed in agriculture, (9) change in retail sales per capita, (10) change in value of farm land and buildings, (11) ratio of bank deposits to aggregate income, (12) change in total payroll per capita, and (13) change in the value of manufacturing per capita.

III. THE ANALYSIS OF THE 150 COUNTY TENNESSEE VALLEY REGION

Computational Procedure

The computational procedure [18, pp. 150-154] utilized in the program can be specified as follows. Given the observations in the form

$$X_{ij}$$

where $i = 1, 2, \ldots n$ counties and $j = 1, 2, \ldots m$ variables, the data were centered and normed such that

$$W = (w_{ij}) = (X_{ij} - X^{\bullet}_{j}) \div \left[\sum_{i=1}^{n} (X_{ij} - X_{ij})^2 \right]^{1/2} \tag{1}$$

where

$$X^{\bullet}_{j} = \frac{1}{h} \sum_{i=1}^{n} X_{ij} \tag{2}$$

or the mean value of the jth variable in the ith county.

The correlation matrix R of the standardized variables w_{ij} was computed

$$C = (c_{jk}) = \sum_{i=1}^{n} (w_{ij}) \cdot (w_{ik}) \quad k = 1, 2, \ldots m \tag{3}$$

Next, the eigenvalues and corresponding eigenvectors of the correlation matrix were computed so as to solve a system of linear simultaneous equations specified as

$$(R - \lambda I)v = 0 \tag{4}$$

where

R = the correlation matrix

λ = eigenvalue

v = the matrix whose columns are the corresponding eigenvectors.

Denoting λ_j as the jth eigenvalue and setting $V = (v_{kj})$ the data (W's) were transformed into their orthogonal components by the matrix multiplication

$$Z = (z_{ij})$$

$$= \sqrt{n-1} \sum_{k=1}^{m} (w_{ik})\ (v_{kj}) \tag{5}$$

The results obtained from equation (5) were used to rank order each standardized county by the size of each principal component separately.

Results of the Analysis

The data representing the 13 selected variables for the 150 county study region were utilized in a principal component analysis program [18, pp. 150-154]. The results of this analysis are reported in Table IV-1 and indicated that the first principal component accounted for 40 percent of the total variation of the 13 variable multivariate system. In other words, 40 percent of the variation in the multivariate system

TABLE IV-1

COMPONENT COEFFICIENTS FOR THE FIRST PRINCIPAL COMPONENTS OF THE MEASURE VARIABLES FOR THE 150 COUNTY TENNESSEE VALLEY REGION

Variable Number	Description	Component Coefficient
1	Rate of population change	0.3747
2	Change in median income of all families	0.3843
3	Change in median income of rural families	0.2981
4	Change in median income of rural farm families	0.3019
5	Change in total property tax	0.1567
6	Change in median value of owner occupied housing units	0.2515
7	Change in the percent of dwellings structurally sound with all plumbing facilities	0.1817
8	Change in the percent employed in agriculture	0.2066
9	Change in retail sales per capita	-0.2091
10	Change in value of farm land and buildings	0.2849
11	Ratio of bank deposits to aggregate income	0.2093
12	Change in total payroll per capita	0.3578
13	Change in the value of manufacturing per capita	0.1568
	Variance	5.1568
	Percentage of total variance	40.0

of 13 responses was accounted for by a simple weighted average of the response values. Thus, it would appear that the variation in the measure variables of economic change can be expressed along a single continuum rather than in a multidimensional space.

As Table IV-1 indicates, the first principal component of the 13 measure variables was the linear compound given by

$$\begin{aligned} C_1 = {} & 0.3747\,X_1 + 0.3843\,X_2 + 0.2981\,X_3 + 0.3019\,X_4 \\ & + 0.1567\,X_5 + 0.2515\,X_6 + 0.1818\,X_7 + 0.2066\,X_8 \\ & - 0.2091\,X_9 + 0.2849\,X_{10} + 0.2093\,X_{11} + 0.3578\,X_{12} \\ & + 0.2678\,X_{13} \end{aligned} \qquad (1)$$

where

C_1 = the first principal component of the 13 measure variables, and

$X_1 \ldots X_{13}$ = the 13 measure variables of economic change.

Not only is this linear compound, C_1, more parsimonious, but the relative importance of each original variable in the newly derived component is indicated by the coefficients of the 13 variables in equation (1). It would appear that in the linear index of economic growth represented by C_1, the most important variables were (1) change in median income of all families, (2) rate of population change, and (3) change in total payroll per capita (Table IV-1).

As a check on the representativeness of the newly derived component as an index of economic growth, the correlation coefficients for the first principal component and the measure variables were computed (Table IV-2). The resulting component correlation coefficients were

TABLE IV-2

COMPONENT CORRELATION COEFFICIENTS FOR THE FIRST PRINCIPAL COMPONENT OF THE MEASURE VARIABLES FOR THE 150 COUNTY TENNESSEE VALLEY REGION

Variable Number	Description	Component Correlation Coefficient
1	Rate of population change	0.8509
2	Change in median income of all families	0.8727
3	Change in median income of rural families	0.6769
4	Change in median income of rural farm families	0.6856
5	Change in total property tax	0.3558
6	Change in median value of owner occupied housing units	0.5711
7	Change in the percent of dwellings structurally sound with all plumbing facilities	0.4126
8	Change in the percent employed in agriculture	0.4692
9	Change in retail sales per capita	-0.4748
10	Change in value of farm land and buildings	0.6470
11	Ratio of bank deposits to aggregate income	0.4753
12	Change in total payroll per capita	0.8125
13	Change in the value of manufacturing per capita	0.6081

of sufficient magnitude to justify interpreting the first principal component as a general index of economic growth. Thus, the linear compound of the 13 measure variables as defined in equation (1) became the index of economic growth which was utilized in the remainder of this study.

The first principal component was provisionally identified with economic growth and the 150 counties were arranged in order according to the magnitude of C_1 as given in equation (1). The results of this ranking are presented in Table IV-3. The ranked counties were then grouped into fast growth, medium growth, slow growth, and no growth groups according to the values of C_1 which they bore. The results agreed with general knowledge about the geographical distribution of economic growth in the Tennessee Valley region.

More specifically, the fast growth counties included the first 13 counties listed in Table IV-3. To be included in this group a county had to have a value of C_1 which was equal to or greater than 0.3143. The second group or the medium growth counties was composed of 21 counties. This group included those counties which had a value of C_1 less than 0.3143 but equal to or greater than 0.1057. The third group of counties was designated as slow growth counties and contained 25 counties. To belong in this group a county had to have a value of C_1 which was less than 0.1057 but greater than or equal to zero. The fourth group contained 90 counties and was labeled no growth counties. To be classified as a no growth county a county had to have a value of C_1 which was less than zero. Thus, as Table IV-3 indicates the majority

TABLE IV-3

THE RANK ORDER OF EACH COUNTY IN THE TENNESSEE VALLEY REGION ORDERED BY THE SIZE OF THE FIRST PRINCIPAL COMPONENT OF ECONOMIC GROWTH

Size of the First Principal Component	County Number	County Name
0.6718	124	Roane (Tennessee)
0.5308	11	Madison (Alabama)
0.4611	31	Marshall (Kentucky)
0.4364	70	Davidson (Tennessee)
0.4267	84	Hamilton (Tennessee)
0.4114	24	Walker (Georgia)
0.3863	133	Sullivan (Tennessee)
0.3859	30	McCracken (Kentucky)
0.3657	130	Shelby (Tennessee)
0.3611	2	Colbert (Alabama)
0.3572	25	Whitfield (Georgia)
0.3305	83	Hamblen (Tennessee)
0.3143	16	Catoosa (Georgia)
0.2966	14	Morgan (Alabama)
0.2892	57	Bradley (Tennessee)
0.2857	43	Henderson (North Carolina)
0.2799	98	Knox (Tennessee)
0.2749	49	Transylvania (North Carolina)
0.2676	105	McMinn (Tennessee)
0.2504	42	Haywood (North Carolina)
0.2265	21	Rabun (Georgia)
0.2188	104	Loudon (Tennessee)
0.2166	67	Coffee (Tennessee)
0.1966	17	Dade (Georgia)
0.1930	8	Lauderdale (Alabama)
0.1655	126	Rutherford (Tennessee)
0.1520	111	Maury (Tennessee)
0.1489	141	Washington (Tennessee)
0.1422	38	Buncombe (North Carolina)
0.1381	5	Etowah (Alabama)
0.1332	56	Blount (Tennessee)
0.1330	52	Anderson (Tennessee)
0.1118	121	Polk (Tennessee)
0.1057	114	Montgomery (Tennessee)
0.0793	26	Calloway (Kentucky
0.0769	101	Lawrence (Tennessee)
0.0743	53	Bedford (Tennessee)

TABLE IV-3 (continued)

Size of the First Principal Component	County Number	County Name
0.0726	109	Marion (Tennessee)
0.0688	146	Wilson (Tennessee)
0.0615	128	Sequatchie (Tennessee)
0.0605	96	Jefferson (Tennessee)
0.0600	102	Lewis (Tennessee)
0.0599	13	Marshall (Alabama)
0.0594	134	Sumner (Tennessee)
0.0552	110	Marshall (Tennessee)
0.0517	33	Alcorn (Mississippi)
0.0504	150	Washington (Virginia)
0.0490	145	Williamson (Tennessee)
0.0465	77	Franklin (Tennessee)
0.0319	129	Sevier (Tennessee)
0.0256	117	Obion (Tennessee)
0.0254	27	Graves (Kentucky)
0.0204	20	Lumpkin (Georgia)
0.0177	140	Warren (Tennessee)
0.0173	81	Greene (Tennessee)
0.0131	73	Dickson (Tennessee)
0.0108	28	Livingston (Kentucky)
0.0102	125	Robertson (Tennessee)
0.0079	91	Henry (Tennessee)
-0.0022	48	Swain (North Carolina)
-0.0030	22	Towns (Georgia)
-0.0177	62	Cheatham (Tennessee)
-0.0178	92	Hickman (Tennessee)
-0.0237	47	Mitchell (North Carolina)
-0.0242	66	Cocke (Tennessee)
-0.0256	136	Trousdale (Tennessee)
-0.0274	45	Macon (North Carolina)
-0.0287	108	Madison (Tennessee)
-0.0331	44	Jackson (North Carolina)
-0.0337	113	Monroe (Tennessee)
-0.0352	10	Limestone (Alabama)
-0.0352	29	Lyon (Kentucky)
-0.0456	50	Watauga (North Carolina)
-0.0464	103	Lincoln (Tennessee)
-0.0467	1	Blount (Alabama)
-0.0502	23	Union (Georgia)
-0.0533	15	Winston (Alabama)

TABLE IV-3 (continued)

Size of the First Principal Component	County Number	County Name
-0.0533	147	Grayson (Virginia)
-0.0541	123	Rhea (Tennessee)
-0.0548	61	Carter (Tennessee)
-0.0555	74	Dyer (Tennessee)
-0.0583	137	Unicoi (Tennessee)
-0.0633	144	White (Tennessee)
-0.0691	94	Humphreys (Tennessee)
-0.0728	59	Campbell (Tennessee)
-0.0756	32	Trigg (Kentucky)
-0.0775	4	DeKalb (Alabama)
-0.0785	88	Hawkins (Tennessee)
-0.0794	19	Gilmer (Georgia)
-0.0806	87	Hardin (Tennessee)
-0.0820	82	Grundy (Tennessee)
-0.0836	120	Pickett (Tennessee)
-0.0868	41	Graham (North Carolina)
-0.0880	69	Cumberland (Tennessee)
-0.0881	93	Houston (Tennessee)
-0.0888	143	Weakley (Tennessee)
-0.0897	115	Moore (Tennessee)
-0.0907	7	Jackson (Alabama)
-0.0920	122	Putnam (Tennessee)
-0.0925	54	Benton (Tennessee)
-0.0962	34	Itawamba (Mississippi)
-0.1010	39	Cherokee (North Carolina)
-0.1030	12	Marion (Alabama)
-0.1035	138	Union (Tennessee)
-0.1051	18	Fannin (Georgia)
-0.1051	60	Carroll (Tennessee)
-0.1060	37	Avery (North Carolina)
-0.1070	80	Grainger (Tennessee)
-0.1074	127	Scott (Tennessee)
-0.1078	107	Macon (Tennessee)
-0.1146	78	Gibson (Tennessee)
-0.1164	139	Van Buren (Tennessee)
-0.1266	6	Franklin (Alabama)
-0.1270	51	Yancey (North Carolina)
-0.1275	90	Henderson (Tennessee)
-0.1290	58	Campbell (Tennessee)
-0.1319	72	DeKalb (Tennessee)

TABLE IV-3 (continued)

Size of the First Principal Component	County Number	County Name
-0.1349	142	Wayne (Tennessee)
-0.1363	76	Fentress (Tennessee)
-0.1397	71	Decatur (Tennessee)
-0.1445	63	Chester (Tennessee)
-0.1479	36	Tishomingo (Mississippi)
-0.1580	119	Perry (Tennessee)
-0.1620	149	Scott (Virginia)
-0.1624	40	Clay (North Carolina)
-0.1628	3	Cullman (Alabama)
-0.1635	79	Giles (Tennessee)
-0.1662	135	Tipton (Tennessee)
-0.1676	9	Lawrence (Alabama)
-0.1722	131	Smith (Tennessee)
-0.1773	35	Prentiss (Mississippi)
-0.1775	65	Clay (Tennessee)
-0.1785	46	Madison (North Carolina)
-0.1793	118	Overton (Tennessee)
-0.1813	116	Morgan (Tennessee)
-0.1824	55	Bledsoe (Tennessee)
-0.1954	86	Hardeman (Tennessee)
-0.1955	68	Crockett (Tennessee)
-0.1958	132	Stewart (Tennessee)
-0.1968	112	Meigs (Tennessee)
-0.1992	97	Johnson (Tennessee)
-0.2108	106	McNairy (Tennessee)
-0.2130	89	Haywood (Tennessee)
-0.2538	64	Claiborne (Tennessee)
-0.2639	95	Jackson (Tennessee)
-0.2828	99	Lake (Tennessee)
-0.2856	85	Hancock (Tennessee)
-0.3046	148	Lee (Virginia)
-0.3214	75	Fayette (Tennessee)
-0.3393	100	Lauderdale (Tennessee)

of the counties in the Tennessee Valley region can be classified as no growth counties on the basis of their values for the first principal component.

A clearer picture of the geographical distribution of economic growth in the Tennessee Valley region was obtained by identifying the four groups of counties on a map. This is accomplished in Figure IV-1 and, in general, indicated that (1) economic growth during 1950-1960 in the Tennessee Valley tended to be urban oriented, and (2) the faster growing counties were generally located along an improved waterway or river.

Reference to Figure IV-1 gives a visual confirmation of the hypothesis that regional growth occurs in specific locational matrices which are urban-industrial in composition. For example, every one of the 13 counties designated as a fast growth county was located in or around a metropolitan area. Figure IV-1 also indicates that these fast growth counties serve as nodes for economic growth with outlying counties being classified either as medium growth or slow growth counties. It would therefore seem likely that the sphere of influence of each of the nodal counties or fast growth counties can be estimated with the aid of principal component analysis.

