9

The Measurement of Changes in Demand: Multiple Correlation

The graphic method of multiple curvilinear correlation is, as it were, an F_2 product. It is an offshoot of an offshoot of the standard mathematical method of linear multiple correlation.

The first important offshoot was developed originally in the fertile mind of Mordecai Ezekiel.[1] He was working with the problem of curvilinear regressions (curved lines of relationship between different series of data). The simplest form of the standard mathematical method involves the assumption that the data are related in straight-line fashion. If this assumption is not valid—if the regressions are actually curvilinear—the standard method yields inaccurate results; before accurate results can be obtained, the curves must be represented by mathematical equations incorporated in the basic formula.

But the regressions cannot be determined accurately until the nature of the curvature is known, and the nature of the curvature cannot be determined accurately until the regressions are known. Ezekiel broke through this impasse by the method of successive approximations, starting with mathematically determined straight lines and adjusting them by graphic methods.[2]

[1] Mordecai Ezekiel, "A Method of Handling Curvilinear Correlation for any Number of Variables," *Journal of the American Statistical Association,* Vol. 19, No. 148, p. 441. See also: M. Ezekiel and Karl Fox, *Methods of Correlation and Regression Analysis,* Wiley, 1959, Chap. 14.

[2] Ezekiel and Fox, *op. cit.,* p. 210:

"The linear partial regressions are . . . computed [by the standard mathematical correlation method]. Then the dependent variable is adjusted for the deviations from the mean of all independent variables except one, and a correlation chart, or dot-chart, is constructed between these adjusted values and that independent variable. This provides the basis for drawing in the first approximation curve for the net regression of the dependent variable on that

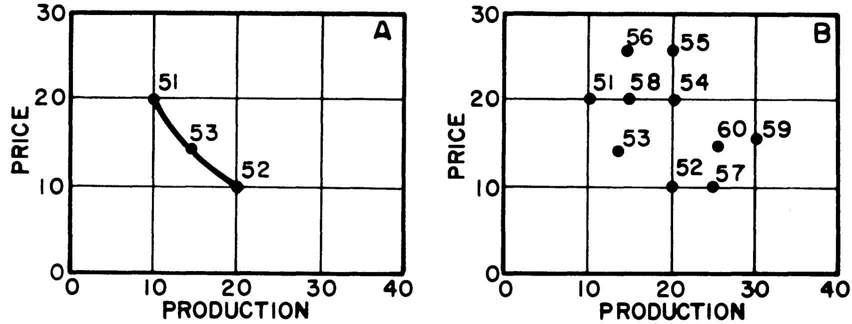

Fig. 9.1 — Hypothetical price and production data.

This first offshoot of the standard mathematical method of linear multiple correlation, therefore, was a hybrid, a combination of mathematics and graphics. Then Louis Bean developed the second offshoot. This second offshoot shed the mathematics inherited from its grandparent completely and became entirely graphic. In effect, Bean said, it is a waste of time to fit straight-line curves mathematically, and then modify them graphically, freehand, to make them fit any curvature existing in the regressions. Don't bother with any mathematics, he said; put the regression lines in freehand in the first place, curves and all.

But how does one know where to draw in the regression lines right off? The essence of Bean's contribution is his simple answer to this question.

In simple correlation, the dependent variable, for example price, is related to one independent variable, for example production. One simply plots price against production in a scatter diagram, and draws in the regression line, straight or curved, wherever the dots indicate that it should go. This is illustrated in section A of Figure 9.1, based upon the first three price and production items in Table 9.1.

In the simplest case of multiple correlation, the dependent variable price is related to two independent variables, for example

independent variable. . . . The dependent variable is then corrected for all except the next independent variable, the corrected values plotted against the values of that variable, and the first approximation curve determined with respect to that variable. This process is carried out for each independent variable in turn, yielding a complete set of first approximations to the net regression curves. These curves are then used as a basis for correcting the dependent factor for the approximate curvilinear effect of all independent variables except one, leaving out each in turn; and second approximation curves are determined by plotting these corrected values against the values of each independent variable in turn. New corrections are made from these curves, and the process is continued until no further change in the several regression curves is indicated."

TABLE 9.1
PRICE, QUANTITY, AND DEMAND SCHEDULE
(Hypothetical Data)

Year	Price	Production	Index of Demand
1951	20	10	10
1952	10	20	10
1953	14	14	8
1954	20	20	20
1955	25	20	25
1956	25	15	20
1957	10	25	15
1958	20	15	15
1959	15	30	25
1960	15	25	20

production and demand. Two scatter diagrams are required here —one to show the regression of price on production (or in more everyday language, the influence of production on price) independent of the influence of demand on price; and the other to show the influence of demand on price independent of the influence of production on price.

