Volume Title: Economics—A Half Century of Research 1920–1970

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume ISBN:

Volume URL: http://www.nber.org/books/annu70-1

Conference Date:

Publication Date: 1970

Chapter Title: Staff Report: Econometrics and Measurement Methods

Chapter Author(s): Various

Chapter URL: http://www.nber.org/chapters/c13559

Chapter pages in book: (p. 107 - 111)

smaller countries compared with the Soviet system relate to the greater relative importance of the financing of foreign trade and the provision of credit to households and to the nonsocialized sector of the economy. Soon after its adoption, the Russian system began to be modified to fit the particular needs of each of the smaller countries embarked upon a program of economic reforms. Credit became one of the main instruments for channeling funds into investment and for achieving a more decentralized and flexible system of production and distribution responding to demands by ultimate consumers. Part III, then, will be a comparative study of the credit systems of the smaller socialist countries. It will show significant differences in policies and techniques among these countries, as well as the way in which the standard system had been adapted and in some cases developed beyond the Soviet example. A separate chapter will deal with foreign monetary relations and operations, including the role of the International Bank for Economic Cooperation and the "transferable ruble."

Yugoslavia, which originally had also slavishly copied the Russian system, has gradually evolved a system which is geared to a competitive socialist market economy. It pioneered some of the ideas embodied in the economic reforms now being undertaken by its neighbors. Because banking and credit in Yugoslavia can no longer be adequately described as just a variant of the Soviet protoype, a separate chapter on Yugoslavia will conclude Part III.

Part IV will discuss the role of credit in the economic reforms in Eastern Europe. A significant part of these reforms concerns the use of financial incentives and the rechanneling of financial flows through the use of credit in preference to the financing of investment by nonreturnable grants. Developments up to the end of 1969 will be covered.

George Garvy

9. ECONOMETRICS AND MEASUREMENT METHODS

Analysis of Long-Run Dependence in Time Series: The R/S Technique

This study is concerned with testing out and improving R/S analysis,[1] a new and extremely promising statistical technique that can be used either to test for the presence of very long-run nonperiodic statistical dependence, or to define and estimate the intensity of such dependence. Very long-run nonperiodic dependence manifests itself by the presence of "cycles," clear cut but of variable periodicity, superimposed upon a variety of "variable trends," "slow cycles," and "long swings"; where the slowest swing period is roughly of the order of magnitude of the total available record. Such behavior is well-known to be characteristic of many economic records.

[1] R/S symbolizes cumulative range divided by the standard deviation.

Since the sources of the R/S analysis are not readily accessible,[2] the basic definitions must be repeated. Let $X(t)$ be a stationary random function with $\sum_{\hat{u}=1}^{t} X(u)$ denoted by $X^*(t)$. For every value of d (called the lag) one defines

$$R(t,d) = \max_{0 \le u \le d} \{X^*(t+u) - X^*(t) - (u/d)[X^*(t+d) - X^*(t)]\} - \min_{0 \le u \le d} \{X^*(t+u) - X^*(t) - (u/d)[X^*(t+d) - X^*(t)]\}$$

and

[2] Two articles by myself and J. R. Wallis: "Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long-Run Statistical Dependence," *Water Resources Research*, October 1969, pp. 967-988; and "Computer Experiments with Fractional Gaussian Noises," *Water Resources Research*, February 1969, pp. 228-267.

$$S^2(t, d) = d^{-1} \sum_{u=1}^{d} X^2(t+u) - d^{-2}\left[\sum_{u=1}^{d} X(t+u)\right]^2$$

and one forms the expression

$$\text{Ros}(d) = E[R(t, d)/S(t, d)].$$

R/S is a very useful statistic because the dependence of the function Ros(d) on the lag happens to separate the effects due to the marginal distribution of $X(t)$ and the effects due to the presence or the absence of long-run statistical dependence.

When the variables $X(t)$ are statistically independent, one has Ros(d) $\sim Cd^H$, with $H = 0.5$ and C a constant; $\sim$ designates a relation valid asymptotically. When $X(t)$ is a Markov process or a more general finite autoregressive process, the relation Ros(d) $\sim Cd^{0.5}$ still holds asymptotically, but the value of C is different and the asymptotic behavior is more slowly attained than in the case of independence. Independent, Markov, and finite autoregressive processes are all such that their values at instants sufficiently apart in time are near independent. For all those processes one has: Ros(d) $\sim Cd^{0.5}$, where C is affected by the precise rule of interdependences, but the exponent is $H = 0.5$ irrespective of the marginal distribution of $X(t)$. H is the same whether $X(t)$ is Gaussian or long tailed, including the cases where $X(t)$ has infinite variance.[3]

But the law Ros(d) $\sim Cd^{0.5}$ fails for random processes that generate sample functions characterized by slow swings, because for such processes the interdependence of values very far apart cannot be neglected. In the simplest cases, one has Ros(d) $\sim Cd^H$, where H lies between 0.5 and 1. In such cases, the value of H can be used to measure the intensity of long-run dependence, namely, the degree of "tendency to slow swinging."

