CHAPTER 13

The Significance of the Results of Price Analyses

When price analysts investigate prices, they necessarily investigate prices that have happened. Most price analysts, however, do their work not merely because they want to explain what has happened in the past, but because they believe that their explanation will have some usefulness in meeting current and future problems. For example, a price analyst discovers by study of past statistics that the demand for potatoes had (he cannot, strictly speaking, say "has") an elasticity between —0.3 and —0.4. He does this not because he is a historian, but because he believes that this finding will be useful in the solution of current potato production and marketing problems.

How well founded is this belief? How likely is it that the quantitative relations revealed by the analysis of past statistics of prices, production, income, etc., will be valid guides to action in the present and future?

Let us illustrate the problem by an extreme case, and then proceed to more typical cases. Suppose that an investigator were analyzing the price of eggs, and had only two annual price data to work with; eggs were 30 cents a dozen in 1940 and 40 cents in 1941. If he plotted these prices against any other variable that changed in value from one year to the other, he would get a perfect positive or negative correlation. He could thus "explain" the price of eggs in terms of any other variable he chose. In this case the explanation would be so obviously absurd that nobody would consider it, because the number of variables is equal to the number of observations and there are no degrees of freedom left.

But suppose the investigator had data for three years. Some of the innumerable economic series available would still, purely by chance, have a high correlation with the price series. If he had data for four years, fewer series would correlate highly with the prices, and data for five and more years would correlate highly with still fewer series. Statisticians have worked out tables showing, for random data, how high the correlation must be for any given number of variables and of items in each series, in order to be adjudged "significant" or "highly significant" and not merely the result of chance.[1]

Thus, a correlation of plus or minus 1.0 between two series, with only two items in each series (for instance, annual data covering only two years) would not mean a thing as an explanation; it would have no real significance; it would not be statistically significant. Tests of significance show that in the case of two series, each three years long, the correlation would have to be 0.997 or higher before it could be considered significant. If the series were each four years long, the correlation would have to be 0.950 or higher, and so on up.

The application of tests of significance to economic data, especially to time series, may give an unwary investigator a confidence in his results which is entirely unwarranted. A series of monthly prices, two years long, would have twenty-four items. A correlation coefficient between it and some other monthly series in excess of 0.404 would be adjudged significant by the application of statistical tests; yet in actual fact the correlation might have no more real significance than the correlation that would result if the monthly data were made into annual data, in which case there would be only two items in each series and the correlation would be perfect.

Other illustrations bring out the point further. Mr. Yule's classic table and chart twenty years ago[2] showed a high correlation (0.9512) between the annual data showing the proportion of Church of England marriages to all marriages and the standardized mortality per 1,000 persons for the same years, over a period of 45 years. For that number of years, any correlation over 0.290 would be adjudged statistically significant. Yet, as he pointed out, all he had there was in "nontechnical language, a fluke"—a purely chance correlation between two trends, both declining without any causal relation between them. The one series was not in any sense an explanation of the other.

Another illustration is the course of prices during a business

---

[1] George Snedecor, *Statistical Methods*, The Iowa State College Press, 1946, p. 149. The meaning of "significant" here is that the correlation coefficient would be as high as (or higher than) the specified figure, in 5 per cent of a large number of such cases taken at random. The term "highly significant" is similar, but applies to the 1 per cent level.

[2] G. Udny Yule, "Why Do We Sometimes Get Nonsense Correlations Between Time Series?" *Journal of the Royal Statistical Society*, Vol. 89, No. 1, 1926, pp. 1–64.

cycle. The annual prices of butter from 1929 to 1936 show a high correlation with the prices of cranberries, but nobody would claim that the one was an explanation of the other. Both were affected by the same decline and recovery of demand. The correlation coefficient is highly (statistically) significant, but not economically significant.

MOST ECONOMIC DATA ARE NOT RANDOM IN CHARACTER

The development of statistical tests of significance, therefore, has not helped the economic statistician very much. For tests of significance, and established statistical methods generally, are designed for use with data that have several important characteristics. These characteristics are: (1) The population must be homogeneous, (2) the distributions of the values of the variables must be approximately normal, (3) each observation must be independent of the others, and (4) the sample must be selected from the parent universe at random.

If the conditions just given are met, even if only approximately, the standard tests of significance of the results of the analysis of a sample measure how likely it is that the characteristics of the sample are true of the population as a whole. But economic data, especially economic time series, clearly do not meet these conditions: (1) The population from which the sample (the data for a certain period of years) is drawn is not homogeneous. A price analyst, investigating the factors determining the price of barley in the United States before 1918, could not rely on tests of significance of his results, because the advent of prohibition in 1918 changed the population. (2) The condition that the data must be normally distributed may be reasonably closely met, although it is more likely that the logarithms of such economic data as prices have normal distributions, than it is that the original data are normally distributed. (3) Each observation is usually not independent of the others. This is true both of successive items in one price series, and of corresponding observations (in time) in different price series. The price of corn in February is not independent of the price of corn in January and March, for all three of these prices are determined (in a given demand situation) by the size of the same corn crop. Similarly, in a given supply situation, the prices of different goods are related to each other at any one time (they are all high or low) according to the prosperity or depression of the country as a whole.