In handling a multiple correlation problem of this sort, the first thing to do is to plot the dependent variable price with one of the independent variables, say production, in a simple scatter diagram. The price and production data from Table 9.1 are thus plotted in Section B of Figure 9.1.

The dots in this Section B are scattered about with no evidence of any relationship. But this may be because the influence of production on price is obscured by the coexisting influence of demand on price. What we want is the *net* influence of production on price—the influence of production on price independent of the influence of demand on price.

This word "independent" is the key to the graphic method. One way to determine the influence of production on price independent of the influence of demand on price is to choose two years in which the values of the demand variable are identical. Any change in price from one of these years to the other then must show the influence of production on price independent of the influence of demand on price, since demand did not change from the one year to the other. A line connecting these two years would then be a preliminary indication or estimate of the influence of production on price independent of the influence of demand on price.

Inspection of Table 9.1 shows that there are several pairs of years in which the values of the demand variable are identical (within each pair). The years 1951 and 1952, for example, both carry demand values of 10. The dots for these two years may therefore be connected by a light line, as shown in Section A of Figure 9.2.

This is a beginning. If now another pair of years can be found in which the demand values are identical, another line can be drawn in connecting these two years. Two such years are 1954 and 1956. This provides a second estimate of the influence of production independent of the influence of demand. Additional pairs of years, in each of which the demand values are identical, may also be connected, providing additional estimates.

By now the chart looks like a piece of prehistoric bedrock with scratches on it showing the direction in which a glacier passed over it. A long heavy line can now be drawn in freehand, passing through the dots with a slope representing the general average slope of various short lines on the chart. It should go through the general mean. This line is shown in Section A of Figure 9.2. It is a first approximation to the net regression line desired.

The determination of the net influence of demand on price is then simple. If the heavy line just drawn in shows the net influence of production on price, the vertical distances of the individual dots above and below this line must show the net influence of demand on price. The way to reveal this net influence clearly is to take these vertical distances or residuals and plot them against demand in a second chart. In this chart, the demand scale runs along the bottom, like the production scale in the first chart. A horizontal line is drawn across the middle of the chart (about half way up). This line is regarded as zero on the vertical scale. The vertical distances of individual dots above or below the preliminary regression line in the first chart are then plotted above or below the horizontal line

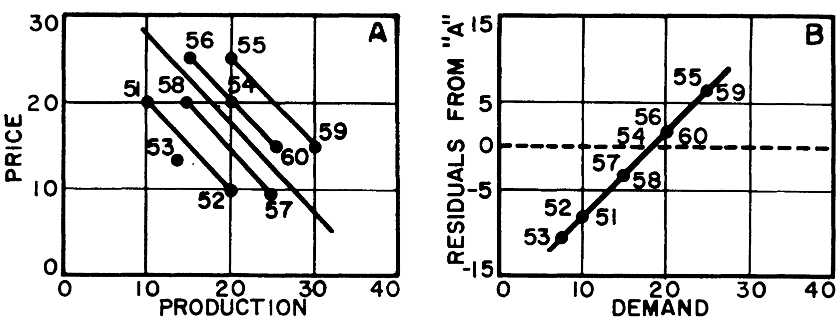

Fig. 9.2 — Hypothetical price and production data. Net regression curves.

across the middle of the second chart against the respective demand readings along the bottom of the second chart. A line drawn through the dots in the second chart then represents the net influence of demand on price—net, because the dots were plotted from the regression line on the first chart which showed the net influence of A on X.

If the first approximation line in the first chart is correct, and if the dependent variable price is completely determined by the two independent variables, production and demand, the line drawn through the dots in the second chart will pass through all of them. This is the situation shown in Section B of Figure 9.2.

If, however, some scatter still remains, either the first approximation line in the first chart was not correctly placed, or one or more additional independent variables need to be taken into account.

The test to determine whether additional variables are needed will also show whether the first approximation line was correctly placed. It consists in taking the residuals from the second chart (the vertical distances above and below the net regression line) and plotting them in red or in some other distinguishing manner, above or below the first approximation regression line in the first chart. (Or this first regression line may be traced on a new clean sheet of graph paper, with the same scales as the original chart). If they fall uniformly about this first approximation line, then a second approximation is not required; what is needed is one or more additional variables. If, however, the dots do not fall uniformly about the first approximation line, but have a different curvature or slope, this indicates that the first approximation line was incorrectly drawn. A second approximation line must be drawn, passing as nearly as possible through the red dots, and the residuals of the *original black dots* from that line plotted in a new second chart (or in different color in the old one) against demand. This may show that the regression line in the second chart needs to be revised. The process is repeated back and forth until the scatter is reduced to the minimum. If some scatter still remains, it means that the study should be extended to include one or more additional variables. In that case the residuals remaining in the second chart should be plotted against a third variable, and so on.

