

A MULTIVARIATE ANALYSIS OF ECONOMIC GROWTH
IN THE TENNESSEE VALLEY REGION, 1950-1960*

Glenn Chappell

East Carolina University

ABSTRACT

The objective of this paper was to identify the factors that generated economic growth during the 1950-1960 decade for the counties making up the Tennessee Valley region. Principal component analysis was applied to one set of data in order to derive an index of economic growth. Factor analysis was employed as a method for identifying and summarizing the structural relationships between the many variables thought to affect regional economic growth so as to combine highly related 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 derived from the principal component analysis.

The results of this study indicated the importance of the interrelationships between social, political, and economic factors in explaining economic growth.

INTRODUCTION

In this paper multivariate analysis was used to provide insight into the socioeconomic variables which interact to affect regional economic growth. The techniques of principal component analysis, factor analysis and regression analysis were applied to data representing the socioeconomic environment of the Tennessee Valley region in an experimental attempt to identify important interrelationships in regional economic growth. The analysis indicated that the original number of explanatory variables could be reduced to a smaller number of independent factors in terms of which the whole set of socioeconomic variables could be understood. The analysis also underscored the importance of the interrelationship between regional economic growth and various socie-economic factors.

PROCEDURE

The Data Base

Data on 60 selected socioeconomic variables were gathered from secondary sources. Twenty-three of the 60 variables measured the magnitude of economic change which occurred in the Tennessee Valley region between 1950 and 1960. These variables were used in an experimental attempt to construct an index of economic growth. The remaining 37 variables were observed as of 1960 and were used to explain the economic changes which occurred in the region.

The Multivariate Approach

The multivariate statistical approach used in this paper consisted of a sequence of interdependent steps in analyzing data. In the first step of this approach correlation analysis was applied to the 60 variables representing the socioeconomic environment of the region. The object of this analysis was to eliminate those variables exhibiting similar relations. Next, principal component analysis³ was utilized to transform the variables representing the magnitude of economic change into an index of economic growth.

*The research reported in this study was completed while the author was a Ph. D. candidate at the University of Tennessee in the Department of Agricultural Economics and Rural Sociology. The author wishes to acknowledge the helpful comments of Joe A. Martin.

In the third step of this procedure, factor analysis was employed to reduce the original number of explanatory variables to a small number of independent factors in terms of which the whole set of variables could be understood.⁴ In this paper the derived factors were utilized as independent variables in a multiple regression analysis of the variations in the index of economic growth derived from principal component analysis.

ANALYSIS

Principal Component Analysis

A multivariate index of economic growth for the counties in the Tennessee Valley region was constructed by subjecting 13 of the 23 variables⁵ measuring economic change to a principal component analysis [3, pp. 150-154]. This analysis produced an index represented by a linear compound of the 13 variables which accounted for 40 per cent of the total variation of the 13-variable system. Not only was this linear compound more parsimonious, but the relative importance of the original variables in the newly derived component were indicated by their coefficients.⁶ The most important variables in the index of economic growth were (1) change in median income of all families, (2) rate of population change, and (3) change in total payroll per capita.

The index of economic growth was computed for each individual county in the Tennessee Valley region.⁷ With the variation in the 13 measure variables expressed along a single continuum for each individual county, the relative economic performances of each county in the region could be compared. This comparison was made by rank-ordering the counties in the region on the basis of the index, since those counties with a relatively larger index were growing at faster rates than those with smaller index values. The counties were then divided into four groups⁸ with the first group representing the fastest and the fourth group representing the slowest growing counties. There were 13 counties in the fastest growth group, 21 counties in the second or moderately fast growth, 25 counties in the moderate growth group and 91 counties in the fourth or slow growth group. Thus it would appear from this classification that a majority of the counties in the Tennessee Valley region could be considered to be slowly growing counties in comparison with other counties in the region.

With the initial classification of the 150-county region into statistically significant group, hypotheses concerning regional economic growth could be formulated for each group. For example, an inspection of the geographical distribution of economic growth suggested that regional economic growth did not appear everywhere at the same time. In contrast to the static equilibrium and stationary system of regional economic growth which are inherent in much of the theoretical treatment of growth, economic growth in the Tennessee Valley region appears to have manifested itself at points or poles of growth. The intensity of regional economic growth at the various poles varied and spreaded through different channels to other areas in the region [10]. The analysis indicated that the fastest growth group of counties and to a certain extent the moderately fast growth group acted as growth poles in the Tennessee Valley region. Counties in these groups actually set in motion certain forces which resulted in growth being communicated from the faster growing, urbanized counties to the more slowly growing areas in the region [5, pp. 183-200]. For example the economic progress of growth poles such as Nashville and Memphis absorbed some of the disguised unemployed of the more slowly growing, surrounding areas and thereby improved the marginal productivity of labor and per capita consumption levels in these counties. Likewise, the increases in purchases and investments that were made by the growth pole areas in the surrounding, more slowly growing counties had a favorable impact on regional economic growth.

