A Statistical Approach for Identifying Socioeconomic Structure in Rural Communities

by

Don Blayney, Research Associate

and

Gerald Marousek, Agricultural Economist

Dept. of Agric. Econ. and Applied Stat.

University of Idaho

Moscow

A.E. Research Series No. 227 (Revised)

Presented at the annual meeting of the Western Agricultural Economics Association, New Mexico State University, July 22, 1980.

A Statistical Approach for Identifying Socioeconomic Structure in Rural Communities

ABSTRACT

Factor analysis was used to identify structural relationships in data for Idaho communities. Six significant factors were identified, supporting recognized associations between rural development and natural resources, education, employment opportunities and government actions. Factor analysis, in conjunction with regression, links the structural and spatial facets of communities targeted for development.

A Statistical Approach for Identifying Socioeconomic Structure in Rural Communities

Passage of the Rural Development Act of 1972 signaled the emergence of rural development as an important research priority. Many rural areas have experienced outmigration, exhaustion of natural resources, and significant technological changes in agriculture. Communities in such areas are faced with shrinking economic bases, reduced tax revenue and lower quality public services. While many rural communities are losing population others are growing because of industrial, recreational, mineral or energy development. The adjustment problems of small rural communities are even more pronounced in regions where population is sparse or great distances separate population centers.

Knowledge of the economic structure of a community is important to both the public and private sectors. Whether a community is growing, stagnant or declining, public officials and private investors need information concerning factors associated with development potential in order to direct effort and funds into appropriate channels.

Literature Review

Some writers, such as Harman [6], have focused primarily on the intricate mathematical relationships in factor analysis. But most analysts, including Rummel [13], are more concerned with application and interpretation of results.

Factor analysis has been used to analyze data in both social and natural sciences. Developed in close conjunction with psychological theories by Spearman, Thurstone and others the technique is often mistaken for those theories. However, the mathematical techniques inherent in the procedure are not limited to psychological applications [6, pp. 3-4].

Economists have used a number of techniques in the application of multivariate analyses. Duncan and Leistritz [4] provide an overview of concepts and economic applications. Edwards, et al [5] and March and Smith [9] developed and evaluated economic indices using factor analysis.

Applications of factor analysis to economic development problems have been reported by Jonassen and Peres [7], Leuck [8], Dorf [3] and Chappell [1]. Reinschmiedt and Jones [12] used the technique in conjunction with other techniques to identify rural development potentials in Texas.

A recent study in Idaho [10], which is one of the fastest growing states in the nation as well as one of the most rural,[1/] revealed a set of 15 socioeconomic variables that were significant in discriminating among the state's six economic planning regions. The 15-variable set was derived from 38 original variables.

Objective

The objective of this study was to apply factor analysis as a means of grouping variables and identifying significant factors associated with development potential in Idaho. The paper describes the methodology used, the results of its application, and interpretation of the results.

Methodology

The principal purpose of factor analysis is to reduce a set of variables to (usually) a smaller number of categories or "factors" by linear methods, [6, p. 4]. Three concepts, patterned variation, vector and vector space, and dimension are helpful in understanding the technique [13, p. 12].

There are three steps in factor analysis: 1) preparation of the correlation matrix, 2) extraction of initial factors, and 3) rotation to a terminal solution. Major options at each step are: R-type vs. Q-type factor analysis in step 1, defined vs. inferred factors in step 2, and orthogonal vs. oblique in step 3 [11, p. 469]. The model chosen for this study used the R-type, defined factor analysis and orthogonal rotation.

R-type analysis of the correlation matrix focuses on correlations between the variables in the data set rather than on those between the observations (or units) on which the variables were measured. This option was selected as the most appropriate method of grouping the variables to reveal their underlying structure.

The defined factor or principal components analysis option transforms the data into a set of uncorrelated components or "factors." Each component is defined as the best linear combination of the variables accounting for the variance in the data. The first is the single best summary of linear relationships exhibited in the data, the second is the best summary of residual variance left after the first factor is removed, and so on until all variance in the data is accounted for. As many components as variables could be defined but generally a smaller number of components is retained for further consideration.

Although rotation of factors creates arbitrariness, it is sometimes necessary to obtain meaningful factors and (hopefully) the simplest factor structure. Thurstone's factor rotation rules are summarized by

Harman [6, p. 98]. Orthogonal rotation was selected in this analysis because independent factors or underlying components were desired. Also, oblique rotation assumes a known a priori relationship exists.

The qualitative aspect of factor analysis should not be confused with the method's mathematical properties. The subjectivity is a function of the interpretive process; it does not detract from the validity of the analysis itself.