EMPIRICAL ILLUSTRATION

The preceding example, based upon hypothetical data with straight-line relationships and perfect correlation among the variables, serves to illustrate the principles of the graphic method under the simplest conditions. A second illustration based upon actual

TABLE 9.2

FARM PRICES OF CORN (PER BUSHEL) AND RELATED VARIABLES, 1936–51 *

Period Beginning	Price Received by Farmers (Nov.–May): Corn	Price Received by Farmers (Nov.–May): Livestock and products	Supply of Feed Concentrates Per Animal Unit †
	Cents	*Per cent of 1910–14 prices*	*Tons*
1936	106	123	0.65
1937	51	114	.89
1938	44	108	.88
1939	55	107	.87
1940	58	122	.90
1941	74	159	.90
1942	90	194	.90
1943	112	196	.85
1944	107	206	.91
1945	115	215	.92
1946	138	278	.99
1947	220	305	.86
1948	120	285	1.04
1949	118	258	1.06
1950	155	327	1.03
1951	167	318	.97

* Source: F. V. Waugh, "Graphic Analysis in Economic Research," USDA, AMS, Agr. Handbook No. 84, June 1955, p. 37. Computed from data in Richard J. Foote, "Statistical Analyses Relating to the Federal Livestock Economy," USDA Tech. Bul. 1070, 1953, p. 6.

† Year beginning in October.

empirical data—"real data"—will now be used. The data for this illustration are given in Table 9.2.

The object here is to explain the variations in the price of corn. "You have to have an idea (hypothesis) to test before you can test it." Our present analysis starts with the hypothesis suggested by economic theory, that variations in the price of corn are caused by changes in demand and supply of corn.

There are several kinds of changes in demand. Two of the most important are:

1. The change in general demand that results from such things as changes in the general price level, changes in population, changes in per capita income, etc.
2. The change in specific demand for the specific product considered, independent of the change in general demand.

Similarly, there are several different kinds of changes in supply.

An analysis involving four variables—two demand variables and two supply variables—becomes a little difficult to explore by graphic methods. For our purposes here, we will use a little ingenuity and reduce the four variables to two which reflect both general and specific changes in demand and supply.

The prices of livestock and livestock products are carried up and down with general changes in demand; they also reflect changes in the specific demand for corn as a feed for livestock (about 90 per cent of the corn crop is fed to livestock). Accordingly, the average United States farm price of corn from November to May (this 7-month average price is used because it is not directly affected by the size of the preceding and succeeding corn crops as a 12-month average price would be) can be plotted against an index of the prices of livestock and livestock products, in order to reveal the effects of this combined reflector of general and specific demand on corn prices. This is shown in Section A, the upper part of Figure 9.3. In this chart, the prices of corn, regarded as the dependent variable, are plotted up the side. The prices of livestock, regarded as the independent variable, are plotted along the bottom. The data are given in Table 9.2.

The dots in the chart are scattered in a general southwest-northeast direction, indicating that there is some positive relationship between changes in demand (as reflected by livestock prices) and the price of corn. But the dots do not lie closely about any single positive line. Evidently, some other factor was at work, causing variations in the price of corn, in addition to changes in demand.

Economic theory suggests that this other factor is likely to be changes in supply. Here again we can combine two factors into one by expressing the supply of feed concentrates (corn is the principal feed) as supply per animal unit. This makes economic sense, because a large supply of corn would depress the price of corn more if the numbers of livestock were small than if they were large. This factor, the supply of feed concentrates per animal unit, is given in the right-hand column of Table 9.2.

Now we come to the essence of the graphic method of multiple correlation analysis. If we can find two different years when the values of this second independent variable, feed supply, were identical, we can say that changes in supply were not exerting any effect on corn prices, because there were no changes in supply. Any changes in corn prices from the one year to the other, then, must be entirely due to changes in demand. They must show the pure, or net, effect of changes in demand, independent of any change in supply, because the supply was constant.

Looking down the right-hand column of Table 9.2, we can find two years when the supply of feed per animal unit was almost exactly the same. The years are 1948, when the supply was 1.04, and 1950, when it was 1.03. We can connect the dots for these two years by a light line. This line shows the net effect of changes in the price of livestock on the price of corn, independent of changes in the supply of corn and other concentrate feeds.