The second indication derived from the mapping of the four groups of counties concerned the apparent relationship between economic growth and location along a river. Reference to Figure IV-1 indicates more clearly this relationship. It would appear from this figure that the fastest growing counties in the Tennessee Valley were generally located

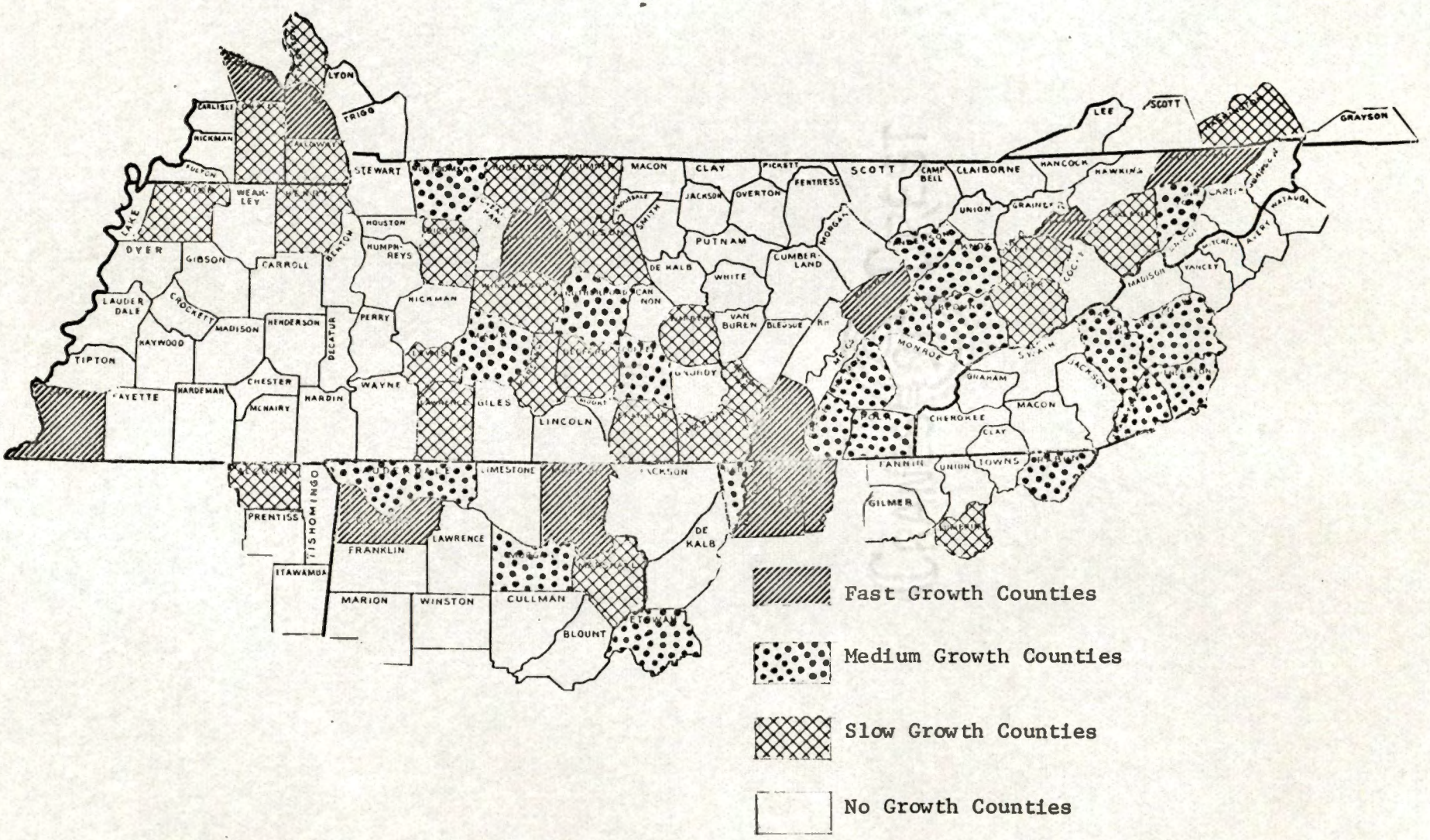


Figure IV-1. The geographical distribution of the fast growth, medium growth, slow growth, and no growth counties in the Tennessee Valley region, 1950-1960.

along a river or waterway which had received some type of installation such as a dam or steam generator plant. This leads to the hypothesis that there is some positive relationship between economic growth in the Tennessee Valley region and natural resource investments in the form of dams and steam generator plants.

CHAPTER V

USE OF DISCRIMINANT ANALYSIS TO EVALUATE THE ECONOMIC GROWTH OF THE 150 COUNTY TENNESSEE VALLEY REGION

I. INTRODUCTION

In this chapter multiple discriminant analysis was applied to detect the differences in the value of the index of economic growth among the counties under study. As was reported in Chapter IV the indices of economic growth were arrayed in order of magnitude and divided into four groups. If it can be shown that the selected groups were significantly different then the derived classification functions will facilitate the separation of the counties under study into groups exhibiting similar characteristics of economic growth. Thus, discriminant analysis was used to make a multivariate comparison of the four groups. In this approach it had two phases: (1) the establishment of significant group differences and (2) the study and explanation of these differences. In this study the technique was utilized primarily to evaluate several different socio-economic variables which may affect group performance and to establish significant group differences.

II. DATA REQUIREMENTS

Twenty-five independent variables for the system under study were selected from more than 37 socio-economic characteristics which were thought to affect regional economic growth. The following variables

were selected for the year 1960: (1) percent urban population, (2) percent rural farm population, (3) percent non-white population, (4) percent of population 65 and over, (5) percent of population voting in the 1960 Presidential election, (6) percent of families with incomes under $3,000, (7) percent of families with incomes over $10,000, (8) median school years completed, (9) percent of migrants from different county, (10) percent of population in labor force, (11) percent of employment in white collar occupations, (12) percent of employment working outside county of residence, (13) average size of household, (14) per capita property taxes, (15) number of towns and municipalities, (16) non-worker to worker ratio, (17) value of mineral industries shipments and receipts, (18) percent of farms with sales under $2,500, (19) percent of farms with sales $10,000 and over, (20) average per farm value of farm products sold, (21) farm operator level-of-living index, (22) local government general expenditures on natural resources, (23) percent of land in capability Classes I-IV, (24) location of TVA steam generator plants (dummy), and (25) location of TVA dams (dummy).

III. COMPUTATIONAL PROCEDURE

The data representing the 25 selected variables for the 150 county Tennessee Valley region were utilized in a multiple discriminant analysis program for several groups [18, pp. 196-203]. The computational procedure of this program can be outlined as follows:

Given the observations in the form

$$X_{ijk}$$

where

$i = 1, \ldots g$, number of groups

$j = 1, \ldots n_i$, sample size of the ith group

$k = 1, \ldots m$, number of variables

and denoting the means of variables considered within each group as

$$X_{i.k}$$

the matrix S_i, which represents the sum of cross products of deviation from the means can be computed. This matrix can be described by

$$S_i = (s^i_{kk'}) \qquad k' = 1, \ldots m \tag{1}$$

where

$$s^i_{kk'} = \sum_{j=1}^{ni} (X_{1jk} - X_{i.k})(X_{ijk'} - X_{1.k}) \tag{2}$$

Based on the matrices S_i, the pooled dispersion matrix D was computed such that

$$D = \frac{\sum_{i=1}^{g} s_i}{\left(\sum_{i=1}^{g}\right) - g} \tag{3}$$

with the common mean

$$X_{..k} = \sum_{i=1}^{g} n_i X_{i.k} \Big/ \sum_{i=1}^{g} n_i \tag{4}$$

The pooled dispersion matrix is then inverted to obtain $D_{kk'}$-1 which was used to calculate the Generalized Mahalanobis D^2 statistic, V or

$$V = \sum_{k=1}^{m} \sum_{k'=1}^{m} (D_{kk'}^{-1}) \sum_{i=1}^{g} n_i (X_{i.k} - x_{..k}) (X_{i.k} - X_{..k'}) \qquad (5)$$

The D^2 statistic can be used as a chi-square with m(g-1) degrees of freedom to test the hypothesis that the mean values are the same in all groups [18, p. 201]. If the null hypothesis of no difference is rejected then the independent variables can be said to be capable of discriminating among groups. However, if it is not rejected, then the data indicate no significant group differences and alternative variables should be selected and the above processes repeated.

Upon failure to reject the hypothesis of no difference among means, the second phase was to calculate the (i*)th discriminating function such that

$$f_{i*}^{mx1} = \sum_{l=1}^{m} Z_l C_{i*1} + C_{i*o} \qquad (6)$$

where

i^* = 1, . . . g, the number of functions

l = 1, . . . m, the number of variables

Z_l = observation for each variable

C_{i*1} = classification function coefficient

C_{i*o} = constant

and

$$C_{i*1} = \sum_{k=1}^{m} d_{lk} X_{1.k}$$

$$C_{i*o} = -1/2 \sum_{k=1}^{m} \sum_{k'=1}^{m} d_{kk'} X_{1.k} X_{1.k'}$$

where

$$(d_{k1}, d_{k2}, \ldots d_{km}) = \text{kth row of } D^{-1}.$$

The (i*)th discriminating function can then be used to evaluate each data point such that for each observation,

$$P_i = \frac{e^{(f_{i*} - \max f_{i*})}}{\sum_{i*} e^{(f_{i*} - \max f_{i*})}} \qquad (7)$$

results in probabilities of classification into each of the i groups. The final classification into a particular group is determined by the selection of the highest probability among each of the groups for every observation. If the experimental groups are widely separated, then the diagonal of largest probabilities of the frequency matrix will contain a large percentage of the frequencies compared to the off-diagonal elements. This gives a visual test of the ability to discriminate among groups.

IV. THE ANALYSIS OF THE 150 COUNTY TENNESSEE VALLEY REGION

Multiple discriminant analysis was used to examine the group classification of counties on the basis of a set of 25 socio-economic characteristics of those counties. An examination of the mean scores by group for these variables revealed no significant patterns or trends between groups (Table V-1). Thus, a precursory examination of the data did not indicate that the four groups were different from each other.

TABLE V-1

MEAN SCORES OF 25 SELECTED SOCIO-ECONOMIC VARIABLES BY GROUP
FOR THE 150 COUNTY TENNESSEE VALLEY REGION

Variable Number	Description	Mean Scores for Groups* 1	2	3	4
1	Percent urban population	34.000	19.052	16.596	20.718
2	Percent rural farm population	26.292	21.395	25.344	32.544
3	Percent non-white population	10.185	50.490	4.880	9.673
4	Percent of population 65 and over	8.323	9.800	9.348	9.920
5	Percent of population voting in the 1960 Presidential election	22.100	31.857	39.348	29.757
6	Percent of families with incomes under $3,000	46.208	46.210	50.692	52.659
7	Percent of families with incomes over $10,000	5.531	4.224	4.440	3.855
8	Median school years completed	8.746	8.376	8.460	8.264
9	Percent of migrants from different county	14.915	15.419	12.044	12.479
10	Percent of population in labor force	32.515	31.786	32.248	32.173
11	Percent of employment in white collar occupations	28.823	27.214	27.476	25.284
12	Percent of employment working outside county of residence	17.654	21.410	13.952	15.682
13	Average size of household	3.654	3.533	3.592	3.596
14	Per capita property taxes	105.308	100.810	111.560	97.341
15	Number of towns and municipalities	5.846	2.571	2.520	3.011
16	Non-worker to worker ratio	1.088	1.131	1.162	1.066
17	Value of mineral industries shipments and receipts	649.692	292.333	709.040	787.132
18	Percent of farms with sales under $2,500	18.700	14.581	17.720	18.777
19	Percent of farms with sales $10,000 and over	9.923	9.005	3.548	4.990
20	Average per farm value of farm products sold	2190.769	827.048	725.360	1507.253
21	Farm operator level of living index	65.000	70.381	60.720	70.077
22	Local government general expenditures on natural resources	16.769	7.048	9.920	12.088
23	Percent of land in capability Classes I-IV	64.092	50.495	34.024	54.952

TABLE V-1 (continued)

Variable Number	Description	Mean scores for groups*			
		1	2	3	4
24	Location of TVA steam generator plants (dummy)	0.154	0.048	0.000	0.099
25	Location of TVA dams (dummy)	0.231	0.143	0.360	0.198

*Sample size for groups 1, 2, 3, and 4 was 13, 21, 25, and 91, respectively.

However, the calculated Mahalanobis D^2 statistic of 276.9773 when tested as a chi-square with 75 degrees of freedom allowed rejection at the .05 significance level of the hypothesis that the mean values were the same in all the four groups for these 25 variables.

The distinctive feature of D^2 is that it is a measure of distance rather than a criterion for testing the hypothesis of zero distance. This provided the additional information that the four groups were widely separated since D^2 calculated (276.9793) is greater than D^2 tabulated (96.2000) by a large margin. It must be remembered that distance, being a scalar quantity, does not exhaust the information in a comparison among three or more groups unless the latter happens to be collinear. The D^2 statistic allowed transition from the multi-group concept of configuration to a two-group concept of distance, which is similar to canonical reduction.

The multiple discriminant analysis program was used to compute the matrix of classification function coefficients and constants (Table V-2). Comparisons between the classification function coefficients and the original mean values indicated that certain variables underwent considerable change in relative importance (Tables I-1, page 6, and V-2). For instance, variables 8, 13, 16, 24 and 25 increased greatly while variables 2 and 7 decreased. This indicates that median school years completed (8), average size of household (13), non-worker to worker ratio (16), location of TVA steam generator plants, and location of TVA dams were more important in discriminating among groups than the raw data indicated. On the other hand, percent rural farm population (2)

TABLE V-2

CLASSIFICATION FUNCTION COEFFICIENTS FOR 25 SELECTED SOCIO-ECONOMIC VARIABLES BY GROUP FOR THE 150 COUNTY TENNESSEE VALLEY REGION

Variable Number	Description	Function Coefficient for Group 1	2	3	4
1	Percent urban population	-0.5243	-0.5141	-0.5329	-0.4930
2	Percent rural farm population	-2.5774	-2.6155	-2.6353	-2.5809
3	Percent non-white population	-0.1048	-0.0918	-0.0975	-0.0985
4	Percent of population 65 and over	17.8609	19.5337	18.8989	19.6111
5	Percent of population voting in the 1960 presidential election	3.3177	3.4423	3.6246	3.4705
6	Percent of families with incomes under $3,000	4.0330	3.9658	4.1313	4.1042
7	Percent of families with incomes over $10,000	-14.8255	-14.2762	-14.1886	-14.3720
8	Median school years completed	117.7836	115.4467	116.4134	115.0909
9	Percent of migrants from different county	6.9068	6.9110	6.8987	6.9011
10	Percent of population in labor force	41.8459	40.7987	41.1052	40.7381
11	Percent of employment in white collar occupations	4.5392	4.4304	4.4478	4.5351
12	Percent of employment working outside county of residence	-0.6337	-0.6406	-0.6957	-0.6822
13	Average size of household	570.1316	568.4603	571.6893	573.8665
14	Per capita property taxes	-0.5466	-0.5703	-0.5563	-0.5876
15	Number of towns and municipalities	-0.2171	-1.4460	-1.1405	-1.2918
16	Non-worker to worker ratio	417.3268	403.5153	401.5978	395.9311
17	Value of mineral industries shipments and receipts	0.0034	0.0032	0.0034	0.0035
18	Percent of farms with sales under $2,500	3.6070	3.6020	3.5531	3.5950
19	Percent of farms with sales $10,000 and over	-0.6217	-0.5168	-0.8448	-0.8447
20	Average per farm value of farm products sold	0.0121	0.0116	0.0123	0.0121
21	Farm operator level of living index	1.3026	1.3967	1.4237	1.4818
22	Local government general expenditures on natural resources	-0.2534	-0.2413	-0.2418	-0.2352
23	Percent of land in capability Classes I-IV	0.7156	0.6555	0.6218	0.6438
24	Location of TVA steam generator plants (dummy)	14.5173	13.6474	12.7203	15.3088
25	Location of TVA dams (dummy)	27.1768	27.1481	27.3521	27.8332
Constant		-2778.4655	-2713.7009	-2748.3676	-2733.5870

and percent of families with incomes over $10,000 (7) were less important as discriminators.

The classification matrix of the original 150 counties provided a visual test of the discriminant functions (Table V-3). This matrix was constructed from the evaluation of the discriminant functions for each group. In the derivation of the classification matrix each county within a group was assigned four probabilities by the program, one for each classification function. The largest of the four probabilities determined within which group the individual county was classified. The results of this classification are presented in Tables V-4, V-5, V-6, and V-7.

In the classification matrix (Table V-3) a perfect discrimination would occur when there was agreement of group and function. In this case all the nondiagonal elements of the matrix would be zero. In this analysis, fewer misclassifications occurred in the extreme groups, suggesting that the fast growth and no growth groups were more easily determined. The fast growth group had 12 out of 13 classifications in agreement while the no growth group had 69 out of 91 classifications in agreement.

V. SUMMARY

The discriminant analysis technique provided a method for testing the significance of the four groups into which the 150 counties were divided by the principal component analysis of Chapter IV. The four groups were compared on the basis of 25 selected socio-economic

TABLE V-3

CLASSIFICATION MATRIX FOR THE 150 COUNTY TENNESSEE VALLEY REGION

Group	Function 1	2	3	4	Total[a]
I	(12)	1	0	0	13
II	1	(10)	4	6	21
III	1	1	(18)	5	25
IV	3	9	10	(69)	91
Grand Total					150

[a]The probabilities for each of the 150 counties in the Tennessee Valley region are given in Tables V-4, V-5, V-6, and V-7, pages 90 through 95.

