The Empirical Determination of Demand Relationships

William J. Baumol

William Baumol is Professor of Economics at Princeton and New York Universities. This piece is taken from his book *Economic Theory and Operations Analysis*, 1965.

I. Why Demand Functions?

Demand functions, as they are, defined in economic analysis, are rather queer creatures, somewhat abstract, containing generous elements of the hypothetical, and, in general, marked by an aura of unreality. The peculiarity of the concept is well illustrated by the fact that only one point on a demand curvy can ever be observed directly with any degree of confidence because by the time we can obtain the data with which to plot a second point, the entire curve may well have shifted without our knowing it. A more fundamental but related source of our discomfort with the idea is the fact that the demand relationship is defined as the answer to the set of hypothetical questions which begin, "What would consumers do if price (or advertising outlay, or some other type of marketing effort) were different than it is in fact?" We are, then, dealing with information about potential consumer behavior in situations, which consumers may never have experienced. And, since we have very little confidence in the constancy of consumer tastes and desires, all of these data are taken to refer to possible events at just one moment of time-e.g., consumer reactions to alternative possible prices if any of them were to occur tomorrow at 2:47 P.m.

In view of all this, there should be little wonder that people with an orientation toward applied economics occasionally become somewhat impatient with the economic theorist's demand function. Yet no matter how ingenious the circumlocutions, which may have been employed, they have been unable to find an acceptable substitute for the concept. For the demand function must ultimately play a critical role in any probing marketing decision process, and there is really no way to get away from it.

For example, to decide on the number of salesmen which will best serve the interests of the firm, it is first necessary to know what difference in consumer purchases would result from alternative sales force sizes. But this is precisely the sort of odd and hypothetical information, which goes to make up the demand relationship. It is for exactly the same reason that many large and reputable firms in diverse fields of industry are conducting ambitious research programs whose aim is the determination of their advertising-demand curves, that is, the relationship between their advertising outlays and their sales. So far, these efforts have met with varying degrees of success, and it must be admitted that many of them have not come up with very meaningful results. For the empirical determination of demand relationships is no simple matter and there are many booby traps for the amateur investigator and the unwary. It is no trick at all, on looking over a small sample of the published demand studies, to come up with horrible examples of just about every available type of misstep.

This chapter is designed primarily to point out some of the pitfalls, which threaten the investigator of demand relationships. Its aim is to warn the reader to proceed with extreme caution in any such enterprise. No cut-and-dried solutions are offered to the problems, which are discussed. This is true for two reasons. First, because many of the methods for dealing with these difficulties are highly technical matters of specialized econometric analysis and so are completely outside the scope of this volume. Second, and more important, solutions are not listed mechanically because there simply are no panaceas; the problems must be dealt with case by case as they arise, and the effectiveness with which they can be handled is still highly dependent on the skill, experience, and judgment of the specialist investigator.

If after reading the chapter the reader is left somewhat worried and uncomfortable, it will have accomplished its purpose. However, it should be emphasized that the problems which are raised, serious and difficult though they be, are not totally intractable and beyond the power of our statistical techniques.

II. Interview Approaches to Demand Determination

Before turning to statistical methods for the finding of demand functions, it is appropriate to say a few words about a more direct method for dealing with the problem-the consumer interview approach. In its most blatant and naive form, consumers are simply collared by the interviewer and asked how much they would be willing to purchase of a given product at a number of alternative product price levels.

It should be obvious enough that this is a dangerous and unreliable procedure. People just have not thought out in advance what they would do in these hypothetical situations, and their snap judgments thrown up at the request of the interviewer cannot inspire a great deal of confidence. Even if they attempt to offer honest answers, even if they had thought about their decisions in advance, consumers might well find that when confronted with the harsh realities of the concrete situation, they behave in a manner, which belies their own expectations. When we get to the effects of advertising on demand, the problems of such a direct interview approach become even more apparent. What is the consumer to be asked - how much more of the company's product he would buy if it were to institute a 1 per cent increase in its spot announcements to its television budget?

Much more subtle and effective approaches to consumer interviewing are indeed possible. Indirect, but far more revealing questions can be asked. Consumers may, for example, be asked about the difference in price between two competing products, and if it turns out that they simply do not know the facts of the matter, one may be led to infer that a lower product price may have a relatively limited influence on consumer behavior just because few consumers are likely to be aware of its existence. A clever interview designer may in this way build up a strategy of indirect questions, which gradually isolates the required facts.

Alternatively, consumers may be placed in simulated market situations, so-called consumer clinics, in which changes in their behavior can be observed as the circumstances of the experiment are varied. An obvious approach to this matter is to get groups of housewives together, give them small amounts of money with which they are offered the opportunity to purchase one of, say, several brands of dishwasher soap which are put on display at the clinic, and observe what happens as the posted prices on the displays are varied from group to group. Here again, much more subtle variants in experimental design are clearly possible.