In summary, principal component analysis was applied to a set of data designed to measure the magnitude of economic change which occurred in the Tennessee Valley region between 1950 and 1960. The results of this analysis provided an index of economic growth which is used in a subsequent section of this paper in order to obtain a quantitative explanation of economic growth in the Tennessee Valley region. The principal component analysis also suggested the hypothesis that regional economic growth was communicated from

the faster growing more urbanized counties to the slower growing counties in the region.

The Factor Analysis

Data on 25 of the 37⁹ independent variables were used in a general factor analysis [3, pp. 169-184]. This analysis produced a matrix of common factor coefficients which provided the basis for identifying or naming the factors. The usual procedure employed by factor analysts is to use the items having "high" loadings to identify the factors. In an effort to reduce subjectivity in the selection of the variables making up the common factors only the significant variables with loadings of 0.230 and above were included in the common factors derived in this study.

Once variables were chosen for inclusion in the common factors, the factors had to be identified or named by providing a reasonable explanation of the underlying forces which they may be interpreted to represent. Therefore, the next step in this analysis was to identify the factors which were specified in the results of the statistical analysis.

The first Factor. The signs of the loadings in Factor 1 indicated a tendency for counties characterized by strongly oriented rural economies to have low incomes. This positive association of low income and rural characteristics was not unexpected. Schultz [11], for example, has argued that economic development occurs in a specific location matrix, that each matrix is primarily urban-industrial at the core and that factor markets and forces of economic development operate best near the core. Thus, a positive change in Factor 1 was associated with increases in (1) percent rural farm population, (2) percent of farms with sales under \$500, (3) percent of families with incomes under \$3,000, and (4) percent of the population 65 and over. Also associated with an increase in Factor 1 were decreases in (1) percent of employment in white collar occupations, (2) percent of migrants from different county, (3) farm operator level-of-living index, (4) percent of families with incomes over \$10,000, (5) percent urban population, (6) percent of farms with sales \$10,000 and over, and (7) per capita property taxes. This factor, therefore, was interpreted to represent the extent of rural poverty.

Factor Two. The pattern of associations incorporated in Factor 2 was strongly suggestive of the broad differences in income levels of the counties in which agriculture was commercialized and those in which agriculture was of a less intensive type. Studies [1], [8], and [9] have shown that historically, industrial development and economic growth stem from basic initial differences among counties in agricultural resources. Thus, an increase in this factor was interpreted to represent a movement along a scale which ranged from intensive agriculture to extensive agriculture. Such an interpretation was consistent with the combination of variables included in this factor. In particular, an increase in Factor 2, which implied a move away from commercial agriculture, was composed of decreases in every significantly loaded factor except percent of population voting in the 1960 Presidential election. Hence, Factor 2 implied a most difficult situation of low income levels due to the submarginal nature of agricultural resources.

Factor Three. Factor 3 was based upon six socioeconomic variables: percent of population 65 and over, percent of population voting in the 1960 Presidential election, percent of land in capability Classes I-IV, percent of population in labor force, percent of employment working outside county of residence, and number of towns and municipalities. These were all variables which described variations among counties in the maturity of the labor force.

At one end of the scale were labor forces which were relatively young with no strong preferences for the home community. Political involvement was low and mobility great. At the other end of the scale the labor force was characterized by the older worker whose close ties to the "home place" and preference for the home community made him less mobile and more active in community affairs. Thus, factor 3 was interpreted as expressing the relationship between labor force maturity and immobility.

Factor Four. The pattern of associations incorporated in Factor 4 was strongly suggestive of broad differences in levels of economic development between counties in the Tennessee Valley which experienced an intensive pro-

gram of natural resource investments and those which experienced a submarginal program. An increase in this factor was interpreted to represent a movement along a scale which ranged from intensive public resource investments to submarginal investment activity. Such an interpretation was consistent with the combination of variables constituting this factor. Thus, a positive change in Factor 4 was composed of (1) decreases in the location of TVA dams, local government expenditures on natural resources, and location of TVA steam generator plants; (2) decreases in the number of towns and municipalities, percent of families with incomes over \$10,000, percent urban population, percent of employment in white collar occupations, and median school years completed, and (3) increases in percent of families with incomes under \$3,000 and percent rural farm population. There relationships suggested that in counties where public investments in natural resources have been small or sub-marginal, the level of economic development has been low. Therefore, Factor 4 was interpreted to represent submarginal investments in natural resources.