Similarly, the values of the feed variable were nearly identical in 1946 and 1951. We can connect those years too, by a line which turns out to be parallel with the line connecting 1948 and 1950. We can do the same thing for 1940, 1941, and 1942, when the supply stood unchanged at .90.

A heavy straight line is then drawn in through the dots with approximately the average slope of these light lines. In the estimation of this average slope, each light line should be given an importance proportional to its length. The reason for this is that if the dots in one pair are only an inch apart, let us say, and one of the dots has been pulled up or down by some other influence a distance of half an inch, the slope of the light line connecting the pair will be very much affected. But if the pair had been four inches apart, the half inch displacement of one of the dots would have only a slight effect on the slope of the line connecting them. The longer the light line, the more likely it is to show the net influence of the variable correctly; the longer lines, therefore, should be given more weight than the shorter.

In Figure 9.3 the heavy straight line fits the slope of the drift lines pretty well; in some cases, a curve would be better. Perhaps the left-hand half of the line in Figure 9.3 could be curved upward a little from the straight sloping line shown, to fit better with the drift lines in the left-hand part of the chart, but for our purposes here we will use the simple straight line shown.

Plotting the Residuals

The next step is to plot the residuals from the heavy sloping line in Section A of Figure 9.3 against the second independent variable, the supply of feed. The theory behind this plotting is that since the heavy sloping line in Section A measures the influence of changes in demand on the price of corn, the residuals (residual differences) from that line reflect changes in supply. The dot for 1936, for example, is about 4.5 points above the heavy sloping line in Section A; it is accordingly plotted 4.5 points higher than the horizontal zero line in Section B, against the value of the feed supply variable that year, 0.65. The same sort of thing is then done for the other years.

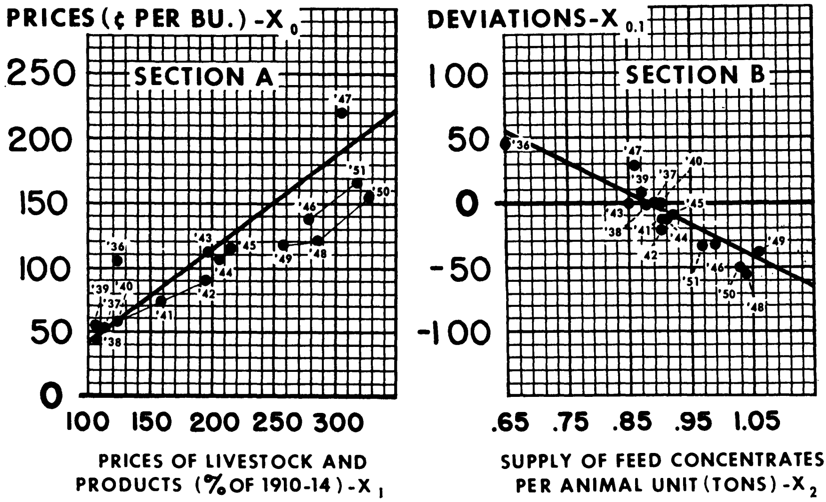

Fig. 9.3 — November–May corn prices received by farmers in relation to specified factors. (Selection A left, Section B, right.)

A heavy sloping line is then drawn in through these dots as shown. There is still some scatter of the dots along this line, and it would be possible to plot the residuals or departures of these dots from the heavy sloping line against another independent variable in order to get a more complete explanation of variations in the price of corn. But for the present expository purposes, the two steps shown in the two sections of Figure 9.3 are sufficient.

The scatter of the dots about the heavy line in the second section of a graphic analysis of this sort could result from an incorrect slope of the heavy line in the first section. In order to test whether this is true, the residuals from Section B of Figure 9.3 should be plotted back against the heavy sloping line (as so much above or below the line) in red or some other distinguishing manner, as explained on page 136. If these dots fall about a line with a somewhat steeper or flatter slope, or with some curvature, a new heavy line should be drawn through the dots in Section A, and residuals from that new line should be plotted against the second independent variable in Section B. This procedure continues until no closer approximations can be made.

X-Ray Vision

One of the great advantages of the graphic method of analysis is that it reveals the anatomy of the subject, like an X-ray photograph, and shows more clearly than any blind application of mathematical methods just what went on during the period covered by the analysis.