But even the best of these procedures has its limitations for our purpose, which is the determination of the precise form of a demand relationship. Artificial consumer clinic experiments inevitably introduce some degree of distortion because subjects cannot be kept from realizing that they are in an experimental situation. In any event, such clinics are rather expensive and so the samples involved are usually extremely small-too small for confidence in any inferences, which are drawn about the magnitudes of the parameters of the demand relationships for the body of consumers as a whole. And large sample interviews, which approach the determination of consumer demand patterns by subtle and indirect questions are often highly revealing, but they rarely can supply the quantitative information required for the estimation of a demand equation.

III. Direct Market Experiments

A second alternative approach, which is sometimes considered as a means for finding demand relationship information is the direct market experiment. A company engages in a deliberate program of price or advertising level variation. Suppose it increases its newspaper advertising outlay in one city by 5 per cent, in another city it increases this outlay by 10 per cent, and in still a third metropolis a 10 per cent reduction is undertaken.

In some ways such a direct experimental approach must always be the most revealing. It gives real answers to our formerly hypothetical questions and does so without subjecting the consumer to the artificial atmosphere of the interview situation or the consumer clinic. However, direct experimentation has its serious limitations as well.

1. It can be very expensive or extremely risky for the firm. Customers lost by an experimental price increase may never be regained from competitive products, which they might otherwise never have tried, and a 10 per cent increase in advertising outlay for any protracted period may be no trivial matter.

2. Market experiments are almost never controlled experiments, so that the observations, which they yield are likely to be colored by all sorts of fortuitous occurrences-coincidental changes in consumer incomes or in competitive advertising programs, peculiarities of the weather during the period of the experiment, etc.

3. Because of the high cost of the experiments and because it is often simply physically impossible to try out a large number of variations, the number of observations is likely to be unsatisfactorily small. If, for example, it is desired to determine the effects of varied advertising outlay in a national periodical, the company cannot increase the size of its ads, which are seen by Nashville readers and simultaneously reduce those, which are seen in Lexington, Kentucky. This difficulty has been eased to some extent by the fact that a number of national magazines now put out several regional editions, but by and large the problem remains: market experiments usually supply information only about a very limited number of alternatives.

4. For similar reasons, market experiments are often of only relatively brief duration. Companies cannot afford to permit them to run long enough to display much more than impact effects. And yet the distinction between impact effects and long-run effects of a change are often extremely significant, as was so clearly demonstrated by the sharp but very temporary drop in cigarette sales when the first announcement was made about the association between smoking and the incidence of cancer. How often has a rise in the price of a product caused a major reduction in purchases for a few weeks, with customers then gradually but steadily drifting back?

Market experiments do have a role to play in demand relationship determination. They can be important as a check on the results of a statistical study. Or they can provide some critical information about a few points on the demand curve in which past experience is entirely lacking. In some special circumstances experimentation is particularly convenient and has been used in the past, apparently with a considerable degree of success. For example, some mail-order houses have employed systematic programs in which a few special experimental pages were bound inconspicuously into the catalogues distributed to customers within restricted geographic regions, thus permitting observation of the effects of price, product, or even catalogue display variations. However, it should also be clear that market experiments cannot by themselves be relied upon universally to provide the demand information needed by management. Economics is just not a subject which lends itself readily to experimentation, largely because there are always too many elements beyond the control of the investigator and because economic experimentation is often inherently too expensive, risky, and difficult.

IV. Standard Statistical Approaches

The third, and generally most attractive, approach to demand function determination attempts to squeeze its information out of sources such as the accumulated records of the past (a time-series analysis), or a comparative evaluation of the performance of different sectors of the market (a cross-sectional analysis). The available statistics on sales, prices, advertising outlays of the most relevant varieties, and other marketing data are gathered together and then analyzed with the aid of 'the standard statistical techniques.

The basic procedure is simple enough; in fact, as we shall see presently, it is often far too simple. Suppose, for example, that the following data on company sales and advertising outlays have been accumulated:

TABLE I.

Year	1950	1951	1952	1953	1954	1955	1956	1957
Sales .... (millions of dollars)	67	73	54	62	70	75	79	83
Advertising .... (millions of dollars)	12	15	13	14	18	17	19	15

Once the figures have been plotted, the pattern formed by the dots can be used in an obvious manner to fit a straight line (see Figure 1) or a curve to them. This line is then taken as the desired advertising-demand curve. Its slope can be used as a measure of advertising effectiveness, that is, it measures the marginal sales productivity of an advertising dollar, 0 sales/ advertising outlay. This line can be determined impressionistically simply by drawing in a line that appears to fit the dots fairly well, or any one of a variety of more systematic methods can be used.