This study extends the analysis of the 15-variable set cited above [10] and recorded in the left hand column of Table 1. In the present study, county rather than regionally aggregated data were analyzed. The variables from the earlier study were treated as "regional location" parameters. Since each county is affected by its spatial relationship to other counties, this analysis identifies locally significant factors within the regional context. Names are then assigned to these factors. Factor analysis is the most substantive multivariate technique available to achieve these goals [11, p. 445].

The Statistical Analysis System (SAS) program Proc Factor was used.

Program Details

The Proc Factor procedure includes several technical aspects important in interpreting results. First, the principal components method requires unities on the main diagonal of the correlation matrix being analyzed, as verified in Table 1.

Secondly, the eigenvalue of each factor, a value related to operations performed on the correlation matrix, divided by total variance (15 unit variances in this study) measures the variance accounted for by each initially extracted factor. The eigenvalues are also used to determine the number of factors retained after the initial extraction All factors with eigenvalues greater than or equal to one are retained; thus there are six factors in this study (Table 2). Table 2 also shows the individual and cummulative variances accounted for by the factors.

Finally, orthogonal rotation has some interesting properties. While unrotated factors are ordered by descending amounts of the variance accounted for, the ordering of rotated factors may be completely different [13, p. 386]. Thus, for example, Factor I in Table 2 and Factor I in Table 3 are not necessarily related. A measure of the contribution to common variance associated with each rotated factor is useful in evaluating its importance. The varimax rotation, generally accepted as coming closest to Thurstone's simple structure goal as summarized in Harman [6, p. 98], maintains the independence of the underlying factors.

Results

Results of the factor analysis of the data for Idaho counties are summarized in Tables 1, 2 and 3. Each table is associated with one stage of the three-step outline previously described. Table 1 is the 15-variable correlation matrix which was factor analyzed.

Factor loadings of the variables resulting from the defined factor (principal components) analysis are shown in Table 2. These values are interpeted as correlation coefficients in naming the factors. Table 3 records factor loadings using the varimax rotation procedure [11, p. 474].

Two criteria may be used in evaluating the factor loadings: selection of some minimum "cut-off" value or a selection based on the relative weighting of the loadings [12, p. 8]. The minimum "cut-off" process was used in this study. A value of .50 (signs ignored) was selected, based on other empirical work [5] and statistical significance criteria [14, p. 557]. Thus there are eight significant loadings in Factor 1 of the principal components analysis (Table 2, signs ignored). The variables associated with these loadings are related to agriculture (variables 11, 12, 13), demography (variable 22), land use and ownership (variables 25, 26) and employment (variables 28, 30). Since the majority of the loadings are resource-related Factor I was named GENERAL RESOURCES. The significant loadings in Factor II are associated with fiscal variables 7, 9 and 10 thus it was called FISCAL. One loadings is significant in Factor III, percentage of sales tax revenue, thus the name STATE REVENUE was applied. Factor VI also contained one significant loading, percentage of welfare expenditure, and was named WELFARE.

Factors IV and V illustrate problems that may arise in a factor analysis study. In factor IV, there are two significant loadings, median education level and percentage of basic employment. Each loading has the same absolute value. Since no common term seemed appropriate to describe the two significant variables, Factor IV was identified as EDUCATION-EMPLOYMENT. Factor V contains no significant loadings although it was retained in the factoring process. These problems indicate the need to clarify the interpretation of factor loadings, which was done by rotation. The naming of the factors is based on a descriptive process [13, p. 473].

Since the rotation was orthogonal, evaluation of the factor loadings in Table 3 is the same as Table 2. Although the values changed, significant loadings in Factors II, IV, and VI were associated with the same variables as in the principal components analysis. Thus the names FISCAL, EDUCATION-EMPLOYMENT and WELFARE were retained for these factors.

Factor I contained only 4 significant loadings after rotation, those associated with agricultural sales, percentage of forest land, percentage of unemployment, and lumber as a percentage of manufacturing employment. This factor was renamed FOREST because the higher factor loadings were associated with that resource. Factor III gained one significant loading, percentage of federal land, to go along with percentage of sales tax revenue. This factor was called STATE-FEDERAL to reflect the roles of those exogenous forces. Although Factor IV was not changed under the rotation, Factor V was. Three variables, agricultural acreage, crops as a percentage of agricultural sales, and percentage of the population over 25 were significant. A fourth loading, agricultural sales, approached significance; thus this factor was named AGRICULTURE.

Variance levels in the rotated factor loadings result in ordering of the factors, in descending importance: I (FOREST), II (FISCAL), V (AGRICULTURE), III (STATE-FEDERAL), IV (EDUCATION-EMPLOYMENT), VI (WELFARE).