And finally, (4) the sample (the period of years chosen) is usually not selected at random. It generally begins either when the data first became available, or just after World War I or some other sort of bench mark, and runs up to World War II, or in some cases up to the present time.

WHAT CAN BE DONE?

Is there any way to render economic time series more amenable to statistical analysis? A careful student of this question brings in a rather discouraging report.[3]

"This problem has been dealt with somewhat satisfactorily mainly in two different ways. The first is the Variate Difference method as proposed by 'Student' and O. Anderson.* This method is essentially based on the assumption that the systematic or non-random part of the time series is such that it can be wholly or partly eliminated by finite differencing. It is a very well-known fact that a polynomial can be entirely eliminated by forming enough differences. But the Variate Difference Method demands such behavior in a restricted neighborhood only. It is not necessary that the *whole series* behaves like a polynomial over the entire range.† The two authors mentioned above developed this idea statistically by the large sample approach (standard errors). I tried to give recently an extension of it which may be applicable even in the case of short series.‡ It gives exact tests of significance, but is not 'efficient' in the sense of Fisher's criterion. I propose to make selections from the data and to utilize only part of the available material, in order to create artificial independence. Some of the available information is hence lost and the method is not efficient.

"The other approach is from the point of view of serial correlation and was first investigated by Yule.§ A recent book by Wold‖ presents a very extensive treatment of this interesting subject, which

[3] Gerhard Tintner, "The Analysis of Economic Time Series," *Journal of the American Statistical Association*, XXXV, March, 1940, pp. 95–96.

* "Student:" "The Elimination of Spurious Correlation Due to Position in Time or Space," *Biometrika*, X, 1914, pp. 179 ff. O. Anderson: *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929, *See also* G. Tintner, The Variate Difference Method, Bloomington, Indiana, 1940.

† G. Tintner, *op. cit.*, pp. 7, 106.

‡ G. Tintner, *op. cit.*, pp. 73 ff, 124 ff. *See also:* "On Tests of Significance in Time Series," *Annals of Mathematical Statistics*, X, 1939, pp. 139 ff.

§ G. U. Yule: "Why Do We Sometimes Get Nonsense-Correlations Between Time Series," *Journal of the Royal Statistical Society*, Vol. 89, 1926, pp. 123 ff.

‖ H. Wold: *A Study in the Analysis of Stationary Time Series*, Uppsala, 1938.

is closely related to the investigation of differences and also to harmonic analysis. The practical statistical, as distinguished from the probability, aspect of this theory has been, however, very much neglected. It is still far from a solution which is reasonably satisfactory for the application of modern statistics to practical problems. The mathematical difficulties involved are very great.

"But even if the problem of separation of the non-random from the random part of the time series has been solved we are still left with a problem of an entirely different nature. It is often desired by economic statisticians to analyze the non-random part of economic time series further into its components.¶ The components which have been distinguished conventionally are: The seasonal with a period of twelve months, the business cycle with a period of between three and ten years, and the trend and longer waves with long periods. (Kondratieff.**) Several procedures have been proposed for the separation of those components. But the methods of analysis up to now seem not to be very satisfactory."

The use of Fourier series has been recommended for dealing with seasonal variation,[4] but this method is not flexible enough to be used with most other less regular kinds of variation.

Other workers have dealt with the problem of homogeneity. "In analyzing poultry prices, the Division of Statistical and Historical Research has recently begun to test the variables used in correlation analyses for homogeneity. The test used is based on standard methods, but, so far as is known, little use of it has been made in testing agricultural time series. The variance of the means of groups of four consecutive time units in a time series is tested for homogeneity by the standard methods of analysis of variance. The variance of the standard deviations of these groups is tested by comparing the actual variance of the group standard deviations with the estimated variance, computed from the standard deviation of the sample as a whole. In each case the F test is used to determine whether the group means and group standard deviations vary more

---

¶ W. C. Mitchell: *Business Cycles, The Problem and Its Setting,* New York, 1927. S. Kuznets: "Time Series" in *Encyclopaedia of the Social Sciences,* New York, 1935. J. A. Schumpeter: *Business Cycles,* New York, 1939, I, pp. 193 ff.

