Multiple Regression in Legal Proceedings
Author(s): Franklin M. Fisher
Source: *Columbia Law Review*, Vol. 80, No. 4 (May, 1980), pp. 702-736
Published by: Columbia Law Review Association, Inc.
Stable URL: https://www.jstor.org/stable/1122137
Accessed: 29-08-2019 20:54 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/1122137?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Columbia Law Review Association, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *Columbia Law Review*

Exhibit P-454
This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

Multiple Regression in Legal Proceedings

Franklin M. Fisher *

Multiple regression analysis is a device for making precise and quantitative estimates of the effects of different factors on some variable of interest. It is not a new tool, going back in its origins to Carl Friedrich Gauss, an extremely important mathematician born about 200 years ago. Nevertheless, the practical use of multiple regression has grown very substantially over the past twenty-five years or so. This growth is due partly to the development of modern statistical methods, partly to increasing availability of decent statistical data, and perhaps most of all to the development of the electronic computer. Some of the increasing use of multiple regression and related techniques has occurred in connection with legal proceedings of various kinds, although lawyers and judges have often tended to view such use with general (and occasionally healthy) distrust.

In light of the increasing prominence of multiple regression analysis, it is important for lawyers to understand what it is, how it works, and what it properly can be used for. Perhaps the single most important legal use of multiple regression thus far has been the analyses of the deterrent effects of the death penalty on murder, cited by the Solicitor General in his amicus brief before the Supreme Court in the death penalty cases.[1] The fact that the studies relied on by the Solicitor General were, in my opinion, fatally flawed [2] only adds to the importance of understanding the methodology involved. On a less grand level, multiple regression studies have figured in a number of other legal proceedings, and while the ones with which I am most familiar have been regulatory proceedings, there is no reason why multiple regression should not be used in other litigation as well.[3]

This Article first explains, on a basic level, the concept of multiple regression analysis, its basic properties, and the fundamental assumptions upon which its validity rests.[4] I will also discuss methods of measuring the

---

* Professor of Economics, Massachusetts Institute of Technology. A.B. 1956, M.A. 1957, Ph.D. 1960, Harvard University.

This Article was adapted from a paper delivered before the Association of the Bar of the City of New York (Special Committee on Empirical Data in Legal Decision Making) in May 1979. I am indebted to Michael O. Finkelstein for helpful criticism but retain the usual responsibility for error.

1. Fowler v. North Carolina, 428 U.S. 904 (1976); Woodson v. North Carolina, 428 U.S. 280 (1976); Jurek v. Texas, 428 U.S. 262 (1976); Proffitt v. Florida, 428 U.S. 242 (1976); Gregg v. Georgia, 428 U.S. 153 (1976). See generally Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates (A. Blumstein, J. Cohen & D. Nagin eds. 1978) [hereinafter cited as Deterrence and Incapacitation], which contains, among other things, some devastating discussion of the studies involved. (In particular, see the paper by Klein, Forst & Filatov, The Deterrent Effect of Capital Punishment: An Assessment of the Estimates, id. at 336.)

2. See text accompanying notes 38-46 infra.

3. For an excellent discussion of proceedings using multiple regression studies, see Finkelstein, Regression Models in Administrative Proceedings, 86 Harv. L. Rev. 1442 (1973).

4. I have been responsible for several multiple regression studies used in legal proceedings and, because I know them best, it is those studies on which I shall draw for examples for

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

accuracy and reliability of estimates generated by multiple regression. The second part of the Article explores in greater depth the proper use of multiple regression in legal proceedings by focusing on three areas in which multiple regression studies might play a role—the examination of wage discrimination, the determination of antitrust damages, and the evaluation of punishment as a deterrent to crime.

I. Multiple Regression Analysis

A. *Uses of Multiple Regression*

The two primary uses of multiple regression analysis are best illustrated through an examination of actual situations in which multiple regression studies were employed. Consider the following two cases:

> 1. For many years after the disappearance of coal-burning locomotives, there was a perennial labor dispute concerning the preservation of the jobs of railroad firemen. Whatever the merits of that dispute (ultimately resolved, I believe, through negotiation), one of the issues in it concerned the question of whether the presence of a fireman on trains contributed to railroad safety. A study of that issue, using multiple regression, was presented in testimony before a Presidential emergency board in 1970.[5]
>
> 2. Cable television systems (CATVs) have been the subject of repeated rulemaking proceedings by the Federal Communications Commission. Among the issues involved in such proceedings is the effect of the entry and activity of CATVs on the profits and growth of broadcast television stations. This issue involves such questions as the influence of CATVs on the viewing audience reached by particular broadcast stations and the effect of changes in a station's audience on the revenue it receives.[6] In general, as one would expect, cable operators have claimed such effects to be small and broadcast stations have insisted they are large. The problem has been studied repeatedly by multiple regression methods, most recently in a study of the relationship between audience size and revenues authored in part by me and submitted to the FCC in 1978-79.[7]

---

much of this Article, hoping thereby to put some more interest into what otherwise might degenerate into a fairly dry and technical discussion.

5. The study is most conveniently reported in Fisher & Kraft, The Effect of the Removal of the Firemen on Railroad Accidents 1962-1967, 2 Bell J. Econ. & Management Sci. 470 (1971).

6. These are important questions for the FCC since they bear on the extent to which regulation of cable television is needed to foster the growth of new UHF stations or to maintain the profits that subsidize the public service and other programming of local broadcast stations.

7. Charles River Associates, The Audience-Revenue Relationship for Local Television Stations (1978) (FCC Docket No. 21284); Charles River Associates, The Value of Different

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

In the first case, the issue is whether or not a particular variable (presence or absence of firemen) has *any* effect on some other variable (railroad safety). In the second case (the audience-revenue relationship), there is not much doubt that audience size affects television revenue—viewer attention is what stations sell to advertisers, and all parties are vitally interested in audience statistics; the problem is rather one of measuring the effect. These two uses of multiple regression are what statisticians call "testing hypotheses" on the one hand and "parameter estimation" on the other. In the first type, one wishes to be able to state whether or not something is true. In the second, one is more interested in the precise magnitude of the effects involved. Obviously, the two questions are closely related.

There is a third, but less widespread, use to which multiple regression analysis can be put: to *forecast* the values of some variable. A multiple regression analysis shows how certain independent variables affect a dependent variable. From that analysis, and from a forecast of the values of the independent variables (obtained from some other source), one can generate a forecast of the dependent variable. This type of "unconditional forecast" is not always useful—which is fortunate, since such unconditional forecasts tend to be relatively inaccurate. Far more often what is of interest is a "conditional forecast"—a prediction of what will happen to the dependent variable if another variable is changed or, looking retrospectively, what would have happened to the dependent variable had the value of an independent variable been different.

Consider the two examples already described. The question in the case of the railroad firemen did not really involve predicting the number of railroad accidents. Rather, it involved trying to decide whether the number of those accidents would be significantly greater if the railroad firemen were no longer employed. Similarly, while prediction of television station revenue would be desirable for some regulatory ends, the primary issue in the audience-revenue study was systematic measurement of the effects on revenue of changes in the size and socio-economic characteristics of a station's audience.

The firemen example best brings out the problems involved in making such predictions. By their nature, railroad accidents involve random, chance events. Even the accident *rate* (however measured) is subject to such chance fluctuations. Simply determining whether the presence or absence of firemen makes a significant difference to the railroad accident rate may be easier than predicting the rate itself with great precision. One of the distinctive characteristics of multiple regression analysis is that it is able to provide information about the effects of the variable of interest (in this case the employment of

---

Day Parts in the Audience-Revenue Relationship for Local Television Stations (1979) (FCC Docket No. 21284). The first study of the problem was, I believe, the one that I gave as written testimony to the FCC in 1964. It is most conveniently reported in Fisher, Ferrall, Jr., et al., Community Antenna Television Systems and Local Television Station Audience, 80 Q.J. Econ. 227 (1966). Both studies were done on behalf of the National Association of Broadcasters.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

the firemen) on the dependent variable (here, railroad accidents) without necessarily being able to predict the dependent variable itself with great accuracy.

In a way, one might describe multiple regression as a method used to extract a systematic signal from the noise presented by data. There are two primary problems involved in extracting such a signal. First, it is typically the case that the factor whose influence one wishes to test or measure is not the only major factor affecting the dependent variable—for example, the amount of traffic on the railroads has something to do with accidents as well. Second, even if one can somehow account for the effects of the other important systematic factors, there typically remain chance components.

If we could make controlled experiments, it would be relatively easy to quantify the relationship being investigated. A controlled experiment in the audience-revenue case, for example, would vary station audiences and the other variables expected to influence revenue one at a time, holding everything else constant and observing the resulting revenue. Obviously, this is impossible—there is no way we can tinker with station audiences. This means that we must be content with analyzing, as it were, the experiments performed by nature, in which more than one of the variables deemed likely to affect revenue move at the same time.

Moreover, even if we could control station audiences and hold constant the variables that we believed to be important, we would not know enough about the audience-revenue relationship to be sure of holding constant all the variables that actually affect the revenues of an individual station. It may be, for example, that the personality and effectiveness of the stations' sales representatives or the advertising policies or publishing quality of competing newspapers affect revenue. These variables are hard to measure, let alone hold constant.[8]

Inability to perform well-controlled experiments is not uncommon. Indeed, it occurs even when one is making so-called controlled experiments in the natural sciences. The difference there is that one can be fairly sure that the uncontrolled effects that one does *not* know about in detail are extremely small. When dealing with observations from the economic system (or, indeed, from any system in which the experiments are performed by nature rather than by the experimenter), there is likely to be a nontrivial, residual element of unexplained effects on the variable of interest, even after one has taken account of the major systematic effects. Multiple regression is a way of dealing with these difficulties.

