R E P O R T R E S U M E S

ED 014 149 EA 000 867

POPULATION AGE DISTRIBUTIONS AND PUBLIC EDUCATION EXPENDITURES.

BY- HENDRIX, VERNON ALKIN, MARVIN

PUB DATE FEB 67

EDRS PRICE MF-$0.25 HC-$0.64 14P.

DESCRIPTORS- POPULATION DISTRIBUTION, *AGE DIFFERENCES, *COMMUNITY SUPPORT, *FINANCIAL SUPPORT, ELEMENTARY SCHOOLS, SECONDARY SCHOOLS, RESOURCE ALLOCATIONS, *MODELS, STATISTICAL ANALYSIS, EXPENDITURES, *PUBLIC SCHOOLS, NEW YORK CITY,

A LINEAR MODEL WAS DEVELOPED TO DETERMINE THE RELATIONSHIP BETWEEN THE AGE CHARACTERISTICS OF A COMMUNITY AND THE AMOUNT OF LOCAL PUBLIC FUNDS ALLOCATED PER PUPIL IN ELEMENTARY AND SECONDARY SCHOOLS. THE MODEL TESTED THE HYPOTHESIS THAT THE LEVEL OF LOCALLY RAISED EXPENDITURES FOR EDUCATION IN EACH STATE COULD BE PREDICTED BETTER WITH KNOWLEDGE OF THE PERCENTAGE OF EACH STATE'S TOTAL POPULATION IN EACH OF NINE AGE BRACKETS THAN WITHOUT THIS INFORMATION. THE STUDY WAS RESTRICTED TO FORTY-EIGHT STATES FOR THE YEARS 1930, 1940, 1950, AND 1960. THE INDEPENDENT VARIABLES CONSISTED OF THE PERCENT OF THE POPULATION OF EACH STATE WITHIN NINE SELECTED AGE INTERVALS FOR EACH OF THESE YEARS. THE DEPENDENT VARIABLE WAS DEFINED AS LOCALLY DERIVED CURRENT EXPENDITURES PER PUPIL IN ADA FOR ELEMENTARY AND SECONDARY SCHOOL INSTRUCTION. THE MODEL, WITH GROSS ECONOMIC CHARACTERISTIC VARIABLES AND COHORT EFFECT VARIABLES CONTROLLED, YIELDED A MULTIPLE R OF .8194. THUS, THE INDEPENDENT VARIABLES EXPLAINED ABOUT 67 PERCENT OF THE VARIANCE IN THE DEPENDENT VARIABLE. RESTRICTED MODELS WERE ALSO DEVELOPED BY OMITTING PARTICULAR VECTORS FROM THE MODEL. PUBLIC EDUCATIONAL EXPENDITURES WERE FOUND TO BE RELATED TO POPULATION AGE CHARACTERISTICS OF THE STATES. THIS PAPER WAS PRESENTED AT THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION (NEW YORK, FEBRUARY, 1967). (HW)

Population Age Distributions and Public Education Expenditures

Vernon Hendrix and Marvin Alkin*

(A paper presented at the American Educational
(Research Association, February, 1967, New York)

Introduction

An integral part of the theory of public finance is the necessity for providing an opportunity to express preferences for the amount of funds to be allocated to the public sector of the economy, and to determine true preferences within that sector. However, the various kinds of taxes utilized for different functions within the public preferences than others. Thus, there is a danger that citizens may express their general displeasure for what they feel to be high taxes in that segment of the public sector which is most accessible and where views are most easily expressed. Very often, that accessible segment of the public sector is property taxes and the local school mechanism.

This peculiar susceptibility of locally derived funds for public education accents the importance of recognizing the factors which may be considered as the determinants of local educational expenditures. A better understanding of these determinants is invaluable in making judgments as to the adequacy of the local financial mechanism in a changing society and the feasibility of specific changes in the support structure.

