AN ANALYSIS OF RATES OF CHANGE IN COMMUNITY PER CAPITA INCOME BY DISCRIMINANT ANALYSIS

Steve Murray

Statistical methods for estimation, hypothesis testing, and confidence statements are based typically on exact specification of the response variates. In the applied sciences another kind of multivariate problem is common in which an observation must be assigned in some optimal fashion to one of several populations. Classification rules based on an index called the linear discriminant function provide a method for such assignment.

Use of the linear discriminant function is relatively new to regional economics. Previously it has been used in such disciplines as botany to classify a new specimen as belonging to one of several recognized species of a flower, in educational psychology to develop rules for admitting applicants to college programs, in routine banking to aid credit officers in evaluating loan applications, and in agricultural economics to determine producer plans for changes in hog marketings and to identify factors associated with watershed development [3, 4, 5, 7, 8]. The linear discriminant function is used to identify characteristics that distinguish between communities in Arkansas in which per capita incomes are growing rapidly and those in which incomes are growing more slowly. The same set of variables used to account for differences between slow- and fast-growing cities in Arkansas is applied to Oklahoma to test the validity of the model.

Implicit in the development of a successful classification scheme is the conclusion that the variables included will continue in the future to be related as in the past. If the classification variables are merely associated with community growth (or no growth), the results can be used for prediction. If, in addition, the classification variables are judged to cause community growth, the results also can be used for prescriptive purposes. Such information may be valuable to planners and government officials.

THE LINEAR DISCRIMINANT FUNCTION FOR TWO GROUPS

The technique of discriminant analysis is based on the assumption that a linear function $Y = B_1X_1 + B_2X_2 + ... + B_nX_n$ exists which will distinguish between the elements of a population. The discriminant model utilizes coefficients B_1, B_2, ..., B_n chosen in such a way that the ratio of between-group sum of squares is maximized. Therefore, the index Y represents the optimal discriminator between the two groups. Factors X_1, X_2, ..., X_n represent quantifiable determinants of income changes.

Several computational approaches are available to use in the discriminant procedure [6]. In this article the classification criterion developed by the discriminant procedure is determined by the measure of the generalized square, or Mahalanobis distance (denoted as D^2 [X]). It can be based either on the individual within-group covariance matrices or on the pooled covariance matrix. If a chi-square test for homogeneity confirms that no difference exists at the specified level between the covariance matrices of the respective samples, the pooled covariance matrix can be used to develop the classification rule [3]. In the development of the rules which follow, the test showed no difference at the .10 level; thus, the pooled covariance matrices were used.

Some authors [9, p. 97] refer to a test of significance of the discriminant function. The approach developed by the Indian school of statistics is concerned instead with the calculation of misclassification probabilities for the assignment of an individual observation. As the percentage of misclassified observations increases, one deduces that the associated discriminant is more likely to be due to chance. That the rules developed here successfully classify a higher percentage of individuals implies that the function is not random.

Steve Murray is Research Assistant, Department of Agricultural Economics, Oklahoma State University, Stillwater.

*Journal Article J3396 of the Oklahoma Agricultural Experiment Station.

By use of the generalized squared distance, the probability of an observational unit falling into one category or the other can be calculated according to the formula

$$\Pr\{\pi_i|\underset{\sim}{X}\} = \frac{\exp\{-\frac{1}{2} D_i^2(\underset{\sim}{X})\}}{\sum_{j=1}^{2} \{-\frac{1}{2} D_j^2(\underset{\sim}{X})\}} \qquad i = 1, 2$$

Classification results presented in Tables 2 and 3 were obtained through application of this formula, which can be shown to be the same as the usual formulation of the linear discriminant rule.

DATA

Data was obtained by accessing the Ozarks Regional Commission's Regional Resources Management Information System, which provides a consistent set of detailed social and economic information about each incorporated city and town in this region [11]. The data stored on the system are collected by the staff of the multicounty planning agency serving the particular community. The data base was supplemented with income and population data supplied by the Office of Revenue Sharing of the U. S. Department of the Treasury.

THE VARIABLES

Approximately 100 variables measuring the effect of some social, economic, or spatial force within the community were available for use in the model. The variables were sorted into five categories—spatial, labor market, demographic, natural resource, and governmental—to aid in choosing variables for the model. Each category was related to principles of a generally recognized theory of development [10, Ch. 3]. The degree of specificity within the set of candidate variables was broad. Some, such as community population and distance to the nearest major metropolitan area, were very general whereas others, such as the number of freight trains conducting daily switching operations, were narrow. Many variables had sound theoretical bases for inclusion in the predictive model, but econometric models tend to break down if too many variables are included. Thus, results of previous multiple regression analysis along with variance-covariance matrices helped to narrow the group.