B. *How Multiple Regression Works*

1. *An Overall View.* In multiple regression, one first specifies the major variables that are believed to influence the dependent variable. In our

8. Similarly, in the case of the firemen, even if we could experiment with firemen employment, we could not hold railroad traffic constant. Moreover, other variables affecting safety (the ones we call "chance") are never known precisely.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

examples, this means specifying the important or systematic influences that may affect station revenue or railroad safety. There inevitably remain minor influences, each one perhaps very small, but creating in combination a non-negligible effect. These minor influences are treated by placing them in what is called a random disturbance term and assuming that their joint effect is not systematically related to the effects of the major variables being investigated—in other words by treating their effects as due to chance.[9] Obviously, it is very desirable to have the random part of the relationship small, particularly relative to the systematic part. Indeed, the size of the random part provides an indication of how correctly one has judged what the systematic part is. Multiple regression thus provides a means not only for extracting the systematic effects from the data but also for assessing how well one has succeeded in doing so in the presence of the remaining random effects.

The relationship between the dependent variable and the independent variable of interest is then estimated by extracting the effects of the other major variables (the systematic part). When this has been done, one has the best available substitute for controlled experimentation. The results of multiple regressions can be read as showing the effects of each variable on the dependent variable, holding the others constant. Moreover, those results allow one to make statements about the probability that the effect described has merely been observed as a result of chance fluctuation.

2. *Estimating Multiple Regressions.* Suppose that the relationship to be examined is to include only two variables, the dependent variable (Y) and one independent variable (X). Suppose further (for simplicity of exposition) that it is believed that the relationship between these variables is a straight line.[10] Such a relationship could be expressed mathematically as:

$$(1) \quad Y = a + bX$$

or, diagrammatically, as in Figure 1. The problem for the investigator is to discover the values of the parameters, a and b (i.e., the intercept and slope of the line). If the relationship really were exact—if there were no random influences at all—this would be extremely easy to do. One would need only to observe two points with different values of X. Since two points determine a line, it would require only routine arithmetic calculation to find the line they determine.

In real life, however, the relationships to be fitted are not exact. Rather

9. The disturbances (the random or unsystematic part) will then affect the dispersion of the true values of the dependent variable around the values that would be predicted from the systematic part alone.

10. I chose the straight line case as the easiest to understand, but the theory is not so restricted. There is nothing to prevent one or more of the variables in equation (1) from being a square, a logarithm, or the ratio or two other variables. Many (not all) mathematical relations can be cast into the form of equation (1) by transforming or redefining the variables. Furthermore, most nonlinear relationships can be at least approximated by straight lines.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

there are random influences on the dependent variable, as described above. Hence, the correct relationship is not equation (1) but rather:

$$(2) \quad Y = a + bX + u$$

where u represents the random influences. Different values of u will produce different values of Y which will be either above or below the line; indeed, they will produce a scatter of points such as that shown in Figure 1. The task for the investigator is to cut through the noise generated by these random influences and extract the "signal," namely, the line around which the points

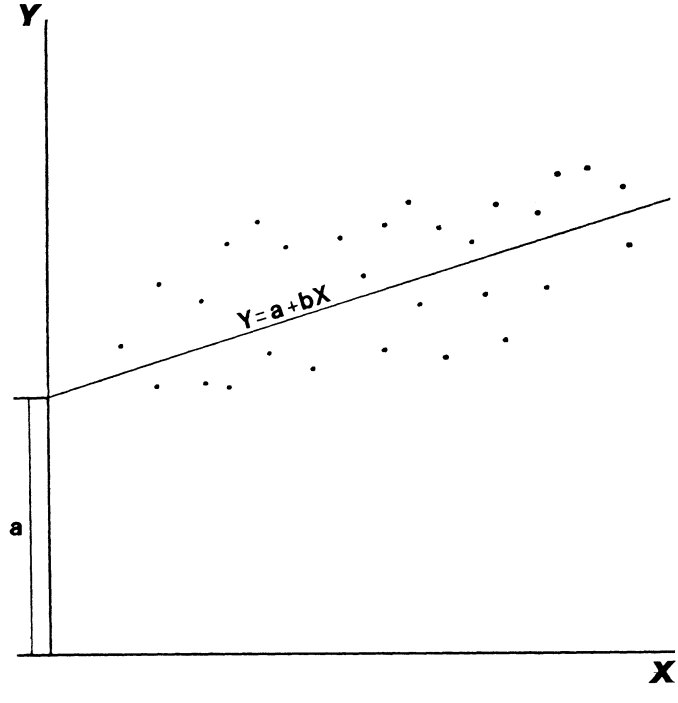

Figure 1

are scattered. This is done by picking the line that best fits the scatter of points in the sense that the sum of the squared deviations between predicted and actual Y values is minimized.[11] This is called "least squares regression." (The adjective "multiple" is used when there is more than one X.)

11. Using the sum of squared deviations gives equal weight to positive and negative deviations. Further, in a multi-dimensional diagram (not drawn) it can be shown that there is a sense in which minimizing the sum of squared deviations amounts to minimizing the

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

3. *Assumptions of Least Squares Regression.* In practice, least squares regression is not done diagrammatically but numerically (generally by computer), resulting in numerical estimates of a and b. The relation that such estimates are likely to have to the true values of a and b depends on the assumptions one is willing to make about the random disturbance term, u, and its relationship with the independent variable, X, whose effect on Y (represented by b) is to be measured.

There are essentially three major assumptions involved: (a) that the effects of the random disturbance term are independent of the effects of the independent variable; (b) that the values of the random term for different observations are not systematically related and that the average squared size of the random effect has no systematic tendency to change over observations; and (c) that the sum of random effects embodied in the disturbance term is distributed normally, in the "bell curve" generally characteristic of the distribution of the sum of independent random effects.[12]

The validity of these assumptions bears on the effectiveness and reliability of least squares analysis. Various properties of multiple regression depend on the accuracy of the assumptions, different properties involving different assumptions. Moreover, the dependence is cumulative: if the early assumptions are invalid, the properties associated with the later assumptions are not likely to be present. In situations where the assumptions may fail, the use of multiple regression analysis is likely to be inappropriate.[13]

a. *Independence of the Disturbance Term.* The fundamental assumption of least squares regression is that the uncontrolled effects of the random disturbance (u) are in an appropriate technical sense independent of the controlled effects of the independent variable (X). (Alternatively, this can be expressed as the assumption that the disturbance term has a zero mean whatever the value of X. In repeated samples, the disturbance term for any given X is neither positive nor negative on the average.) If this were not so, then attempting to determine the effects of X on Y could not be done simply by observing different X's and trying to average out the effects of u. In such a case, movements in X would be systematically associated with movements in u and, without a great deal of care, the estimates of b would include not merely the direct effects of X on Y, but also the associated effects of movements in the disturbance term, u.

When is such an assumption likely to fail? The simplest case to understand occurs when some large and systematic factor, other than X, has been

distance between the point which represents the actual values of the dependent variable and the point which represents the values one would predict from the regression. Average squared values are the standard statistical measure of dispersion.

12. The word "normal" here is a term of art referring to the shape of the distribution. The name indicates that the distribution involved is characteristic of many random variables. Most important, if a random variable is composed of the sum of other random variables acting independently, that sum tends to be distributed normally. This makes the assumption of normality the obvious one unless there is a compelling reason to depart from it.

13. As a general rule, there are methods of testing for and dealing with the failure of such assumptions, but they involve the more advanced tools of econometrics rather than least squares regression.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

left out of the analysis; this is called misspecification. In the revenue-audience study, for example, it turns out that average household income as well as size of audience affects television station revenue. Suppose, however, that we had not thought of this but had simply tried to estimate the effect of audience size on revenue. (Here, revenue would be Y and audience size would be X.) In effect, this would mean that we were placing household income in the disturbance term. Yet, if household income across television markets is positively associated with audience size, then part of what we would attribute to larger audience size would in fact be attributable to higher income. In other words, we would have failed to control for income levels and the lack of such control would matter.

Obviously, the assumption that one has controlled for all the important influences is basic to any attempt to measure those influences correctly. There are, however, other ways in which the assumption of independence between random disturbance and included factors can be violated. In general, this will happen when there exist relations between the dependent and independent variables in addition to the relation being estimated. I shall discuss specific examples of such cases in part II.

If the assumption of independence between u and X is warranted, then least squares estimates of the parameters (a and b) will have some desirable properties. First, the estimates will be *unbiased*—they will be correct on the average. This means that if one did the calculations for a sample of a particular size, and were then to repeat the procedure on numerous samples of the same size, each time obtaining different estimates for a and b, the average of the estimates so obtained would be the true values of a and b. To put it a little differently, least squares estimates have no tendency to err systematically on either the high side or the low side.

Further, if the assumption of independence is correct, least squares estimates will be *consistent*. The property of consistency means that, as the sample size increases, the probability of obtaining least squares estimates that differ from the true values by more than any given amount goes to zero. Thus, as more data become available it will become easier to extract the true values of a and b from the noise presented by the random part.

b. *Behavior of the Disturbance Term.* Consistency is the minimal property that one wants an estimator to have. But there are many consistent estimators and, in some situations, even many unbiased ones. Moreover, unbiasedness assures only that the estimator is right on the average; it does not indicate how far off it is likely to be on any given sample. Similarly, consistency guarantees only that one will get close to the true values of the parameters if one knows enough; it cannot determine how much one needs to know to get close. It is clearly desirable to have measures of reliability—that is, measures of how far off one can generally expect estimates to be. Moreover, within the class of unbiased or consistent estimators, it is obviously desirable to choose the one likely to be most reliable.

With an additional assumption, least squares regression turns out to be

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

such an estimator and will itself generate estimates of its reliability. This assumption concerns the nature of the random disturbance term (u), rather than an assumption concerning its relation with X. The assumption can be divided into two parts.

First, it is assumed that if one had information about the value of u for some observations, one would not thereby gain any information about its value for other observations. For example, if the observations are on the variables over time, an unusually high and positive value for u should not be followed by a tendency for u to be high the next year. Rather, successive values of u should be independent of each other. One can see why this is likely to matter. Least squares regression is a generalized form of averaging. Averaging is an excellent way to take care of random noise, provided that one is averaging over independent events. If the random disturbances from different observations are not mutually independent, however, then the averaging involved in least squares regression will not defuse the random effects. In such a case one could do better by expressly assuming that a high disturbance term in one period indicates something about the value of the disturbance term in the following period, and then using this information to attempt to factor the disturbance out of the equation.