There has been some attention directed to examining the determinants of educational expenditures by several scholars including Hirsch, James, and Miner.[1] Of the variables which may be considered as determinants, most of the explained variance is attributed to differences in income, years of schooling of parents, proportion of pupils in secondary grades, etc.

* This paper was prepared jointly by professors Hendrix and Alkin. Professor Hendrix's name is listed first as he read the paper at the conference.

One set of variables describing communities which has not been examined systematically[2] and which bears further attention is age characteristics. Is there a relationship between the per cent of the community in each age category and the amount of local public funds which has been expended per elementary and secondary child in ADA?

The Study

The primary purpose of this study is to examine the nature and extent of relationships between population age distributions and locally derived financial support for public elementary and secondary education. The study should be considered as a continuation of the inquiry into the determinants of educational expenditures. Specifically, our major hypothesis is that the level of locally raised expenditure for education in each state can be predicted better from knowledge of the percentage of each state's total population in each of nine age brackets than without this information. This study is considered as a pilot endeavor--an initial attempt at probing a series of variables which have not previously been systematically considered as determinants of school expenditures. It is expected that any variables found to be of interest would be included in a later and more comprehensive model.

In developing the research design, several immediate questions needed to be answered: What point (or points) in time should be selected for performing the analyses? What level of aggregation should be chosen for analyzing the local financial data (e.g., districts, counties, states, etc.)? What particular variables might be selected to operationally define the conceptualized dimensions of concern (persons of a given age, persons within age brackets, local tax effort, etc.)? What other control variables might be related to the dependent variable and how should they be operationally defined.

If one point in time is selected only a limited number of subjects are available. Also, certain variables of interest that would conceivably operate over a period of time could not be included. While attempting to make the results as general as possible, abiding by the constraints of time and money imposed on the study, and insuring the comparability of operationally defined variables, it was decided to use states as the basic unit and rely upon U.S. government publications for data, specifically... the decennial census publications. Given these concerns and limitations, the available operational definitions then were restricted to forty-eight states for the time points of 1930, 1940, 1950, and 1960.

The independent variables for each of these years consisted of the percent of the population of each State within nine selected age intervals. The indicated age groups were chosen from those available in the census publications to correspond with the ten-year intervals and to permit separate analysis for particular age groups. These groups were pre-school age, elementary age, secondary age, and "65 years old or over" which are represented approximately by the first three and the ninth age groups respectively. (See Figure I)

The dependent variable was defined as locally derived, current expenditures per pupil in average daily attendance (ADA) for public elementary and secondary school instruction. The locally derived funds were chosen as the desired criterion measure because it was considered to be a more accurate measure of demand, or willingness to support public education, than the total current expense data which includes state and federal apportionments. The dependent variable data available for any given state can be thought of for purposes of analysis, as a nine by four matrix. This is illustrated in Figure I where the data for the four years are represented in columns and the data foreach age group are represented in rows.

As the problem developed, general linear models seemed most appropriate for analysis of the data.[3] The nature of these models may be most conveniently described in terms of vectors. Since there are 192 data points for the dependent variable (48 states in each of four years) each vector would contain 192 elements. For purposes of illustration these have been collapsed in Figure 1, to four elements. This represents the data for one state. The four elements in the first vector then, consist of the dependent variable data points for state "k" for each of the four years. Vectors three through eleven contain data on the percent in each age group. E.g., the first element in vector three indicates the percent of the population less than five years in age in 1930 for state "k." Vector four would have in its first element the percent of population from five to fourteen years in age for 1930 in state "k." The age groups are then conceptualized as nine independent variables.

Several control variables immediately suggest themselves. There are obviously changes in the economic profile of the nation from one year to another as well as changes in the relative national position of Education as compared to expenditures in other "industries." In short, it is completely safe to assume that there have been significant gross increases in the mean of the dependent variable for the 48 states from 1930 to 1940, 1940 to 1950, etc. Therefore, in order to control for this, four additional vectors were generated. The elements in vector twelve then contain "one" if the associated element in the dependent variable vector is for 1930, otherwise they are zero. The elements in vector thirteen similarly contain "one" if the associated elements in the dependent variable vectors are from 1940, otherwise they are zero. In this way, the gross economic characteristics associated with each of the four years and which relate to the values of the dependent variable may be controlled, thus permitting independent assessments of the relationships of each independent variable with the dependent variable.