TABLE V-4

EVALUATION OF THE DISCRIMINATING FUNCTION FOR GROUP I FOR EACH COUNTY IN THE 150 COUNTY TENNESSEE VALLEY REGION

County Number	Value of the Discriminant Function 1	2	3	4	Largest Probability	Function Number for Largest Probability
1	0.89853	0.06474	0.00718	0.02955	0.89853	1
2	0.99998	0.00001	0.00001	0.00000	0.99998	1
3	0.97555	0.01940	0.00084	0.00421	0.97555	1
4	0.99932	0.00010	0.00020	0.00038	0.99932	1
5	0.99723	0.00051	0.00116	0.00111	0.99723	1
6	0.35169	0.42528	0.11236	0.11067	0.42528	2
7	0.99117	0.00041	0.00009	0.00833	0.99117	1
8	0.99532	0.00215	0.00103	0.00150	0.99532	1
9	0.98854	0.00453	0.00020	0.00673	0.98854	1
10	0.99228	0.00097	0.00014	0.00660	0.99228	1
11	0.91380	0.01867	0.04147	0.02607	0.91380	1
12	0.97022	0.00855	0.01483	0.00640	0.97022	1
13	0.87784	0.06035	0.00433	0.57480	0.97784	1

TABLE V-5

EVALUATION OF THE DISCRIMINATING FUNCTION FOR GROUP II FOR EACH COUNTY IN THE 150 COUNTY TENNESSEE VALLEY REGION

County Number	Value of the Discriminant Function 1	2	3	4	Largest Probability	Function Number for Largest Probability
1	0.97480	0.00470	0.00394	0.01656	0.97480	1
2	0.00116	0.99754	0.0-044	0.00086	0.99754	2
3	0.26241	0.73748	0.00001	0.00009	0.73748	2
4	0.00486	0.60633	0.05570	0.33311	0.60633	2
5	0.01403	0.80780	0.14566	0.03251	0.80780	2
6	0.00031	0.43524	0.52588	0.03856	0.52588	3
7	0.00628	0.98841	0.00026	0.00505	0.98841	2
8	0.00788	0.30164	0.52844	0.16205	0.52844	3
9	0.00074	0.04647	0.94278	0.01001	0.94278	3
10	0.00010	0.16651	0.64014	0.19326	0.64014	3
11	0.00000	0.99984	0.00009	0.00007	0.99984	2
12	0.00032	0.98895	0.00474	0.00599	0.98895	2
13	0.00043	0.39575	0.04755	0.55627	0.55627	4
14	0.00010	0.23113	0.08999	0.67678	0.67878	4
15	0.00002	0.96060	0.01560	0.02379	0.96060	2
16	0.01997	0.95271	0.02502	0.00230	0.95271	2
17	0.00010	0.38909	0.06131	0.54951	0.54951	4
18	0.00103	0.43997	0.23156	0.32744	0.43997	2
19	0.00002	0.18505	0.32363	0.49127	0.49127	4
20	0.02237	0.07314	0.04999	0.85450	0.85450	4
21	0.03852	0.35024	0.07497	0.53627	0.63627	4

TABLE V-6

EVALUATION OF THE DISCRIMINATING FUNCTION FOR GROUP III FOR EACH COUNTY IN THE 150 COUNTY TENNESSEE VALLEY REGION

County Number	Value of the Discriminant Function 1	2	3	4	Largest Probability	Function Number for Largest Probability
1	0.15132	0.21785	0.04943	0.58140	0.58140	4
2	0.57314	0.06871	0.05752	0.30063	0.57314	1
3	0.00053	0.00168	0.97265	0.02514	0.97265	3
4	0.02343	0.02522	0.94478	0.00658	0.94478	3
5	0.00003	0.08536	0.87499	0.03962	0.87499	3
6	0.00044	0.22202	0.76701	0.01052	0.76701	3
7	0.00005	0.00981	0.93974	0.05040	0.93974	3
8	0.00008	0.01353	0.97878	0.00761	0.97878	3
9	0.00005	0.34976	0.59812	0.05208	0.59812	3
10	0.00002	0.00492	0.97382	0.02125	0.97382	3
11	0.00000	0.03012	0.91345	0.05642	0.91345	3
12	0.00174	0.01317	0.96108	0.02402	0.96108	3
13	0.00021	0.02785	0.88342	0.08852	0.88342	3
14	0.00000	0.00934	0.98431	0.00635	0.98341	3
15	0.00003	0.00451	0.99290	0.00256	0.99290	3
16	0.00297	0.00588	0.97868	0.01247	0.97868	3
17	0.00010	0.06696	0.88517	0.04777	0.88517	3
18	0.02932	0.00291	0.96495	0.00283	0.96495	3
19	0.00168	0.12296	0.21831	0.65705	0.65705	4
20	0.00005	0.49763	0.04347	0.45885	0.49763	2
21	0.00127	0.06283	0.24349	0.69240	0.69240	4
22	0.00481	0.18345	0.58502	0.22583	0.58502	3
23	0.00598	0.27712	0.11526	0.60164	0.60164	4
24	0.00185	0.04696	0.30858	0.64261	0.64261	4
25	0.09145	0.03073	0.66127	0.21656	0.66127	3

TABLE V-7

EVALUATION OF THE DISCRIMINATING FUNCTION FOR GROUP IV FOR EACH COUNTY IN THE 150 COUNTY TENNESSEE VALLEY REGION

County Number	Value of the Discriminant Function 1	2	3	4	Largest Probability	Function Number for Largest Probability
1	0.07676	0.06898	0.13390	0.72035	0.72035	4
2	0.00015	0.06343	0.53662	0.39979	0.53662	3
3	0.00014	0.33748	0.03137	0.63100	0.63100	4
4	0.00160	0.04899	0.06952	0.87989	0.87989	4
5	0.01399	0.01372	0.19845	0.77384	0.77384	4
6	0.00148	0.00881	0.14691	0.84280	0.84280	4
7	0.00042	0.05407	0.44659	0.49892	0.49892	2
8	0.13039	0.34942	0.18467	0.33552	0.34942	4
9	0.30088	0.15918	0.00827	0.53168	0.53168	4
10	0.00135	0.11790	0.20915	0.67159	0.67159	4
11	0.00204	0.00552	0.02372	0.96871	0.96871	4
12	0.00010	0.08157	0.28262	0.63571	0.63571	4
13	0.00705	0.01815	0.16547	0.80933	0.80933	4
14	0.05374	0.20281	0.07065	0.67280	0.67280	4
15	0.00515	0.91268	0.00677	0.07540	0.91268	2
16	0.01903	0.00043	0.00021	0.98033	0.98033	4
17	0.00221	0.16838	0.24974	0.57966	0.57966	4
18	0.02698	0.06087	0.09295	0.81920	0.81920	4
19	0.17751	0.03403	0.01184	0.77662	0.77662	4
20	0.00022	0.02965	0.05591	0.91422	0.91422	4
21	0.00053	0.03450	0.02714	0.93782	0.93782	4
22	0.06059	0101671	0.18043	0.74228	0.74228	4
23	0.01127	0.21110	0.04449	0.73314	0.73314	4
24	0.00804	0.12343	0.28492	0.58360	0.58360	4
25	0.10608	0.07853	0.65435	0.16105	0.65435	3
26	0.01095	0.04937	0.14159	0.79809	0.79809	4
27	0.96700	0.00485	0.00052	0.02764	0.96700	1
28	0.02841	0.04705	0.45954	0.46500	0.46500	4
29	0.14064	0.02387	0.02491	0.81057	0.81057	4
30	0.00201	0.00085	0.00070	0.99645	0.99645	4
31	0.00022	0.02903	0.11018	0.86058	0.86058	4
32	0.00009	0.12875	0.06340	0.80777	0.80777	4
33	0.00008	0.19904	0.08146	0.71943	0.71943	4
34	0.00055	0.64248	0.07226	0.28470	0.53248	2
35	0.00094	0.27903	0.03520	0.68482	0.68482	4
36	0.00010	0.03616	0.05615	0.90759	0.90759	4
37	0.24228	0.05630	0.25555	0.66388	0.66388	4
38	0.00010	0.08297	0.55303	0.36390	0.55303	3
39	0.00012	0.04305	0.31731	0.63952	0.63952	4

TABLE V-7 (continued)

County Number	Value of the Discriminant Function 1	2	3	4	Largest Probability	Function Number for Largest Probability
40	0.00295	0.00593	0.03553	0.95559	0.95559	4
41	0.03750	0.01985	0.01293	0.92972	0.92972	4
42	0.00247	0.10581	0.22300	0.66872	0.66872	4
43	0.00019	0.12774	0.79057	0.08149	0.79067	3
44	0.01318	0.09781	0.13971	0.74931	0.74931	4
45	0.00359	0.29440	0.09736	0.60465	0.60465	4
46	0.03932	0.22322	0.15695	0.58051	0.58051	4
47	0.05592	0.11285	0.15593	0.67529	0.67529	4
48	0.00018	0.01105	0.05241	0.93636	0.93636	4
49	0.01437	0.08435	0.08170	0.81958	0.81958	4
50	0.00065	0.02421	0.07188	0.90326	0.90326	4
51	0.00035	0.19662	0.07931	0.72371	0.72371	4
52	0.00019	0.00144	0.00434	0.99403	0.99403	4
53	0.00013	0.00830	0.07734	0.91423	0.94123	4
54	0.00015	0.02860	0.13520	0.83604	0.83604	4
55	0.00000	0.03092	0.00215	0.96693	0.96693	4
56	0.03506	0.20072	0.51083	0.25338	0.51083	3
57	0.02774	0.66082	0.18293	0.12851	0.66082	2
58	0.08109	0.53466	0.03579	0.34846	0.53466	2
59	0.00007	0.07184	0.06378	0.86431	0.86431	4
60	0.00063	0.09434	0.72368	0.18135	0.72368	3
61	0.00004	0.01272	0.42340	0.56384	0.56384	4
62	0.00095	0.85511	0.00690	0.13703	0.85511	2
63	0.05781	0.14059	0.20629	0.59531	0.59531	4
64	0.00005	0.00583	0.00969	0.98443	0.98443	4
65	0.00364	0.02191	0.05820	0.91625	0.91625	4
66	0.00266	0.26665	0.00449	0.72620	0.72620	4
67	0.00047	0.19896	0.03065	0.76993	0.76993	4
68	0.00195	0.65745	0.20624	0.13436	0.65745	2
69	0.00140	0.66928	0.14747	0.18185	0.66928	2
70	0.00189	0.01889	0.62730	0.35192	0.62730	3
71	0.95182	0.00145	0.00801	0.03872	0.95182	1
72	0.00007	0.02717	0.09931	0.87345	0.87345	4
73	0.00352	0.59926	0.10421	0.29302	0.59926	2
74	0.00450	0.03182	0.73417	0.22951	0.73417	3
74	0.00329	0.05391	0.00379	0.93901	0.93901	4
76	0.03094	0.00074	0.01723	0.95109	0.95109	4
77	0.00007	0.01141	0.05858	0.92994	0.92994	4
78	0.00007	0.12284	0.52951	0.34759	0.52951	3
79	0.00138	0.08605	0.10013	0.81243	0.81243	4
80	0.00062	0.00768	0.14220	0.84951	0.84951	4

TABLE V-7 (continued)

County Number	Value of the Discriminant Function 1	2	3	4	Largest Probability	Function Number for Largest Probability
81	0.19789	0.04092	0.06936	0.69183	0.69183	4
82	0.00124	0.12819	0.00808	0.86249	0.86249	4
83	0.00545	0.07308	0.56265	0.35881	0.56265	3
84	0.00008	0.10521	0.01340	0.88131	0.88131	4
85	0.00006	0.05675	0.02413	0.91906	0.90906	4
86	0.00013	0.15364	0.02529	0.82094	0.82094	4
87	0.00032	0.26351	0.04202	0.69414	0.69414	4
88	0.00024	0.14548	0.18195	0.72233	0.72233	4
89	0.00270	0.09620	0.01299	0.88811	0.88811	4
90	0.34472	0.25028	0.06550	0.33950	0.34472	1
91	0.01083	0.15684	0.00607	0.82627	0.82627	4

characteristics of economic growth observed as of 1960. The results of the discriminant analysis indicated that the four groups were significantly different from each other. The analysis also indicated that the most important variables in discriminating among the four groups were (1) median school years completed, (2) average size of household, (3) non-worker to worker ratio, (4) location of TVA steam generator plants, and (5) location of TVA dams. Thus, the discriminant analysis confirmed the results obtained from the principal component analysis of Chapter IV. In addition to the above results, the analysis also provided classification function coefficients which can be used to predict the behavior of other counties for which data on the same variables are available. Thus, the discriminant analysis indicated (1) that it was possible to discriminate between the four groups of counties representing different rates of economic growth and (2) that the performance of other counties could be predicted with the classification functions provided by the discriminant analysis.

CHAPTER VI

FACTOR ANALYSIS OF ECONOMIC GROWTH IN THE TENNESSEE VALLEY REGION

I. INTRODUCTION

The multiplicity of variables which have an impact on the economic growth of a region's counties creates a vast complexity of relationships that requires simplification if the essential factors of economic growth are to be understood. Otherwise, the lack of precision of conceptualization and measurement reduces the researcher's effectiveness.

In the study of regional economic growth, numerous hypotheses may be suggested. The problem is to select the most meaningful hypotheses. In this study, the technique of factor analysis provided an objective method for formulating hypotheses and in selecting dependent variables. As Cattell [11, p. 362] has stated:

> Factor analysis is superior to most methods of exploration in comprehensively revealing such nexuses of interaction. It makes no assumptions about the direction of causal action, or about what is endogenous or exogenous to the system. If certain variables are in fact independent and outside the system, this will be shown by zero loadings in the factors that comprise the system. If nature does not know about the experimenter's favorite hypothesis which assigns pivotal importance to a set of supposed independent variables in a regression equation, the factor analysis will quickly show the fallacy of the supposed regression equation. If the experimenter has set up a criterion which he believes is influenced by such and such factors, the inclusion of the criterion in the factor analysis will quickly show whether in fact these factors need to be included in the regression equation for that criterion, and so on.

Keeping in mind the above comments, factor analysis was used to select common factors which accounted for the interrelationships between variables. The selected factors then became the dependent variables of the system. Inferences of causality and of relationships among the many variables were then made possible by the identification of these functional unities.

II. DATA REQUIREMENTS

Since the factor analysis technique does not require the selection of a dependent variable, many independent variables were selected from the socio-economic variables which were thought to affect regional economic growth. This analysis was accomplished in two parts. First, the same 25 independent variables used in the discriminant analysis were selected to analyze the economic growth of the 150 county Tennessee Valley region. Secondly, this set of 25 independent variables was increased to 38 variables by the inclusion of the 13 measure variables utilized in the principal component analysis of Chapter IV and were utilized in a general factor analysis program. The details of the computational procedure are presented in the following section.

III. COMPUTATIONAL PROCEDURE

The computational procedure of this program [18, pp. 169-184] can be outlined as follows:

Given the observations on the variables in the form X_{ij} where $i = 1, \ldots n$ and $j = 1, \ldots p$ variables, the means,

$$\bar{X}_j = \sum_{i=1}^{n} X_{ij}/n \tag{1}$$

the standard deviations,

$$sj = \sqrt{\frac{\sum_{i=1}^{n} (X_{ij}-\bar{X}_j)^2}{n-1}} \tag{2}$$

and the correlation coefficients,

i-1,2, . . . p

j=1,2, . . . p

$$\Gamma ij = \frac{\sum_{\alpha=1}^{n} (X_{\alpha i}-\bar{X}_i) \quad \sum_{\alpha=1}^{n} (X\alpha_j=\bar{X}_j)}{\sqrt{\sum_{\alpha=1}^{n} (X\alpha_i-\bar{X}_i)^2 \quad \sum_{=1}^{n} (X\alpha_j-\bar{X}_i)^2}} \tag{3}$$

were calculated in order to provide a correlation matrix indicating the degree of relationship between variables [18, pp. 174-175]. From this correlation matrix the eigenvalues and eigenvectors were derived and then utilized to obtain the coefficients (aij) of each factor where

$$a_{ij} = \sqrt{\lambda_j} \quad V_{ij} \tag{4}$$

and

λ_j = jth eigenvalue i = 1, . . . p variables

V_{ij} =eigenvector j = 1, . . . q factors

with the number of eigenvalues which are greater than unity determining the number of factors to be rotated [14, p. 130]. Once this determination has been made the factors are rotated in order to maximize the

difference between, and minimize the difference within, factors which explain the maximum variance of the individual variables. The communalities (h_j^2)

$$h_i^2 = \sum_j a_{ij}^2$$

were computed as measures of the proportion of variance of the variable explained by the extracted factors and provided a guide in the interpretation of the contribution made by each variable in each extracted factor. The computational procedure was ended with the output of the rotated factor matrix. Since all variables were included in each extracted factor, each factor was reduced to its simplest form by referral to statistical tables which gave the minimum loading that would provide an acceptable level of significance.