Second, it is assumed that there is no systematic tendency for the random disturbance (u) to be either big or small.[14] To put it differently, one assumes that the chances of a large random effect versus a small one are the same for all observations.[15] Again, one can see why this will matter. If some observations tended to have larger random effects than others, then the observations with large random effects would contain less reliable information than would the observations with small random effects. In any averaging procedure, one would want to give more weight to the latter. Since least squares regression will treat all observations equally, it will not take this into account.[16]

These assumptions will be violated if, when dealing with a series of observations over time, the disturbance term includes the effects of variables that behave systematically over time. Certainly, this is a serious possibility

---

14. We have already assumed that the random disturbance term has no systematic tendency to be high or low—that is, that it has a mean, or expected value, of 0 for all values of X. ("Expected value" is to be thought of as the population mean. Roughly speaking it is the average value one expects to obtain if one takes a large enough sample.) That assumption involves the algebraic sign of the random disturbance term. The present assumption, on the other hand, has to do with the absolute magnitude of the disturbance term, regardless of sign. Put more precisely, the dispersion of a random variable is measured by the average or expected value of the squared deviation from its mean. This is called the "variance." Its square root is called the "standard deviation." The assumption previously made in the text was that the mean of the random disturbance term is not systematically related to X. The assumption now being made is that the variance or standard deviation of the disturbance term is not so related and is, in fact, the same for all observations.

15. Technically, this is the property that the variance of the disturbance term should be the same for all observations.

16. There are ways of taking this failure of assumption into account: not surprisingly, the technique involved is called "weighted least squares," a variety of "generalized least squares."

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

in econometric models. Similarly, if the observations are of individual entities, such as firms, it may very well be that the effects of particular uncontrolled events (such as political events) will be larger for large firms than for small ones. In such a case, the second part of the assumption would be violated. As with all the assumptions of least squares regression, however, one would want to be sure that the violations are really important before abandoning regression analysis. In the cases posited above, small departures from the assumptions would have small effects. Furthermore, the properties of least squares associated with the assumptions are so strong as to make least squares regression superior to the alternative estimators that would result from trying to cure such small departures.

Given the validity of the assumptions under discussion, least squares estimates will be *efficient*. This means that, within a wide class of unbiased and consistent estimators, least squares estimates will have the smallest variation. Thus, if one could take repeated samples, the variation of the least squares estimates around the true values of a and b would be less than the variation of other unbiased and consistent estimators; in short, the least squares estimates will be more reliable.

c. *Normality of Distribution.* The last assumption of least squares imposes greater restrictions on the random disturbance, u, than the ones already discussed. The assumption is that u, for all values of X, follows the normal distribution (bell curve),[17] with a mean of zero, as already assumed. This, however, is not as restrictive as it may appear. As a general proposition of statistics, the normal distribution is characteristic of large averages of independent random effects. To the extent that the error term is made up of the sum of small random effects, that sum will tend to be distributed normally.[18]

The normality assumption, in addition to bolstering least squares' property of efficiency, implies the ability to make precise probability statements concerning how far off the least squares estimates are likely to be.[19]

4. *Multiple Independent Variables.* In practice, one does not usually work with relationships involving only two variables, but rather with relationships in which a dependent variable is influenced by many independent ones (railroad traffic as well as firemen employment; audience income as well as audience size). Denoting the independent variables as $X_1, X_2, \ldots, X_k$, the relationship to be estimated (assuming linearity) [20] can be expressed as:

$$(3)\quad Y = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k + u$$

17. See note 12 supra.

18. See note 12 supra. The "normal" distribution is completely characterized by its mean and variance. It is hard to construct practical examples in which one would be inclined to question normality without also questioning the other assumptions about the random disturbance term. Hence, while there are tests for departure from normality, they are hardly ever used.

19. See text accompanying notes 24-28 infra.

20. Again, I have chosen a linear form here. Least squares theory runs mostly in terms of such forms, but this is not as restrictive as it might appear, since many nonlinear forms

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

If there is only one independent variable, this is the case already considered, the case of a straight line. When there are two independent variables, one is fitting a plane to a scatter of points in space. When there are more than two independent variables, one is fitting a hyperplane (the generalization of a plane to more than three dimensions), but the principles are still the same, although the visualization is no longer immediate. Least squares still retains all the properties listed for the simple case above.

Least squares regression takes advantage of the fact that the independent variables seldom move in perfect step together but rather move (as the name suggests) independently. By determining how the dependent variable changes when the independent variables move in a variety of different ways, the effect of each of the independent variables is extracted.

This kind of systematic extraction of the effects of each variable is important. Examination of raw data leads to facile, and sometimes erroneous, conclusions. Over time, for example, removal of firemen and increased numbers of accidents both occurred. That these events were causally connected cannot be concluded if both are also associated with increases in a third variable (railroad traffic) that plausibly affects railroad accidents. Only by systematically using the fact that railroad traffic, while associated with fireman employment in the data, is not perfectly so associated, can one find out about the independent effect of the firemen. Not controlling for railroad traffic would place it in the disturbance term of equation (3) and violate the fundamental assumption of least squares that disturbance terms and independent variables are independent.

---

can be cast into a linear form similar to equation (3) by appropriate transformations of the variables.

The basic assumption involved in linearity is that the effect of each independent variable on the dependent variable is independent of the level of the other independent variables. Thus, in the firemen example, linearity would imply that the effect of the presence of firemen on the number of railroad accidents was the same at high levels of traffic as at low levels. It would also imply that the effect was the same regardless of whether there were other crew members substituting for the firemen. Obviously, these are not assumptions on which one necessarily wants to rely.

Fortunately, it is not necessary to rely on them. If one thought, for example, that two of the variables—say X_1 and X_2—interacted, then one could define a new variable X_3 as the product of X_1 and X_2. Least squares regression would then proceed as if X_3 were simply a different variable, but its coefficient would tell you something about the importance of such interaction.

To take a different example, it is often not very plausible to suppose (as linearity does) that the effect on the dependent variable of changing an independent variable by one unit should be the same in absolute terms for all levels of the independent variable. It is frequently more plausible to assume that a one percent change in an independent variable has a constant percentage effect on the dependent variable. Such cases can be treated within the framework of linearity by entering into equation (3) not the original variables themselves, but rather their logarithms. This is frequently done and has the advantage, as well, of assuming that the effect of the random error on different observations is likely to be of the same size in percentage rather than absolute terms, a matter that came up above in the discussion of one of the least squares assumptions. See text accompanying notes 14-16 supra.

In general, the choice of the form in which to enter the variables or, more generally, the form of the relationship requires serious thinking about the way in which the relationship being estimated is likely to work. As with deciding which variable to include in the relationship in the first place, this must be done in large part by thinking about the problem rather than by hoping that the data will provide the answer. In any case, relations such as equation (3) are substantially more general than might appear at first sight.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

As this description suggests, it is very important that variables do in fact move somewhat independently. Suppose, for example, that in the revenue-audience study one wished to investigate the separate effects on station revenue of audiences close to a station (X_1) and audiences located farther away (X_2). Suppose, as is not the case, that whenever the nearby audience increased by ten percent, as one went from station to station, the far-away audience also increased by ten percent. Then, although one would be able to determine the influence on revenue of the *total* audience, one could not find out the separate effects on revenue of the two subdivisions of that audience. No "experiment performed by nature" would have separated those effects in any way.

Such an extreme situation is not generally encountered in practice; rather what is encountered is something close to it. Suppose that every time the nearby audience went up by ten percent, the far-away audience went up by amounts that varied only slightly up or down from ten percent. In that case, it would be possible to estimate the separate effects generated by each subdivision of audience size, but one would be very uncertain about the estimate. Nature would not be performing experiments calculated to separate those effects with any high degree of accuracy. Such a circumstance is called *multicollinearity*—so called because it involves an additional linear relationship between the variables on the right hand side of the equation.

Obviously, the less multicollinearity is present, the better able one will be to separate out the effects of interest. Unless multicollinearity is perfect, however, multiple regression will be able to separate the effects to some extent and, again, will do so more precisely than any other method, producing estimates with the properties discussed above as well as measures of the reliability of these estimates. The effects of multicollinearity will show up in such reliability measures (standard errors), as discussed below.[21]

5. *Erroneous Inclusion or Exclusion of Variables.* The discussion thus far has presumed that the true systematic relationship is the one being estimated. To put it another way, we have already seen in discussing unbiasedness that multiple regression retains the desirable properties associated with it only if one has in fact included all the variables likely to have a large effect on the dependent variable and can safely assume that the remaining effects are not correlated with the independent variables included. In the audience-revenue study it was thus necessary to control for household income and not place it in the disturbance term. It is therefore important to

21. See text accompanying notes 24-28 infra.

Note that the problem here occurs when two of the independent variables move together in an approximately linear fashion. If they move together nonlinearly, there will not be so severe a problem. If what is involved is not another relation between two or more of the independent variables but another relation between the *dependent* variable and an independent variable, then the basic assumption of least squares will be violated and we will have a situation involving simultaneous equations as discussed below. See text accompanying notes 34, 35 & 43-45 infra.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

proceed by including at some stage all the variables that one might think could possibly have a significant effect on the dependent variable. In general, one does this by first examining those variables that one thinks are actually important and then asking what happens when additional variables are included.

Note that this must be done by specifying *in advance* what variables are thought to be important. To proceed by first looking at the data and then including those factors that appear correlated with the dependent variable is a recipe for spurious results. It leads to a situation where no true test of the estimated relationship can be made. In addition, it is likely to leave out variables that truly belong in and thus lead to invalid as well as untested results. The measurement provided by least squares regression is a way of making theoretical assumptions precise or of testing them; it is not a substitute for thought.

I mention this emphatically because a number of packaged computer programs that are sometimes used involve what is known as "step-wise regression." Such programs build up multiple regressions in ways similar to the following. First, the program finds the independent variable in the list most correlated with the dependent variable and does a regression involving it. It then looks at the sample deviations from the regression (the differences between actual and predicted values) and asks whether those deviations are correlated with another independent variable. If so, it puts in the variable most correlated with those errors and so forth. This is not recommended. In the first place, even if none of the independent variables have anything to do with the dependent variable, proceeding in this fashion is very likely to produce the appearance of a high correlation in a particular sample. Second, variables that in fact belong in the relationship but that are correlated with the independent variables used early in the procedure tend never to get in. In general, such computer programs suffer from the same problems as attempts to look by eye at bilateral relationships that in fact involve the influence of many variables: they are likely to attribute the effects of the omitted variables to the included ones and result in biased estimates.