Further examination of Figure 1 suggests another control variable which can be designated as the "cohort effect." For example, accepting the roughness of the data due to the categorizations, the persons constituting the fifth age group (25-34 years of age) in 1930 constitute the sixth age group (35-44 years of age) in 1940, the seventh age group in 1950, etc. In other words, if there are certain peculiarities about the persons 25-34 years of age in 1930, which are related to the dependent variable, these might affect the dependent variable in later years. Omitting the first five year age group, and collapsing age groups three and four to correspond with the ten-year intervals, ten cohort groups can be identified. This is indicated in Figure 3.

Examination of Figure 3 indicates that any effects attributable to cohorts one through four cannot be accounted for since these cohort groups are present in all four years and thus are represented in all 192 data points. Similarly, it would be useless to attempt to account for any effects attributable to cohort groups 7 and 10 since these are present only in 1930 and 1960 respectively. It is reasonable to consider cohort groups 5, 6, 8, and 9. In Figure 2, vector 16 indicates the presence of cohort group 6, in the data points of a given year vector 17 indicates the presence of cohort group 5, vector 18 indicates the presence of cohort group 8, and vector 17 indicates the presence of cohort group 9. Given the limitations of the current study, these are the only cohort groups that can be included in the model.

As indicated later, the addition of these cohort vectors does not increase the efficiency with which the dependent variable vector can be predicted, since each of the cohort vectors are linear combinations of the already present year vectors. In fact, the maximum amount of prediction, given the indicated

variables, could be obtained with the nine age group vectors and any three of the year vectors. In order to allow for interactions among the variables, the model (Full Model A) includes all year and cohort vectors.

The authors wished to include in the model some variables generally regarded as indicative of socio-economic status. Most of the commonly used measures were not available for all four years. Thus, while not totally satisfied with the variables available, two additional vectors were formed to be included in a second model, Full Model B. These represented other variables to be controlled and were, respectively, percent owner occupied housing and percent of the population in urban areas.[4]

Results

The analysis of the two full models yielded quite interesting results. Full Model A has as its predictor variables from nine age-group categories, four vectors representing membership in a year and four cohort vectors. This model yielded a Multiple R of .8194. (Table I) Thus, the independent variables "explain" about 67% of the variance in the dependent variable. Full Model B, which utilized the same variables as the preceding model with the addition of two covariates, had a Multiple R of .8286, with 69% of the variance explained by the independent predictors.

Restricted models were developed by omitting particular vectors. This is the same as testing the hypothesis that the coefficients for the omitted vectors are equal to zero. The extent to which the restricted model produced a smaller multiple R^2 may be used, with appropriate degrees of freedom, to obtain an F statistic. A number of restricted models have been formed from Full Model A to demonstrate this point (see Table I). These indicate, for Full Model A, that 20% of the variance is accounted for by the age vectors (restricted models 4, 5, and 6) and that 3% of the variance is accounted for by the year and/or cohort vectors (restricted models 1, 2, and 3).

It is proposed that statistical significance tests for the coefficients are not necessary although they are indicated in Table I for the multiple R^2's. In this study, a sample has not been drawn from a population in the usual sense of hypothesis testing. Rather, the population has been analyzed. (Comparable data are not available for Hawaii and Alaska, thus they are not part of the population.) It might be argued that the population is actually infinite since there can be an infinite number of time points at which measurements are taken. Since these time points are not selected randomly and since the mean effects for different time points can be statistically controlled, this would not seem to be a debilitating argument. Also, as here formulated, tests for each age group are not possible.

Results

The large value of the Multiple R in each of the models summarized in Table I is an important finding. The hypothesis of the study--that locally derived public educational expenditures are related to the age characteristics of States--seems to be confirmed. At any rate, the age variables seem to be worthy of far greater attention in future research on the determinants of educational expenditures.