A good example of this was shown in Figure 8.2 in the preceding chapter. This figure showed the relation between prices and quantities for several meats. The charts show that if the series were thrown into a calculating machine, they would yield very low correlation coefficients. But the charts also show that the demand curves shifted during the period covered by the data, and that when these shifts are taken into account, logical explanations can be offered for the behavior of the data.[3]

APPRAISING THE RESULTS OF GRAPHIC PRICE ANALYSES

The validity of price analyses of this sort should be judged by three criteria, (1) the closeness of fit of the dots about the lines of net regression, and the conformity of the results (2) with economic theory, and (3) with what is already known about the characteristics of the commodity. None of these tests alone is adequate. An analysis that yielded a positively sloping demand curve, no matter how closely the dots lay about it, would at least call for further investigation, if not rejection; so would an analysis that yielded a demand curve with the expected negative slope but a wide scatter of the dots. A careful worker tests his analyses by these three criteria, subconsciously, as he goes along.

Some controversy arose at one time over the question whether the investigator should follow "the procedure of determining in advance to fit a specific type of curve or set of curves," or should "more nearly allow the data to determine the shape of the curves by fitting a set of curves of minimum residuals." Most of this controversy seems unnecessary; no sensible person follows either procedure alone. It is easy to carry the spirit of determining what to do in advance so far as to make it unnecessary to make the analysis at all, if the investigator already knows all he will permit himself to learn from it. On the other hand, Bean exposed himself to later criticism because in his earlier work he let his curves follow his data too closely to conform well with economic theory.

---

[3] A similar analysis covering an earlier period is given in G. S. Shepherd, J. C. Purcell, and L. V. Manderscheid, "Economic Analysis of Trends in Beef Cattle and Hog Prices," Agr. Exp. Sta. Res. Bul. 405, Jan. 1954.

In presenting their results, some workers show a final chart in which the prices estimated from the regression lines are plotted along with the actual prices in an ordinary time chart (with the price scale running up the side and time along the bottom). This shows nothing about the amount of the difference between the estimated and actual prices that is not already shown in the final regression chart. The scatter of the dots about the line in the final regression chart is the same as the "scatter" of the actual prices about the estimates in the time chart.

It is worth while for purposes of further analysis, however, to plot the residuals from the final regression chart against time treated as an additional variable. This will reveal any serial correlations (cycles or waves) in the residuals, and any trends that may exist. It may be that the residuals gradually rise, or fall, with the passage of time, or show some sort of cyclic movement. If so, plotting them against time will reveal it. If the residuals do rise or fall, the investigator may be tempted to use time as an additional independent variable. But that would be a mistake. Time of itself does not cause residuals to rise or fall; some variable associated with time does it, and the variable itself should be isolated and used, not merely time as such. Otherwise the results may be disastrous, since a variable — such as the displacement of horses by tractors — that moves in one direction over a certain period of time may cease to move, or reverse the direction of its motion, in another.

THE ORDER OF PRESENTATION

The results attained will be the same no matter what the order in which the variables are worked through. The results will be attained more quickly if the variables are taken in the order of their importance, for in that case the first approximation lines are most likely to be accurate. But the lines will be the same whatever the order of the variables may be.

For purposes of presentation, however, one order may be better than another. In some cases attention is being focused in the discussion upon the effect of one particular variable, for instance quantity, upon price. In that case it may be advisable for presentation purposes to use that variable last (even if it is the most important) because the average nontechnical reader is more convinced by the close scatter of the dots around the final variable than by the wider scatter around the earlier ones, though there is actually no real difference between them.

INTERCORRELATION AND GROSS AND NET REGRESSION

Whenever there is any correlation between two independent variables[4] — intercorrelation, it is called — the average slopes of the light lines connecting pairs of years in each of which the value of the next variable are equal will be flatter or steeper than the slope of the group of dots as a whole. Where this happens, the demand curve should be drawn in with reference only to the light lines, not to the group of dots as a whole. For the objective is to ascertain, not gross regression (i. e., simple regression, the regression of X on A[5]) but net regression (i. e., partial regression, the regression of X on A independent of its regression on other variables). The group of dots as a whole shows the gross regression of the dependent variable X upon the first independent variable A; but what we are trying to find is the *net* regression of X upon A after the influence of other independent variables has been taken into account.

A more detailed statement of what gross and net regression are may be helpful here. Regression means, roughly, dependence; we may speak of the dependence of X on A, but it is more direct to speak of the influence[6] of A upon X. When statisticians speak of the regression of X upon A, they mean, in everyday language, the influence of A upon X. Now the gross influence of A upon X is actually the gross influence of a rather extended phrase, namely "A and everything else correlated with A by chance or otherwise." And this phrase is not merely a qualification seldom required; it is the rule, rather than the exception. In our economic world, so characterized by interdependence, it is only rarely that A is not correlated with other variables that have an influence upon X.