Conclusions

Development economists, including Clark [2], have shown that change in economic structure is a prerequisite for, as well as a means of measuring, economic development. With respect to rural communities, the structure of the economy and the relative importance of each element within the structure are important in determining development potential and selecting strategies for development. Factor analysis is a statistical method which can satisfy these requirements by suggesting policy targets from an ordered set of factors.

The results of this study indicate that 1) natural resources, 2) fiscal status, 3) role of the state and federal governments, and 4) education level and employment opportunities are the ordered general structural elements in Idaho rural communities. The importance of natural resource - based activities, education and skill training, and employment opportunities have been commonly associated with economic development of rural areas. The impacts of government actions, as evidenced by the fiscal, state-federal and welfare factors, are recognized by community personnel as playing a role in development. This study confirms the importance of these elements in Idaho rural communities and the need to consider them in designing development programs.

The methodology suggests that multivariate indices may be useful in themselves, as contrasted with the use of factor analysis to select individual variables. Regression analysis uses dummy variables to represent location parameters, but their interpretation is ambiguous. With variables serving as location parameters, the multivariate factors represent the structure or mix of those variables in communities. Use of the factors as input variables is an improvement of regression analysis because it identifies more realistic location parameters. However, care must be taken in selecting other variables in the regression framework to avoid possible statistical complications.

TABLE 1. CORRELATION MATRIX USED IN FACTOR ANALYSIS OF IDAHO RURAL COMMUNITY STRUCTURE STUDY[1]

Variable Name	Variable Number	7	8	9	10	11	12	13	17	22	25	26	28	30	37	38
% education expend.	7	1.00														
% welfare expend.	8	.41	1.00													
% health expend.	9	-.73	-.37	1.00												
% police expend.	10	.62	.11	-.65	1.00											
agricultural acreage	11	.08	.13	.10	-.21	1.00										
crops/agric. sales	12	.16	-.03	-.03	.18	.42	1.00									
agricultural sales	13	.12	.16	.09	-.07	.53	.28	1.00								
median education	17	-.01	-.03	.01	-.03	.17	.14	.15	1.00							
% population over 25	22	.08	-.09	-.22	.30	-.36	-.42	-.34	-.39	1.00						
% federal land	25	-.16	.00	-.11	-.19	-.12	-.52	-.36	-.04	.23	1.00					
% forest land	26	-.03	.13	-.05	.07	-.35	-.21	-.52	-.27	.22	.26	1.00				
% unemployment	28	-.10	-.23	.04	-.02	-.33	-.34	-.44	-.27	.30	.34	.52	1.00			
lumber/manu. employ.	30	.04	-.14	-.22	.20	-.31	-.05	-.52	-.28	.22	.27	.64	.45	1.00		
% basic employment	37	-.07	.04	-.15	-.09	-.03	-.10	-.03	-.37	.25	.07	-.08	.07	.09	1.00	
% sales tax revenues	38	.04	-.05	-.04	-.14	.09	-.26	.07	-.07	.22	.44	-.04	.20	.13	.12	1.00

[1]N = 44 counties

TABLE 2. FACTOR LOADINGS USING PRINCIPAL COMPONENTS METHOD, IDAHO RURAL CC UNITY STRUCTURE STUDY

Variable	(var. number)	Factor I	Factor II	Factor III	Factor IV	Factor V	Factor VI
		Factor Loadings of the Six Principal Components					
% education expend.	(7)	-.06	.88*	.13	.14	.05	.08
% welfare expend.	(8)	-.13	.48	.27	.20	.26	-.69*
% health expend.	(9)	-.18	-.86*	-.19	-.12	.03	-.05
% police expend.	(10)	.14	.83*	-.20	-.05	-.24	.25
agricultural acreage	(11)	-.63*	-.04	.28	.07	.49	.18
crops/agric. sales	(12)	-.57*	.24	-.40	-.20	.40	.31
agricultural sales	(13)	-.74*	.05	.33	-.09	.09	.06
median education	(17)	-.44	-.05	-.16	.64*	-.27	.13
% population over 25	(22)	.62*	.19	.27	.29	-.37	.10
% federal land	(25)	.55*	-.22	.43	.47	.08	.01
% forest land	(26)	.69*	.04	.36	.16	.39	-.25
% unemployment	(28)	.71*	-.17	-.08	.05	.16	.17
lumber/manu. employ.	(30)	.69*	.14	-.29	.02	.42	.24
% basic employment	(37)	.20	.01	.45	-.64*	.15	-.09
% sales tax revenue	(38)	.24	-.12	.68*	.24	.15	.44
eigenvalues		3.77	2.64	1.68	1.34	1.15	1.06
portion		25.1%	17.6%	11.2%	8.9%	7.7%	7.1%
cumulative portion		25.1%	42.7%	53.9%	62.8%	70.5%	77.6%