In order to form hypotheses about the relative importance of each age group category as a predictor of the criterion variable, the regression coefficients were examined for each of the models. In the case of the age variables of this paper, each regression coefficient represents the dollar change in locally derived public school expenditures which would be associated with a one percentage point increase in that age category.

Percent population aged 5-19, 25-34, and 45-54 appears to be negatively associated with local support. This is not surprising since an increase in school aged persons (5-19) and persons aged 25-34 (their parents) is likely to lower the ADA derived financial variable. The 45-54 age group normally does not have children in school and thus are less willing to support public education.

Percent population aged 0-4, 20-24, 35-44, and greater than 55 appears to be positively associated with local support although these are more moderate relationships.

Summary

More rigorous study is indicated. The separation of "year" and "cohort" variables, the inclusion of additional control variables, and a refinement of the criterion variable would be obvious points of departure. At least, before age can be dismissed as an independently functioning determinant of public educational expenditures, more refined examinations are required.

FIGURE I

Representation of Data Points for State k

Variable	Year 1930	1940	1950	1960
Percentage 0-4	X 11k	X 12k	X 13k	X 14k
5-14	X 21k	X 22k	X 23k	X 24k
15-19	X 31k	X 32k	X 33k	X 34k
20-24	X 41k	X 42k	X 43k	X 44k
25-34	X 51k	X 52k	X 53k	X 54k
35-44	X 61k	X 62k	X 63k	X 64k
45-54	X 71k	X 72k	X 73k	X 74k
55-64	X 81k	X 82k	X 83k	X 84k
65+	X 91k	X 92k	X 93k	X 94k
Local Expenditures	Y 1k	Y 2k	Y 3k	Y 4k