The opposite of building regressions up one variable at a time is to put many variables in and then see whether some of them should come out. This is a somewhat better method. Whereas there is a major effect from excluding a variable whose true coefficient is far from zero, the effect of erroneously including a variable whose true coefficient is zero is of very little consequence. Such a variable can be thought of as actually present in the relationship, with the zero coefficient simply indicating that the variable has little or no effect. The multiple regression technique then estimates that coefficient along with the other true coefficients; thus, the regression technique must extract one more parameter from the same number of observations. This is equivalent to having one less observation with which

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

to extract the nonzero parameters.[22] If the sample size is large (there are more than five hundred television stations in the United States, for example), there will be only a very small effect on the estimates of the remaining coefficients and on the prediction of the dependent variable (unless the inclusion of the extra variable adds to multicollinearity). The reliability measures and the measures of "goodness-of-fit" [23] will take full account of the slight reduction in information involved. Where possible, therefore, it may be best to start with an overly complex model and build down.

Nevertheless, it is important to realize that such building down cannot be done without an antecedent theory; the use of computer programs that do "backwards step-wise regression" is not recommended. Without some theory about which variables are likely to matter, throwing a great number of variables into the hopper is likely to lead to spurious results. If one tries enough combinations of variables, then, in a particular sample, one will tend to get some relationship that appears to fit well. Therefore, a properly done study begins with a decent theoretical idea of what variables are likely to be important. It then can proceed to test well-defined hypotheses about additional variables. But a study that casts about for a good-looking relationship by trying all sorts of possibilities is very likely to come up with relationships where none exist.

This leads directly to two comments relevant to lawyers. First, when having a study done by an expert, one should not be too insistent about covering every possibility at once. Rather, one should make sure that the expert proceeds by estimating a reasonable model including the major variables and then goes on to test other possibilities. If one insists that all possible variables are likely to be of equal importance, one is likely to end up with a rather doubtful result.

Second, when faced with an opposing expert who has done a regression study, one should find out how the expert decided on the variables he included and how many different combinations of variables and models he tried before settling on the one that is being presented. If the basic model was tried relatively early and variations were then tried simply to see if anything else seemed to matter, the study may be sound. If, however, the basic model being presented is the end result of vast amounts of computer work, particularly mindless and mechanical computer work, then one may have a legitimate point of attack.

---

22. This is because, for the purpose of assessing reliability of the regression estimate, what matters is the number of "degrees of freedom"—the excess of the number of observations over the number of parameters to be estimated. The following conveys some idea of what is involved. One can always fit a line to two observations, but there are no degrees of freedom and no way of assessing the reliability of the result. If one has a third observation, then one cannot always fit a line exactly but some notion of reliability can be gained from observing how close one comes in fact. Add another variable with a coefficient to be estimated, however, and one is estimating a plane that can be fitted precisely to three observations. Thus, the addition of another coefficient to be estimated has the same effect as the removal of one observation.

23. See text accompanying notes 24-28 infra.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

C. *Measuring "Goodness-of-Fit"*

As I have already mentioned several times, least squares regression not only estimates the effects of the variables involved in the model but also measures the certainty or accuracy of such estimates. In addition, it provides overall measures of how well the model fits the data as a whole. There are several different measures involved and because they each measure different things, it is important to be clear on the differences among them.

1. *Standard Errors of Coefficients and t-Statistics.* Associated with the estimated value of each regression coefficient (a and b in the above equations) is a figure known as the standard error [24] of that coefficient, which measures the coefficient's reliability. In general, the larger the standard error, the less reliable or the less accurate is the estimated value of the coefficient.

Speaking somewhat loosely, in large samples the chances are nineteen out of twenty that the true coefficient lies within approximately two standard errors of the estimated coefficient. The chances are ninety-nine out of one hundred that it lies within approximately two and one half standard errors of the coefficient.[25] (In small samples the bounds tend to be wider.) Thus, for example, if the estimated coefficient is ten with a standard error of two, the chances are nineteen out of twenty that the true coefficient lies between six and fourteen and ninety-nine out of one hundred that it lies between five and fifteen. To say that the chances are nineteen out of twenty that the true coefficient lies between six and fourteen, however, does not mean that the true coefficient is equally likely to be in any part of that range. The single most probable figure is ten. The probability of matching the correct figure decreases as one moves away from ten and, as the slight difference between the six-to-fourteen and five-to-fifteen ranges indicates, that probability decreases very fast as one moves substantially away from the middle estimate.

It is conventional to use the standard error of an estimated coefficient to make a statistical test of the hypothesis that the true coefficient is actually zero—i.e., that the variable to which it corresponds really has no effect on the dependent variable. Essentially, such statements are constructed by asking how likely it is that ranges of the sort just described will include zero. This is done by taking the ratio of the estimated coefficient to its standard error. Such a ratio is called a t-statistic.

---

24. As explained in note 14 supra, the two basic measures of dispersion of a random variable are its variance, the average square deviation around its mean, and its standard deviation, the square root of the variance. The standard error of a statistic (here, the standard error of a regression coefficient) is, in a rough sense, its expected standard deviation. More precisely, it is the square root of the average squared deviation that one would expect to obtain if one used the same estimating procedure over and over again. It is a convenient measure of the reliability of the statistic with which it is associated since the probability that the statistic differs from the true value by any given amount depends directly on the number of standard errors that the amount represents.

25. This will depend on the normality assumption, discussed at text accompanying notes 17-19 supra.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

In large samples, a t-statistic of approximately two means that the chances are less than one in twenty that the true coefficient is actually zero and that we are observing a larger coefficient just by chance. In such a case, the coefficient is said to be "significant at the five percent level." A t-statistic of approximately two and one half means that the chances are only one in one hundred that the true coefficient is zero; in that case, the coefficient is "significant at the one percent level." [26] (In small samples, t-statistics must be larger for a given significance level.) In the numerical example given, the t-statistic would be five (ten divided by two) and the probability that the true coefficient is zero is extremely small. The coefficient would be significant at much better than the one percent level.

Significance levels of five percent and one percent are generally used by statisticians in testing hypotheses. That is, given a significance level of five percent (or one percent for a stricter researcher) it is safe to assume that the true coefficient is not zero and that therefore the variable being tested has some effect on the dependent variable in question. Some lawyers might question whether the use of such levels imposes too severe a standard. Why reject the hypothesis that a certain coefficient is zero only if the probability that the results obtained are due to chance is five percent or less? Where the hypothesis involved is of legal importance (for example, when a nonzero coefficient would indicate the presence of sex discrimination in wages), would it not make more sense to use a "preponderence of the evidence" standard and require only significance at fifty percent?

Such an approach, however, would reflect a flawed understanding of what significance levels really mean. In particular, a significance level of fifty percent would not correspond to a "preponderence of the evidence" standard. The significance level tells us only the probability of obtaining the measured coefficient value *if* the true value is zero; it does *not* give the probability that the coefficient's true value *is* zero, nor does subtracting the significance level from one hundred percent give the probability that the hypothesis is not true. Because, even with a large sample, it is quite possible to obtain results differing from a coefficient's true value, it is conventionally thought that there must be a very high probability that the coefficient is not zero before it can be conclusively claimed that the variable associated with the coefficient has a definite effect on the dependent variable.

This does not mean that only results significant at the five percent

26. The examples of significance given in the text are for what is known as a "two-tailed test." For example, the significance level of five percent associated with a t-statistic of about two is the probability of obtaining an estimated coefficient as large as that actually obtained, *either positive or negative,* if the true coefficient is actually zero. In many situations, for example, there is no issue as to whether or not a particular coefficient is positive or negative; rather, the only issue may be whether it is positive or zero. In such a circumstance, the appropriate test is a "one-tailed test" in which five percent would represent the probability of observing some *positive* coefficient if the true value were really zero. The t-statistic required for significance at a given level on a one-tailed test is less than that required for the same level on a two-tailed test. In the case of five percent, for example, what is required is approximately 1.6 rather than 2.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

level should be presented or considered. Less significant results may be suggestive, even if not probative, and suggestive evidence is certainly worth something. In multiple regressions, one should never eliminate a variable that there is firm theoretical foundation for including just because its estimated coefficient happens not to be significant in a particular sample.

Nevertheless, the computation of the standard errors of the coefficients or the corresponding t-statistics is a matter of considerable importance. It is routinely done by all professionals, with the five and one percent significance levels generally accepted as the point at which the zero hypothesis is rejected. Failure to report such measures of reliability is a clear signal that the study is suspect.

2. *The Standard Error of Estimate.* Another statistic often reported with the results of least squares regression is the "standard error of estimate" or "standard error of the regression." This is not to be confused with the standard errors of the coefficients. The standard error of estimate is one of the summary measures reflecting the degree to which the estimated regression line or plane fits the data. In terms of the discussion given earlier, it is an estimate of how widely the points are scattered around the line.

More precisely, the standard error of estimate describes the average deviation of the actual values of the dependent variable in the sample from the values that would be predicted from the regression.[27] Thus a standard error of zero would correspond to a perfect fit. The larger the standard error of estimate, the poorer is the fit, in the sense that the more important is the random component not being explained.

The size of the standard error of estimate will depend upon the units in which the variables are measured. For example, if we were to measure the dependent variable in pennies rather than in dollars, the standard error of estimate would also be in pennies rather than in dollars and would therefore be multiplied by one hundred. To judge whether the standard error of estimate is large or small, therefore, one must compare it with something else. One such comparison involves computation of the correlation coefficient, discussed below. Other comparisons involve looking at, for example, the mean value of the dependent variable and determining what percentage of that value the standard error is. In general, the standard error of estimate can be used to make probability statements about how far off forecasts from the model are likely to be. Around the mean of the sample (if the sample is of considerable size), forecasts are likely to be off by more than approximately two standard errors of estimate only once in twenty times.[28]

---

27. It is in fact not computed as an arithmetic average. Rather, it is the square root of the average squared deviation in the sample (with an adjustment for degrees of freedom, see note 22 supra).