IV. FACTOR ANALYSIS OF THE 25 INDEPENDENT VARIABLES

The general factor analysis program extracted seven orthogonal factors based on the criterion that all factors with eigenvalues greater than one be included as final factors. The resulting seven factors and the communalities (h^2), which are the summation of the squared row factor loadings are presented in Table VI-1. The communalities are a measure of the proportion of variance of the variable explained by the extracted factors. As Table VI-2 indicates, a very substantial proportion of the variance was explained by the seven factors. The explained variance varied from a high of 0.9069 for variable number 6, percent of

TABLE VI-1

THE ROTATED ORTHOGONAL FACTOR LOADINGS MATRIX FOR THE 25 INDEPENDENT SOCIO-ECONOMIC VARIABLES USED IN THE ANALYSIS OF THE 150 COUNTY TENNESSEE VALLEY REGION

Variable Number	Description	Factors							h^{2**}
		1	2	3	4	5	6	7	
1	Percent urban population	-0.4596*	-0.1831	-0.0157	-0.4691*	0.1728	0.4746*	0.3391*	0.8352
2	Percent rural farm population	0.8211*	-0.1701	0.1404	0.2456*	-0.1052	0.0024	-0.2187	0.8421
3	Percent non-white population	0.1852	-0.6976*	-0.0438	-0.2296	0.0369	0.1424	0.2162	0.6440
4	Percent of population 65 and over	0.3084*	-0.0326	0.8679*	0.1738	0.0212	0.0136	-0.1148	0.8934
5	Percent of population voting in the 1960 Presidential election	-0.0376	0.6171*	0.3747*	0.0607	0.3002*	-0.3545*	-0.0313	0.7430
6	Percent of families with incomes under $3,000	0.7431*	-0.0899	0.1003	0.3198*	0.0251	-0.4480*	-0.1815	0.9069
7	Percent of families with incomes over $10,000	-0.5420*	-0.0976	-0.1723	-0.4793*	0.2090	0.3405*	0.3240*	0.8273
8	Median school years completed	-0.5336*	0.0473	0.0989	-0.4354*	0.1235	0.4641*	0.3122*	0.8143
9	Percent of migrants from different county	-0.6421*	-0.0802	-0.0260	-0.1083	-0.2017	0.2040	-0.2048	0.5554
10	Percent of population in labor force	-0.1160	-0.0574	0.2904*	-0.0951	0.1219	0.8741*	-0.0686	0.8937
11	Percent of employment in white collar occupations	-0.6648*	0.0200	0.0083	-0.4593*	0.2289	0.1273	0.3623*	0.8532
12	Percent of employment working outside county of residence	-0.1825	0.1970	-0.2633*	0.2231	-0.6771*	-0.0575	-0.211-	0.6936
13	Average size of household	0.2184	-0.0649	-0.8061	0.1034	-0.1322	-0.4095*	-0.0062	0.8976
14	Per capita property taxes	-0.2449*	0.0265	-0.0202	-0.1151	0.7031*	0.1548	0.0130	0.5928
15	Number of towns and municipalities	-0.0510	-0.3813*	0.2563*	-0.5021*	-0.3060*	0.0041	0.3107*	0.6500
16	Non-worker to worker ratio	0.2250	0.1390	0.0077	0.0892	-0.0879	-0.9025*	0.0378	0.9016
17	Value of mineral industries shipments and receipts	-0.1322	0.0522	-0.0772	0.0507	0.0691	-0.0563	0.7552*	0.6070
18	Percent of farms with sales under $2,500	0.8189*	0.0995	-0.0384	-0.1085	-0.0946	-0.0654	0.0160	0.7072
19	Percent of farms with sales $10,000 and over	-0.4329*	-0.6176*	-0.1243	0.1153	0.2559*	-0.0636	-0.1937	0.7046
20	Average per farm value of farm products sold	0.0082	-0.8054*	0.0286	0.0101	0.1742	-0.0919	-0.1266	0.7045
21	Farm operator level of living index	-0.5639*	-0.3434*	0.1744	-0.0105	0.0273	0.5108*	0.0992	0.7380
22	Local government general expenditures on natural resources	-0.0781	-0.0411	-0.0498	-0.6645*	0.2104	0.0892	-0.0062	0.5041
23	Percent of land in capability Classes I-IV	-0.2028	-0.7015*	0.2926*	-0.0583	-0.1706	0.3752*	-0.0095	0.7922
24	Location of TVA steam generator plants (dummy)	-0.1007	-0.1531	-0.0447	-0.5508*	0.2796*	0.2339*	0.3407*	0.5879
25	Location of TVA dams (dummy)	-0.1464	0.0238	-0.0296	-0.6983*	-0.0645	-0.0180	-0.2943*	0.6016

*Significant loadings of 0.230 and above at $\alpha = .01$ and 123 degrees of freedom.

**Communalities.

TABLE VI-2

DISTRIBUTION OF COMMUNALITY COEFFICIENTS (h^2) FOR THE FACTOR ANALYSIS OF THE 25 INDEPENDENT SOCIO-ECONOMIC VARIABLES USED IN THE ANALYSIS OF THE 150 COUNTY TENNESSEE VALLEY REGION

Percent of Variance Explained	Number of Variables	Percent of Variables
90-99	2	8
80-90	8	32
70-79	6	24
60-69	5	20
50-59	4	16
49 and below	0	0
Total	25	100

families with incomes under $3,000 in 1960, to a low of 0.5041 for variable number 22, local government general expenditures on natural resources. The data on the number and percent of variables with the different percentages of variance explained are presented in Table VI-2.

Identification of Factors

The identification of factors was accomplished by inspecting the variables with large loadings on a given factor and discovering what they had in common which is not shared by variables not having large loadings on that factor. The naming of factors is therefore highly subjective but as Cattell [11, p. 338] has stated:

> The definition of a factor as an empirical construct follows remarkably closely the procedures stated by Bacon and refined by Mill for arriving at the essential nature of anything. One observes where it is conspicuously present as a positive influence, where it is conspicuously present as a negative influence, and where it is generally absent . , . . These presences and absences suffice to give a reasonably clear picture--an empirical construct--of the dimension with which we are concerned.

Factor loadings were included in each of the seven factors, 1, 2, 3, 4, 5, 6, and 7 for all 25 independent variables. However, only factor loadings of an acceptable level of significance were included in the final factors. The included factor loadings are indicated by the asterisk in Table VI-1. The final or basic factors are analyzed and named in the following sections.

Factor 1. The variables loaded significantly on Factor 1 are shown in Table VI-3. The positive loadings on variables 2, 18, 6, and 4 when associated with the negative loadings on variables 11, 9, 21, 7, 8, 1, 19, and 14 presented a logical and consistent functional unity identified as the Extent of Poverty. These factor loadings may be interpreted in the following manner. As variable 2, percent rural farm population, increases by one unit, the Extent of Poverty increases by 0.8211 of one unit. Similarly, as variable 11, percent of employment in white collar occupations, increases by one unit, the Extent of Poverty decreases by 0.6648 of one unit. Extent of Poverty emerged as the most important factor because it had the largest eigenvalue.

Factor 2. The variables loaded significantly on Factor 2 are shown in Table VI-4. This factor had significant loadings on seven of the 25 variables. It is apparent that a large number of the variables are indicative of commercial agriculture. The high negative loadings on variables 20, 23, 3, and 19 coupled with smaller negative loadings on variables 15 and 21 suggested that this factor was measuring a functional unit named Sub-Marginal Agriculture.

Factor 3. The variables loaded significantly on Factor 3 are shown in Table VI-5. The relatively high positive loading on variable 4, the percent of population 65 and over along with smaller positive loadings

TABLE VI-3

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 1* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 25 INDEPENDENT VARIABLES

Variable Number	Description	Factor Loadings
2	Percent rural farm population	0.8211
18	Percent of farms with sales under $2,500	0.8189
6	Percent of families with incomes under $3,000	0.7431
11	Percent of employment in white collar occupations	-0.6648
9	Percent of migrants from different county	-0.6421
21	Farm operator level-of-living index	-0.5639
7	Percent of families with incomes over $10,000	-0.5420
8	Median school years completed	-0.5336
1	Percent urban population	-0.4596
19	Percent of farms with sales $10,000 and over	-0.4329
4	Percent of population 65 and over	0.3084
14	Per capita property taxes	-0.2449

*Factor 1 was interpreted as Extent of Poverty.

TABLE VI-4

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 2* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 25 INDEPENDENT VARIABLES

Variable Number	Description	Factor Loadings
20	Average per farm value of farm products sold	-0.8054
23	Percent of land in capability Classes I-IV	-0.7015
3	Percent non-white population	-0.6976
19	Percent of farms with sales $10,000 and over	-0.6176
5	Percent of population voting in the 1960 Presidential election	0.6171
15	Number of towns and municipalities	-0.3813
21	Farm operator level-of-living index	-0.3434

*Factor 2 was interpreted as Sub-Marginal Agriculture.

TABLE VI-5

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 3* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 25 INDEPENDENT VARIABLES

Variable Number	Description	Factor Loadings
4	Percent of population 65 and over	0.8679
5	Percent of population voting in the 1960 Presidential election	0.3747
23	Percent of land in capability Classes I-IV	0.2926
10	Percent of population in labor force	0.2904
12	Percent of employment working outside county of residence	-0.2633
15	Number of towns and municipalities	0.2563

*Factor 3 was interpreted as Labor Force Maturity.

on variables 5, 23, 10, and 15 suggested that this factor was measuring the maturity of the labor force. Therefore, this factor was named Labor Force Maturity.

Factor 4. Factor 4 had significant loadings on 10 of the 25 variables as shown in Table VI-6. The relatively high negative loadings on variables 25, 22, 24, and 15 along with the lower negative loadings on variables 7, 1, 11, and 8 indicated that Factor 4 was measuring some relationship between investments in natural resources and economic development. The positive loadings on percent of families with incomes under $3,000 and percent rural farm population suggested that rural poverty occurred where there had been an absence of natural resource investments of any great value. As a result of these relationships Factor 4 was named Inadequate Natural Resource Investments.

Factor 5. The variables loaded significantly on Factor 5 are shown in Table VI-7. This factor had significant loadings on 6 of the 25 variables. The relatively high loading on variable 14, per capita property taxes along with smaller positive loadings on variables 5, 24, and 19 combined with the negative loadings on variables 12 and 15 suggested that this factor was measuring the relationship between urbanization and work opportunities. The high positive loading on per capita property taxes along with the high negative loading on the percent of employment working outside the county of residence indicated that work opportunities were greatest in metropolitan areas. As a result of this relationship, Factor 5 was identified as Extent of Metropolitan Work Opportunities.

TABLE VI-6

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 4* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 25 INDEPENDENT VARIABLES

Variable Number	Description	Factor Loadings
25	Location of TVA dams (dummy)	-0.6983
22	Local government expenditures on natural resources	-0.6645
24	Location of TVA steam generator plants (dummy)	-0.5508
15	Number of towns and municipalities	-0.5021
7	Percent of families with incomes over $10,000	-0.4793
1	Percent urban population	-0.4691
11	Percent of employment in white collar occupations	-0.4593
8	Median school years completed	-0.4354
6	Percent of families with incomes under $3,000	0.3198
2	Percent rural farm population	0.2456

*Factor 4 was interpreted as Inadequate Natural Resource Investments.

TABLE VI-7

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 5* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 25 INDEPENDENT VARIABLES

Variable Number	Description	Factor Loadings
14	Per capita property taxes	0.7031
12	Percent of employment working outside county of residence	-0.6771
15	Number of towns and municipalities	-0.3060
5	Percent of population voting in 1960 Presidential election	0.3002
24	Location of TVA steam generator plants (dummy)	0.2796
19	Percent of farms with sales $10,000 and over	0.2559

*Factor 5 was interpreted as Extent of Metropolitan Work Opportunities.

Factor 6. The variables significantly loaded on Factor 6 are shown in Table VI-8. This factor had significant loadings on 11 of the 25 variables. The variables loaded significantly show a considerable degree of homogeneity of content since all are related to the productive capacity of the labor force. The positive loadings on variables 10, 21, 1, 8, 23, 7, and 24 along with the negative loadings on variables 16, 6, 13, and 5 indicated that Factor 5 was measuring the productivity of the labor force. The relatively high negative loading on variable 16, non-worker to worker ratio, along with the relatively high positive loading on percent of the population in the labor force resulted in the identification of this factor as Labor Force Productivity.

Factor 7. As Table VI-9 indicates Factor 7 had significant loadings on eight of the 25 variables. The relatively high positive loading on variable 17, value of mineral industries shipments and receipts, along with the positive loadings on the remaining variables except number 25, location of TVA dams, indicated that Factor 7 was measuring the economic contributions of mineral industry locations in each county. Accordingly, Factor 7 was identified as Extractive.

Discussion of Results

The above factors are independent as possible, given the power of the statistical technique. Nevertheless, there was some overlapping due to the interdependent nature of the socio-economic forces at work in the

TABLE VI-8

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 6* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 25 INDEPENDENT VARIABLES

Variable Number	Description	Factor Loadings
16	Non-worker to worker ratio	-0.9025
10	Percent of population in labor force	0.8741
21	Farm operator level-of-living index	0.5108
1	Percent urban population	0.4746
8	Median school years completed	0.4641
6	Percent of families with incomes under $3,000	-0.4480
13	Average size of household	-0.4095
23	Percent of land in capability Classes I-IV	0.3752
5	Percent of population voting in the 1960 Presidential election	-0.3545
7	Percent of families with incomes over $10,000	0.3405
24	Location of TVA steam generator plants (dummy)	0.2339

*Factor 6 was interpreted as Labor Force Productivity.

TABLE VI-9

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 7* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 25 INDEPENDENT VARIABLES

Variable Number	Description	Factor Loadings
17	Value of minerals industries shipments and receipts	0.7552
11	Percent of employment in white collar occupations	0.3623
24	Location of TVA steam generator plants (dummy)	0.3407
1	Percent urban population	0.3391
7	Percent of families with incomes over $10,000	0.3240
8	Median school years completed	0.3122
15	Number of towns and municipalities	0.3107
25	Location of TVA dams (dummy)	-0.2943

*Factor 7 was interpreted as Extractive.

region. For example, of the 25 variables listed in Table VI-1, page 101, five appeared as significant variables in four different factors and five others appeared as significant variables in three different factors. The five variables that appeared as significant in four factors were percent urban population, percent of population voting in the 1960 Presidential election, percent of families with incomes over $10,000, median school years completed, and location of TVA steam generator plants. The five variables that appeared as significant in three factors were percent of families with incomes under $3,000, percent of employment in white collar occupations, percent of farms with sales $10,000 and over, farm operator level-of-living index, and percent of land in capability classes I-IV. One variable, number of towns and municipalities, appeared as significant in five different factors. The analysis indicated that along with economic variables, social and political variables were also important in investigating the economic growth of the Tennessee Valley region. This might indicate a need for more research emphasis upon social and political forces when investigating economic growth.

Relationship of the Seven Factors to the Index of Economic Growth

The present section attempts to analyze the statistical relationship between the factors identified in the preceding section and the index of economic growth that was derived in Chapter IV by principal component analysis. The general hypotheses tested was that some portion of the variation in the index of economic growth would be explained by variations in the seven derived factors.

The analysis of the relationship between factors and the index of economic growth was completed by utilizing the following procedure: (1) factor scores were computed[1] to indicate the degree to which the 150 individual counties possessed each of the seven factors; and (2) the computed factor scores were then treated as independent variables in a linear multiple regression, using the index of economic growth derived by principal component analysis as the dependent variable. For purposes of comparison, multiple regression analysis of the 25 independent variables and of the seven factors, was run on the dependent variable represented by the derived index of economic growth.