It is indeed impossible, practically speaking, to show *the* net influence of A upon X. What we speak of as the net influence of A upon X, with the net influence of B taken into account, is actually the net influence of "A and everything else correlated with A by chance or otherwise, *except* B." What we speak of as the net influence of A upon X, with the influence of B and C taken into account, is actually the net influence of "A and everything else correlated

[4] This sounds like a contradiction in terms, for one might think that if two variables were correlated, they could not be independent. But that would be a mistake. Two variables may be completely independent, completely free of any causal relationship to each other, and yet show some degree of correlation, either because they are both influenced by a third variable, or merely by chance.

[5] Where X = the dependent variable, and A, B, etc., = the independent variables.

[6] The word influence is more accurate than the word effect. An influence may be more or less offset by another influence; this is frequently the case in economics. An influence is exerted, but not necessarily registered, whereas an effect is not an effect until it is registered, i.e. effected.

with A by chance or otherwise, except B and C." And so on for additional variables.

In a world full of complex interrelationships, therefore, successive net regressions, as more and more independent variables are taken into account, should be expected to be different, not only from the gross regression, but from each other. They may even be different in sign. One almost hesitates to use concrete illustrations, for so many other intercorrelations are involved than the one selected for the illustration. But consider the net influence of hog prices upon corn prices, before the influence of hog numbers has been taken into account (and assuming that there are no changes in the general price level to complicate the picture). Hog prices are negatively correlated with hog numbers. If the net influence of hog numbers is greater than the net influence of hog prices, then the net influence of hog prices alone before hog numbers are taken into account would be more than offset by the effect of hog numbers, and would appear actually negative. But the addition of hog numbers as an additional variable would change the influence of hog prices (change the slope of the hog price regression curve) to its proper sign, positive.

This means that absolute net relationships are unattainable, because we can only ascertain absolute net influence if we take all other influences into account—literally hundreds of them. And this is a practical impossibility. But from a practical point of view, absolute net influences can be closely approximated. Economic reasoning and published studies in the field of agricultural economics both indicate that serviceably accurate results can be attained in most cases by the use of a relatively small number of variables. Practically all of the published studies use only two or three independent variables. While interdependence is ubiquitous, its quantitative importance diminishes rapidly after the most influential variables have been taken into account; and these most influential variables are usually few in number.

Some of the problems of intercorrelation are not as baffling as they appear at first sight. A concrete case will illustrate this. An analysis of the United States average farm price of corn shows that the elasticity of the demand for corn is about —0.5. The bulk of this corn is No. 3 Yellow. But if the price of No. 2 Yellow corn were used as an additional independent variable, the elasticity of the demand for No. 3 corn would become almost infinitely great; that is, the regression of No. 3 Yellow corn prices on corn production would be practically a horizontal straight line. This results from the fact that the price of No. 2 Yellow corn is so highly intercorrelated with the dependent factor that there is not much left over for the other independent variables to explain.

What does this result mean? What is the real or true elasticity of demand for No. 3 Yellow corn—is it —0.5, or is it practically infinity?

The answer is, both. In both cases, the coefficient of elasticity shows what happens to prices when production changes, "the other independent variables being held constant," as it is often expressed, or more accurately, "independent of the accompanying variation of the other variables." If the price of No. 2 Yellow corn were in actual fact "held constant," it is clear that changes in corn production would have very little effect on the price of No. 3 Yellow corn. Or to use the more accurate phrase above, there is very little fluctuation in the price of No. 3 Yellow corn independent of the fluctuation in the price of No. 2. When the price of No. 2 corn is included as one of the variables, the regression of the price of No. 3 corn should be practically zero, as in fact it is. This is merely an extreme illustration of the fact that the addition of another independent variable changes the so-called "net" regression of the dependent on the independent variables whenever (as usually happens) the additional variable is correlated with any of the other independent variables.

CORRELATION AND CAUSATION

One must clearly keep in mind the difference between correlation and causation. Two series may be highly correlated, and this correlation may be used (wrongly) to demonstrate that the one is the cause of the other. But actually the causation may run the other way; or there may be no causal relationship between the two whatever.

About all that can be deduced from a correlation coefficient is that the higher the coefficient the more likely it is that the relationship between the two variables is not due to chance, but is due to some definite relationship, such as cause and effect, between them. For data that are random in character, the standard statistical tests of significance put this statement in precise numerical form.[7] Since most economic time series are not random, however, these tests of significance have only a restricted validity in economics. This matter is discussed at some length in Chapter 13.

