TEST OF HYPOTHESIS

The working hypothesis which was enunciated in the preceding pages will now be tested in the light of available national data. In many respects, availability more than appropriateness of data determines their application. Time series data for selected years during 1900–1958 will be used and subjected to multiple correlation and regression analyses.

By insisting on intervals of at least 2, and in many instances 7 or 8 years, it is hoped that much otherwise existing serial intercorrelation is eliminated.[24] But with only data for 17 years at hand, one is somewhat handicapped in including all the independent variables that appear appropriate. Care must be exercised that no unduly large number of degrees of freedom is lost. Because of this consideration, it was decided to use as dependent variable, daily total current expenditures plus debt service per pupil in average daily attendance. In a sense, the dependent variable is already adjusted for possibly two independent variables—number of pupils in average daily attendance and length of school year. While there are some risks in taking such a step, they were considered outweighed by the advantages of retaining more degrees of freedom.

In order to shed light on such questions as, "Why has the public education sector had such large expenditure increases?" and "What are the underlying relationships for the near term?" two working hypotheses have been enunciated and empirically tested. The first hypothesis is as follows:

X_{1a}—daily total current expenditure plus debt service for public primary and secondary education per pupil in average daily attendance is a function of—

X_2—percent of public high school enrollment relative to total public school enrollment;

X_3—percent of pupils (5 to 19 years old) in average daily attendance living in urban areas;

X_4—average annual salary for member of instruction staff; and

X_6—number of principals, superintendents, and consultants per 1,000 pupils in average daily attendance.

In brief,

$$X_{1a}=f\,(X_2,\ X_3,\ X_4,\ X_6).$$

The following multiple regression equation was obtained for 17 selected years during 1900–1958:

$$X'_{1a}-0.164999+\frac{0.002067\ X_2}{(0.0257)}-\frac{0.005288\ X_3}{(0.1069)}+\frac{0.000441\ X_4}{(0.9895)}+\frac{0.022391\ X_6}{(0.0271)}.^{25}\quad(1.1)$$

The coefficient of multiple determination, adjusted for degrees of freedom lost—$R^{*2}_{1.2346}$—is 0.998 and is statistically highly significant. Thus, about 99.8 percent of the variation of daily total current ex-

[24] Although no detailed checks were made, it appears that also the other two important assumptions underlying the method are met. Thus, the values of the dependent variable of the population appear to be normally distributed around the least square line, and also the standard deviations of these normal population distributions appear about equal.

[25] The figures in brackets are partial correlation coefficients. Since there are 17–5 or 12 degrees of freedom, coefficients are statistically significant at an α of 0.05, when they are larger than 0.532.