The Regression Analysis.[2] The regression of the index of economic growth on the seven identified factors resulted in the following regression equation, where the variables are arranged in the order in which they entered into the equation:[3]

$$Y = 0.0000053 - 0.020056X_1^* + 0.01843X_2^* + 0.016765X_6^* - 0.017422X_3^* - 0.0075402X_4 - 0.0088316X_7 + 0.0013345X_5$$

---

[1]Computer programs are currently available for efficiently performing this computation. This study utilized the BMDO 3 M program in the BMD Biomedical Computer Program series outlined in [18, pp. 169-184].

[2]The model used in this study to evaluate the relationship between the hypothetical factors and the index of economic growth was the Stepwise Regression model as outlined in [18, pp. 233-247].

[3]The asterisk (*) indicates statistical significance at the .01 level. The variable X_4, Inadequate Natural Resource Investments, was significant at the 0.15 level; X_5, Extent of Metropolitan Work Opportunities, was significant at the 0.85 level; and X_7, Extractive, was significant at the 0.20 level.

where

Y = the index of economic growth

X_1 = Extent of Poverty

X_2 = Sub-Marginal Agriculture

X_3 = Labor Force Maturity

X_4 = Inadequate Natural Resource Investments

X_5 = Extent of Metropolitan Work Opportunities

X_6 = Labor Force Productivity

X_7 = Extractive.

The R^2 for the above equation was .7633 indicating that just over 76 percent of the variation in the index of economic growth was explained by the equation. This R^2 value was significant at the .01 level. For comparison the index of economic growth was regressed on the same 25 independent variables utilized in the discriminant analysis. For this regression equation the R^2 was .8242. However, only one of the independent variables had a significant student t-value. Thus, while the regression of the index of economic growth on the seven factors reduced R^2 by .0609, a gain from 4 percent to 57.1 percent of the significant student t-values of the independent variables was achieved. This implies that the factor analysis technique permits a simplification of confounding effects, thus revealing the underlying relationships important to economic growth in the Tennessee Valley.

The coefficients in the above regression equation can be interpreted as follows. As the Extent of Poverty increased by one unit, the index of economic growth, Y, decreased by 0.020056 of one unit indicating

a decline in economic growth. Similarly, as Labor Force Maturity, Inadequate Natural Resource Investments, and Extractive increase by one unit, the index of economic growth decreases 0.017422, 0.0075402, and 0.0088316, respectively. On the other hand, as Sub-Marginal Agriculture, Labor Force Productivity, and the Extent of Metropolitan Work Opportunities increase by one unit, the index of economic growth increases by 0.018142, 0.016765, and 0.0013345 of one unit, an increase in the economic growth of the region.

All of the above signs for the coefficients were in agreement with a priori reasoning. The relatively small values of the coefficients reflect the complexity of the relationships involved in economic growth and probably indicate that several other factors outside the above system of 25 independent variables are involved. Therefore in the next section this system will be enlarged to 38 variables and the factor analysis and regression analysis will be repeated.

V. FACTOR ANALYSIS OF THE 38 SOCIO-ECONOMIC VARIABLES

The previous analysis indicated that several factors outside the system were involved in the explanation of economic growth in the Tennessee Valley region. Therefore in this section additional variables were added to the 25 independent variables in an attempt to improve the explanation of economic growth. More specifically, the system of variables was increased to 38 by the addition of the 13 measure variables utilized in the analysis conducted in Chapter IV. These 38 socio-economic indicators were then subjected to factor analysis and the

factors were utilized in a regression analysis of economic growth following the same procedures as outlined in the preceding sections of this chapter.

The general factor analysis program extracted 10 orthogonal factors based on the criterion that all factors with eigenvalues greater than one be included as final factors. These 10 factors and the communalities are presented in Table VI-10. As Table VI-11 indicates, a very substantial proportion of the variance as measured by the communalities was explained by the 10 factors. The explained variance varied from a high of 0.9235 for variable number 19, percent of families with incomes under $3,000, to a low of 0.3970 for variable number 30, value of minerals industries shipments and receipts (Table VI-10). The data on the number and percent of variables with the different percentages of variance are presented in Table VI-11.

Identification of Factors

The identification of factors in this section followed the same general procedure as outlined in the previous factor analysis. Factor loadings were included in each of the 10 factors, 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 for all 38 socio-economic variables. As in the previous analysis only factor loadings of an acceptable level of significance were included in the final factors. The included factor loadings are indicated by the asterisk in Table VI-10. The final or basic factors are analyzed and named in the following sections.

TABLE VI-10

THE ROTATED ORTHOGONAL FACTOR LOADINGS MATRIX FOR THE 38 SOCIO-ECONOMIC VARIABLES USED IN THE ANALYSIS OF THE 150 COUNTY TENNESSEE VALLEY REGION

Variable Number	Description	Factors										h^{2**}
		1	2	3	4	5	6	7	8	9	10	
1	Rate of population change	0.7135*	-0.2598*	-0.0285	0.0285	-0.0903	-0.2405*	0.0812	-0.3445*	-0.0312	0.0412	0.8214
2	Change in median income of all families	0.8231*	-0.0251	0.0974	0.0974	-0.0580	-0.2394	0.0619	-0.3148*	-0.1266	0.0270	0.8852
3	Change in median income of rural families	0.0526*	-0.0689	0.1249	0.1249	0.0954	-0.2014	0.0719	-0.2600*	-0.5538	0.0603	0.7273
4	Change in median income of rural farm families	0.5689*	-0.0460	0.0616	0.0616	-0.2031	-0.1540	0.1343	-0.1984	-0.4647*	0.1754	0.6993
5	Change in total property tax	0.0956	-0.0494	-0.0011	-0.0011	-0.2104	0.0279	0.0726	-0.1384	0.0297	-0.0965	0.7345
6	Change in median value of owner occupied housing units	0.7571*	-0.2540*	0.0147	0.0147	0.0501	-0.0409	-0.0283	0.1270	0.0574	-0.2186	0.7207
7	Change in the percent of dwellings structurally sound with all plumbing facilities	0.6072*	0.2186	-0.0151	-0.0151	0.3756*	-0.1452	0.2298	-0.0330	0.1889	0.2066	0.7237
8	Change in the percent employed in agriculture	0.1018	-0.1383	-0.1382	-0.1382	-0.7465*	-0.1571	0.1604	-0.2312	-0.0821	-0.0065	0.7764
9	Change in retail sales per capita	-0.0274	0.0569	-0.0254	-0.0254	0.3242*	0.1456	-0.1779	0.5975*	0.1464	0.3393*	0.6720
10	Change in value of farm land and buildings	0.1484	-0.0067	0.0225	0.0225	-0.0555	-0.9118*	0.0405	-0.2063	-0.0183	-0.1238	0.9170
11	Ratio of bank deposits to aggregate income	0.2404*	-0.0035	-0.0044	-0.0044	-0.1088	-0.0792	0.8842*	-0.1022	-0.0286	-0.0753	0.9007
12	Change in total payroll per capita	0.4948*	-0.0675	0.0083	0.0083	-0.1817	-0.5697*	0.0648	-0.1521	-0.0809	-0.1426	0.7500
13	Change in the value of manufacturing per capita	0.1685	0.0220	0.0252	0.0252	-0.0637	-0.9067*	-0.0012	-0.1235	0.0599	-0.0802	0.8811
14	Percent urban population	0.4694*	-0.0204	-0.0845	-0.0845	-0.3526*	-0.1721	-0.0473	-0.4723*	0.2311	-0.2612*	0.8519
15	Percent rural farm population	-0.6865*	0.0024	0.0759	0.0759	0.4680*	0.2504*	-0.1551	-0.0546	0.1190	0.0519	0.8245
16	Percent non-white population	0.0406	-0.0028	-0.0441	-0.0833	0.0454	0.0252	0.9153*	-0.0523	0.0530	0.1719	0.8862
17	Percent of population 65 and over	-0.3849*	-0.1269	0.8065*	0.0469	0.1699	0.0812	-0.0791	-0.0514	0.0909	-0.0371	0.8709
18	Percent of population voting in the 1960 Presidential election	-0.0609	-0.0069	0.4338*	0.3775*	-0.0937	0.0857	0.0304*	-0.3777*	-0.4633*	-0.2253	0.7595
19	Percent of families with incomes under $3,000	-0.7222*	-0.2100	0.0077	0.0088	0.3313	0.2397	-0.1363	0.4047*	0.0751	-0.0494	0.9235
20	Percent of families with incomes over $10,000	0.5933*	0.3605*	-0.1647	-0.0621	-0.3551*	-0.1005	0.1132	-0.3494*	0.0784	-0.2645*	0.8601
21	Median school years completed	0.6463*	0.3144*	0.1250	0.1113	-0.2529*	-0.0254	0.0379	-0.4198*	0.1378	-0.2462*	0.8652
22	Percent of migrants from different county	0.7046*	0.0568	0.0765	-0.1850	0.0970	0.0713	0.0118	-0.1220	0.0525	0.2239	0.6221
23	Percent of population in labor force	0.1378	0.1302	0.3308*	-0.0383	0.1318	-0.1522	0.0664	-0.8312*	0.0077	-0.0602	0.8864
24	Percent of employment in white collar occupations	0.6688*	0.3190*	0.0501	-0.0163	-0.4523*	-0.0468	0.0923	-0.1188	0.0960	-0.3083*	0.8829
25	Percent of employment working outside county of residence	0.2181	-0.1720	-0.1537	0.1084	0.1352	0.2125	0.1113	0.1286	-0.1758	0.7009*	0.7271
26	Average size of family	-0.1696	-0.0713	-0.0869*	-0.0454	0.0138	0.0466	-0.0006	0.4089*	-0.0976	0.1563	0.8904
27	Per capita property taxes	0.2937*	0.2259	0.1200	-0.1440	0.0721	-0.1251	-0.1111	-0.0388	-0.2210	-0.5698*	0.5806
28	Number of townships and municipalities	0.1478	0.2059	0.1292	-0.0805	-0.0859	-0.1483	0.0910	-0.0462	0.7557*	0.0214	0.6987
29	Non-worker to worker ratio	-0.2613*	-0.1253	-0.0516	0.1114	-0.0360	0.1572	-0.0395	0.8627*	-0.0062	-0.0191	0.8612
30	Value of minerals industries shipments and receipts	0.0765	0.0775	-0.0021	0.0832	-0.5968*	-0.0104	-0.0518	0.0837	0.1080	-0.0244	0.3970
31	Percent of farms with sales under $2,500	-0.5440*	0.0545	-0.2004	0.3996*	0.3092*	0.2123	-0.0066	0.0516	0.2867*	-0.0926	0.7328
32	Percent of farms with sales $10,000 and over	0.1858	0.0143	-0.0581	-0.8674*	0.0079	-0.0125	0.0708	-0.0365	-0.0186	-0.0953	0.8064
33	Average per farm value of farm products sold	-0.1862	0.0878	-0.0595	-0.7458*	0.0527	0.0894	0.0211	-0.0835	0.2357	-0.1042	0.6867
34	Farm operator level-of-living index	0.3547*	0.1089	0.2431*	-0.3134*	-0.2926*	-0.1181	0.0092	-0.5651*	-0.0810	0.1526	0.8165
35	Local government general expenditures or natural resources	-0.5137*	0.1405	0.0270	0.1085	-0.3572*	-0.0180	-0.1863	-0.2673*	-0.1761	0.3048*	0.6541
36	Percent of land in capability Classes I-IV	0.1201	0.6924*	0.0003	-0.0512	-0.0259	0.0120	0.0764	-0.0871	0.0163	-0.1332	0.5287
37	Location of TVA steam generating plants (dummy)	0.1327	0.7138*	-0.0184	-0.0374	-0.2252	0.0270	0.0588	-0.2412*	0.0774	-0.2047	0.6898
38	Location of TVA dams	0.1284	0.6542*	0.0271	-0.0244	0.0338	-0.3500*	-0.1029	0.0942	0.0493	0.2864*	0.6733

*Significant loadings of 0.241 or above at $\alpha = .01$ and 110 degrees of freedom.

**Communalities.

TABLE VI-11

DISTRIBUTION OF COMMUNALITY COEFFICIENTS (h^2) FOR THE FACTOR ANALYSIS OF THE 38 SOCIO-ECONOMIC VARIABLES USED IN THE ANALYSIS OF THE 150 COUNTY TENNESSEE VALLEY REGION

Percent of Variance Explained	Number of Variables	Percent of Variables
90-99	3	8
80-89	15	39
70-79	9	24
60-69	8	21
50-59	2	5
49 and below	1	3
Total	38	100

Factor 1. Factor 1 emerged as the most important factor because it had the largest eigenvalue. The 21 variables loaded significantly on Factor 1 are shown in Table VI-12. Positive loadings on variables 2, 6, 1, 22, 24, 21, 7, 20, 4, 3, 12, 14, 34, 27, and 11 when associated with negative loadings on variables 19, 15, 31, 35, 17, and 29 presented a logical and consistent functional unit that was named Level-of-Living.

Factor 2. Factor 2 had significant loadings on 10 of the 38 variables as shown in Table VI-13. The relatively high positive loadings on variables 5, 37, 36, and 38 coupled with the lower positive loadings on variables 20, 14, 21, 24, 12, and 8 indicated that Factor 2 was measuring some relationship between natural resource investments and economic growth. As a result Factor 2 was identified as Natural Resource Investments.

Factor 3. The variables loaded significantly on Factor 3 are shown in Table VI-14. This factor had significant loadings on seven of the 38 variables. The significantly loaded variables in this factor showed a considerable degree of homogeneity of content since many were indicators of the relative maturity of the population. The relatively high positive loading on variable 17, percent of population 65 and over, and the lower positive loadings on variables 18, 23, and 34 along with the relatively high negative loadings on variable 26, average size of household, and the lower negative loadings on variables 1 and 6 led to the identification of Factor 3 as Population Maturity.

TABLE VI-12

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 1* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
2	Change in median income of all families	0.8231
6	Change in median value of owner occupied housing units	0.7571
19	Percent of families with incomes under $3,000	-0.7222
1	Rate of population change	0.7135
22	Percent of migrants from different county	0.7046
15	Percent rural farm population	-0.6865
24	Percent of employment in white collar occupations	0.6668
21	Median school years completed	0.6463
7	Change in the percent of dwellings structurally sound with all plumbing facilities	0.6072
20	Percent of families with incomes over $10,000	0.5933
4	Change in median income of rural farm families	0.5689
31	Percent of farms with sales under $2,500	-0.5440
35	Local government general expenditures on natural resources	-0.5137
3	Change in income of rural farm families	0.5026
12	Change in total payroll per capita	0.4948
14	Percent urban population	0.4694
17	Percent population 65 and over	-0.3849
34	Farm operator level-of-living index	0.3547
27	Per capita property taxes	0.2937
29	Non-worker to worker ratio	-0.2413
11	Ratio of bank deposits to aggregate income	0.2404

*Factor 1 was interpreted as Level-of-Living.

TABLE VI-13

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 2* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
5	Change in total property tax	0.8021
37	Location of TVA steam generator plants (dummy)	0.7138
36	Percent of land in capability Classes I-IV	0.6924
38	Location of TVA dams (dummy)	0.6542
20	Percent of families with incomes over $10,000	0.3605
14	Percent urban population	0.3509
21	Median school years completed	0.3144
24	Percent of employment in white collar occupations	0.3190
12	Change in total payroll per capita	0.2978
8	Change in the percent employed in agriculture	0.2448

*Factor 2 was interpreted as Natural Resource Investments.

TABLE VI-14

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 3* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
26	Average size of household	-0.8069
17	Percent of population 65 and over	0.8065
18	Percent of population voting in 1960 Presidential election	0.4338
23	Percent of population in labor force	0.3308
1	Rate of population change	-0.2598
6	Change in median value of owner occupied housing units	-0.2540
34	Farm operator level-of-living index	0.2431

*Factor 3 was interpreted as Population Maturity.

Factor 4. Factor 4 had significant loadings on five of the 38 variables as indicated in Table VI-15. The very high negative loadings on percent of farms with sales $10,000 and over and average per farm value of farm products sold and the lower negative loading on variable 34 along with positive loadings on variables 31 and 18 suggested an agricultural sector characterized by low income farms. Accordingly, Factor 4 was identified as Sub-Marginal Agriculture.

Factor 5. Factor 5 had significant loadings on 13 of the 38 variables as shown in Table VI-16. Negative loadings on variables 8, 30, 24, 35, 20, 14, 34, and 21 along with positive loadings on variables 15, 7, 19, 9, and 31 indicated that Factor 5 was measuring some relationship between agricultural employment and the degree of geographic isolation. Accordingly, Factor 5 was identified as Agricultural Employment.