The most widely employed and best known of these techniques is the method of least squares,' in which the object is to find that line which makes the sum of the (squared) vertical deviations between our dots and the fitted line as small as possible, where the deviations are defined as the vertical distances such as *AB* or CD in Figure 1. The idea is inherently attractive. We wish to minimize deviations because a line, which involves very substantial deviations from the dots representing our data surely does not represent the information in a very satisfactory way. But if, in our addition process, a large negative deviation such as *AB* (that is, a case where the line underestimates the vertical coordinate of our dot) happens to be largely cancelled out by a positive deviation, CD, the sum of the deviations can turn out to be small. This is surely not what we want in looking for a line, which does not deviate much from the dots. One can avoid ending up with a line, which fits the facts rather badly but in which the positive and negative deviations add up to a rather small number, by squaring all the deviation figures before adding them together. Since the square of a negative real number as well as that of a positive real number is always positive, large, squared negative deviations cannot offset large, squared positive deviations, and the sum of squares deviations will never add up to a small number unless our line happens to fit the dots closely.[2]

There exist still more sophisticated techniques for fitting our advertising demand curve from the data. Although it is often too complex and expensive to employ in practice, professional statisticians usually consider the method of maximum likelihood as their ideal. This method requires some information about the probability distribution of the random elements, which influence sales. From this probability distribution the statistician determines a likelihood function

$$L=f\,(x_t, ,y_{t,}, a, b)$$

where x_t and y_t represent, respectively, advertising expenditures and sales in year *t,* and *a* and *b* are

[2] Other devices (such as the absolute value or the fourth power of the deviation) might accomplish the objective discussed. The reason one chooses to minimize the sum of the *squares is* that under very simple assumptions such estimates have several extremely desirable technical properties, among them, that these estimated parameter values are "best" in the sense that they minimize variance of the estimate and are unbiased.
To find the straight-line equation which satisfies our least squares requirement we employ the symbol x_t to represent sales in year *t* and y_t to represent advertising outlay in that year and let the equation of the line to be fitted be written $y_{ct} = a + bx_t$ where the subscript c in y_{ct} *is* there to remind us that in our equation the y is a figure calculated from the formula rather than observation. Now we proceed as follows:

Step 1. Define a deviation from our line as

$$y_t - y_{ct} = y_t - (a + bx_t) = y_t - a - bx_t.$$

Step 2. Define a squared deviation as

$$(y_t - y_{ct})^2 = y^2_t + a^2 + b^2x^2_t - 2ay_t - 2bx_ty_t + 2abx_t.$$

Step 3. Add the squared deviations

$$\Sigma(y_t - y_{ct})^2 = \Sigma y^2_t + na^2 + b^2\,\Sigma x^2_t - 2a\,\Sigma y_t - 2b\,\Sigma x_ty_t + 2ab\,\Sigma x_t$$

where, since *a is a constant,* $\Sigma a^2 = a^2 + a^2 + a^2 \ldots$ (*n* equal terms) = na^2.

Step 4. Find the values of *a* and *b* (the parameters of our equation) which minimize the sum of the squared deviations. Vie do this with the aid of the usual calculus procedure by taking partial derivatives with respect to *a* and *b* and setting them equal to zero, thus:

$$\frac{\partial \Sigma (y_t - y_{ct})^2}{\partial a} = 2an - 2\, \Sigma\, y_t + 2b\, \Sigma\, x_t = 0$$

and

$$\frac{\partial \Sigma (y_t - y_{tc})^2}{\partial a} = 2b\, \Sigma\, x^2_t - 2\, \Sigma\, x_t y_t + 2a\, \Sigma\, x_t = 0$$

These last two equations contain, in addition to *a* and *b,* only known statistical figures x_t and y_t. The equations can therefore be solved simultaneously to obtain the desired parameter values, *a* and *b,* i.e., they determine for us the least squares line $y_{ct} = a + bx_t$. These two equations are usually referred to as the *normal equations* of the least squares method in this most elementary (two-variable straight line) case. The procedure employed in fitting many variable equations or curvilinear equations is a simple and obvious extension of that which has just been described. the constants in our advertising demand equation $y_t = a + bx_t$.. *This* likelihood function is defined as an answer to the following type of question: "Given any specific values of the parameters in our equation, say a = 5 and *b = 63,* how likely is it that the demand situation would have generated the statistics $x_{1950} = 12$, $y_{1950} = 67$, etc.?"