The dependent variable for the model is the rate of change in community per capita income between 1969 and 1972. Analysis of determinants of rates are inherently more complex than are analyses of determinants of a steady-state variable [2] and accordingly less work has been done in economic dynamics. The number of variables included in the discriminant function was purposely kept small, and very specific or detailed variables were omitted. The ultimate hypothesis tested was that changes in community per capita income are determined by phenomena captured by the variables of base year city population, base year per capita income, the proportion of the county (in which the city is located) population living on farms, the median educational level, and the dropout rate within the city school system.

Historically, incomes in lagging areas of the United States have tended to increase faster on a percentage basis than have incomes in more prosperous areas although the real dollar gap generally widens [1]. Economic theory suggests that young and marginal workers in depressed areas will be the first to migrate in response to the prospect of better jobs or increased public welfare benefits in cities. Marginal workers who leave may have larger than average families. Community per capita incomes increase because the marginal family no longer holds down the community average. For the model, characteristics including educational levels, dropout rates, and rural residency were used to differentiate communities with large numbers of marginal workers from the more affluent communities.

THE MODEL

The initial discriminant model was developed by examining the income growth processes of every community in Arkansas with a population of 2,500 to 100,000. Communities were separated by quintile—the 14 communities with the slowest growing incomes were assigned to the first quintile and those with the fastest growing incomes were put in the fifth quintile (see Table 1). To sharpen the distinction between communities (and because previous regression analysis had suggested problems in predicting growth rates of communities) the discriminant analysis was applied only to the first and fifth quintiles.

The observations in the slow-growing set were assigned *a priori* to group 1 and the observations in the fast-growing set were assigned to group 2. Posterior probabilities of group membership then were calculated for each observation according to the rule

$$\mathrm{PR}\{\pi_i|\underset{\sim}{X}\} = \frac{\exp\{-\frac{1}{2} D_i^2(\underset{\sim}{X})\}}{\sum_{j=1}^{2} \{-\frac{1}{2} D_j^2(\underset{\sim}{X})\}} \qquad i = 1, 2$$

TABLE 1. DESCRIPTIVE INCOME AND POPULATION STATISTICS FOR 75 ARKANSAS COMMUNITIES BY QUINTILE

Quintile	Variable	Mean	Standard Deviation
1	INCOME_CHANGE (%)	.0468	.0236
	INCOME_1969 ($)	2417	369
	POP_1970 (no. of inhabitants)	11098	8996
2	INCOME_CHANGE (%)	.0758	.0063
	INCOME_1969 ($)	2383	299
	POP_1970 (no. of inhabitants)	12316	15194
3	INCOME_CHANGE (%)	.0982	.0080
	INCOME_1969 ($)	2240	313
	POP_1970 (no. of inhabitants)	10238	9544
4	INCOME_CHANGE (%)	.1209	.0057
	INCOME_1969 ($)	2152	307
	POP_1970 (no. of inhabitants)	15679	19557
5	INCOME_CHANGE (%)	.1645	.0292
	INCOME_1969 ($)	2071	249
	POP_1970 (no. of inhabitants)	5060	2332

Classification results are presented in Table 2. Twenty-seven of 28 cities classified correctly by use of the information from the five variables.

TABLE 2. CLASSIFICATION OF PER CAPITA INCOME GROWTH RATES OF ARKANSAS CITIES

Classification by discriminant function	Slow-growing	Fast-growing	Total
Slow-growing (number of observations)	13	1	14
%	92.86	7.14	100.00
Fast-growing (number of observations)	0	14	14
%	0.00	100.00	100.00
Total	13	15	28
%	46.43	53.57	100.00

Alternatively, the usual form of the linear discriminant function could have been used to classify the communities. The linear form is

$$279.4644 = -.0003\ \text{POP 70} - 0.0163\ \text{INCOME} + 281.9381\ \text{FARM} + 10.9889\ \text{DROPOUT} + 26.3678\ \text{EDUCATION}$$

To apply the rule for a sample community, observations on the five variables are used in the right side of the formula. If the resulting value is less than 279.4644 the community is placed in the slow income growth category. If the value is greater, the community is in the fast growth category.