Factor 6. The variables loaded significantly on Factor 6 are shown in Table VI-17. The relatively high negative loadings on variables 10 and 13 and the lower negative loadings on variables 12 and 38 along with the positive loading on variable 15 indicated a strongly oriented rural economy. The positive loading of variable 15, the percent rural farm population, resulted in the identification of Factor 6 as Ruralism.

TABLE VI-15

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 4* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
32	Percent of farms with sales $10,000 and over	-0.8674
33	Average per farm value of farm products sold	-0.7458
34	Farm operator level-of-living index	-0.4134
31	Percent of farms with sales under $2,500	0.3996
18	Percent of population voting in the 1960 Presidential election	0.3775

*Factor 4 was interpreted as Sub-Marginal Agriculture.

TABLE VI-16

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 5* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
8	Change in percent employed in agriculture	-0.7465
30	Value of minerals industries shipments and receipts	-0.5968
15	Percent rural farm population	0.4680
24	Percent of population in labor force	-0.4523
7	Change in the percent of dwellings structurally sound with all plumbing facilities	0.3756
35	Percent of land in capability Classes I-IV	-0.3572
20	Percent of families with incomes over $10,000	-0.3551
14	Percent urban population	-0.3526
19	Percent of families with incomes under $3,000	0.3313
9	Change in retail sales per capita	0.3242
31	Percent of farms with sales under $2,500	0.3092
34	Farm operator level-of-living index	-0.2926
21	Median school years completed	-0.2529

*Factor 5 was interpreted as Agricultural Employment.

TABLE VI-17

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 6* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
10	Change in value of farm land and buildings	-0.9118
13	Change in the value of manufacturing per capita	-0.9067
12	Change in total payroll per capita	-0.5697
38	Location of TVA dams (dummy)	-0.3500
15	Percent rural farm population	0.2504

*Factor 6 was interpreted as Ruralism.

Factor 7. Factor 7 was difficult to identify because as Table VI-18 indicates, only two of the 38 variables had significant loadings. In order to identify this factor those variables in Table VI-10, page 119, which were not loaded significantly were also considered. The relatively high positive loadings on variable 16, percent non-white population and variable 11, ratio of bank deposits to aggregate income, indicated a plantation-type social structure characterized by a highly dependent population of low wage earners. As a result of these relationships, Factor 7 was identified as Extensiveness.

Factor 8. Factor 8 had significant loadings on 15 of the 38 socio-economic variables as shown in Table VI-19. The relatively high positive loading on variable 29, non-worker to worker ratio, along with the positive loading on variables 9, 26, 19, and 18 combined with the relatively high negative loading on variable 23, percent of population in labor force along with the negatively loaded variables 34, 14, 21, 1, 20, 2, 35, 3, and 37 indicated that Factor 8 was measuring the state of welfare of the population. Accordingly, Factor 8 was designated as Welfare Needs.

Factor 9. The variables loaded significantly on Factor 9 are shown in Table VI-20. This factor had significant loadings on 5 of the 38 variables. The high positive loading of variable 28, number of towns and municipalities, and the lower positive loading on variable 31 along with negative loadings on variables 3, 4, and 18 suggested a rurally

TABLE VI-18

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 7* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
16	Percent non-white population	0.9153
11	Ratio of bank deposits to aggregate income	0.8842

*Factor 7 was interpreted as Extensiveness.

TABLE VI-19

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 8* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
29	Non-worker to worker ratio	0.8627
23	Percent of population in labor force	-0.8312
9	Change in retail sales per capita	0.5975
34	Farm operator level-of-living index	-0.5651
14	Percent urban population	-0.4723
26	Average size of household	0.4089
21	Median school years completed	-0.4198
19	Percent of families with income under $3,000	0.4047
18	Percent of population voting in the 1960 Presidential election	0.3777
1	Rate of population change	-0.3545
20	Percent of families with incomes over $10,000	-0.3494
2	Change in median income of all families	-0.3148
35	Local government general expenditures on natural resources	-0.2673
3	Change in median income of rural families	-0.2600
37	Location of TVA steam generator plants (dummy)	-0.2412

*Factor 8 was interpreted as Welfare Needs.

TABLE VI-20

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 9* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
28	Number of towns and municipalities	0.7557
3	Change in median income of rural families	-0.5538
4	Change in median income of rural farm families	-0.4647
18	Percent of population voting in the 1960 Presidential election	-0.4633
31	Percent of farms with sales under $2,500	0.2867

*Factor 9 was interpreted as Local Government Ineptness.

oriented community in which local governments were small and poor. As a result of these relationships Factor 9 was named Local Government Ineptness.

Factor 10. Factor 10 had significant loadings on nine of the 38 variables as indicated in Table VI-21. The relatively high positive loading of variable 25, percent of employment working outside county of residence, along with the negative loadings on variables 27, 24, 20, 14, and 21 indicated a preference for the home community. Accordingly, Factor 10 was identified as Immobility.

Discussion of Results

While the above factors are independent as possible, given the power of the statistical technique, there was some overlapping similar to that in the first factor analysis. For example, of the 38 variables listed in Table VI-10, page 119, five appeared as significant variables in four different factors; four appeared as significant variables in five different factors; and three appeared as significant variables in three different factors. The five variables that appeared as significant in four factors were rate of population change, percent urban population, percent of employment in white collar occupations, percent of farms with sales under $2,500, and local government general expenditures on natural resources. The four variables that appeared as significant in five different factors were percent of population voting in the 1960 Presidential election, percent of families with incomes over $10,000, median school years completed, and farm operator

TABLE VI-21

SOCIO-ECONOMIC VARIABLES USED TO IDENTIFY FACTOR 10* AND FACTOR LOADINGS DERIVED FROM THE FACTOR ANALYSIS OF THE 38 VARIABLES INCLUDED IN THE STUDY

Variable Number	Description	Factor Loadings
25	Percent of employment working outside county of residence	0.7009
27	Per capita property taxes	-0.5698
9	Change in retail sales per capita	0.3393
24	Percent of employment in white collar occupations	-0.3083
35	Local government general expenditures on natural resources	0.3048
20	Percent of families with incomes over $10,000	-0.2645
14	Percent urban population	-0.2612
21	Median school years completed	-0.2436
38	Location of TVA dams (dummy)	0.2864

*Factor 10 was interpreted as Immobility.

level-of-living index. The three variables that appeared as significant in three factors were change in retail sales per capita, percent rural farm population, and location of TVA dams. As was the case in the first factor analysis the second factor analysis indicated that along with economic variables, social and political variables were also important factors in the economic growth of the Tennessee Valley region. Thus, it would appear that future research should place more emphasis upon social and political forces as well as economic forces in studies of regional economic growth and development.

Relationship of the 10 Factors to the Index of Economic Growth

In this section, multiple regression analysis was used to test the general hypothesis that some portion of the variation in the index of economic growth derived by principal component analysis in Chapter IV would be explained by variations in the 10 derived factors. The analysis of the relationship between factors and the index of economic growth was completed utilizing the same procedure that was employed in the previous analysis of the 25 independent variables

The Regression Analysis. The regression of the index of economic growth on the 10 identified factors resulted in the following regression equation, where the variables are arranged in the order in which they entered into the equation.[4]

---

[4]The asterisk (*) indicates statistical significance at the .01 level. The variables X_2 and X_{10}, Natural Resource Investments and Immobility, were significant at the 0.25 and the 0.70 levels of significance, respectively.

$$Y = 0.000005354 - 0.024748X_6^* + 0.0079191X_1^* - 0.011300X_9^* + 0.0096339X_7^* - 0.010402X_3^* + 0.0066783X_4^* + 0.0062459X_5^* - 0.0042086X_8^* + 0.0018649X_2 - 0.00064529X_{10}$$

where

Y = the index of economic growth

X_1 = Level-of-Living

X_2 = Natural Resource Investments

X_3 = Population Maturity

X_4 = Sub-Marginal Agriculture

X_5 = Agricultural Employment

X_6 = Ruralism

X_7 = Extensiveness

X_8 = Welfare Needs

X_9 = Local Government Ineptness

X_{10} = Immobility

The R^2 for the above equation was .9767 indicating that almost 98 percent of the variation in the index of economic growth was explained by the equation. This R^2 value was significant at the .01 level. Eight of the 10 regression coefficients had significant student t-values at the .01 level.

The coefficients in the above regression equation can be interpreted as follows. As Level-of-Living increased by one unit, the index of economic growth, Y, increased by 0.0079191 of one unit indicating an

increase in the level of economic growth. Similarly, as Ruralism, Local Government Ineptness, Population Maturity, Welfare Needs, and Immobility increased by one unit, the index of economic growth decreased 0.024748, 0.011300, 0.010402, 0.0042086, and 0.00064529, respectively. On the other hand, as Extensiveness, Sub-Marginal Agriculture, Isolationism, and Natural Resource Investments increased by one unit, the index of economic growth increased by 0.0096339, 0.0066783, 0.0062459, and 0.0018649 of one unit, respectively.

All of the above signs for the coefficients were in agreement with a priori reasoning. As was the case in the analysis of the 25 independent variables, the relatively small values of the regression coefficients reflect the complexity of the relationships involved in economic growth. These small coefficients probably also indicate that there are several more factors outside the set of 38 variables which are important in explaining economic growth in the Tennessee Valley.

VI. SUMMARY

In this chapter factor analysis was used as an objective method for formulating hypotheses about regional economic growth and in selecting dependent variables which constituted the basic or functional unities of a complex system of interrelated variables. The derived factors were then utilized as independent variables in a multiple regression analysis of the variations in the index of economic growth that was derived by principal component analysis in Chapter IV.

The first factor analysis was carried out on a set of 25 independent variables. In this analysis seven factors were derived and identified as (1) Extent of Poverty, (2) Sub-Marginal Agriculture, (3) Labor Force Maturity, (4) Inadequate Natural Resource Investments, (5) Extent of Metropolitan Work Opportunities, (6) Labor Force Productivity, and (7) Extractive.

The regression of the index of economic growth on the seven identified factors yielded a regression equation with an R^2 of 0.7633. In this equation Extent of Poverty, Sub-Marginal Agriculture, Labor Force Maturity and Labor Force Productivity had significant regression coefficients at the .01 level of significance. Inadequate Natural Resource Investments, Extent of Metropolitan Work Opportunities, and Extractive were significant at much lower levels of significance.

Because the regression coefficients were relatively small in the first regression analysis it was thought that there was variables outside the set of 25 variables that might be important in explaining the variation in the level of economic growth as indicated by the derived index of economic growth. Therefore, a second factor analysis was completed on a system of 38 variables which included the original 25 variables from the first factor analysis plus 13 more from the analysis conducted in Chapter IV.

The second factor analysis derived 10 factors from the 38 variables: (1) Level-of-Living, (2) Natural Resource Investments, (3) Population Maturity, (4) Sub-Marginal Agriculture, (5) Agricultural Employment, (6) Ruralism, (7) Extensiveness, (8) Welfare Needs, (9) Local Government Ineptness, and (10) Immobility.

In a regression equation these 10 factors explained 97.67 percent of the variation in the index of economic growth and had eight regression coefficients with significant student t-values. Again the values of the regression coefficients were relatively small indicating that because of the complexity of the interrelationships in regional economic growth it is difficult to identify and include all the relevant variables in an analysis. In conclusion, while the selection of original variables is still a subjective process, and while important dimensions may be left out, factor analysis did facilitate the generation of inferences of causality and of relationships among the many variables important to regional economic growth. Also, the factor analysis emphasized the importance of the interrelationships between economic, social, and political factors in explaining economic growth.

CHAPTER VII

SUMMARY AND CONCLUSIONS

Despite large investments in natural resource projects, economic growth in the Tennessee Valley region during the decade from 1950 to 1960 was not evenly spread across the region. While some areas of the region grew at rates equal to or above the national rate of economic growth, other areas within the region suffered from poverty, unemployment, and lagging rates of economic development. Thus, while it is apparent that economic growth has taken place in some areas of the region, it is not so apparent why some areas within the region have prospered while others have not. In an attempt to provide information that will give some insight into the problem of ascertaining how effective various kinds and amounts of natural resource investments are in stimulating regional economic growth this study was undertaken.

In this study it was hypothesized that regional economic growth is a function of the socio-economic characteristics of the region in question. Therefore, in order to investigate the economic growth of the Tennessee Valley region various socio-economic variables were observed and analyzed in order to fulfill the objectives of this study. The overall objectives of this study were (1) to ascertain the factors that determine the rate of economic growth in the counties making up the Tennessee Valley region, and (2) to assess the multistatistical approach as a method of analysis in regional research. Specific objectives of the

study were (1) to determine the relationship of various types of socio-economic changes with the level and pace of economic development and growth in the Tennessee Valley region, (2) to estimate the relationships between various kinds and amounts of investments in natural resources and the economic growth of the Tennessee Valley region, and (3) to discover the causes of disparity in the rate of economic growth between counties in the Tennessee Valley region.

A review of the literature and research applicable to the study of regional economic growth revealed that there is no one generally accepted theory of regional economic growth. Because the processes involved in regional growth are extremely complex, the theoretical work to date suggested that some consideration should be given to alternative methods of investigating the relationships involved in regional growth.

Observations on the various socio-economic variables thought to affect regional growth were collected from secondary sources for the 150 county study region. The statistical methodology chosen to analyze the data so as to provide meaningful and testable hypotheses was multivariate analysis. In the present research the multivariate statistical analysis employed included the following statistical techniques: (1) principal component analysis--a technique for transforming the original variables into new variables which may reduce the dimensions of the original observations and serve as an index of economic growth; (2) factor analysis--a technique for explaining the possible interactions of the socio-economic variables so as to provide a simpler, more compact explanation of the regularities apparent in the data; (3) multiple discriminant

analysis--a technique for making multivariate comparisons of several groups of counties on the basis of their performance as it relates to economic growth; and (4) multiple regression analysis--a technique used to analyze the relationship between the extracted factors derived by the factor analysis and the index of economic growth derived from the principal component analysis.

To test the hypothesis that regional economic growth is a function of the socio-economic characteristics of the region, data were collected from various secondary sources and subjected to the above mentioned statistical techniques. In selecting the data to be analyzed the following criteria were utilized: (1) the data had to be readily available from secondary sources; (2) the data had to be available for the same time period, 1950 to 1960; and (3) the data had to be available by county for all counties in the study region.

Initially, 60 socio-economic variables were selected to represent the socio-economic environment of the region. The first 23 of these variables were designed to measure the magnitude of economic change which occurred in the region between 1950 and 1960. The remaining 37 variables were observed as of 1960 and were entered into the data analysis as independent variables.

Principal component analysis was used to reduce the multiple dimensions of regional economic growth into a single dimension that was subsequently utilized as an index of economic growth. The technique of principal component analysis removes those variables exhibiting small variances and concentrates attention on only those variables with large

variances. The analysis results in a set of principal components which are derived from the original data matrix through a linear transformation of the variables. The first principal component in this set is that linear combination of the original variables which contributes a maximum to their total variance. Thus, the first principal component of the original variables, which explains more variation than any other component, becomes the index of economic growth.

The results of the principal component analysis of the data representing 13 of the first 23 measure variables[1] indicated that the first principal component accounted for 40 percent of the total variation in the multivariate system. Of the variables making up the first principal component or the index of economic growth, the variables having the most weight were (1) change in the median income of all families, (2) rate of population change, and (3) change in total payroll per capita.

On the basis of this index, the 150 counties in the study area were arrayed in order according to the size of the first principal component. The ranked counties were then divided into four groups representing different rates of economic growth. There were 13 counties in the fast growth group, 21 in the medium growth group, and 91 in the no growth group, and 25 in the slow growth group.

The ranked counties displayed certain economic characteristics. First, the results of the analysis indicated that economic growth in

---

[1]These 13 variables were selected through correlation analysis.

the Tennessee Valley region during the 1950-1960 period tended to be urban oriented. Secondly, the analysis indicated that the faster growing counties in this period were located along waterways or rivers which had received natural resource investments in the form of dams and/or steam generator plants. These results suggested the hypothesis that economic growth during the 1950-1960 period in the Tennessee Valley was directly affected by natural resource investments in such installations as dams and steam generator plants.

Multiple discriminant analysis was used to detect the differences in the value of the index of economic growth among the counties under study. The discriminant analysis provided a method for testing the significance of the four groups of counties identified by the principal component analysis by examining the group membership of the counties on the basis of 25 independent variables.