(Note that these values of sales and advertising are in fact our observed statistical figures taken from Table I.) Considering all possible values of a and *b,* we can then employ the differential calculus to find the a and *b* combination, which maximizes the value of the likelihood, L. We will then have found the a and *b* which provide, in this sense, the best possible explanation of the observed facts, i.e., we will have found that equation $y_t = a + bx_t$ whose parameters a and *b* are most likely to be the correct values of the true but unknown parameters, given the facts which were actually observed by the data collector.

It is of interest to note that in some special cases the least squares method turns out to be identical with maximum likelihood. That is, in these fortunate circumstances the least squares calculation becomes equivalent to the maximum likelihood procedure. We shall presently discuss one of the things which may go wrong if the least squares method is employed in situations where it does not yield the same results as the maximum likelihood calculation.

Having described now in highly general and impressionistic terms the methods which are most commonly employed by the statistician to determine relationships, let us now see some of the problems to which they give rise.

V. Omission of Important Variables

Clearly, sales are affected by other variables in addition to the company's advertising expenditure. Prices, competitive advertising, consumer income variations, and other variables also play an important role in any demand relationship. If, therefore, we try to extract from our statistics a simple equation relating sales to advertising outlay alone, and in the process we ignore all other variables, our results are likely to be very badly distorted. We may ascribe to the company's advertising outlays sales trends, which are really the result of the behavior of other economic changes. The behavior of other variables can thus conceal and even offset the effects of advertising. To show how serious the results can be, consider the illustrative demand equation

(1) $$S = 50 + 4A + 0.02Y$$

where S represents sales, A advertising expenditure, and Y consumer income. The values given in Table II can easily be seen to satisfy the equation precisely, and any standard estimation procedure based on such information can be expected to yield the correct equation.

TABLE II.

Date	*1956*	*1957*	*1958*
Y	*3,000*	*4,000*	*3,500*
A	*2*	*3*	*2.5*
S	*118*	*142*	*130*

But the standard calculation shows that a *two*-variable, straight, least squares line which gives us a (perfect!) correlation between S and A alone (ignoring Y) and which is based on these same values will yield the equation

(2) $$S - 24A + 70.$$

This equation asserts that each added dollar of advertising expenditure brings in $24 in sales, instead of the true $4 return shown by equation (1). in addition, because of the perfect correlation there is, in this case, no residual unexplained variation in S which is left to be accounted for by a subsequent correlation between S and Y, i.e., *this incorrect procedure appears to show that consumer income has absolutely no influence on demand!* This advertising coefficient has been inflated by usurping to itself the influence of Y on sales. Incidentally, if, instead of proceeding as we just did, we had started off by finding a least squares equation relating sales to consumer income alone, we would have obtained from the same statistics the equation

$$S - 0.024Y + 46$$

which this time overvalues the influence of income on sales and ascribes absolutely no effectiveness to advertising.

It is clear, then, that more than two variables must usually be taken into account in the statistical estimation of a demand relationship. And, in fact, this is ordinarily done, the estimation usually employing what is called a least squares *multiple regression* technique. However, it should be remembered that even if we include five variables in our analysis but omit a sixth rather important variable, precisely the same difficulties will be encountered. That is, the omission of any important variable, however defined, from the statistical procedure can lead to serious distortions in its results.

This might appear to constitute an argument for the inclusion in the analysis of every variable, which comes to the statistician's mind as a factor of possible importance, just as a matter of insurance. Unfortunately, however, we are not at liberty to go on adding variables willy-nilly. The more variables whose influence we want to take into account the more data we require as a basis for the estimation. If we only have statistical information pertaining to three points in time, it is ridiculous to try to disentangle the influence of fifteen variables. In fact, the statistician requires many pieces of information for every variable he includes in his analysis, if he is to estimate his relationship with a clear conscience.

However, large masses of marketing data are not easily come by. Records are often woefully incomplete; additional data can sometimes be acquired only at considerable expense, and in any event, statistics, which go too far back in time, are apt to be obsolete and irrelevant for the company's current circumstances. We must; therefore, very frequently be contented with skimpy figures, which force us to be extremely niggardly in the number of variables which we take into account, despite the very great dangers involved.

VI. Inclusion of Mutually Correlated Variables

Another difficulty which, to some extent, can help to make life easier as far as the problem of the preceding section is concerned arises when a number of the relevant variables are themselves closely interrelated. For example, one encounters advertising effectiveness studies in which income and years of education per inhabitant are both included as variables. Now, education is itself very closely related to income level both because higher-income families can afford to provide more education and larger inheritances to their children and because a more educated person is often in a position to earn a higher income.

