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?

This is a problem of inference, the basic problem in statistics. Most of our statistics (except those in the Census) are derived from samples of whatever "population" we are talking about, not from the whole population; in many cases, "the whole population" hasn't happened yet. A manufacturer tests a mixture of ingredients and processes once, and it works all right; a second time, and it fails; 10 times, and it fails only twice. What percentage will fail if he goes into production? Does he need a still larger sample? The problem of inference is to determine what we can infer about the whole population from the information we get from a sample, and how confidently we can infer it.

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 1960 and 40 cents in 1961. 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 precise meaning of the term "significant" here is this: In repeated samples taken at random from a population with a bivariate normal distribution, where the true correlation (for the whole population) is 0, the confidence intervals for the correlation coefficients would not include the true correlation of 0 in 5 per cent of such samples, purely because of accidents of sampling (sampling error or variation).

The term "highly significant" has a similar meaning, but applies to the 1 per cent level.

Why do statisticians set the limits of significance at 5 per cent and 1 per cent? Why do they need 100 to 1 or 20 to 1 odds? Why not 60 to 40 or even only 51 to 49?

The importance of a clear answer to this kind of question is shown in a recent specific case. A report of a multiple correlation

[1] George Snedecor and William G. Cochran, *Statistical Methods*, Iowa State Univ. Press, 6th ed., 1967, p. 557.

study of the effect of futures trading was presented, purporting to show "that the variation of prices is reduced by an active futures market." The speaker's conclusions were challenged because most of the coefficients on which the conclusions were based were not statistically significant. This interesting colloquy then took place:[2]

> SPEAKER: ". . . . the tests Mr. Shepherd referred to are tests of statistical significance. 'Significance' is very different in what it means in the ordinary usage of the English language and in tests of statistical significance in which we rule out items on the basis of extraordinarily rigorous standards.
>
> "Now, I am a short-odds player myself. You don't have to give me 20-to-1 before you get me to bet on a point. And I should say that to take the very arbitrary levels of significance the statistical fraternity uses and say that something is of no value because it doesn't meet those particular standards, impresses me only very negligibly."
>
> SHEPHERD: "Insignificantly."
>
> SPEAKER: "Yes."
>
> (Laughter)

The speaker's opinions typify a common misunderstanding of the whole concept of statistical significance. Statistics is the science of drawing inferences from data, not the science of betting on horse races, and such. An investigator compares the yields of two varieties of wheat. On those two plots (or more if the trial is replicated) at that station that year, the one variety yields 5 bushels an acre more than the other. Can he release the new variety for general distribution with a statement that it will outyield the other 5 bushels an acre?

Clearly, he would not be justified in doing this. For what he has is only the results of one sample — on the one type of soil at his station, fertilized to the degree he used, with the particular weather he had that year. From that one experiment, he cannot infer that the same results will be obtained in the whole population of various other soils, weather, etc., over the country as a whole in other years. The statistical fraternity has not established "very arbitrary levels of significance;" they have worked out mathematically the validity of inferences concerning the whole population that can be made from a small number of samples. They have established tests of significance to indicate how valid a particular inference based on one sample concerning the parameter of a whole population may be. Determining betting odds is one thing; determining what infer-

[2] Futures Trading Seminar, History and Development, Vol. 1, Mimir Press, 1960, p. 193.

ences concerning a whole population can safely (reliably) be drawn from one sample is quite another.

Another real life example illustrates what significance means. A candidate for the Master's Degree recently worked out a thesis in which he ran a multiple correlation analysis of county average data showing net farm income per farm, by counties in Iowa (99 of them). The independent variables he used were county averages of capital inputs, value of land, and man-days of labor, per farm. He got a multiple correlation coefficient of about 0.8. He tested this for significance, and found that it was highly significant (i.e., at the 1 per cent level).

He was asked, during the oral examination, why he tested his coefficient for significance. He was not inferring a parameter of a whole population from a statistic (a characteristic derived from a sample); to begin with he had the whole population. Whatever relations he found, for the whole population, were the relations for that population, and that was that. A test of significance has no significance for a parameter derived from a whole population.

The candidate could have attempted to defend himself by regarding his Iowa data as a sample of farms in the United States as a whole, and drawing inferences concerning relations for United States farms regarded as the population. But if he had, he would have been in hot water on another count — his sample was not a representative random sample of the whole United States. Or he might have regarded his data for one year as a sample of data for all years, and inferred relations for other years (for Iowa) as the population. But the data for the one year would not have been a representative sample of data for all years, or even for a reasonably long period of time, say 100 years; the sample would not have been random, and the population would not have remained constant over those years. Tests of significance, far from being too rigorous when applied to economic data, actually are not rigorous enough. They are likely to overstate the actual significance of the coefficients rather than understate it.

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 devised more than thirty years ago[3] 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 cycle. In any five years, prices of butter might 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

[3] 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.

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?

The Durbin-Watson test for serial independence of disturbances[4] is:

$$d = \frac{\Delta^2}{S^2}$$

where

$$\Delta^2 = \sum_{t=2}^{T} (\tilde{u}_t - \tilde{u}_{t-1})^2$$

$$S^2 = \sum_{t=1}^{T} \tilde{u}_t^2$$

$\tilde{u}$ = the residual of the fitted relation for time t.

Thus Yule's original "nonsense correlation" example covered forty-five years, for which by ordinary tests the correlation of 0.95

[4] This test was originally presented in *Biometrika*, Vol. 37, p. 409, and Vol. 38, p. 159. An application of this test is discussed in C. Hildreth and F. G. Garrett, *A Statistical Study of Livestock Production and Marketing*, Cowles Commission for Research in Economics, Monograph No. 15, pp. 77–79.

See also, A. A. Harlow, "Factors Affecting the Price and Supply of Hogs," USDA Tech. Bul. 1274, 1962, Appendix.

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.

What this means is that a test of significance attributes more significance to relations among nonrandom economic series than really exists. The actual significance is less than the statistical significance based on random data as shown in significance tables in statistical test books.

There are one or two less technical observations about economic time series that should be made. 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 farm employment (the number of workers on farms) decreased 50 per cent in the 20 years from 1947 to 1967, the change took place fairly steadily, at from 2 to 3 per cent per year. Any forecasts which left even this important change out of account would have been only 2 to 3 per cent wrong per year. 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. Changes in factors that are not included in an analysis change a population and render tests of significance unreliable for that reason. If the number of workers on farms 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.[5]

[5] For useful observations on this subject, see Mordecai Ezekiel and Karl Fox, *Methods of Correlation Analysis and Regression Analysis: Linear and Curvilinear,* 3rd ed., Wiley, 1959, Chap. 20.

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.

five

ANALYSES WITH RESPECT TO PLACE AND FORM