As to the nature of the relation between two variables, a correlation coefficient gives no answer. The relation may be one of cause and effect, as in the case of corn production and corn prices (though

[7] See, for example, George W. Snedecor and William G. Cochran, *Statistical Methods,* Iowa State Univ. Press, 6th ed., 1967, Table A 11, p. 557.

the correlation shows nothing as to which is cause and which is effect). Or both variables may respond to a third causal factor, as when the prices of two unrelated agricultural products are both affected by industrial prosperity or depression. Or the relationship may result entirely from chance.

ADVANTAGES AND LIMITATIONS

The graphic method has several advantages over the standard method of mathematical correlation analysis, and several limitations. The advantages will be considered first.

1. The graphic method enables the investigator to see just what he is doing. With the mathematical method, he merely feeds the data into the machine and comes out with some numerical coefficients. He does not know without additional testing whether his multiple correlation coefficient, for example, is 0.8 rather than some higher figure because the relationships are curvilinear, because one or two exceptional years were far out of line, or because additional variables are needed. But with the graphic method, he can see just what the curvilinearity is, just how many and which years are exceptional, and whether additional variables are needed, or not. These are ponderable advantages.

These advantages have been arithmetically demonstrated by Elmer Working.[8] He set up four different pairs of variables and plotted each pair in a scatter diagram to show the relation between each pair of items. The four scatter diagrams are shown in Figure 9.4. Two of the relationships shown are curvilinear, one is linear, and one is based upon a very abnormal distribution. Yet the mathematical coefficients—standard deviations and correlation coefficients—are practically identical. This example illustrates how graphic methods would protect the investigator from errors he might not otherwise have discovered (although the fact that the correlations in this example are "significant" but just barely "highly significant" should put him on his guard).

2. In the second place, the graphic method usually saves a good deal of time and energy. In many cases where problems of the same number of observations and variables were treated by both methods, the graphic method proved just as accurate but consumed from one-fourth to one-third of the time. The greatest saving comes in connection with rather short series, say from 15 to 20 items, such as are common in economic price analysis. The work calls for no more training — in fact, probably for less training — on the part of the

[8] E. J. Working, "Graphic Methods in Price Analysis," *Journal of Farm Economics*, Vol. 21, No. 1.

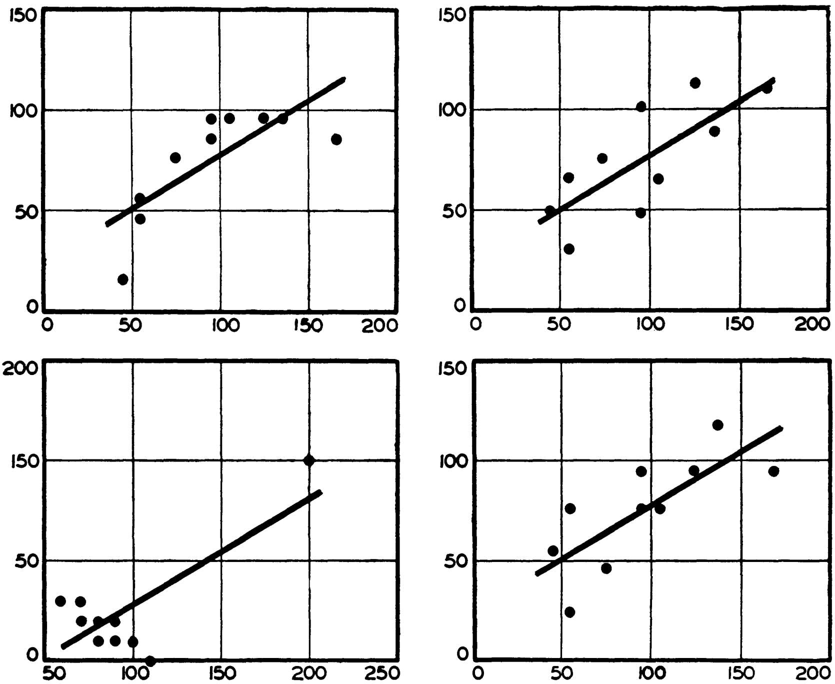

Fig. 9.4 — Linear regression lines fitted to four pairs of variables. Hypothetical data provided by Elmer Working.

worker than is required for the formal mathematical method, and it requires no more complicated computing machinery than a pencil and ruler. In a large research department with plenty of mechanical computing equipment, this consideration is not very important; but in other cases it is.