When $X(t)$ is an empirical time series of length T, the definitions of $R(t, d)$ and $S(t, d)$ remain meaningful and the expression Ros $(d) =$

$$(T\text{-}s) \sum_{t=1}^{T\text{-}s} R(t, d)/S(t, d)$$ can be considered an

estimate of Ros(d). R/S testing consists in testing whether or not the departure of Ros(d) from Ros^ $(d) \sim Cd^{0.5}$ is statistically significant. R/S estimation consists in estimating from Ros (d) the value of the exponent H that best represents Ros^ (d) in the form Cd^H.

One of the main weaknesses of conventional econometrics has been that its tools lose part or all of their validity when applied to time series whose variance is very large or infinite. The robustness of the statistic Ros(d) with respect to the marginal distribution is therefore extremely valuable. The main thrust of this study is to perfect R/S analysis for small samples and to apply it to an increasing variety of time series.

Benoit B. Mandelbrot

Analysis of Time Series

During the first few months of my tenure as a postdoctoral Research Fellow at the Bureau, I have completed two papers dealing with the analysis of economic time series. The first is "Spectral Analysis and the Detection of Lead-Lag Relations." This paper is concerned with invalid attempts by some economists to infer timing relationships between pairs of economic series directly from phase statistics calculated from the cross spectrum of the series. The paper points out fundamental differences between the engineering and economic definitions of lead and lag that have caused some confusion in the economist's interpretation of phase statistics. Assumptions about the model linking the time series play an essential role in the correct interpretation of phase statistics. This conclusion is illustrated with several explicit

[3] B. Mandelbrot, "The Variation of Certain Speculative Prices," *Journal of Business*, October 1963, pp. 394-419.

models. The stringent conditions required for the existence of a simple relationship between phase statistics and the economist's concept of lead and lag are briefly discussed.

The second paper is "Dynamic Equivalents of Distributed Lags." Much econometric work using distributed lags starts with loose qualitative notions that, in some manner, the effect of one variable on another is spread over time. Without further development, this approach sometimes leads economists to estimate lagged structures without much attention to dynamic considerations that make theoretical sense. This paper argues that qualitative characteristics of distributed lags can often be represented by functions that are mathematically equivalent to simple dynamic mechanisms (such as linear differential equations). These may be more fruitful for further theoretical and empirical work than the initial distributed lag formulation. Hence a careful analysis of distributed lags and equivalent dynamic systems may be a useful approach to developing better dynamic models in economics. After an introductory section, the paper discusses the formal relationship between distributed lags and other linear dynamic systems. A third section makes use of a simple example to illustrate theoretical advantages that may arise from analyzing a distributed lag system by some dynamically equivalent system. In the final section, several examples are further developed to indicate the usefulness of this approach. If distributed lags are constructed from exponentials (possibly complex) and polynomials of time, the equivalent dynamic forms reduce to differential equations (difference equations in discrete time) that may have relatively simple theoretical interpretations. Lags generated by so-called "rational polynomial generating functions" are equivalent to this class. Hence this analysis has some important implications for interpreting distributed lag coefficients of dynamic structures estimated from completely ad hoc rational polynomial generating functions.

John C. Hause

Papers on Statistical and Economic Methodology

Provisional plans have been made to publish as NBER Technical Papers two collections of papers which I wrote while a Research Fellow at the Bureau. The first collection will be on multicollinearity and measurement errors and will include the following papers (two of which have been previously published as journal articles):

1. "Multicollinearity in Regression Analysis: An Experimental Evaluation of Alternative Procedures," read at the Joint Statistical Meetings of the American Statistical Association (Section on Physical and Engineering Sciences) and the Biometric Society in August 1969
2. "On Multicollinearity in Regression Analysis: A Comment," published in *Review of Economics and Statistics,* September 1969
3. "On the Correlations Between Estimated Parameters in Linear Regression"
4. "A Note on Regression on Principal Components," published in *The American Statistician,* October 1966
5. "A Note on Regression on Principal Components and Constrained Least Squares"
6. "On Errors of Measurements in Regression Analysis"

The second Technical Paper will be devoted to missing observations in regression analysis and will include: (1) a modified and expanded version of my paper in the *Journal of the Royal Statistical Society,* No. 1, 1968, (2) "Estimation of Regression Equations when a Block of Observations is Missing," *1968 Proceedings of the Business and Economic Statistics Section* of the American Statistical Association; and a third paper entitled "On the Use of Auxiliary Information for Estimating Missing Observations in Regression," prepared jointly with Neil Wallace. Most of the theory in the last paper has been developed, but modifications are necessary in light of some results obtained by a Monte Carlo study designed to evaluate the small-sample properties of our suggested estimators.

Another research project under way is a study of the comparative properties of forecasting and estimation of time series models with the first difference transformation vs. "zero difference transformation." This is being carried on jointly with Professor P. J. Verdoorn of the Netherlands Central Planning Bureau and Rotterdam University. We plan to have it ready for publication by the end of the academic year.

