

Fitting the production function to the historical data

The variables to be used to derive a production function for the American economy for the period 1909–58 now are complete. The six variables are shown on chart IV. The equation was fitted by the classic method of least-squares to obtain the regression of output (Oa) on the other variables, the equation being initially in the form given at the end of chapter II with modifications as will be seen below. The data were first translated into logarithms and punched on cards suitable for feeding as input to an IBM 650 electronic computer.

Seven separate runs of the data were made on the computer, each run yielding a correlation matrix, regressions between different combinations of the variables and various statistical measures such as averages, standard deviations, R^2 , and a computation of the residuals for selected equations together with the Durban-Watson statistic to test for autocorrelation.

This procedure could be used only because of the generosity and cooperation of the Bureau of Labor Statistics and the Board of Governors of the Federal Reserve System, who made time available on their IBM 650 computers.

Two difficulties in finding the structural relationship involved in the aggregate production function have already been noted: a high degree of correlation of the dependent variable (Oa) and four of the dependent variables (Lp , K/Lp , k , and X) with the time trend (t); and the aggregation problems arising from combining data for a large number of enterprises with varying characteristics into a simple model for the economy as a whole. In addition, the usual complexities associated with time series of economic data had to be dealt with in the analysis. As is well known, in time series the data for each year are not independent observations arising from random samples of the unknown universe but instead values for each year depend on what happened in the preceding year or years. Furthermore, it was clear that both the variables and the equations were subject to error so that the structural parameters could not be determined accurately simply by a direct fit by the least-squares method.

The procedure evolved to overcome these familiar problems was as follows:

(1) Seven successive runs of the data were made on the IBM 650, with several forms of the equation being fitted on each run.

(2) After each run the coefficients for each variable in each form of the equation were tabulated, along with R^2 , Durbin-Watson statistic, σ 's and r 's between variables; and charts were prepared showing—

(a) actual and computed values for the more promising equations;

(b) scatters of residuals from the equations with some or all of the variables; and

(c) interrelationships between values of the coefficients in the different equations—particularly the dependence of the coefficients of the other variables on the value of the coefficient for time (t).

(3) The several values of the coefficients for each variable found on each run were tested in the light of—

(a) their standard errors;

(b) results of other research reported in the literature;