It may nevertheless be true that education and income do have different consequences for advertising effectiveness. For example, an increase in income without any change in educational level could increase the person's willingness to purchase more in response to an ad, whereas more education not backed up by larger purchasing power might have the reverse effect. But, in general, there, is no statistical method whereby these two consequences can be separated, because, for the bulk of the population, whenever one of these variables increases in value, so does the other. Hence, the statistics which can merely exhibit directions of variation might show that, other things remaining equal, whenever sales increased, income also increased, and so (as a consequence?) did education.

In such circumstances if we include both the income and the educational level variables in the statistical demand-fitting procedure, the chances are that the mechanics of the procedure will provide a perfectly arbitrary ascription of the sales changes to our two causal variables. And sometimes the results may turn out completely nonsensical because the standard computational procedure has no way to apply common sense in imputing the total sales change to the separate influences of education and income changes. Therefore, if in a demand relationship there occur several variables, which are themselves highly correlated, it is usually wise to omit all but one of any such set of variables in a statistical study. If this is not done, another powerful source of nonsense results is introduced.

VII. Simultaneous Relationship Problems

The difficulties, which have so far been discussed, while they can be extremely important and are often overlooked in practice (with rather sad consequences) may, by and large, be considered rather routine and in retrospect, fairly obvious matters.

We come now to a far more subtle and perhaps a far more serious problem, which was only brought to our attention in 1927 by E. J. Working and which has only received serious and systematic attention quite recently, largely as a result of the work of the Cowles Foundation. The problem in question, in a sense, follows from the difficulty, which was discussed in the previous section. If there is a close correlation between two variables, it is likely to mean that they are not independent of one another and that there is at least one other relevant equation in the system, which expresses the relationship between them. For example, in our illustrative case there *f* might be an equation indicating how income level is ordinarily increased by a person's education. We then end up having to deal with not just a single demand equation, but with a system of several equations in which a number of the variables interact mutually and are determined simultaneously.

Economics is characterized by such simultaneous relationships. The standard example is the price determination process in which a supply equation is involved as well as our demand relationship. Similarly, simultaneous relationships constitute the core of national income analysis. National income depends on the demand for consumers' goods which helps determine the level of profitable production. But the consumption demand equation, in turn, involves national income (as a measure of the public's purchasing power) as a variable. To mention another simultaneous relationship example, the coal mining industry is a customer for steel whose volume of demand depends on coal sales, but the demand for coal itself depends heavily on the amount of coal to be used in producing steel. It is possible to expand the list of simultaneous relationships in economics indefinitely.

The empirical data, which are generated by such a set of equations, are the information source on which the statistician must base his estimates of the relationships. But since these data are the result of a number of such relationships, the difficult problem arises of separating out the relationships from the observed statistics.

Unless steps are taken to make sure that the influences of the several simultaneous relationships on the data can be and have been separated, there is not the slightest justification for the use of any estimation procedure, such as that depicted in Figure I, to compute a statistical relationship. Yet it will readily be recognized how frequently this completely fallacious procedure is employed in practice in the form of simple or multiple correlations computed without any attempt to cope with the simultaneous relationship problem. Let us see now how serious are the distortions, which can be expected to result.

VIII. The Identification Problem

In rather general terms our basic problem can conveniently be divided into two parts:

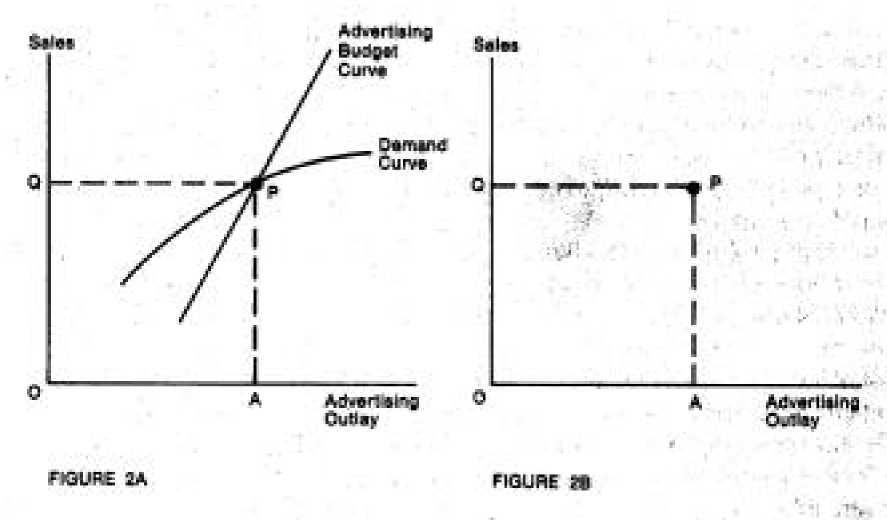
1. In some circumstances the simultaneous relationships (equations) will be so similar in character that it will be impossible to unscramble them (or at least some of them) from the statistics. Such relationships are said to be *unidentifiable.* Presently it will be shown how such an unhappy situation can arise, and it will be indicated that it is unfortunately not unheard of in marketing problems. Clearly, in such a case, we are wasting our time in a statistical investigation of the equation in question. There do exist some mathematical tests, which show whether or not an equation is *identified* (i.e., whether or not it is in principle possible to separate it from the other relationships in the system). These tests should always be applied before embarking on the type of statistical investigation under discussion. It must be emphasized that if an equation happens not to be identified, it is impossible even to approximate the true equation from statistical data alone. Market experiments or other substitute approaches must be employed to obtain this information.