Factor Five. Factor 5 was based on six variables: per capita property taxes, percent of employment working outside county of residence, number of towns and municipalities, percent of population voting in 1960 Presidential election, location of TVA steam generator plants, and percent of farms with sales \$10,000 and over. The interrelationships among these variables suggested that this factor was measuring the relationship between urbanization and employment 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 the extent of metropolitan work opportunities.

Factor Six. Factor 6 had significant loadings on 11 of the 25 socioeconomic variables. The significantly loaded variables showed a considerable degree of homogeneity of content since all were related to the productive capacity of the labor force. The relatively high positive loadings on percent of population in labor force, farm operator level-of-living index, percent urban population and median school years completed along with negative loadings on non-worker to worker ratio, percent of families with incomes under \$3,000, and average size of household indicated that Factor 6 was measuring labor force productivity.

Factor Seven. Factor 7 had significant loadings on eight of the 25 variables. The relatively high positive loading on value of minerals industries shipments and receipts along with positive loadings on the remaining variables except location of TVA dams, indicated that Factor 7 was measuring the economic contributions of the mineral industry location in each county. Accordingly Factor 7 was identified as extractive.

Relationships of the Seven Factors to the Index of Economic Growth

The analysis of the relationship between factors and the index of economic growth was completed by utilizing the following procedure: (1) factor scores were computed¹¹ to indicate the degree to which the 150 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 the principal component analysis as the dependent variable.

The Regression Analysis.¹² The regression of the index of economic growth on the seven factors resulted in the following regression equation, where the variables are arranged in the order in which they entered the equation:¹³

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

where

Y = the index of economic growth

X_1 = Extent of Rural Poverty

X_2 = Sub-Marginal Agriculture

X_3 = Labor Force Maturity and Immobility

X_4 = Submarginal 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.

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 were involved.

CONCLUSIONS

In this paper multivariate analysis was used to analyze regional economic growth. Through principal component analysis, a multivariate index of economic growth was constructed. Factor analysis was then employed to find a small number of factors suggesting common causes among the variables analyzed. The use of factor analysis facilitated the determination of dependency and isolated relationships which would have been difficult to hypothesize before the analysis. Finally, the index of economic growth was regressed on the derived factors in an attempt to explain differentials in economic growth within the Tennessee Valley region.

The results of the analysis served to emphasize the importance of the interrelationship between regional economic growth and the socioeconomic environment. 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 index of economic growth was explained by the extracted factors, the principal component analysis used to derive the index left a large proportion of the total variation in the 13-variable system unexplained. Research is now underway that is aimed at defining a set of variables which will yield a more satisfactory index of economic growth.

Also, while the use of factor analysis results in more reliable predictions, it does not permit the researcher to determine how much each individual variable contributed to the explanation of the variation in economic growth between counties. To overcome this problem the present analysis is being extended to include a regression analysis of the independent variables which were significant in each of the extracted factors. This procedure should specify which socioeconomic variables are important in explaining regional economic growth.

Finally, the results of this study demonstrated the potential usefulness of the multistatistical approach as a method for systemization and rationalization of facts that will allow the expression and testing of meaningful hypotheses about regional economic growth. What is now needed is more experience in the use of the multivariate approach.

FOOTNOTES

¹The Tennessee Valley region as defined in this paper includes the 125 counties in the Tennessee River Watershed in addition to 76 counties which are served by distributors to TVA electric power.

²A Complete list of the variables included in this study is available from the author upon request.

³A mathematical specification of principal component analysis may be found in [6] and [12].

⁴For a mathematical specification of factor analysis see [4] and [12].

⁵Ten of the 23 variables were eliminated because they exhibited similar relations in the correlation analysis.

⁶The coefficients of the component are the elements of the characteristic vector associated with the largest characteristic root of the correlation matrix of the observations [7, p. 224]. The coefficients are available from the author upon request.

⁷The results of this computation may be obtained from the author upon request.

⁸Multiple discriminant analysis was used to test the significance of the four groups identified by the principal component analysis. For the details of the results of this test see [2, pp. 79-96].

⁹Twelve of the independent variables were eliminated from further consideration because they exhibited similar relations.

¹⁰The rotated factor loadings matrix is available from the author upon request.

¹¹Computer programs are currently available for efficiently performing this computation. For further details see [3, pp. 169-184].

¹²A stepwise regression program as outlined in [3, pp. 233-247] was used in this study.

¹³The asterisk (*) indicates statistical significance at the .01 level.