** N. D. Kondratieff: "The Long Waves in Economic Life," *Review of Economic Statistics,* XVII, 1935, pp. 105 ff.

[4] Alexander Sturges, "The Use of Fourier Series in the Analysis of Seasonal Variation," *Report of Fifth Annual Research Conference on Economics and Statistics,* held at Colorado Springs, Colorado, July 3–28, 1939, Univ. of Chicago, 1939.

than would be expected by random sampling from a homogeneous population. In the poultry price work the method has been applied largely to time series which have been transformed into first difference logarithms. Production series appear to be more homogeneous than do price series, although there is much variation in results between the different variables for each type of series."[5]

The most recent contribution to the solution of the problem of rendering economic time series amenable to statistical analysis has been made by an astronomer.[6] He uses what he describes as the Bartels technique. His article may be summarized as follows:

If the items in a universe u are independent of one another, the standard deviation of the means of random samples, each sample consisting of h items, is:

(1) $$\sigma(h) = \frac{\sigma(u)}{\sqrt{h}}$$

In terms of variance: $$\sigma(h)^2 = \frac{\sigma(u)^2}{h}$$

Then the ratio r

(2) $$r = \frac{h\sigma(h)^2}{\sigma(u)^2}$$

should be constant and equal to unity if the sample is large enough.[7] (The standard deviation of u is not usually known, but the standard deviation of the largest possible sample is taken as the best approximation to it.)

If the items in a universe are *not* independent of one another, then as larger and larger samples are taken (as h increases) the value of the ratio will stabilize at some figure greater than unity. The fact that stability is reached at some figure greater than unity shows that the items are not independent, and the figure at which the value of the ratio stabilizes shows how many items are required in order to eliminate the influence of the serial correlation on the standard deviation of the means of the samples.

Thus Yule's original "nonsense correlation" example covered

[5] "Statistical News and Notes," *Journal of the American Statistical Association*, XXXIV, No. 205, March, 1939, p. 377.

[6] L. R. Hafstad, "On the Bartels Technique for Time-Series Analysis," *Journal of the American Statistical Association*, June, 1940, pp. 347–61.

[7] Using $\sigma(h)^2 = \frac{\sigma(u)^2}{h}$ in equation (2) we obtain $r = \frac{h\sigma(u)^2}{h} / \sigma(u)^2 = 1$

forty-five years, for which by ordinary tests the correlation of 0.95 would be rated highly significant. But application of the procedure described above shows that the size of the sample required to bring the ratio to stability is about fifteen. The forty-five years, therefore, are equivalent only to three independent items; and for series as short as three a correlation coefficient of 0.95 is not significant.

The procedure just outlined has been published so recently that it has not been tested or criticized by other workers. No doubt that will come shortly. Meanwhile, one or two less technical observations may be made about economic time series. While change is the order of the day in economics, so that populations (of economic data) are not homogeneous, it is also true that some of these changes are gradual, not sudden; they are evolutionary, not revolutionary. Thus, while tractors and trucks have displaced half the horses and a quarter of the mules in the country since the time of World War I, the change did not take place all at once, but at the rate of only 1 or 2 per cent per year. Any forecasts which left this important and obvious change out of account would have been only 1 or 2 per cent wrong per year—and forecasts are not usually required to predict changes more than one or two, or at least only a few years ahead. When, as in this case, the direction and extent of a change can be foreseen for several years ahead, its influence can be taken into account. An analysis which includes all the factors that change in the future is really dealing with a homogeneous population. It is changes in factors that are not included in an analysis that change a population and render tests of significance unreliable for that reason. If the number of horses and mules are included as a factor in a price analysis, then (1) future changes in these numbers will not destroy the validity of the analysis, and (2) in this case at least the future changes in this factor can be forecast with some degree of accuracy.

Finally, it must be recognized that there are large random elements in economic data, particularly agricultural economic data. Crop production series meet the requirements for random data rather closely, in those cases where acreage does not change greatly from year to year, since yields fluctuate from year to year chiefly in response to changes in the weather, which are random in character. Fluctuations in *demand* may be cyclic rather than random in character, but that part of a statistical price analysis which deals with the relation between production and price is related to random changes (in yields) and therefore approaches the requirements for random data laid down earlier in this chapter, and is more nearly amenable to statistical analytical methods.[8]

The foregoing considerations mean that the significance of economic analyses depends, not so much upon objective statistical tests, as upon the conformity of the analysis with economic theory on the one hand and with the characteristics of the commodity concerned on the other. It is not sufficient for a price analyst to be familiar with economic theory and statistical methods, although that is indispensable; in addition, he must know a good deal about the particular commodity or service concerned.

[8] For useful observations on this subject, *see* Mordecai Ezekiel, *Methods of Correlation Analysis*, Wiley & Sons, 1941, pp. 349–58.

A group of economists at the University of Chicago is attacking the problem from a new angle that looks promising, but their work has not yet been developed to the stage of general application.