2. Even if an equation turns out to be identified, precautions must be taken to insure that a statistically estimated equation is not distorted by the presence of the simultaneous relationships. We will see in the next section that an ordinary least-squares procedure is likely to lead to precisely this sort of distortion.

In this section we deal with the first of these, the identification problem-the circumstances under which it is, at least in principle, possible to unscramble our simultaneous relationships statistically. To illustrate, let us consider what is involved in finding statistically an advertising-demand curve such as the one, which Figure I attempted to construct in a rather primitive fashion. Now while sales are doubtless affected by advertising, as the advertising-demand function assumes, this function is often accompanied by a second relationship in which what we might call the direction of causation is reversed. It is well known that a firm's advertising budget is frequently affected by its sales volume. In fact, many businesses operate on a rule of thumb, which allocates to advertising expenditure a fixed proportion of their total revenues. For such a business, then, we will have two advertising expenditure demand relationships: (1) the demand function which shows how quantity demanded, Q, is affected by a firm's advertising budget, A : Q = f (A) and (2) the budgeting equation which shows how the firm's advertising decisions are affected by the demand for its product: A = g (Q).

Both of these relationships may actually be of interest to the businessman. The first, as already stated, is directly relevant to his own optimal expenditure decision. The second, if obtained from industry records, will give him vital information about the behavior patterns of his competitors.

The firm's actual sales and its actual advertising expenditure will, of course, depend on both its advertising budgeting practices (the budgeting equation) and on the demand-advertising relationships. In Figure 2 the graphs of two such hypothetical relationships are depicted.



In Figure 2A we show the two curves which the statistician is seeking. We make ourselves, as it were, momentarily omniscient and thus have no difficulty envisioning the true relationships. However, the information available to the statistician is much more restricted, as we shall now see. In our situation the actual advertising expenditure, A, and the volume of sales, Q, are determined, as for any simultaneous equation, by the point of intersection, P, of the two curves.

We now can describe two cases of non-identification:

Case 1. *Neither curve identified.* If the two curves were to retain their shape from year to year, that is, *i f neither o f them ever shifted,* all the intersection points P would coincide or at least lie very close together (Figure 2B). There would only be a single observed point, as in the figure, or the tightly clustered points would form no discernible pattern, and so the shape of neither curve could even approximately be found from the data. We see then, though it may be a bit surprising, that curves, which never shift are from this point of view the worst of all possibilities.

Case 2. *One of the curves not identified* (but the other curve identifiable). This is a case frequently encountered in practice when the demand curve of one firm is investigated. The data form a neat and simple pattern, but what they describe is the firm's inflexible advertising budgeting practices rather than the nature of the demand for its product. In such circumstances what happens is that the budget curve never shifts but the demand curve does. There will then be a number of different intersection points, such as P, P, and P", but they will always describe only the shape of the advertising budget line (Figure 3) . The reader can well imagine how often statistical attempts to find the advertising demand curve have produced neat linear relationships (and spectacularly high correlation coefficients), though what the triumphant investigator has located (without his knowing it) is a totally different curve from the one he was seeking. The situation, which we have just examined, is really ideal from the point of view of the statistician, *provided the relationship, which is not shifting happens to be the one he is seeking.* But the question remains: how is he to know when one relationship is standing still, and even if he somehow knows this, how does he determine which one it is? We will see that in the answers to these questions lies the key to the solution of the identification problem.