The saving of time is greatest where the regressions are curvilinear. With the graphic method, it is just as easy to put in a curve freehand as a straight line. But with the mathematical method, the process is complex and laborious. First, the investigator must run the regular straight-line multiple correlation computation. Next, he must test each regression mathematically for curvilinearity. If it exists, he has then to select the mathematical curve most nearly appropriate to the data, add one or more extra terms to his equation, and run the multiple correlation computation again. Finally, he comes out with the coefficients expressing the relationships numerically.

3. Economic data do not usually follow any mathematical formula. They do not arrange themselves, like snowflakes, in geometrical patterns. There is no reason, for example, that the demand

curve for wheat should be a straight line, a parabola, a hyperbola, or any other mathematical curve. It is determined by the physiological reactions of consumers, the distribution of income, the effect of different volumes on the intervening charges between producer and consumer, and so forth. The total effect of these things is unlikely to follow any simple mathematical formula.

Accordingly, mathematical curves cannot be expected to fit economic data very accurately. The final result is merely a compromise between the characteristics of the curve chosen by the investigator, and the characteristics of the data. The investigator has to use his judgment in selecting the type of curve to fit, and his choice of curve determines in considerable part the results he gets. The mathematical method, therefore, is not purely objective; it has a large element of subjectivity in it. The results obtained depend to a considerable extent upon the mathematical curve selected.

The graphic method is similarly subjective. The curves are drawn in freehand, and some judgment is required, as for example in deciding how much weight to give one or two extreme items. No two investigators will draw two curves in exactly alike, any more than they will agree on the exact length of a bar they may measure. So exact tests of significance, standard errors, correlation coefficients, betas, etc., cannot be computed. This appears to be a fairly important weakness. Yet tests of significance are so inapplicable to economic time series that it is doubtful whether the weakness is as great as it seems.

With mathematical methods, if two different investigators choose the same mathematical curves, their results will agree out to as many decimal places as they may wish. When the differences resulting from choosing different mathematical curves may run into whole numbers, however, the identicality of results out to several decimal places is more misleading than confirmatory. For mathematical straight lines or curves are unlikely to fit the data any more exactly than two graphic workers fit their data freehand. And the inexactitude of the freehand line is at least clearly shown in the charts, while that of the mathematical method is covered up by figures running with a great profession of accuracy out to several decimal places.

To summarize, then: Practically all of the real issues involved in the use of either method—representativeness of sample, serial correlation, intercorrelation, multiplicative relationship, etc.—are common to both. Any careful user of the mathematical method would use scatter diagrams (i. e., make an informal use of graphic methods) in deciding whether to use straight lines or curves in his formulae; for him, the graphic method is a useful exploratory tool.

Conversely, any graphic worker who wished to take the time could well go ahead after he had completed his graphic analysis and express his results in mathematical form. The differences of opinion as to the merits of the two methods then reduce merely to differences in the emphasis to be given to each. The mathematical statistician regards the graphic method as an exploratory tool, useful in preparing the way for mathematical analysis; the economic statistician, on the other hand, is inclined to regard the graphic analysis as the main job, and publish the mathematical coefficients in a footnote.[9]

This difference in emphasis is nothing to provoke serious controversy. It results primarily from the differences in the kind of data with which mathematicians and economists generally work. The graphic method is most useful in problems: (1) Where the number of items is small, not over twenty or thirty for example; with longer series, the labor of plotting may be as great as the labor of computing the coefficients mathematically. (2) Where the number of variables is small, say three or four; with a larger number, the process of working back and forth becomes complicated. (3) Where the correlation is rather high; this reduces the judgment required in drawing in the curves. These conditions are frequently met in economic problems, and this is probably the reason why the graphic method has been used so widely by agricultural economists.

The simultaneous equation method. A more fundamental question is whether to use the elaborate simultaneous equation method of analysis developed chiefly by the Cowles Commission of Yale University,[10] instead of the single equation method described above.

This question is discussed in the next chapter.

---

[9] Practical applications of the combined use of the graphic and mathematical methods along with other analytical procedures and concepts developed earlier in this book, are made in two bulletins, "Changes in the Demand for Meat and Dairy Products in the United States Since 1910," Iowa Agr. Exp. Sta. Res. Bul. 368, Nov., 1949, by the present author, and "Economic Analysis of Trends in Beef Cattle and Hog Prices," Iowa Agr. Exp. Sta. Res. Bul. 405, Jan., 1954, by the present author *et al.*

[10] *Statistical Inference in Dynamic Economic Models,* Tjalling C. Koopmans, editor, by Cowles Commission Research Staff Members and Guests, Wiley, New York, 1950.