28. Related to the standard error of estimate, but not identical to it, is the standard error of forecast. This is a measure of how reliable forecasts made from the regression equation are likely to be. More precisely, it is the square root of the expected squared difference between the actual value of the dependent variable and its forecast value. The

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

It is very important, however, to realize that a large standard error of estimate does not tell one anything at all about the accuracy with which the effects of the independent variables are measured. Similarly, a large standard error of estimate says nothing at all about the probability that the effects of those variables are really zero and one is observing only chance effects. (Those propositions are assessed by means of the standard errors of the coefficients and the t-statistics as described above.) The standard error of estimate is a way of assessing how important the random part of the model is; it does not tell one how large the effects of such randomness are on one's ability to measure the systematic part.

An example may make this clear. Suppose that a group of workers are all paid the same per-hour wage, w, for each hour worked. Suppose, in addition, that workers are employed for different numbers of hours. Now suppose that at the end of each week each worker takes his pay and engages in a high-stakes roulette game. Then the income of each worker will be the sum of his pay from his job and his winnings or losings in the roulette game.

Now suppose that we are trying to estimate the common per-hour wage, w, from data on the number of hours worked and total income, but that we cannot observe take-home pay directly. We could do this by a regression in which the dependent variable was total income and the independent variable was hours worked; the coefficient of hours worked would be our estimate of the per-hour wage, w. The influence of the roulette game, of course, would be the random part of the model.

How would we measure the accuracy of our estimate of the per-hour wage? This would be measured in terms of the standard error *of the estimated coefficient* (w). If we had a large enough sample, that standard error would be very small. (This is the consistency property of least squares.) Despite this, we would still find a large standard error *of estimate* because no matter what we did, we would be unable systematically to estimate the effects of the unsystematic roulette game. In such a circumstance, we would be entitled to conclude that there were large unsystematic effects that affected our ability to predict total income. However, under no circumstances would we be entitled to conclude from that fact that we had a biased or unreliable estimate of the per-hour wage. Still less would we be entitled to conclude that changing the number of hours worked had no effect on income (i.e., that the true wage was equal to zero) or, to take the most extreme case, that workers should be

standard error of forecast and the standard error of estimate differ for the following reason. Whereas the standard error of estimate measures the extent of deviation in the sample period around the relationship as estimated, forecast errors will involve not only deviation from the estimated relationship but also the fact that the estimated relationship itself may deviate from the true relationship.

The way in which these two standard errors differ is somewhat instructive. In general, one expects to be surest about where the true relationship is for points that fall inside the range of points already observed in the sample. One would be less sure of points less typical of the sample. The standard error of forecast does depend on how far from typical sample values the values of the independent variable for the forecast period happen to fall. It is larger the farther away from the sample are such values. Given the location of the independent variables for the forecast period, however, the standard error of forecast is proportional to the standard error of estimate, which does not vary with such location.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

indifferent about whether or not they are laid off. Statements of this sort would be signaled by very large standard errors of the estimated per-hour wage, the regression coefficient of hours worked, not large standard errors of estimate of the regression.

Thus, a large standard error of estimate of the regression tells you that you do not know *everything*. This is not the same as telling you that you do not know *anything*. This is important in practice. In the case of the firemen what is involved is the difference between being able to predict the number of accidents well and being sure that employment of firemen affected that number. While related, these are not the same thing and they are measured differently.

3. *The Correlation Coefficient.* The most common way of normalizing the standard error of estimate for different units is to compare it (or more properly, its square) with a measure of the total variation of the dependent variable. What such a comparison does is to split the variation of the dependent variable around its mean into a part that is explained by movements of the independent variable (the systematic part) and a part that is not so explained (the unsystematic part). The squared multiple correlation coefficient, R^2, measures the percentage of that variation that is explained by the systematic part.[29]

How should values of R^2 be interpreted? Obviously, a value of zero means that movements in the independent variables do not explain movements in the dependent variable at all. The higher R^2, the greater the association between movements in the dependent and independent variables. A value of unity means that the entire variation in the dependent variable is explained by the model.[30] Beyond that, this commonly used measure must be approached with a fair amount of caution, since R^2 can be affected by otherwise trivial changes in the way in which the problem is set up.[31]

II. The Appropriate and Inappropriate Use of Multiple Regression in Legal Proceedings

So far, this Article on "Multiple Regression in Legal Proceedings" has been primarily about multiple regression. The time has come to talk about

29. The reasons for writing the correlation coefficient as a square need not detain us here.

30. How high a value of R^2 is to be expected depends on the number of degrees of freedom. (See note 22 supra). When one has two observations with which to fit a line, for example, such a fit will always be exact and R^2 always equal to unity. Where the line must fit many observations, then an R^2 near unity would be more impressive evidence that movements in the dependent variable are explained by movements in the independent variables.

31. Thus, for example, suppose that in the audience-revenue relationship, we had decided that the true relationship was logarithmic, with the logarithm of revenue as the dependent variable. Suppose also that one of the independent variables was the log of audience size. Suppose then that we subtracted the log of audience size from both sides, making the dependent variable the log of revenue per viewer (equal to log of revenue minus log of audience size). Obviously, the only substantive thing that this would do would be to subtract one from the coefficient of the log of the audience. But it would also change R^2, which would now measure how much of the variation in the log of revenue *per viewer* we were explaining rather than how much of the variation of the log of revenue itself. The resulting value of R^2 might thus be either higher or lower than the original value.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

legal proceedings. I shall do this by discussing three areas where multiple regression analysis has figured: the examination of wage discrimination against women; the determination of damages in price-fixing cases; and the evaluation of punishment as a deterrent to crime. These three examples will illustrate a number of the technical points already made as well as providing some lessons concerning what multiple regression analysis can and cannot do. I believe multiple regression analysis to be an entirely appropriate tool for the examination of possible discrimination in wages, but I am very dubious about its utility in price-fixing cases and I believe it to be dangerously misleading in the examination of deterrence.

A. *Discrimination in Wages*

In this example, a case is brought against a firm on behalf of a group of its women employees. They charge that the firm discriminates by paying women less than men. The object of the statistical study is to test whether this is indeed so.

Let us suppose that the facts are such that it appears to be so. The wage paid the average female employee is less than that paid the average male employee. To make things simple, let us suppose that we are considering only women and men in similar jobs.[32] The firm defends (or is likely to defend) by claiming that the women are on the average not as qualified as the men. In particular, they are less well educated and have less job experience. They also score lower on certain aptitude tests.

This is obviously a reasonable defense, if in fact it is true. For it to be true, however, it must not only be the case that women, on the average, are less qualified according to these various measures, but also that the difference in qualifications accounts for the difference in pay. If the firm does not pay well-educated men more than less-educated men, then it can hardly claim that this is the basis for the difference between male and female wages.

Multiple regression is well suited to answer this sort of question fairly precisely. Moreover, without a multiple regression study it is difficult to see how it could be decided. The raw comparison of average wages for women and for men may make one suspicious, but it cannot tell one anything definite. Indeed, it can be misleading in either direction. For example, it would be entirely possible in a different setting that women are paid on the average just as much as men but that a multiple regression analysis would show that there is indeed discrimination because women are *more* highly qualified in the measures that account for the variation in male pay.[33]

Returning to the original problem, how can this be set up in a multiple

32. Controlling for job classification is an obvious thing to do and might be done by multiple regression.

33. See Finkelstein, The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases, 80 Colum. L. Rev. 737 (1980).

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

regression framework? We begin by doing something that may seem needlessly cumbersome but will pay off later. We define a variable, S, as follows:

$$(4) \quad S = \begin{cases} 0 \text{ if the employee is a woman} \\ 1 \text{ if the employee is a man} \end{cases}$$

S is what is called a "dummy" variable, used in situations where one wants to examine discrete rather than continuous variations—in particular, classification into categories. Consider the regression equation:

$$(5) \quad Y = a + bS + u$$

where Y denotes the income paid to a particular employee. It is not hard to see that estimating equation (5) by least squares regression is simply another way of computing the difference in average pay between men and women. If $S = 0$, then, on the average, pay will be given by a; this will be the average pay of female employees. On the other hand, if $S = 1$, then, on the average, pay will be given by $(a + b)$; this will be the average pay of male employees. The difference in the averages is thus b, the coefficient of S, and testing whether that coefficient is significantly different from zero tests whether men are indeed paid more than women.

But of course, such a test is only a test of the original proposition, that men, on the average, are paid more than women and that the difference in pay is not accounted for only by random fluctuations. Such a test is better than simply looking at the difference in pay, but we have not yet tackled the problem of controlling for other variables, namely qualifications.

Such controlling is fairly easily done. For example, suppose for a moment that there were only one measure of qualifications (say, aptitude test scores), denoted by A. Consider the following modification of equation (5):

$$(6) \quad Y = a + bS + cA + u$$

Estimation of this equation by multiple regression will give an answer to the question of whether sex affects wages, with aptitude test scores constant. This may be seen diagramatically in Figure 2.