The analysis indicated that the four groups were widely separated, thus supporting the results of the initial grouping of counties based on the index of economic growth derived from principal component analysis. However, the discriminant analysis also indicated that the extreme groups--fast growth and no growth--were more easily determined. The discriminant analysis also indicated that median school years completed, average size of household, non-worker to worker ratio, location of TVA steam generator plants, and location of TVA dams were more important in the discriminating equations than the raw data indicated.

Factor analysis was used as an objective method for formulating hypotheses about regional economic growth and selecting dependent variables which constituted the basic or functional unities of the complex system of interrelated variables in the regional economy. The primary purpose of factor analysis was to reduce the original number of explanatory variables to a number of independent factors in terms of which the whole set of variables could be understood. It accomplished this purpose by identifying and summarizing the structural relationships of a system so as to combine highly interrelated variables into clusters called factors. The derived factors were then utilized as independent variables in a multiple regression analysis of the variations in the index of economic growth that was derived from the principal component analysis.

The factor analysis was applied to two sets of variables. The first factor analysis was applied to the same 25 independent variables used in the discriminant analysis and extracted seven factors. These factors were tentatively identified as (1) Extent of Poverty, (2) Sub-Marginal Agriculture, (3) Labor Force Maturity, (4) Inadequate Natural Resource Investments, (5) Extent of Metropolitan Work Opportunities, (6) Labor Force Productivity, and (7) Extractive.

Following the identification of the seven factors derived by the factor analysis it was assumed that these factors represented certain dimensions of regional economic growth, and thus should explain some part of the variation in economic growth among counties within the Tennessee Valley region. A measure of the degree to which each county exhibited or possessed each of the seven factors was provided by the

calculation of the factor score matrix. These scores became independent variables in regression analysis where the index of economic growth was used as the dependent variable.

The relationship between the seven factors and regional economic growth was analyzed by regressing the index of economic growth derived by principal component analysis on the seven factors. The resulting regression equation had an R^2 of 0.76, indicating that 76 percent of the variation in the index of economic growth was explained by variations in the factors. This R^2 value was significant at the .01 level. In this regression equation four of the seven independent variables had significant student t-values at the .01 level while the other three variables were significant at much lower levels of significance. The first four factors to enter the stepwise equation were Factor 1 (Extent of Poverty), Factor 2 (Sub-Marginal Agriculture), Factor 6 (Labor Force Productivity), and Factor 3 (Labor Force Maturity). These four factors actually explained 99 percent of the total variation explained by the seven factors at the .01 level.

Because the regression coefficients were relatively small in the first regression analysis it was thought that there were variables outside the closed system of 25 variables that might be important in explaining the variations in the index of economic growth. Therefore, a second factor analysis was completed on an expanded system of 38 variables. The second factor analysis extracted 10 factors which were tentatively identified as (1) Level-of-Living, (2) Natural Resource Investments, (3) Population Maturity, (4) Sub-Marginal Agriculture, (5) Agricultural

Employment, (6) Ruralism, (7) Extensiveness, (8) Welfare Needs, (9) Local Government Ineptness, and (10) Immobility.

Following the same procedure as was used with the first factor analysis, the relationship between the 10 factors and the index of economic growth derived in Chapter IV was analyzed by regressing the index of economic growth on the 10 factors. The resulting regression equation had an R^2 of 0.977 and eight regression coefficients with significant student t-values at the .01 level. The first five factors to enter the equation were Factor 6 (Ruralism), Factor 1 (Level-of-Living), Factor 9 (Local Government Ineptness), Factor 7 (Extensiveness), and Factor 3 (Population Maturity). These five factors actually explained 99 percent of the total variation explained by the 10 factors at the .01 level.

The results of the present study supported the following hypotheses: (1) regional economic growth is a function of the socio-economic characteristics of the region, (2) the observed socio-economic characteristics can be analyzed by various multivariate statistical procedures as a means for gaining an understanding of the process of regional economic growth, and (3) when the socio-economic characteristics representing man's reactions to his economic environment are hypothesized to be relevant to regional economic growth, on the basis of a sound theoretical framework, then the resulting factors derived from the data through factor analysis will explain variations in a derived index of economic growth.

The present research led to the following conclusions and implications regarding additional research: (1) The present study appeared to indicate the importance of the interrelationships between economic, social, and political factors in explaining economic growth. It may well be that social and political factors may actually play a more important role in explaining regional economic growth than strictly economic factors. There seems to be a definite need for research aimed at developing a system of social, economic, and political variables that can be utilized in future analyses of regional economic growth. (2) The results of the analyses of the present study appeared to indicate that natural resource investments such as dams and steam generator plants had very little direct impact on county economic growth rates during the period 1950 to 1960.[2] However, variables representing natural resource investments appeared in several of the factors extracted by the factor analysis. This would seem to indicate that natural resource investments during the 1950 to 1960 period played an indirect role in stimulating local economic growth. Additional research is needed that will be aimed at identifying and quantifying the various types of natural resource investments that have been made in the Tennessee Valley region, and measuring the indirect and direct effects of these investments on regional economic growth. (3) The factors

---

[2]This conclusion is supported by the results of Wiebe's study [94] of the role of natural resource investments in stimulating economic growth in the Tennessee Valley. He found that TVA investments did not show a strong impact on the counties where the investment projects were located but perhaps had their greatest impact on the urban growth centers of the region.

identified in the present study are considered to be a contribution to some subset of the regional economic growth process, which will ultimately define the dimensions of regional economic growth. However, the present study was considered exploratory in nature and the processes identified through the multistatistical approach require additional empirical validation as well as the identification and evaluation of additional dimensions of economic growth.

As is the case with much of the research in regional economic growth, the findings of the present study are highly tentative. While a high proportion of the variation in the derived index of economic growth was explained by the extracted factors, the principal component analysis used to derive the index of economic growth left a large proportion of the total variation of the 13 variable multivariate system unexplained. Clearly, additional research is needed that will be aimed at determining the set of socio-economic variables which are representative of the changes that occur in regional economics over a given time period. Nevertheless, the results of the present study should be of value to subsequent researchers who are interested in the factors affecting economic growth in a regional economy. The present research also demonstrates the usefulness of the multistatistical approach as a methodology for systemization and rationalization of facts that will allow the expression and testing of meaningful hypotheses about regional economic growth. It is hoped that the results of the present study will provide some insights and inspiration to those who are interested in continuing research in this important subject area of regional economic growth.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Abdel-Badie, Farid and L. A. Parcher. Regression and Discriminant Analyses of Agricultural Land Prices. Agricultural Experiment Station Processed Series P-579. Stillwater, Oklahoma: Oklahoma State University, December, 1967.

2. Adelman, Irma and Cynthia Taft Morris. "Performance Criteria for Evaluating Economic Development Potential--An Operational Approach," Quarterly Journal of Economics, 72:260-280, May, 1968.

3. _____. Society, Politics and Economic Development: A Quantitative Approach. Baltimore: Johns Hopkins Press, 1967.

4. Airov, Joseph. "The Construction of Interregional Business Cycle Models," Journal of Regional Science, 5:1-20, Summer, 1963.

5. Anderson, T. W. "The Use of Factor Analysis in the Statistical Analysis of Multiple Time Series," Psychometrika, 28:1-25, March, 1963.

6. Andrews, Richard B. "Mechanics of the Urban Economic Base: Historical Development of the Base Concept," Land Economics, 29:161-167, May, 1953.

7. Barlowe, Raleigh. Land Resource Economics. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1958.

8. Barnett, Harold J. and Chandler Morse. Scarcity and Growth, The Economics of Natural Resource Availability. Baltimore: Johns Hopkins Press, 1963.

9. Baumol, William J. Economic Dynamics. New York: The Macmillan Company, 1951.

10. Berry, Brian J. L. "City Size Distributions and Economic Development," Economic Development and Cultural Change, 9:573-587, July, 1961.

11. Cattell, Raymond B. Factor Analysis, An Introduction and Manual for the Psychologist and Social Scientist. New York: Harper Brothers, 1952.

12. Christian, Virgil L. and Claude M. Vaughn. "Electric Power and the Growth of the Tennessee Valley Economy," The Atlanta Economic Review, 17:9-13, October, 1967.

13. Clark, Colin. The Conditions of Economic Progress. Third edition. London: MacMillan and Company, 1957.

14. Cooley, William W. and Paul R. Lohnes. Multivariate Procedures for the Behavioral Sciences. New York: John Wiley and Sons, Inc., 1962.

15. Cotner, Melvin L. "A Policy for Public Investments in Natural Resources," American Journal of Agricultural Economics, 51: 87-99, February, 1969.

16. Cox, P. Thomas. "Statistical Analyses of Growth and Development in Project Areas or Potential Project Areas." Washington: Natural Resource Economics Division, Economic Research Service, United States Department of Agriculture, 1969. (Mimeographed.)

17. Cox, P. Thomas and Daniel D. Badger. Factors Contributing to Success of Upstream Watershed Development in Oklahoma. Agricultural Experiment Station Processed Series P-578. Stillwater, Oklahoma: Oklahoma State University, November, 1967.

18. Dixon, W. J. (ed.). BMD Biomedical Computer Programs. Berkeley: University of California, 1968.

19. Durand, David. "Risk Elements in Consumer Installment Financing," Financial Research Program, Studies in Consumer Installment Financing, 8. New York: National Bureau of Economic Research, 1941, pp. 125-135.

20. Fisher, Allan G. B. "Capital and the Growth of Knowledge," Economic Journal, 43:379-389, September, 1933.

21. Fleming, J. Marcus. "External Economies and the Doctrine of Balanced Growth," The Economics of Underdevelopment, A. N. Agarwala and S. P. Singh (eds.). New York: Oxford University Press, 1963.

22. Friedman, John. "Cities in Social Transformation," Regional Development and Planning, A Reader, John Friedman and William Alonso (eds.). Cambridge: The M. I. T. Press, 1964.

23. Friedman, John and William Alonso (eds.). Regional Development and Planning, A Reader. Cambridge: The M. I. T. Press, 1964.

24. Gurshick, M. A. "Principal Components," Journal of the American Statistical Association, 31:519-528, 1936.

25. Hansen, W. Lee and Charles M. Tiebout. "An Intersectoral Flows Analysis of the California Economy," The Review of Economics and Statistics, 45:409-418, November, 1963.

26. Harman, Harvey H. Modern Factor Analysis. Second edition, Revised. Chicago: The University of Chicago Press, 1960.

27. Harrod, R. F. Towards a Dynamic Economics. London: MacMillan and Company, 1948.

28. Haygood, Margaret Faraman, Nadia Danilevsleyand and Colin O. Brown. "An Examination of the Use of Factor Analysis in the Problem of Subregional Delineation," Rural Sociology, 6:216-233, September, 1941.

29. Henderson, James M. and Richard E. Quandt. Microeconomics Theory, A Mathematical Approach. New York: McGraw-Hill Book Company, 1958.

30. Higgins, Benjamin. Economic Development: Principles, Problems, and Policies. Revised edition. New York: W. W. Norton and Company, 1968.

31. Hirschman, Albert G. The Strategy of Economic Development. New Haven: Yale University Press, 1958.

32. Hoel, Paul G. Introduction to Mathematical Statistics. Third edition. New York: John Wiley and Sons, Inc., 1962.

33. Hoover, E. M. and Joseph Fisher. "Research in Regional Economic Growth," Problems in the Study of Economic Growth. New York: National Bureau of Economic Research, 1949.

34. Horst, Paul. Factor Analysis of Data Matrices. New York: Holt, Rinehart and Winston, Inc., 1965.

35. Hotelling, Harold. "Analysis of a Complex of Statistical Variables into Principal Components," Journal of Educational Psychology, 24:417-441, 1933.

36. _____. "Simplified Calculation of Principal Components," Psychometrika, 1:27-35, March, 1936.

37. Isard, Walter, et al. Methods of Regional Analysis: An Introduction to Regional Science. Cambridge: The M. I. T. Press, 1960.

38. Kalter, Robert J. "A Model to Estimate the Economic Effects of Water-Based Recreation Projects on Local Political Subdivisions." Unpublished Doctoral dissertation, University of Wisconsin, 1966.

39. Kendall, M. G. A Course in Multivariate Statistical Analysis. New York: Hafner Publishing Company, Inc., 1961.

40. _____. "The Analysis of Economic Time Series," Journal of the Royal Statistical Society, Series A, 116:11-34, March, 1953.

41. Lampard, Eric E. "The History of Cities in the Economically Advanced Areas," Economic Development and Cultural Change, 3:81-136, January, 1955.

42. Little, Charles H. and Gerald A. Doeksen. An Input-Output Analysis of Oklahoma's Economy. Technical Bulletin T-124. Stillwater, Oklahoma: Department of Agricultural Economics, Oklahoma State University, February, 1968.

43. Metzler, Lloyd. "A Multiple-Region Theory of Income and Trade," Econometrics, 18:329-354, October, 1950.

44. Morrison, Donald F. Multivariate Statistical Methods. New York: McGraw-Hill Book Company, 1967.

45. Morrow, William Ruthvan. "The Identification and Evaluation of Managerial Behavior." Unpublished Doctoral dissertation, University of Tennessee, December, 1968.

46. North, Douglas D. "Location Theory and Regional Economic Growth," Journal of Political Economy, 63:243-258, June, 1955.

47. Nourse, Hugh O. Regional Economics: A Study in the Economic Structure, Stability, and Growth of Regions. New York: McGraw-Hill Book Company, 1968.

48. Nurkse, Ragnar. Problems of Capital Formation in Underdeveloped Countries and Patterns of Trade and Development. New York: McGraw-Hill Book Company, 1968.

49. Pearson, Karl. "On Lines and Planes of Closest Fit to Systems of Points in Space," Philosophical Magazine, Series 6, 2:559-572, 1901.

50. Perloff, Harvey S., Edhar S. Dunn, Jr., Eric E. Lampard, and Richard F. Muth. Regions, Resources, and Economic Growth. Baltimore: The Johns Hopkins Press, 1960.

HC 103.R45

51. Peters, William S. and George W. Sumners. Statistical Analysis for Business Decisions. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968.

52. Pfouts, Ralph (ed.). The Techniques of Urban Economic Analysis. West Trenton, New Jersey: Chandler-Davis Publishing Company, 1960.

53. Rao, C. Radhakrishna. Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, Inc., 1952.

54. Rosenstein-Roden, P. N. "Problems of Industrialization of Eastern and South-Eastern Europe," The Economics of Underdevelopment, A. N. Agarwala and S. P. Singh (eds.), New York: Oxford University Press, 1963.

55. Schnore, L. F. "The Statistical Measurement of Urbanization and Economic Development," Land Economics, 37:229-245, August, 1961.

56. Schultz, Theodore W. "A Framework for Land Economics--The Long View," Journal of Farm Economics, 33:204-215, May, 1951.

57. _____. "Connections Between Natural Resources and Economic Growth," Natural Resources and Economic Growth, Joseph J. Spangler (ed.). Washington: Resources for the Future, Inc., 1961.

58. _____. Economic Growth and Agriculture. New York: McGraw-Hill Book Company, 1968.

59. _____. The Economic Organization of Agriculture. New York: McGraw-Hill Book Company, 1953.

60. Stone, J. R. N. "On the Interdependence of Blocks of Transactions," Journal of the Royal Statistical Society, Supplement, 8:1-12, 1947.

61. Tang, Anthony M. Economic Development in the Southern Piedmont, 1860-1950. Chapel Hill, North Carolina: The University of North Carolina Press, 1958.

62. Tangri, Shanti. "Urbanization, Political Stability, and Economic Growth," Regional Development and Planning, A Reader, John Friedman and William Alonso (eds.). Cambridge: The M. I. T. Press, 1964.

63. Tennessee Valley Authority. The Tennessee Valley Region, Highlights of Growth and Change. Knoxville: Tennessee Valley Authority, 1968.

64. Thurstone, L. L. Multiple Factor Analysis. Chicago: University of Chicago Press, 1961.

65. Tiebout, Charles M. "Exports and Regional Economic Growth," Journal of Political Economy, 64:160-164, April, 1956.

66. _____. "The Community Economic Base Study," Supplementary Paper No. 16. New York: The Committee for Economic Development, December, 1962.

67. Tintner, Gerhard. Econometrics. New York: John Wiley and Sons, Inc., 1952.

68. _____. "Some Applications of Multivariate Analysis to Economic Data," Journal of the American Statistical Association, 41: 472-500, December, 1946.