Finally, a computer program for Monte Carlo studies entitled "REGEN-Computer Program to Generate Multivariate Observations for Linear Regression Equations," prepared jointly with Sidney Jacobs of the Data Processing Unit, has been reviewed by a staff reading committee. After revision, it will be submitted for publication as a Technical Paper.

Yoel Haitovsky

Experimentation with Nonlinear Regression Programs

Available programs to estimate a nonlinear equation of the form

$$Y = \alpha M^{\eta_1} (1 + r_1)^t X_1 + \beta M^{\eta_2} (1 + r_2)^t X_2$$

have proven unsatisfactory in two respects: they use too much machine time, and some of the estimated coefficient values are unrealistic. We have found it preferable to program our own method for estimating this equation. Our program computes the two modified Cobb-Douglas terms at points in a grid of parameter values for the parameters η_1, r_1, η_2, and r_2. At each grid point selected, the two nonlinear terms enter into a linear regression in which α and β are estimated; this determines what area of the grid is to be searched more closely. The search procedure converges to a best grid point. Substantial computer time is saved by tailoring the program to the specific equation at hand.

In principle, the method could be generalized to handle any nonlinear equation, but this would reduce the saving of machine time. A more promising approach to a general nonlinear regression program proceeds by making a first-order approximation of the regression, using the first partials of the equation with respect to the parameters to be estimated. A linear regression in these partials is run, obtaining a correction to the initial guess of the parameters. By successive approximations, convergence to parameter estimates is achieved.[1] If any of the estimated values lie outside of their expected range, one should repeat the regression with boundary constraints on the parameters. Even if the results fall within the expected range, it may be worthwhile to experiment with alternative boundry constraints.

While this method is used in some existing nonlinear regression programs, we are incorporating several innovations:

1. The nonlinear regression equation will usually be analytic in the parameters. Hence the variance-covariance matrix of partials can be computed by first-order approximation formulas applied to the variance-covariance matrix of the original variables.[2] The original matrix is computed just once, at the beginning of the program; thereafter the program need no longer refer to the individual observations. When the number of observations is large, this procedure will result in substantial savings of computer time. It may be desirable, however, to recompute the exact matrix from the original observations periodically, after a specified number of iterations, to prevent the procedure from going astray.

2. The program operates in a thoroughly conversational manner, via keyboard terminal. It requests data as needed, offers alternative modes of operation, and asks what to do next—all in plain English. Thus the program explains itself to the user as it runs, and requires no previous study of instruction manuals. In the con-

[1] E. J. Williams, *Regression Analysis,* 1967, pp. 60-62.

[2] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics,* Vol. 1, 1963, pp. 231-232.

versational mode, the user sees his answers almost immediately and can decide on the spot what further equations he may want to run on the same or other data.

3. The program language used is APL, which is particularly suited to the conversational mode and which offers a highly condensed notation for mathematical operations. A typical sixty-line page of FORTRAN statements can usually be expressed in two or three lines of APL statements.

Parts of the above innovations have been successfully tested in a prototype APL regression program written for the IBM-1130.

Sidney Jacobs

A Study of the Properties of the Minimum-Sum-of-Absolute-Errors Estimator

In this study, an attempt is being made to determine the sampling distribution of the minimum-sum-of-absolute-errors (MSAE) estimator of the parameters of a linear regression. The estimator is known to out-perform the least-squares estimator in applications where disturbances are characterized by very dense extreme tails, as, for example, where the disturbances follow the symmetric stable distributions with characteristic parameter α very much less than two. Mandelbrot has argued persuasively that such fat-tailed distributions are the appropriate model for many economic processes, which suggests that MSAE may be a good estimator for economists to employ.

Developing a sampling theory for MSAE is made difficult by the fact that there exists no analytic expression for the MSAE estimator. Instead, it is calculated via a linear-programming algorithm. However, by utilizing some results on the distributions of order statistics, it should be possible to characterize the asymptotic distribution of the MSAE estimator. Monte Carlo experiments can then be used to supplement the asymptotic results and enable us to assess how seriously the small-sample distributions seem to depart from the asymptotic distributions.

Reports on some preliminary work on this topic are contained in last year's *Annual Report,* as well as in a forthcoming volume of *Econometrica.* The study is being conducted jointly with Robert Blattberg of the University of Chicago.

Thomas J. Sargent

10. ELECTRONIC COMPUTER SERVICES IN SUPPORT OF ECONOMIC RESEARCH

Introduction

The Bureau's electronic data processing operations encompass a large variety of activities, such as programming, consulting, and other services connected with individual research projects; improvement of data storage and retrieval; and development of programmed approaches to statistical problems. While major services are provided in response to internal demands, we are increasingly attempting to make our resources available to outsiders.

In the supporting operations there are three developments worth pointing out: (1) with the increasing size and complexity of data sets received from government agencies and other sources, data retrieval has become more important and more difficult; (2) the fact that programming is taught in schools and universities has led to an increase in the importance of our consulting functions as compared with our programming activities; (3) easier access to various computer systems, via remote terminals and time-sharing, has increased the necessity for system selection, job channeling, and similar operations. These changes are described below in the report on the activities of the E.D.P. unit.