In Figure 2, employee income is plotted against aptitude test scores. Points denoting male employees are indicated by M; points denoting female employees are indicated by F. I have drawn a case in which male employees are obviously paid more than female employees on the average, but in which, again on the average, female employees score lower on aptitude tests than do male employees. Examination of the average wages without correcting for aptitude tests (equivalent to least squares regression estimation of equation (5)) amounts to drawing a horizontal line in the diagram (horizontal because aptitude is assumed to have no effect in equation (5)) at the level of average male income and another one at the level of average female income. These are relatively far apart. Correcting for aptitude test scores by estimating equation (6), on the other hand, amounts to drawing two

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

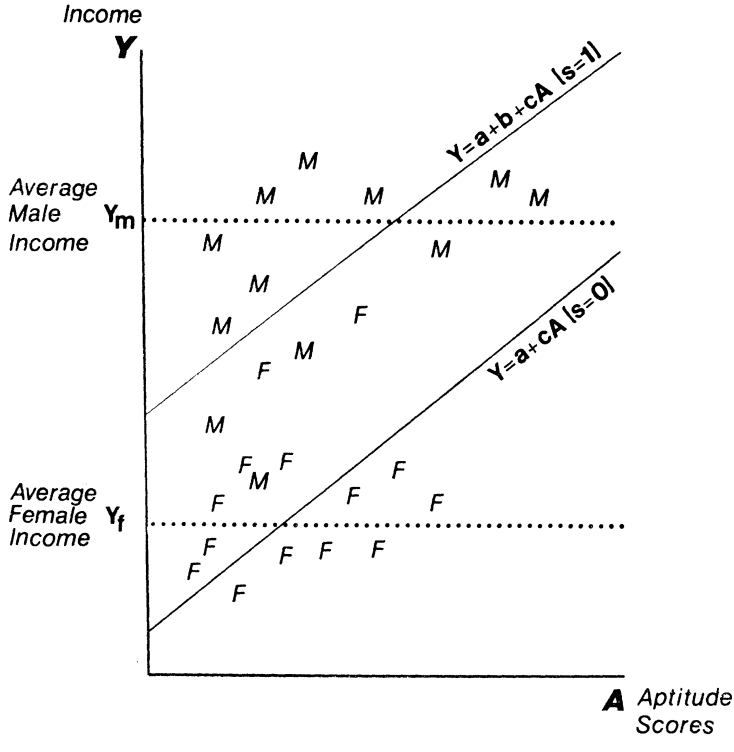

Figure 2

parallel lines through the male and female points respectively. The fact that the lines are parallel indicates the assumption that aptitude tests should have the same effect on wages for males and females if there is no discrimination. The difference in the intercepts is the coefficient of S, a measure of the remaining difference in wages after aptitude scores have been controlled for.

The proposition that males systematically earn more than females even after controlling for aptitude test scores can now be directly tested. This would be done using the t-statistic associated with b (the coefficient of S) to see whether that coefficient is significantly different from zero. (Since no one supposes that women earn systematically *more* than men, the appropriate test would be a one-tailed test.) "Significance at the five percent level" would require a t-statistic of a little more than 1.6.

This example can be extended in a few ways that are worth discussing. In the first place, there is no reason why only one measure of qualifications—aptitude test scores—should be controlled. I chose that case because the resulting diagram was easy to draw. If there are several possible measures of qualifications, then all of them can be included in the regression as new

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

variables. One of the great advantages in this problem is that there are not many variables that plausibly explain wages, and thus interest centers simply on whether sex is one of them. There is little need to thrash about for various different combinations of variables that might be included. Rather, having found an apparent effect in the raw data, the only question is whether that effect is caused by failure to control for other plausible variables.

I have set up the problem in equation (6) as though the only issue was whether a man with given aptitude was paid a fixed number of dollars more than a woman with the same aptitude. This is indicated in Figure 2 by the constant distance between the two sloping lines. According to the equation, women are at a constant dollar handicap whatever their aptitude, and the question is whether or not that handicap is zero. But of course, this may not be the most likely possibility. It is at least as plausible that women are at a constant *percentage* handicap, so that the difference in dollar terms is greatest for women with high aptitudes. This is easily accommodated in the analysis. I shall not attempt to draw the resulting diagram, but all that would be required would be the use of the logarithm of income instead of income itself as the dependent variable in equation (6).

One might also try a somewhat subtler variation. I have set up equation (6) (or its logarithmic equivalent) so that what is tested is the hypothesis that women are at a disadvantage, *given* that aptitude test scores affect wages in the same way for men and for women (the sloping lines in Figure 2 are drawn parallel). This is a good way to do it, but it is not the only way. One could estimate two separate regression equations—one for men and one for women—in which income would be regressed on aptitude. One could then test to see whether the regression coefficients for the two equations were the same *in all respects*. After all, it would be evidence of discrimination if the effect of aptitude tests on wages was not the same for men as for women. It is possible to construct cases in which b in equation (6) turns out to be zero, but in which separately estimated equations would yield significantly different values of b for men and women. On the other hand, trying to examine several things at once (i.e., whether whole sets of coefficients are the same for men and women) will produce less powerful tests than will examining each one of them individually.

Two other features of this example deserve comment. First, I have deliberately used aptitude test scores as a measure of aptitude. It is common knowledge that such tests do not provide perfect measures of ability. However, this may not make any difference in the validity of the regression model. To the extent that true aptitude has different dimensions, the crudeness of aptitude test scores as a measure may be corrected for by the other variables to be introduced into equation (6)—variables such as years of education or work experience. Second, what matters in the current problem is what the employer can observe in distinguishing aptitude. The defendant in this case will look relatively weak if he claims only that he had an unmeasurable way of evaluating aptitude and that all measurable methods are subject to

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

error. In effect, what is important in this problem is not some underlying measure of aptitude but the measure that the employer can see and reward. An argument that aptitude tests are subject to error ought to be challenged by demand for some more reliable but objective measure.

Putting this aside, however, the crudeness of aptitude scores might make a substantial difference if the true variable (aptitude) were measured only by aptitude test scores with a random error. In such a case, it is possible to show that the estimates of c, the coefficient of aptitude test scores in equation (6), would be biased toward zero. This is perhaps what one would expect, since putting in variables that contain a lot of "noise" is likely to result in estimates suggesting that those variables do not have much systematic effect. More important, however, the bias will not be restricted to the coefficient of the variable that is subject to the error. In the present problem, the variable S (describing sex differences) is correlated with the variable A (denoting aptitude test scores), reflecting the fact that, in the sample of employees, women tend to score lower than men on aptitude tests. Such correlation means that the coefficient of S will also be biased and this coefficient is the one that is of interest. Unfortunately, it is not possible to say (without more assumptions) in what direction that coefficient will be biased. Under some circumstances, there are steps that can be taken to guard against the effects of measurement error, but it would take me too far afield to discuss them here.

The final point to be made about this example is that accurate prediction of the dependent variable, income, is not required for successful resolution of the problem. Rather what is involved here is a direct test of the significance of a particular coefficient. The precision of that test (technically its "power") will depend on the standard error of that coefficient and not directly on how well the equation can be expected to do in predicting the dependent variable. Generally, tests like these are likely to be more successful than tests that depend directly on predictions.

What makes the wage discrimination example so suitable for multiple regression is its simplicity and the readiness with which it can be cast into the mold of a test of the significance of a particular regression coefficient. Notice in particular the following feature: whether there is discrimination or not, one would expect the expanded version of equation (6) to fit well. What is being done there is to imbed in a theory of wage determination the difference that discrimination does or does not make. At least at this level, the question of what factors other than discrimination determine wages can be considered without regard to whether or not there is in fact discrimination. Further, the presence or absence of discrimination makes a clearly definable difference in the result one would expect to find. These features stand in contrast to those of the next example.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

B. *Antitrust Damages in Price-Fixing Cases*

In this example, the defendants have lost on the issue of liability in a price-fixing case, and the issue to be decided is the extent of damages. The defendants prepare a study attempting to show that the effect of fixing the price was minimal, in that the price would have been the same (or higher) without the conspiracy.[34] There are a number of ways in which this might be done, but I am very dubious about the usefulness of any of them.

One way to proceed is to take a leaf from the discrimination example just discussed. In that example, the study proceeded by controlling for several variables and, in effect, estimating what income would have been if there were no discrimination. Why not systematically estimate what prices would have been without price fixing? We might think of doing this as follows. Under competition, price is determined by the intersection of supply and demand curves. Let us assume, for simplicity, that there are no close substitutes for the product in question, so that demand depends only on the income of consumers (or the output of industrial customers) as well as on price. Supply will depend on price and on costs, which in turn depend on the prices of the factors of production. This suggests that we ought to be able to explain price by a regression equation involving consumer income and factor prices.

Although one might assume that *quantity* should be included as one of the variables that may have an impact on price, it is more appropriate to treat price and quantity independently since, in fact, the same market forces control both. This is evident from an examination of the specific equations (supply and demand curves) that determine supply and demand in the market.

Quantity, like price, is determined by the intersection of the supply and demand curves. Assuming linearity, for convenience, we can write the demand curve as:

$$(7) \quad Q = a + bP + cY + u$$

Here, Q denotes quantity, P denotes price and Y denotes consumer income. As before, u is a random disturbance. Similarly we can write the supply curve as:

$$(8) \quad Q = d + eP + fW + v$$

Here, W is a measure of factor prices and v is another random disturbance.

Equations (7) and (8) form what are called the "structural equations" of a "simultaneous equation system." Such a system involves the interaction of more than one equation—equations that can be solved simultaneously. The fact that price is determined by the intersection of supply and demand is reflected by the fact that P and Q must have the same value in both equations. We can thus solve both equations together for those two variables

34. Since, under the per se rule, the ineffectiveness of a price-fixing conspiracy is not a defense, such a showing would be irrelevant to the issue of liability.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

by equating the values that each equation predicts for the "quantity" variable. To do this, we create new coefficients (π_0, π_1, π_2, etc.) that depend on all the coefficients of the supply and demand curves. When this is done, the solution for price will look as follows:

$$(9) \quad P = \pi_0 + \pi_1 y + \pi_2 w + u^*$$

u* is the random disturbance, which depends on some of the coefficients in the supply and demand curves, as well as on u and v. (Precisely, it is equal to $(u - v)/(e - b)$.) The exact algebra need not detain us. There will be a similar solution for Q.

Equation (9) and its companion for Q are called the "reduced form" of the model. They show price and quantity directly in terms of those variables that are determined by forces other than those being modelled (Y, W, u, and v). Such reduced-form equations can be estimated by least squares regression.

It would be a mistake, however, to include Q in the equation for P. It does not appear in equation (9) for the very good reason that quantity and price are jointly determined by the same forces, and it cannot be said that one of them determines the other. A regression that includes quantity on one side and price on the other might be interpreted as an attempt to estimate either equation (7) or equation (8) directly, but this cannot be done consistently by least squares. The easiest way to see this is as follows. A movement in the disturbance term in equation (7), u, affects quantity, Q; this is essentially a random shift of the demand curve. But random shifts of the demand curve affect not only quantity but also price. Hence, shifts in u are associated with movements in P, as can be seen directly from equation (9) and the fact that u* depends on u. This means that, in estimating equation (7), the fundamental assumption of least squares—that random disturbances move independently of the independent variables—is violated. Equation (7) can be estimated, but least squares is not the way to do it.