REFERENCES

1. Booth, E. J. R. "Economic Development in Eastern Oklahoma until 1950," unpublished Ph. D. Thesis, Vanderbilt University, 1961.
2. Chappell, Vernon Glenn, "The Identification and Evaluation of Factors Affecting Economic Growth in the Tennessee Valley Region, 1950-1960," unpublished Ph. D. dissertation, The University of Tennessee, 1970.
3. Dixon, W. J., editor, BMD Biomedical Computer Programs, Berkeley, The University of California, 1968.
4. Harman, Harvey H. Modern Factor Analysis, second edition, Revised, Chicago, The University of Chicago Press, 1960.
5. Hirschman, Albert O., The Strategy of Economic Development, New Haven, Yale University Press, 1958.

6. Kendall, M. G., A Course in Multivariate Statistical Analysis, New York, Hafner Publishing Company, 1961.
7. Morrison, D. F., Multivariate Statistical Methods, New York: McGraw Hill Book Company, 1967.
8. Nicholls, William H. "Human Resources and Industrial Development in the Upper East Tennessee Valley, 1900-1950. Quarterly Journal of Economics, 71: 289-316, May, 1957.
9. Southern Tradition and Regional Progress, Chapel Hill: The University of North Carolina Press, 1960.
10. Perroux, Francois, "Note sur la notion de 'pole de croissance'" Materiaux pour une analyse de la croissance economique, cahiers de l'Institut de Science Economique Appliquee, Serie D, No. 8, 1955.
11. Schultz, T. W., "A Framework for Land Economics -- The Long View," Journal of Farm Economics, 33: 204-215, May, 1951.
12. Tintner, Gerhard, Econometrics, New York: John Wiley and Sons, 1952.

The following are comments by Thomas E. Snider, Economist, Federal Reserve Bank of Richmond on "A Multivariate Analysis of Economic Growth in the Tennessee Valley Region, 1950-1960," by Glenn Chappell, East Carolina University.

Studying economic growth on a county basis in an area as large as the Tennessee Valley is a Herculean Undertaking and the author should be commended for his initiative and industriousness.

The stated purpose of the paper was to provide insight into the chaos of multitudinous variables which interact to affect regional economic growth. Multivariate analysis supposedly allowed "... the researcher to suggest more meaningful hypotheses suitable for testing by other intensive analysis" The reader expecting new and more meaningful hypotheses will, however, be disappointed. Two hypotheses are set forth in the paper (1) that regional economic growth is a function of the socioeconomic environment of the region under consideration and (2) "... that regional economic growth was communicated from the faster growing more urbanized counties in the region to the slower growing counties. The first of these is a familiar hypothesis and one which would be readily accepted by students of growth. Similarly, the second is not a new hypothesis but rather a well accepted characteristic of economic growth first discussed in the early 1930's and which most reputable students of economic growth take for granted. [1 and 2]

In some respects this paper illustrates the tendency of some researchers to become preoccupied with the computer and sophisticated statistical techniques to the neglect of economic theory. For example, it is unclear what economic theory is being tested by the study. No hypotheses concerning how the socioeconomic environment affects growth are advanced, and no insights from which a theory might be inferred are revealed. It may just be a characteristic of the analytical tools employed, but at times one is left with the impression that the author loses sight of the distinction between description and explanation. The use of county data to test established theories of interregional development and trade may have been a more fruitful approach than analyzing masses of data with the hope that some hypotheses would emerge.

The index of economic growth developed in the paper is a worthwhile contribution, although one would be more comfortable with the index if it explained a higher proportion of the variation in the 13 variables. The index identifies counties and multicounty areas in the Tennessee Valley where growth has occurred and points out areas where there is a need for economic development. Later parts of the analysis tend to confuse the matter, however, by finding other variables that are also characteristic of the level of economic development. Ostensibly, the purpose of the analysis was to ex-

plain the economic growth of various counties by examining and isolating the causes of economic well being or lack of it. Whether or not this goal was accomplished is unclear. For example, are we seriously expected to believe that having relatively more people over 65 years of age causes a county to be underdeveloped? In the same vein, what significance are we expected to attach to the finding that as poverty increases economic growth decreases? Does poverty affect economic growth or does economic growth affect poverty? The author has found hitherto well-known characteristics of poor counties but that is all that his sophisticated analytical tools provide us with.

There might be considerable merit to the paper if the statistical processes isolated some previously unknown relationships between economic growth and economic data. If the paper accomplished this, it is not easy to discern. Even the principal conclusion of the paper, which is that there is "an important inter-relationship between social, political, and economic factors in explaining economic growth" offers no new insight into the process of regional economic growth.

REFERENCES

1. Hirschman, Albert O. "Interregional and International Transmission of Economic Growth," in The Strategy of Economic Development, Yale University Press 1958.
2. Perroux, Francis. "Note on the Concept of Growth Poles," Economic Appliquee 1955.