It will be shown presently that only where both curves shift over time or from firm to firm or from geographical territory to territory can they ordinarily both be identified. However, in this case the difficult task of unscrambling the two relationships becomes particularly acute. Figure 4 illustrates how three points, A, B, and C, in a diagram similar to Figure 1 might have been generated by three different (shifted) pairs of our curves. It is, noteworthy that the negatively sloping (!) "advertising curve" FP estimated statistically from these points bears not the slightest resemblance to any of the true curves. Nor, since it is merely a recording of points

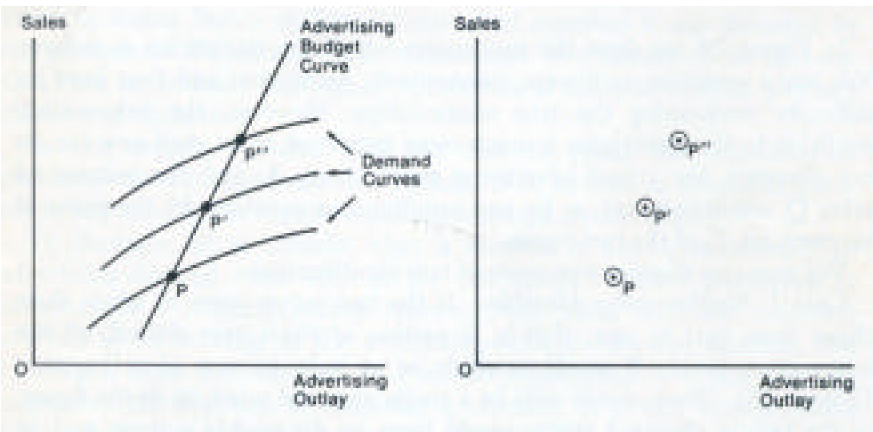


FIGURE 3A FIGURE 3B

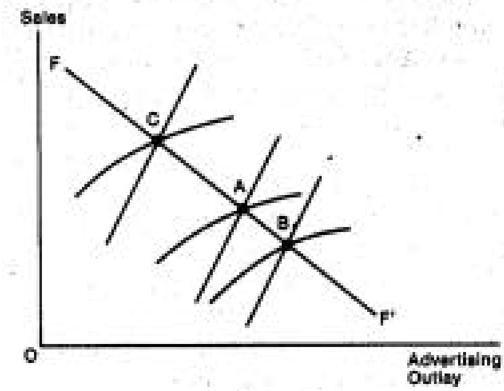


FIGURE 4

of intersection, is there any reason why it should. *The shape of FF' is not even any sort of "compromise" between those o f the budget and advertising demand curves!* We conclude that where simultaneous relationships are present the standard curve-fitting techniques described in Section 4 and Figure 1 may well break down completely. *Their results are likely to bear absolutely no resemblance to the equations, which are being sought!* Such a naive approach may therefore well be worse than no investigation because misleading information is usually worse than no information at all.

Let us now see how one can, in principle, test whether the relationship we are seeking is identified (potentially discoverable by statistical means). First we note that, as the model has so far been described, there is no way of accounting for any shifts in either relationship, which as we have observed, are crucial for our problem. The reason is that only two variables, A and Q, have been considered in the relationships Q = *f* (A) (the demand relationship) and A = g (Q) (the advertising budget equation).

There must, in fact, be some other influences (other variables), which disturb the relationships between Q and A and produce the shifts in their graphs. These additional variables must be taken explicitly into account. As we know, the demand relationship is likely to involve many variables in addition to A. For example, consumer's disposable income is a variable which affects the volume of sales resulting from a given level of advertising expenditure though, very likely, it does not enter the firm's budget calculation explicitly but only indirectly via the effects of income on the sales of the company's product. Similarly, the firm's budget policy may be affected by its past dividend payments, which determine how much it can currently spare for advertising expenditure, but this dividend policy will have little or no effect on the demand curve for its products. Suppose, for the sake of simplicity, that the four variables Q, A, Y (the disposable income), and D (the total dividend payments in the preceding year) are the only ones that are relevant to the problem. Our two relationships then become:

(3) the advertising demand function $Q = f(A, Y)$
and
(4) the advertising budget equation $A = g(Q, D)$.

Here changes in the value of Y are what produce the shifts in the graph of the demand equation, which have been discussed. Similarly, changes in *D* produce shifts in the advertising budget curve.

Now that we have examined how shifts in the two curves are produced we can return to the question of identification. Let us see, intuitively, how the presence of the shift variables in equations (3) and (4) makes it possible, in principle, to separate the relationships from the statistics (i.e., how the shift variables identify the equations). It will be shown now that Y and *D* permit the statistician, at least conceptually, to divide up the statistical information in such a way that he is left with situations like that depicted in Figure 3. Such a situation gives him the information that permits him to infer which of the relationships is shifting and which is standing still. That is, he can determine when one graph is not moving while the other shifts around, so that the resulting dots trace out the graph of the equation which is not shifting, the equation he is trying to estimate. The reader should first be warned, however, that the procedure, which is about to be described is not usually a practical estimation (curve finding) procedure and that other, more sophisticated measures are normally employed fox the purpose.