Ability to generalize the results was checked by performing a discriminant analysis on the set of slowest growing and fastest growing cities in Oklahoma with the same five variables used to develop the classification rule. Classification results are presented in Table 3. Twenty-four of 34 (71 percent) Oklahoma communities were classified correctly by use of the same set of variables as was used for Arkansas.

TABLE 3. CLASSIFICATION OF PER CAPITA INCOME GROWTH RATES OF OKLAHOMA CITIES

Classification by discriminant function	Slow-growing	Fast-growing	Total
Slow-growing (number of observations)	11	6	17
%	64.71	35.24	100.00
Fast-growing (number of observations)	4	13	17
%	23.53	76.47	100.00
Total	15	19	34
%	44.12	55.88	100.00

The results of the study suggest that discriminant analysis can be used to identify characteristics associated with community per capita income growth. Less success would be expected in classifying communities in the middle three quintiles.

The results confirmed the hypothesis that small rural communities with an undereducated population—all characteristics associated with a declining economy—are likely to undergo greater percentage increases in per capita income than are the more prosperous cities. The results are consistent with descriptive statistics shown in Table 4.

The discriminant function presented here suggests that percentage increase in per capita income is associated negatively with base year income and city size; it is associated positively with the proportion of the county population living on farms, the school dropout rate, and median educational levels. Persons wishing to use these results for prescriptive purposes would be advised to raise the educational level of the population.

Government officials might use the coefficients for predictive purposes in formulating policy. For example, the coefficients suggest that smaller cities are associated with slower income growth. Thus, a public works project designed specifically to accelerate income might be more appropriately placed in the smaller community. Other coefficients should be interpreted in the same manner.

TABLE 4. MEAN VALUES OF VARIABLES INCLUDED IN LINEAR DISCRIMINANT MODEL

Variable	Arkansas Income Growth Rate Slow	Arkansas Income Growth Rate Fast	Arkansas Difference Between Groups	Oklahoma Income Growth Rate Slow	Oklahoma Income Growth Rate Fast	Oklahoma Difference Between Groups
1970 City Population	11098	5060	6038	12528	12132	396
City Per Capita 1970 Income ($)	2417	2071	346	2538	2351	187
County Farm Population (%)	6.95	13.93	-6.98	11.37	14.74	-3.37
Dropout Rate (%)	3.22	4.67	-1.45	1.35	1.22	.13
County Median Educational Level (years)	10.09	8.65	1.44	11.15	10.84	.31

REFERENCES

[1] Beale, Calvin. "Quantitative Dimensions of Decline and Stability Among Rural Communities," *in* Larry R. Vining, ed., *Communities Left Behind: Alternatives for Development,* Ames: The Iowa State University Press, 1974, pp. 3-21.

[2] Chiang, Alpha C. *Fundamental Methods of Mathematical Economics,* 2nd ed., New York: McGraw-Hill Book Co., 1974.

[3] Cox, P. Thomas and Daniel D. Badger. "Factors Contributing to Success of Upstream Watershed Development in Oklahoma," Oklahoma State University Processed Series P-578, November 1967.

[4] Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics,* Volume 7, 1936, pp. 179-188.

[5] Johnson, R. Bruce and Albert R. Hagan. "Agricultural Loan Evaluation with Discriminant Analysis," *Southern Journal of Agricultural Economics,* Volume 5, No. 2, December 1973, pp. 57-62.

[6] Morrison, Donald F. *Multivariate Statistical Methods,* 2nd ed., New York: McGraw-Hill Book Co., 1976.

[7] Raikes, Ronald and Michael Trampel. "An Analysis of Producers' Year-to-Year Changes in Slaughter Hog Marketings," *Southern Journal of Agricultural Economics,* Volume 8, No. 2, December 1976, pp. 95-101.

[8] Sexton, Dennis W. and Roy D. Goldman. "High School Transcript as a Set of 'Nonreactive' Measures for Predicting College Success and Major Field," *Journal of Educational Psychology,* Volume 67, No. 1, February 1975, pp. 30-37.

[9] Titner, Gerhard. *Econometrics,* New York: John Wiley and Sons, Inc., 1952.

[10] Tweeten, Luther and George Brinkman. *Micropolitan Development,* Ames: The Iowa State University Press, 1976.

[11] Willis, T. W. and Associates. *Ozarks Regional Commission Community Development Profile System: Uses and Applications,* Farmville, North Carolina, 1970.