69. United States Bureau of the Census. Census of Agriculture: 1959. Vol. I, Counties, Part 31, Alabama. Washington: United States Government Printing Office, 1959.

70. _____. Census of Agriculture: 1959. Vol. I, Counties, Part 31, Georgia. Washington: United States Government Printing Office, 1959.

71. _____. Census of Agriculture: 1959. Vol. I, Counties, Part 31, Kentucky. Washington: United States Government Printing Office, 1959.

72. _____. Census of Agriculture: 1959. Vol. I, Counties, Part 31, Mississippi. Washington: United States Government Printing Office, 1959.

73. _____. Census of Agriculture: 1959. Vol. I, Counties, Part 31, North Carolina. Washington: United States Government Printing Office, 1959.

74. _____. Census of Agriculture: 1959. Vol. I, Counties, Part 31, Tennessee. Washington: United States Government Printing Office, 1959.

75. _____. Census of Agriculture: 1959. Vol. I, Counties, Part 31, Virginia. Washington: United States Government Printing Office, 1959.

76. _____. Census of Governments: 1962. Vol. IV, No. 2, Finance of County Governments. Washington: United States Government Printing Office, 1962.

77. _____. County and City Data Book, 1952, A Statistical Abstract Supplement. Washington: United States Government Printing Office, 1952.

78. _____. County and City Data Book, 1962, A Statistical Abstract Supplement. Washington: United States Government Printing Office, 1962.

79. _____. United States Census of Population: 1950. General Social and Economic Characteristics, Alabama. Washington: United States Government Printing Office, 1950.

80. _____. United States Census of Population: 1950. General Social and Economic Characteristics, Georgia. Washington: United States Government Printing Office, 1950.

81. _____. United States Census of Population: 1950. General Social and Economic Characteristics, Kentucky. Washington: United States Government Printing Office, 1950.

82. _____. United States Census of Population: 1950. General Social and Economic Characteristics, Mississippi. Washington: United States Government Printing Office, 1950.

83. _____. United States Census of Population: 1950. General Social and Economic Characteristics, North Carolina. Washington: United States Government Printing Office, 1950.

84. _____. United States Census of Population: 1950. General Social and Economic Characteristics, Tennessee. Washington: United States Government Printing Office, 1950.

85. _____. United States Census of Population: 1950. General Social and Economic Characteristics, Virginia. Washington: United States Government Printing Office, 1950.

86. _____. United States Census of Population: 1960. General Social and Economic Characteristics, Alabama. Washington: United States Government Printing Office, 1960.

87. _____. United States Census of Population: 1960. General Social and Economic Characteristics, Georgia. Washington: United States Government Printing Office, 1960.

88. _____. United States Census of Population: 1960. General Social and Economic Characteristics, Kentucky. Washington: United States Government Printing Office, 1960.

89. _____. United States Census of Population: 1960. General Social and Economic Characteristics, Mississippi. Washington: United States Government Printing Office, 1960.

90. _____. United States Census of Population, 1960. General Social and Economic Characteristics, North Carolina. Washington: United States Government Printing Office, 1960.

91. _____. United States Census of Population, 1960. General Social and Economic Characteristics, Tennessee. Washington: United States Government Printing Office, 1960.

92. _____. United States Census of Population, 1960. General Social and Economic Characteristics, Virginia. Washington: United States Government Printing Office, 1960.

93. _____. United States Congress, Senate, Committee on Interior and Insular Affairs. Policies, Standards, and Procedures on the Formulation, Evaluation, and Review of Plans for Use and Development of Water and Related Land Resources, 87th Congress, 2nd Session, Senate Document No. 92, 1962.

94. Wiebe, Jacob Edwin. "Effects of Investments in Water Resources on Regional Income and Employment." Unpublished Doctoral dissertation, University of Tennessee, June, 1970.

95. Wilson, Charles MacArthur. "An Interindustry Analysis of Tennessee with Emphasis on Agriculture." Unpublished Doctoral dissertation, University of Tennessee, June, 1968.

APPENDIXES

APPENDIX A

LIST OF VARIABLES SUBJECTED TO MULTIVARIATE STATISTICAL ANALYSIS

TABLE A-1

LIST OF VARIABLES SUBJECTED TO MULTIVARIATE STATISTICAL ANALYSIS

Variable Number	Variable Identification
1	Rate of population change, 1950-1960
2	Change in median income of all families, 1950-1960
3	Change in median income of rural families, 1950-1960
4	Change in median income of rural farm families, 1950-1960
5	Change in per capita income, 1950-1960
6	Change in the ratio of school enrollment to population, 1950-1960
7	Change in the percent of owner occupied dwellings, 1950-1960
8	Change in total property tax, 1950-1960
9	Change in the median value of owner occupied housing units, 1950-1960
10	Change in the percent of dwellings structurally sound with all plumbing facilities
11	Ratio of part-time to commercial farms, 1960
12	Change in the percent employed in manufacturing, 1950-1960
13	Change in the percent employed in agriculture, 1950-1960
14	Change in the percent employed in trades and services, 1950-1960
15	Change in retail sales per capita, 1950-1960
16	Change in the value of wholesale trade per capita, 1950-1960
17	Change in the value of manufacturing payroll per capita, 1950-1960
18	Change in the value of selected services per capita, 1950-1960
19	Change in the value added in mining per capita, 1950-1960
20	Change in the value of farm land and buildings, 1950-1960
21	Ratio of bank deposits to aggregate income
22	Change in total payroll per capita, 1950-1960
23	Change in value added by manufacturing per capita, 1950-1960
24	Population density
25	Percent urban population
26	Percent rural farm population
27	Percent non-white population
28	Percent of population under 5 years
29	Percent of population 21 years and over
30	Percent of population 65 and over
31	Percent of population voting in the 1960 Presidential election
32	Percent of families with incomes under $3,000
33	Percent of families with incomes over $10,000
34	Median school years completed

TABLE A-1 (continued)

Variable Number	Variable Identification
35	Percent of population over 25 who completed less than 5 years of school
36	Percent of population over 25 who completed high school or more
37	Percent of migrants from different county
38	Percent of population in labor force
39	Percent of employment on white collar occupations
40	Percent of employment working outside county of residence
41	Average size of household
42	Per capita property taxes
43	Number of towns and municipalities
44	Non-worker to worker ratio
45	Number of manufacturing establishments
46	Value added by manufacture ($1,000's)
47	New capital expenditures in manufacturing ($1,000's)
48	Number of retail establishments
49	Value of minerals, industries, shipments, and receipts
50	Average size of farms
51	Percent of farms with sales under $2,500
52	Percent of farms with sales $10,000 and over
53	Average per farm value of farm products sold
54	Farm operator level of living index
55	Local government total general expenditures ($1,000's)
56	Local government general expenditure on natural resources ($1,000's)
57	Percent of land in capability classes I-IV
58	Number of new and expanded plants in the TVA area
59	Location of steam generator plants 0 - none 1 - one or more
60	Location of TVA dams 0 - none 1 - one or more

CRANES CREST

APPENDIX B

CORRELATION MATRIX FOR 60 VARIABLES USED IN THE ANALYSIS OF ECONOMIC GROWTH IN THE 150 COUNTY TENNESSEE VALLEY REGION

TABLE A-2

CORRELATION MATRIX FOR VARIABLES 1-10 USED IN THE ANALYSIS OF ECONOMIC GROWTH IN THE 150 COUNTY TENNESSEE VALLEY REGION

Variables	1	2	3	4	5	6	7	8	9	10
1	1.000	.7542	.484	.488	.263	.176	.056	.332	.586	.586
2		1.000	.665	.686	.233	.089	.202	.240	.565	.565
3			1.000	.605	.200	-.018	-.021	.128	.306	.306
4				1.000	.218	-.007	.031	.090	.314	.314
5					1.000	.010	-.015	.035	.139	.139
6						1.000	.167	.215	.153	-.057
7							1.000	.126	.116	-.063
8								1.000	.131	-.029
9									1.000	.354
10										1.000

TABLE A-3

CORRELATION MATRIX FOR VARIABLES 11-20 USED IN THE ANALYSIS OF ECONOMIC GROWTH IN THE 150 COUNTY TENNESSEE VALLEY REGION

Variables	11	12	13	14	15	16	17	18	19	20
11	1.000	-.171	.251	-.068	.025	-.016	.189	.184	.278	-.216
12		1.000	-.659	.104	-.080	-.196	.004	-.212	-.109	-.228
13			1.000	-.435	.079	.293	.225	.338	.135	.322
14				1.000	-.097	-.190	-.293	-.368	.030	-.216
15					1.000	.210	.173	.282	-.035	.095
16						1.000	.242	.222	-.049	.012
17							1.000	.287	-.053	.066
18								1.000	-.062	.212
19									1.000	-.095
20										1.000

TABLE A-4

CORRELATION MATRIX FOR VARIABLES 21-30 USED IN THE ANALYSIS OF ECONOMIC GROWTH IN THE 150 COUNTY TENNESSEE VALLEY REGION

Variables	21	22	23	24	25	26	27	28	29	30
21	1.000	-.0821	.023	.087	.055	.282	.251	-.155	.236	.381
22		1.000	.553	.477	.542	-.494	-.177	.060	.156	-.276
23			1.000	.173	.371	-.340	-.072	.016	.170	-.114
24				1.000	.631	.436	.056	.181	.092	-.308
25					1.000	-.549	.149	.240	.161	-.270
26						1.000	.287	-.105	-.050	.442
27							1.000	.633	-.421	.024
28								1.000	-.762	-.596
29									1.000	.606
30										1.000

TABLE A-5

CORRELATION MATRIX FOR VARIABLES 31-40 USED IN THE ANALYSIS OF ECONOMIC GROWTH IN THE 150 COUNTY TENNESSEE VALLEY REGION

Variables	31	32	33	34	35	36	37	38	39	40
31	1.000	-.037	-.042	-.162	-.055	-.119	-.153	-.020	.058	.053
32		1.000	-.777	.014	.610	-.702	-.470	-.420	-.731	-.118
33			1.000	.041	-.454	.773	.372	.337	.818	-.198
34				1.0-0	.034	.107	-.044	.002	.043	-.097
35					1.000	-.684	-.424	-.442	-.579	.059
36						1.000	.520	.345	.787	-.239
37							1.000	.135	.451	.110
38								1.000	.169	-.140
39									1.000	-.188
40										1.000

TABLE A-6

CORRELATION MATRIX FOR VARIABLES 41-50 USED IN THE ANALYSIS OF ECONOMIC GROWTH IN THE 150 COUNTY TENNESSEE VALLEY REGION

Variables	41	42	43	44	45	46	47	48	49	50
41	1.000	-.244	-.075	.371	-.153	-.146	-.135	-.168	.198	-.087
42		1.000	-.022	-.294	.257	.271	.226	.225	-.260	-.006
43			1.000	-.141	.253	.268	.245	.312	.102	.004
44				1.000	-.282	-.288	-.250	-.286	.230	-.148
45					1.000	.921	.904	.978	.085	-.108
46						1.000	.909	.918	.076	-.100
47							1.000	.915	.075	-.126
48								1.000	.129	-.132
49									1.000	-.136
50										1.000

TABLE A-7

CORRELATION MATRIX FOR VARIABLES 51-60 USED IN THE ANALYSIS OF ECONOMIC GROWTH IN THE 150 COUNTY TENNESSEE VALLEY REGION

Variables	51	52	53	54	55	56	57	58	59	60
51	1.000	-.038	.241	.145	-.106	-.027	.270	-.090	-.144	-.174
52		1.000	.025	-.618	-.131	-.067	-.287	-.095	.001	.118
53			1.000	.221	.076	.033	.401	.076	-.005	-.121
54				1.000	.283	.173	.589	.296	.126	.016
55					1.000	.542	.162	.855	.361	.033
56						1.000	.035	.469	.277	-.038
57							1.000	.1772	.110	-.125
58								1.000	.226	.069
59									1.000	.102
60										1.000

APPENDIX C

IDENTIFICATION OF THE COUNTIES INCLUDED IN THE STUDY REGION

TABLE A-8

IDENTIFICATION OF THE COUNTIES INCLUDED IN THE STUDY REGION

County Number	County Name
Alabama	
1	Blount
2	Colbert
3	Cullman
4	DeKalb
5	Etowah
6	Franklin
7	Jackson
8	Lauderdale
9	Lawrence
10	Limestone
11	Madison
12	Marion
13	Marshall
14	Morgan
15	Winston
Georgia	
16	Catoosa
17	Dade
18	Fannin
19	Gilmer
20	Lumpkin
21	Rabun
22	Towns
23	Union
24	Walker
25	Whitfield
Kentucky	
26	Calloway
27	Graves
28	Livingston
29	Lyon
30	McCracken
31	Marshall
32	Trigg

TABLE A-8 (continued)

County Number	County Name
Mississippi	
33	Alcorn
34	Itawamba
35	Prentiss
36	Tishomingo
North Carolina	
37	Avery
38	Buncombe
39	Cherokee
40	Clay
41	Graham
42	Haywood
43	Henderson
44	Jackson
45	Macon
46	Madison
47	Mitchell
48	Swain
49	Transylvania
50	Watauga
51	Yancey
Tennessee	
52	Anderson
53	Bedford
54	Benton
55	Bledsoe
56	Blount
57	Bradley
58	Campbell
59	Cannon
60	Carroll
61	Carter
62	Cheatham
63	Chester
64	Claiborne
65	Clay
66	Cocke
67	Coffee
68	Crockett
69	Cumberland
70	Davidson

TABLE A-8 (continued)

County Number	County Name
Tennessee	
71	Decatur
72	DeKalb
73	Dickson
74	Dyer
75	Fayette
76	Fentress
77	Franklin
78	Gibson
79	Giles
80	Grainger
81	Greene
82	Grundy
83	Hamblen
84	Hamilton
85	Hancock
86	Hardeman
87	Hardin
88	Hawkins
89	Haywood
90	Henderson
91	Henry
92	Hickman
93	Houston
94	Humphreys
95	Jackson
96	Jefferson
97	Johnson
98	Knox
99	Lake
100	Lauderdale
101	Lawrence
102	Lewis
103	Lincoln
104	Loudon
105	McMinn
106	McNairy
107	Macon
108	Madison
109	Marion
110	Marshall
111	Maury

TABLE A-8 (continued)

County Name	County Name
Tennessee	
112	Meigs
113	Monroe
114	Montgomery
115	Moore
116	Morgan
117	Obion
118	Overton
119	Perry
120	Pickett
121	Polk
122	Putnam
123	Rhea
124	Roane
125	Robertson
126	Rutherford
127	Scott
128	Sequatchie
129	Sevier
130	Shelby
131	Smith
132	Stewart
133	Sullivan
134	Sumner
135	Tipton
136	Trousdale
137	Unicoi
138	Union
139	Van Buren
140	Warren
141	Washington
142	Wayne
143	Weakley
144	White
145	Williamson
146	Wilson
Virginia	
147	Grayson
148	Lee
149	Scott
150	Washington

VITA

Vernon Glenn Chappell, Jr. was born November 6, 1942, in Newport News, Virginia, the son of Mr. and Mrs. Vernon G. Chappell. He graduated from Rich Square High School in Rich Square, North Carolina, in 1961. In the summer of that same year he entered North Carolina State University, located in Raleigh, North Carolina, where he pursued an agricultural economics curriculum.

He graduated from North Carolina State University with a B. S. degree in Agricultural Economics in 1965. While attending North Carolina State University he was initiated into Alpha Zeta, Blue Key National Honor Society, and the Order of Thirty and Three. He was also a member of Lambda Chi Alpha social fraternity.

Upon graduation he took a position with International Minerals and Chemical Corporation in Augusta, Georgia, as a production trainee in fertilizer manufacturing.

In September, 1966, he entered the Graduate School of the University of Tennessee where he was a National Defense Education Act Fellow in the Agricultural Economics Department. In January, 1969, he became a candidate for the Ph.D. degree in Agricultural Economics. His major area of concentration was in General Agricultural Economics, with minor concentrations of interest in Agircultural Marketing and Price Analysis and Economic Development. While attending the University of Tennessee he was initiated into Gamma Sigma Delta.

Vernon Glenn Chappell, Jr., was married in March, 1965, to the former June Tolson of Woodland, North Carolina. He and his wife are the parents of two children, Christopher and Kimberly Anne. Presently, he is employed by East Carolina University as an Assistant Professor in the Department of Business Administration.