Thus, trying to determine what price would have been in a competitive market by regressing price on a set of variables including quantity is doomed to failure. Suppose, however, that we were more sensible and simply regressed price on income and factor price (Y and W), thus estimating equation (9) directly and using that equation to predict price absent the price-fixing agreement.

This is better, but still not adequate. The problem here is that there will not be a clear distinction between the results that one would obtain if the market was affected by the price-fixing scheme and the results that one would obtain if it was not. If the market was not competitive but was seriously affected by price fixing, price was not determined by the intersection of competitive supply and demand curves. Rather, price was determined largely by the price fixers. But the price fixers presumably did not

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

set arbitrary prices but rather set prices to maximize their profits to the extent that they could.

Without going into great detail, it is not hard to see that profit maximization would have required consideration of the position and shape of the demand curve (7) as well as consideration of the costs of production. In the standard terms of economists, profit maximization requires the equating of marginal revenue and marginal cost. Marginal revenue will depend directly on demand and marginal cost directly on factor prices. The price and quantity that equate marginal revenue and marginal cost will, just as in equation (9), depend on income and factor costs. Indeed, for price, one is quite likely to end up with an equation identical to equation (9); the difference that price fixing makes is that the coefficients in equation (9) will be different under price fixing than under competition.

This means, however, that there is no point to estimating equation (9) directly and using it to forecast price. Equation (9) would be valid whether or not there was price fixing and one will not be able to tell whether the predictions that it generates are competitive or noncompetitive. The case was quite different in the wage discrimination example. There the issue was sharply defined as whether a certain coefficient was zero or nonzero. Here the issue might be described as involving differences in a certain set of coefficients (the π's in equation (9)), but we can estimate those coefficients only once and there is thus no way that we can compare the values we obtain with the unknown values that we would have obtained under either the competitive or the noncompetitive hypothesis.

Does this mean there is nothing that can be done? No, but it comes close. We might proceed in a somewhat more sophisticated manner and try to estimate equation (7), the demand curve, which is the same under both regimes. We might then ask what the competitive supply curve would have looked like. Theoretically this could be done, but in practice it is probably impossible. The demand curve (equation (7)) can be estimated. As we have seen, it cannot be estimated by least squares under the hypothesis of competition, but there are other methods of estimating it, and those methods would remain valid, in general, even under a scheme of price fixing.[35] However, in order to find out what price would have been under competitive conditions, it will be necessary to estimate the competitive supply curve. One cannot do that directly from the observations because to do so is to assume that the observations were generated by competitive supply and demand. That, however, is what one wants to prove. Hence, one will have to look elsewhere. In general this will mean estimating the cost curve of the producers and calculating marginal cost. Even if the defendants are

35. If one were willing to admit that the price-fixing agreement did have a substantial impact on price (which, presumably, one is not), least squares estimation of the demand curve might become easier, essentially because prices would have been determined in a controlled manner. On this point, see my study of aluminum demand, F.M. Fisher, A Priori Information and Time Series Analysis 93-117 (1962).

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

willing to give up the information required for this calculation, it is likely to prove extraordinarily difficult to estimate. Once we move away from simple one-product examples, the cost calculations (and indeed the estimation of the various demand curves as well) become quite complicated. What is involved here is a major undertaking requiring a great deal of data, most of it unlikely to be in usable form, and generating only a thin promise at the other end. Indeed, if one is going to look directly at cost information, it might be better to make a direct showing that prices approximated marginal costs. To do that, one would not need to look at demand.

There remains one possibility in this area that looks slightly more promising. Many of the problems just discussed occur because one wants to know how competition would have looked without directly assuming that competition in fact existed. If, however, there is agreement that the price-fixing conspiracy was in effect only for a limited time, then one might consider estimating the reduced form equataion for price (equation (9)) and the companion equation for quantity, using only data from the competitive period. One could then use those equations to forecast price for the price-fixing period and study the difference in results.

This sort of program is feasible, at least in principle.[36] Unfortunately, it is unlikely to pay off in practice. One will be using the estimated equations to forecast out of the sample period. If conditions have changed (and over time they usually do) this is going to mean forecasting away from the mean of the sample. Even if the model is entirely correct, one is not going to be able to make this sort of forecast with a great deal of certainty. One is likely to find that the price at a given moment during the price-fixing period is not significantly higher than that which would be predicted by the competitive model, but that the standard error of that prediction is large. Thus, although it will be possible to test whether the difference in price is significant, it will probably be very hard to decide how much of that difference is due to random error.[37] Furthermore, variations in price in either direction can be explained away, by either plaintiffs or defendants, in terms of shifts in demand or cost conditions. Hence, if what is involved is prediction over a long time, this forecasting may be worth trying, but it is not likely to be useful. As opposed to the other approaches already discussed, however, it does have the merit of providing a clear comparison of the two hypotheses involved.

---

36. There may be some technical problems concerning whether to estimate equation (9) directly by multiple regression or to use sophisticated simultaneous equation techniques to estimate equations (7) and (8) directly, but they need not detain us.

37. This would generally be tested by a so-called "Chow" test. See Fisher, Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note, 38 Econometrica 361-66 (1970). This would also be the test used to determine whether the entire regression of income on aptitude was the same for men and for women in the discrimination example above.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

C. *Punishment as a Deterrent to Crime*

The last topic that I will discuss is the use of studies that purport to examine the effect of punishment as a general deterrent to crime—that is, as a deterrent to persons other than those being punished. I have already mentioned the death penalty studies referred to by the Solicitor General. In addition, there are a number of studies of other categories of crimes and types of punishment. This is not the occasion to discuss these studies in great detail; such discussions can be found elsewhere.[38] However, a consideration of some of the reasons why these studies are unsatisfactory will illustrate points that are generally applicable to the use of multiple regression analysis.

At first glance, the problem seems to be eminently suitable for regression analysis. Nearly any examination of data in which punishment varies also shows crime varying in the opposite direction. Yearly data on murders committed in the United States (a "time series") show the number of murders rising in years with no executions. With respect to other crimes, cross-section data show that jurisdictions with less severe sentences tend to be the jurisdictions with higher crime rates. It plainly appears that there is a negative correlation between severity of punishment and crime rate and that the problem is merely that of assessing the magnitude of the deterrent effect.

Unfortunately, while I agree that there probably is something significant in these data, the problem of measurement turns out to be very severe. This is true for more than one reason. First, there is a problem because we do not have a very good theory of what causes crime, and thus we do not really know what other variables should be controlled for in deriving a crime equation. Second, one has to control not only for other variables in the same equation but also for the presence of additional relations between those variables and crime. Add to this the doubtful nature of much of the data and one has a serious problem.

Let me begin by considering the death penalty studies.[39] The primary study [40] used time-series data on the United States as a whole for the years 1933-1969. This is a sample of thirty-seven observations, although data on some of the variables had to be constructed. However, it turns out that the results depend almost entirely on the years after 1962. This is, perhaps, no surprise; it was primarily in these years and in the early 1970's that many jurisdictions experimented with the abolition of capital punishment. It does mean, however, that there is only a relatively limited amount of data to use in controlling for other effects, despite the seemingly large sample size. Furthermore, these same years coincide with a general upsurge in crime,

38. See, e.g., Deterrence and Incapacitation, supra note 1.
39. For a more detailed discussion and references, see Klein, Forst & Filatov, supra note 1.
40. Ehrlich, The Deterrent Effect of Capital Punishment: A Question of Life and Death, 65 Am. Econ. Rev. 397 (1975).

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

not just in those crimes subject to capital punishment. Therefore, we cannot be sure that the results of the study do not simply depend on poorly understood phenomena concerning the causes of crime.

There are lessons to be learned here. First, when faced with a multiple regression study, one should try to determine whether the results crucially depend on certain of the years chosen or whether they stand up to variations in the sample. If the results do depend on certain years, one should try to decide whether there are other characteristics specially associated with these years that might have affected the results. Second, and perhaps more important, one must try to determine whether enough is known about the phenomenon being investigated (here, the causes of crime) to estimate it in terms of the model selected. If not, there will be other plausible explanations for the results achieved.

The death penalty study also turns out to depend rather crucially on the form of the equation used. There is a big difference in its results depending on whether the equation is estimated in linear or logarithmic form.[41] Of course, if one had reason to believe that the correct form of the equation was one or the other, one would simply use that form. But one does not know which form is "correct." Results that depend on the use of a particular version of the equation may not be valid; they depend on an unsupported assumption.[42] When one is deciding whether to execute a man, it ought to concentrate the mind wonderfully. In such matters, the studies to be relied on ought not depend on particular sample periods or choice of specifications.

Many of the problems with the capital punishment study arise because of the limited nature of the available data. An obvious alternative set of experiments would involve the use of data concerning various crimes and drawn from different jurisdictions, in order to get a large sample and a lot of variation.[43] The trouble here is as follows.

Obviously, there are reasons other than variations in punishment why crime rates vary over jurisdictions. It is therefore necessary to control for such reasons. Some possibilities for such variables are unemployment rate, percentage of urban population, and so forth. Multiple regression might in fact do this.

Unfortunately, there are also reasons why punishment levels vary over jurisdictions. One of the reasons suggested in the literature has to do with crime rates. It is easy to see how this might happen. Jurisdictions

41. See notes 10 & 20 supra.

42. There are ways of testing whether one form is better than another. Often, however, it is hard to tell from the results.