X_{ijk} Y_{jk}

i = (age groups) 1.----9

j = (years) 1.--4

k = (states) 1.----48

FIGURE 2

Vector Model of Data for State k

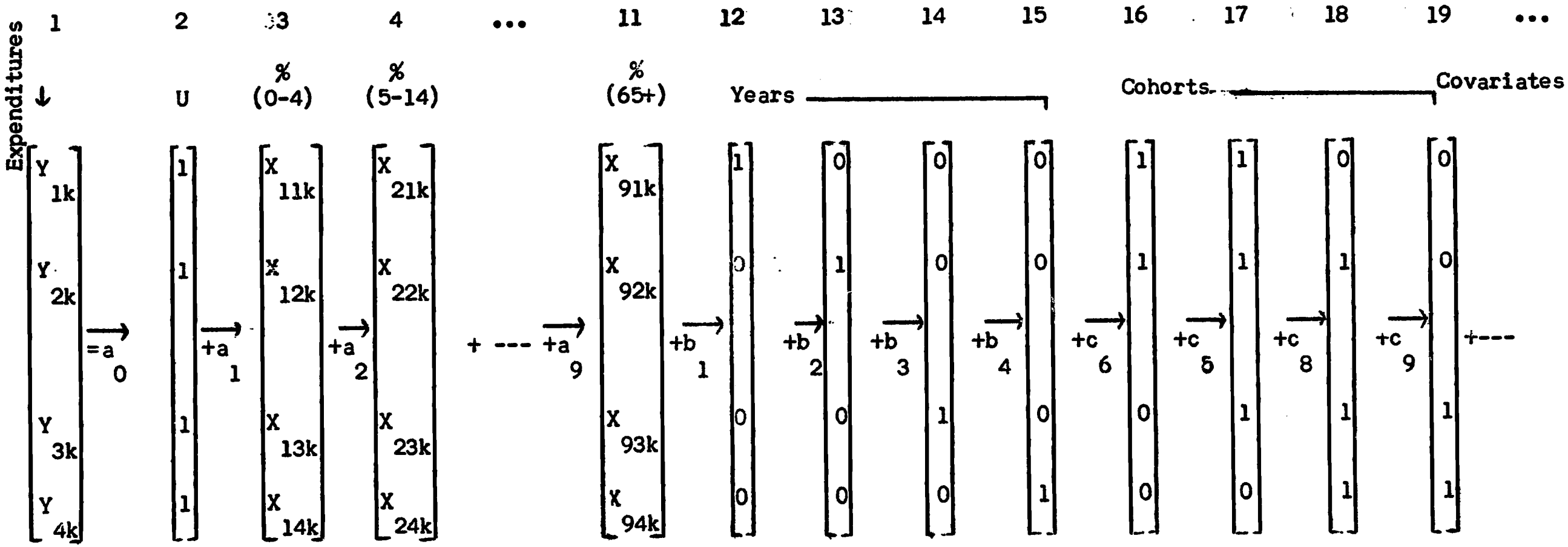

FIGURE 3

Representation of Cohort Groups

Ten Year Age Groups	Years			
	1930	1940	1950	1960
Age 5-14	C 1	C 8	C 9	C 10
Age 15-24	C 2	C 1	C 8	C 9
Age 25-34	C 3	C 2	C 1	C 8
Age 35-44	C 4	C 3	C 2	C 1
Age 45-54	C 5	C 4	C 3	C 2
Age 55-64	C 6	C 5	C 4	C 3
Age 65+	C 7	C 6	C 5	C 4

TABLE I

Full and Restricted Model Analyses

	Predicter Vectors Present						
	Age (9)	Year (4)	Cohort (4)	Mult R	Mult R^2	R^2 Difference From Full Model	F Ratio and d f
Full Model A	X	X	X	.8194	.6715	---	30.67** 12/180
Restricted Model A_1	X	X		.8194	.6715	0	---
Restricted Model A_2	X		X	.8194	.6715	0	---
Restricted Model A_3	X			.8005	.6408	.0307	5.61** 3/180
Restricted Model A_4		X	X	.6822	.4655	.2060	14.71** 8/180
Restricted Model A_5		X		.6822	.4655	.2060	14.71** 8/180
Restricted Model A_6			X	.6822	.4655	.2060	14.71** 8/180

**Significant at .01 level

TABLE II

Age Variables - R^2 Differences and

Regression Coefficients for Each Full Model

Age Category	R^2 Difference when Deleted from Full Model		Regression Coefficient in Full Model	
	A	B	A	B
1 (0-4)	.0030	.0055	12.5	17.3
2 (5-14)	.0063	.0050	-16.7	-14.8
3 (15-19)	.0039	.0035	-27.7	-27.0
4 (20-24)	.0001	.0000	5.2	0.2
5 (25-34)	.0058	.0066	-20.3	-21.9
6 (35-44)	.0013	.0012	6.9	6.6
7 (45-54	.0064	.0059	-15.5	-15.1
8 (55-64)	.0009	.0012	5.5	6.5
9 (65+)	.0002	.0006	1.4	2.9

Notes and References

1. Werner Z. Hirsch, "Determinants of Public Education Expenditures," National Tax Journal, March 1960, pp. 29-40.

   H. Thomas James, Wealth Expenditures and Decision-Making in Education, Stanford University Press, Stanford, California, 1963.

   Jerry Miner, Social and Economic Factors in Spending for Public Education, Syracuse University Press, Syracuse, New York, 1963.

2. Single age categories have been examined by Miner, op. cit; and Sherman Shapiro, "Some Socioeconomic Determinants of Expenditures for Education; Southern and Other States Compared," Comparative Education Review, October 1962, pp. 160-166.

   Hirsch, op.cit. used the variables "high school pupils in ADA as a per cent of all pupils in ADA" and "number of pupils in ADA in public primary and secondary schools."

3. Bottenberg, Robert A., and Ward, Joe H. Jr., Applied Multiple Linear Regression, Lackland Air Force Base, Texas; 6570th Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, PRL-TDR-63-6, March 1963.

4. The "old" definition of urban area was used in order to obtain comparable data in each of the four years 1930, 1940, 1950, and 1960.