In Figure 5 we replot the data of Figure 1. Let us, in addition, determine the level of income for each point, Y, for that particular year. Suppose this information is as shown in Table III.

TABLE III.

Advertising Demand point	1950	1951	1952	1953	1954	1955	1956	1957
Disposable Income Y ($ billions)	360	297	295	307	428	381	420	300

We note that the income values for the points representing 1951, 1952, 1953, and 1957 are fairly close together. Hence, if we are convinced that Y is the only variable which makes for sizable shifts in the advertising demand curve, it is reasonable to assume that all four points lie on (or close to) the same curve, that is, among these points there has occurred little or no shift in the curve. We may therefore use these four points (ignoring the others) to locate a demand curve UU' (for income level approximately $300 billion) as shown. Similarly, we can use points for years

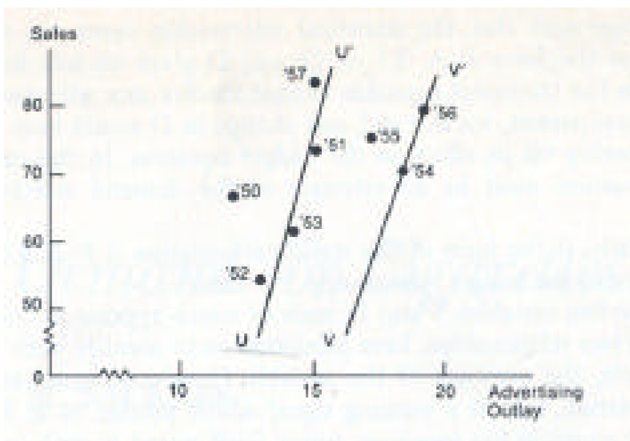


FIGURE 5

1954 and 1956 alone to find the shape of the advertising demand curve W', which pertains to income level approximately $420 billion, etc. In other words, the additional information on the value of Y for each point has permitted us, in principle, to ignore all points which contain information irrelevant to a given advertising demand curve.

We see, then, that if variable Y is present in one equation but not in the other, it permits us, in principle, to discover statistical points over which the budget line has shifted but through which the demand curve remains unchanged. In this same way we were able to trace out a budget line in Figure 3. But it will be remembered that in the situation shown in Figure 3, the demand curve is unidentifiable because the budget curve never shifts. There is no variable such as D in the budget relationship, which will move the budget line about and yet permit the demand curve to stay still. This gives us the following result: *one of a pair of simultaneous relationships will be identified if it lacks a variable, which is present in the other relationship.*

The relevance of the shift variables Y and D for identification can also be seen in another way. Suppose we use some correlation procedure to find a statistical relationship among variables Q, A, Y, and *D.* The system is identified if it is possible in principle for this statistical relationship to be an approximation to either equation (3) (the demand function) or to (4) (the budget function) and if it is possible to find out whether the statistical curve represents (3), (4), or neither. There are three possibilities:

1. The statistical relationship turns out to take the form Q = F(A, Y, *D)* in which all four variables are present (their coefficients are significantly different from zero). In that case we know that the statistics have given us a mongrel function, which resembles neither of the relationships, which we are seeking, for neither of the relationships contains both variables, Y and D.
2. Suppose now that the statistical relationship turns out to have an equation of the form F (A. Y) = Q; i.e., D plays no role in the equation. Then the statistical equation cannot involve any advertising budget function component, for if it did, any change in *D* would have influenced the relationship via its effect on the budget equation. In this case the statistical equation must be an estimate of the demand relationship (3) alone.
3. Similarly, if the form of the statistical equation is F (A, D) = Q, it must represent the budget relationship (4) alone.

Thus the two variables Y and *D,* each of which appears in one and only one of the two relationships, have permitted us to identify both equations. For example, the presence of the variable D, which occurs only in the budget equation, acts as a warning signal, which notifies us at once when the budget equation has somehow gotten itself mixed in with our demand information. We conclude, again, that two simultaneous equations are normally identified (they can, at least in principle, be unscrambled from the statistics) if each equation contains at least one variable, which is absent from the other.

Of course, some more powerful identification criteria exist. For example, an obvious extension of the preceding result is the theorem that in a system of n simultaneous equations, a necessary condition for identification of any one of the equations, say the *i*th, is that every other equation in the system contain at least one variable which is missing from equation i.

This is hardly the place for a systematic discussion of the identification problem, and most of what has been said on the subject has been intended to be intuitive rather than rigorous. However, the reader should have gathered that it is an extremely serious problem and that inadequate attention on the part of the analyst to this problem can easily invalidate his statistical results in their, entirety.