43. Ehrlich has also performed cross-section analyses of murder, but I am less familiar with these than with his study of noncapital crimes. The latter is Ehrlich, Participation in Illegitimate Activities: A Theoretical and Empirical Investigation, 81 J. Pol. Econ. 521 (1973). The following comments are expanded in Nagin, General Deterrence: A Review of the Empirical Evidence, in Deterrence and Incapacitation, supra note 1, at 95, and Fisher & Nagin, On the Feasibility of Identifying the Crime Function in a Simultaneous Model of Crime Rate and Sanction Levels, id. at 361.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

with higher crime rates may adopt "get tough" policies. Alternatively (and this is the suggestion in much of the literature), jurisdictions with high crime rates may overload their punishment facilities and thus may come to tolerate relatively common offenses somewhat more than do jurisdictions with low crime rates. In any event, there is a serious possibility that the variation in punishment levels over jurisdictions can be accounted for, at least in part, by the variation in crime rates. In this circumstance, as in part of the supply and demand example given above, the problem is not merely that one has to control for other variables, but that one has to control for the presence of another equation. To see the kind of problem that arises, consider the following vastly simplified example.

Assume, for the moment, that the *only thing* that affects crime rates is punishment. Assuming linearity, for convenience only, the crime rate equation to be estimated could then be written as:

$$(10) \quad C = a + bS + u$$

Here, C is the measure of crime rate and S is a measure of punishment or

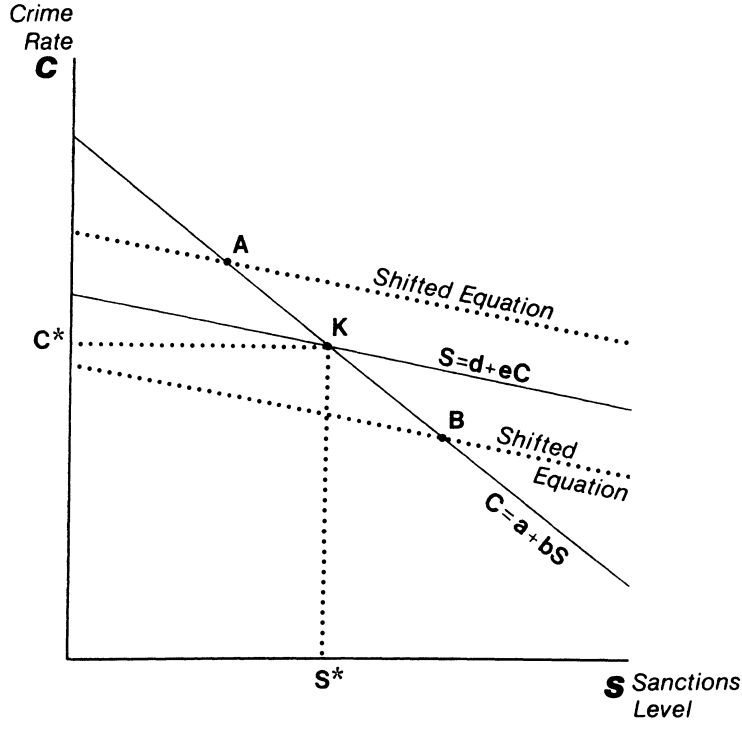

Figure 3

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

sanctions levels. The coefficient b would represent the deterrent effect of increasing sanctions.

Suppose, however, that sanctions also depended on the crime rate and only on the crime rate. Then the equation that shows how sanction levels are determined can be written (again assuming linearity):

$$(11) \quad S = d + eC + v$$

In these equations u and v are random disturbances.

Given the interrelation between these two equations, one could not effectively estimate the crime equation (equation (10)) by least squares regression. The fundamental assumption of least squares regression is that the random disturbance term operates independently of the independent variable. All of the properties of least squares depend on this. In the present instance this would require that u and S be uncorrelated. This cannot be the case, however, because the model itself (just as in the supply and demand example given above) implies that it is not so. An upward shift in u, according to equation (10) itself, will mean an upward shift in the crime rate, C. But an upward shift in the crime rate, C, will, according to equation (11), cause a shift in the sanctions level, S. Hence, shifts in u cannot be independent of shifts of S and least squares regression will fail. (This may also be seen by solving equations (10) and (11) for C and S to obtain the reduced form of the system, as was done in the supply and demand example.)

The problem is worse than this, however. To see this, ignore the random disturbances, for a moment, and suppose that equations (10) and (11) were exact. I have graphed those equations in Figure 3. In such a situation, the crime rate and the sanctions level would be entirely determined by the simultaneous solution of the two nonrandom equations—the intersection of the two lines in Figure 3 at K. (The resemblance to a supply and demand graph is not accidental.) If this were really the case, the only point we would ever observe would correspond to that intersection at sanctions level denoted by S * and crime rate denoted by C *. But if that point were the only one observed, there would be no way of recovering equations (10) and (11). In terms of the graph, we could not tell the true crime function (the more steeply sloped line) apart from the sanctions function (the less steeply sloped line) or, indeed, from any other line that went through that same point; each line could vary, in an infinite number of ways, around the point K.

Even if we put random disturbances back in, we would not get anywhere. The effect of random disturbances would be to produce a cluster of points surrounding the intersection drawn in the graph, but again, it would not be possible to recover the two underlying lines that generated this cluster or to tell the two lines apart even if we could recover them. In this circumstance, the crime and sanctions equations are said to be "not identifiable."

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

This problem, one of *identification,* is a well-studied subject in econometrics.[44] I have deliberately chosen an extreme case. Unfortunately, the identification problem continues in the deterrence studies even when the extreme assumptions are relaxed.

Suppose, for example, that there was some variable that shifted sanctions levels over jurisdictions but did not affect crime rate. This would mean that there would be an additional significant variable in equation (11) that was not also a variable in an expanded version of equation (10). Leaving equation (10) as it is, the effect would be to shift the sanctions equation in Figure 3 up and down. (This is illustrated by dashed lines parallel to the solid line corresponding to the sanctions equation in Figure 3 and marked "shifted equation.") If this happened, we would observe not merely one intersection of the sanctions equation and the crime equation but several intersections, points such as A and B, for example. Those points would all lie on the crime equation and, indeed, as the sanctions equation shifted back and forth because of the presence of the additional variable, the points of intersection would trace out the crime equation.

In such a situation, as the diagram suggests, there is a technique for recovering the crime equation from the data. That technique, however, is not least squares regression, because the correlation between the disturbance term and the independent variable in equation (10) would generate invalid results. Moreover, it will still not be possible to recover the sanctions equation itself.

Because of the identification problem it is necessary to find a variable that shifts one equation of the model but not the equation to be identified. However, it is not only bad practice to attempt to find such variables from the data, it is literally impossible. No amount of manipulation of data generated by the model will reveal such variables; the selection of such a variable must be done as a matter of prior theory.

It is easy to see from Figure 3 why this should be so. If there is a variable shifting the sanctions equation but not the crime equation, then the observed points will be like the points A, B, and K in the diagram. But such a pattern of intersection could also be produced by a variable shifting the crime equation but not the sanctions equation. More generally, it could be produced by shifts in both equations. Only if we *know* from theoretical, nondata-generated considerations that it is the sanctions equation that shifts can we be sure that it is the crime equation that is traced out.

In most situations, such theoretical considerations may readily be found. (For example, consumer income enters demand but not supply curves; factor costs affect supply but not demand.) This is not so in the present case, however. While there are a number of variables that may enter the sanctions equation, it is difficult, if not impossible, to think of such

44. See F.M. Fisher, The Identification Problem in Econometrics (1966) (reissued 1976).

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

a variable that would not also enter the crime equation.[45] The existing studies have tried to get around this by casually assuming that variables such as unemployment influence sanctions levels but not crime. This is plainly wrong. In the present state of our knowledge, we simply do not know enough about the structure of the system generating the observations to be able validly to estimate the crime equation.

This problem has some general implications for the use of regression analysis. First, it is important to be very careful not only about controlling for additional variables, but also about the possibility that one must control for the existence of additional relationships between the dependent and independent variables. If there are such relationships, least squares will not be an appropriate estimator, and it is at least possible that no appropriate estimator will exist (although this is not common). Second, if there is another equation involved, one must find out how the expert really did his estimation. If he explored the data by multiple regression and then, having decided on a model, altered it with another estimation technique, the results are quite suspect.[46]

Finally, one should make sure that the model used is constructed on sound hypotheses based on theoretical considerations generated from outside the model itself. While multiple regression and related econometric techniques are powerful tools for analyzing data, their proper use presupposes an underlying theory of the structure generating those data. While some hypotheses concerning that structure can be tested with these tools, the theory itself cannot be discovered by computer runs and data experimentation. Thus, the expert making the study must not only understand the proper uses of the statistical tools, he also must learn something about the phenomena and hypotheses being investigated.

CONCLUSION

Multiple regression analysis can play a vital role in legal proceedings. Used properly, it is an accurate and reliable method of determining the relationships between two or more variables, and it can be a valuable tool for resolving factual disputes. In order for this to happen, however, multiple

45. On the other hand, it is not difficult to think of variables that enter the crime equation but that would not directly influence the choice of sanctions. Unemployment, for example, is far more likely to influence the crime rate than to influence sanctions. Other examples might include measures of income disparity or expenditures on security systems. If such variables really do influence crime rate, but not sanctions, then including them in the crime equation would shift that equation relative to the sanctions equation. The points of intersection traced out would all lie on the sanctions equation, which would then be identifiable and could be estimated (although still not by least squares).

46. Consider the following all too common procedure. Since multiple regression is easy to do, one experiments with multiple regression until one has a version of the estimated equation that corresponds to one's own predilections. Then one reestimates the equation by an appropriate simultaneous equation technique. If the results look very different from the least squares version one goes on exploring. This is not a way to produce consistent results.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms

regression must be better understood by the legal community; in particular, there must be an understanding of both the potential and the limits of the technique.

It is not necessary that lawyers understand the mechanics of multiple regression in terms of what goes on inside the computer. It is necessary, however, that they understand the regression model and the assumptions being used in any given regression study, how the results of the regression bear on the hypothesis to be tested, and how the results distinguish this particular hypothesis from other hypotheses. The expert constructing the analysis should be able to explain all of this to the attorney who employs him, and an expert who cannot explain such things is likely to fall apart on cross-examination.

Lawyers will increasingly find themselves in a position where it would be profitable to use a regression analysis or where they must confront a regression study produced by an opponent. When that happens, a basic knowledge of multiple regression may be a valuable asset.

This content downloaded from 130.91.93.179 on Thu, 29 Aug 2019 20:54:17 UTC
All use subject to https://about.jstor.org/terms