*significant loading ($\geq$ .50)

TABLE 3. FACTOR LOADINGS USING VARIMAX ROTATION, IDAHO RURAL COMMUNITY STRUCTURE STUDY

Variable	(var. number)	Factor Loadings of the Rotated Components					
		Factor I	Factor II	Factor III	Factor IV	Factor V	Factor VI
% education expend.	(7)	-.05	.86*	.02	-.03	.14	.24
% welfare expend.	(8)	-.08	.27	-.03	.05	.06	.91*
% health expend.	(9)	-.06	-.87*	-.11	-.06	.07	-.22
% police expend.	(10)	.08	.87*	-.22	-.04	-.12	-.16
agricultural acreage	(11)	-.33	-.08	.19	-.02	.77*	.12
crops/agric. sales	(12)	-.05	.19	-.44	-.06	.75*	-.20
agricultural sales	(13)	-.66*	.01	.02	.03	.49	.09
median education	(17)	-.29	.00	.05	-.80*	.07	-.03
% population over 25	(22)	.11	.29	.21	.43	-.59*	-.21
% federal land	(25)	.30	-.12	.74*	-.07	-.28	.16
% forest land	(26)	.87*	-.02	-.03	.02	-.15	.25
% unemployment	(28)	.64*	-.08	.28	.12	-.22	-.19
lumber/manu. employ.	(30)	.85*	.19	.11	.14	.04	.16
% basic employment	(37)	-.08	-.02	.10	.82*	.04	.03
% sales tax revenue	(38)	-.02	.06	.87*	.13	.09	-.14
contribution to common variance		2.65	2.52	1.74	1.55	1.95	1.23

*significant loading (≥ .50)

References

[1] Chappell, Jr., V. Glenn. "Determining Economic Sub-Areas with Principal Components Analysis." Contributed paper at the annual meetings of the American Agricultural Economics Association, College Station, Texas, August, 1974.

[2] Clark, Colin. The Conditions of Economic Progress, Third Edition. London, Macmillan and Co., 1957.

[3] Dorf, Ronald J. "An Econometric Analysis of Structural Difference Among Rural Communities and Their Effect on the Location of Manufacturing Plants." P76-1, Staff Paper Series of Department of Agricultural and Applied Economics, University of Minnesota, January, 1976.

[4] Duncan, Marvin and F. Larry Leistritz. "Multivariate Statistical Analysis: Concepts and Economic Applications." Agricultural Economics Miscellaneous Report No. 10, Department of Agricultural Economics, North Dakota State University, October, 1972.

[5] Edwards, Clark, Robert Coltrane and Stan Daberkow. "Regional Variations in Economic Growth and Development with Emphasis on rural areas." Agricultural Economic Report No. 205, USDA/ERS, Washington, D.C., May 1971.

[6] Harman, Harry H. Modern Factor Analysis, Second Edition, Revised. Chicago, The University of Chicago Press, 1967.

[7] Jonassen, Christen T. and Sherwood H. Peres. "Interrelationships of Dimensions of Community Systems - A Factor Analysis of Eighty-two Variables." Columbus, Ohio State University Press, 1960.

[8] Leuck, Dale J. " An Econometric Model of Manufacturing Employment Growth in Rural Tennessee Counties from 1962 to 1976." Southern Journal of Agricultural Economics, December, 1979, pp. 63-67.

[9] March, Richard A. and Dennis K. Smith. "An Analytic Comparison of Multivariate Regional Rural Development Economic Well-Being Indices." SP-76-7, Staff Paper of Department of Agricultural Economics, Virginia Polytechnic Institute and State University, August, 1976.

[10] Marousek, Gerald and Don Blayney. "Identifying Potential Development in Rural Communities - A Statistical Approach." Research Bulletin No. 111, Agricultural Experiment Station, University of Idaho, September, 1979.

[11] Nie, Norman H. and others. Statistical Package for the Social Sciences, Second Edition. New York, McGraw Hill Book Company, 1975.

[12] Reinschmiedt, Lynn and L.L. Jones. "Rural Development Potentials: A Multivariate Approach." DIR 75-1, SP-9, Staff Paper Series of Department of Agricultural Economics, Texas A & M University, August, 1975.

[13] Rummell, R.J. Applied Factor Analysis, Evanston, Northwestern University Press, 1970.

[14] Snedecor, George W. and William G. Cochran. Statistical Methods, Sixth Edition. Ames, The Iowa State University Press, 1967.

Footnote

1. Idaho population increased 16.5% from 1970 to 1976, the 7th highest growth rate among states; only one of the state's counties is designated a metropolitan area (SMSA) by the Census Bureau.