

QUANTITATIVE METHODS IN ANTITRUST LITIGATION*

DANIEL L. RUBINFELD† AND PETER O. STEINER**

I

INTRODUCTION

The adversary process in antitrust cases involves opposing assertions about (a) what occurred; (b) what caused what occurred; and (c) what would have occurred "but for" some event. The validity of these assertions in turn usually depends upon facts upon which economics and econometrics can shed light. Econometrics is the subfield of economics that applies statistical techniques to make quantitative analyses of variables of interest. While statistical analyses are not the only options available to resolve competing claims, they are surely among the more useful of such tools. Like any technique whose use has grown rapidly, the opportunity for such use carries with it the opportunity for misuse, even abuse. In this regard this article will contend both that the constructive uses of quantitative methods in antitrust litigation are varied and substantial, and that there are a number of important dangers associated with use of these methods.

This article explains, on a basic level, how the seemingly arcane tools of statistics and econometrics can be of practical value to attorneys trying antitrust cases, as well as to legal scholars studying antitrust law. The outline of this article follows closely the three assertions described previously. Section II focuses on the description of what occurred. The use of statistics in systematic description is a complex process and, in fact, raises a real danger of incomplete as well as biased description. The value, as well as the dangers, of descriptive statistics are illustrated through an examination of two cases, one of which will be of use throughout the text. The first case, involving the mailing machine industry, illustrates a situation in which descriptive statistics can be dispositive. The second case, involving the drug ampicillin, will provide an illustration of some not uncommon abuses of descriptive statistics. Section II makes it clear that in many situations statistical analyses of two, or more than two, variables may be essential.¹

Copyright © 1984 by Law and Contemporary Problems

* This research was sponsored by the Law and Economics Program of the University of Michigan. We wish to thank Martin Adelman, Joseph Brodley, Frank Easterbrook, Thomas Kauper, and Richard Lempert for their helpful comments.

† Professor of Law, Professor of Economics, University of California at Berkeley.

** Professor of Economics, Professor of Law, and Dean of the College of Literature, Science, and the Arts, University of Michigan.

1. A third case, involving the plywood industry, is more complex, however, and is discussed in section IV of the article. Simple economic theory can be used to evaluate some plaintiff theories in the plywood case, while a more complete econometric analysis is necessary in order to deal with the more sophisticated theories.

Section III is devoted to an elementary presentation of multiple regression analysis, a technique which provides quantitative estimates of the effects of a number of different explanatory factors on one or more variables of interest. Multiple regression also determines the reliability of statistical results, and it can be used for purposes of forecasting.

Hypothesis testing is particularly useful for dealing with the question of whether an antitrust violation has occurred. Hypothesis testing involves specific statistical tests which allow one to draw inferences about which of two or more competing explanations of an agreed set of events is the most credible. In section IV an example of its usefulness is presented in the context of the ampicillin case. The question of violation is also addressed in section IV using multiple regression as a forecasting device in the context of the plywood case.

Section V considers the measurement of impact and damages once a violation has been established. The measurement issue involves an evaluation of the difference between what did occur and what would have occurred but for the violation. This analysis starts with an illustration of some of the special problems associated with statistical methods as they relate to measuring impact and damages. First, a case study relating to the uranium industry is used to show why the proper specification of the appropriate variables and their relationship to each other is vital to a proper analysis of impact and damages. Second, "but for" prices are determined empirically in the context of the ampicillin case through the use of econometric forecasting techniques. Finally, the use of anecdotal evidence and samples to obtain damage estimates is evaluated with the focus on a case involving the possible resale price maintenance of Datsun cars.

Section VI contains some brief comments about the implications of the analysis for policy. It raises a number of questions concerning the role of statistical and economic experts in antitrust litigation, especially with respect to the presentation of complex materials.

II

VIOLATION: THE USE OF ECONOMIC THEORY AND STATISTICS

While purely economic analysis may be dispositive in some situations, that case is rare. The reasons may be suggested by an example. Suppose a group of sellers agreed that all would offer the same cash discount of two percent for thirty days, and they never vary that discount. This is the extent of their agreement, and each is fully free to charge any base price he wishes. It is a matter of simple economic theory that, for any seller facing a buyer who will pay cash, this agreement places no limit on the actual price negotiated and is therefore a theoretically ineffective restriction insofar as price fixing is concerned.² But this argument is not dispositive. Suppose quoted prices can be readily verified by competitors, while the size

2. Let (1) $P_A = P_{Q1} - D_{Q1}$ (Price Actual = Price Quoted - Discount Quoted)

(2) $P_A = P_{Q2} - .02 P_{Q2} = .98 P_{Q2}$.

If both buyers and sellers are informed, they will negotiate about P_A , and there is no limitation on their ability to do so by the restriction to (2) rather than (1). In addition, there may be a convenience in negotiating in one dimension rather than two.

of cash discounts cannot. Then eliminating variation in discounts may eliminate a source of undetected, and thus effective, price competition. In this case, then, an empirical question becomes relevant: Would price be different if competition in cash discounts were eliminated? Theory cannot tell us the answer.

Far more common are the cases in which the need for economic measurement is explicitly required. In such situations, economic theory provides, at best, a framework for thinking about alternative explanations, each of which is the credible inference from some (different) factual constellation.

In the cases considered in this article, no *a priori* solution exists. However, data can be developed which bear on the underlying questions. Indeed, in every one of these cases one or both of the authors has attempted to do just that, as consultants for one of the parties.³

On occasion, a simple statistical description of the data may also be conclusive. Often, however, allegedly dispositive theory and statistics are misleading and a more complex description or analysis may yield quite different results. Some more complex applications of statistics are analyzed in the following sections. For the moment, some of the uses and dangers of simple economic theory and statistics are considered in two actual antitrust cases. This analysis begins with an examination of the "mailing machines case,"⁴ and focuses primarily on the use of economic theory to state competing hypotheses which can be easily distinguished using direct empirical evidence. In the ampicillin case,⁵ the dangers of simplicity are examined in depth.

A. Mailing Machines

*Pacific Mailing Equipment Corp. v. Pitney Bowes*⁶ originally involved a number of complex issues of law and economics, but in the form in which it was finally tried before a jury it had been simplified (by the judge's prior rulings) to a direct confrontation between two competing hypotheses to explain a single undisputed fact. Although the use of economic theory could pose, but not solve, the question of which hypothesis was correct, it provided the framework into which simple statistical showings were sufficient to conclusively resolve the question. Its discussion here is intended to show that sophisticated analyses involving a number of interrelated variables are not always required.

1. *Background.* Mailing machines are used in conjunction with postage meters which print prepaid postage on mail before it is deposited into the United

3. Since this article is not concerned with the merits of the litigation, the author's partisan retention does not prejudice a discussion of the use of quantitative methods. For the record, however, the authors' involvements were as experts for the defense in the mailing machines case (Steiner, with Paul Courant), plywood (both), ampicillin (both), automobiles (both), and for the plaintiff in the uranium case (Steiner). The reason for relying on cases in which at least one of the authors has been involved is that only in such cases are the authors likely to have enough in-depth acquaintance with both the issues and the data available at the time of litigation.

4. *Pacific Mailing Equip. Corp. v. Pitney Bowes*, 499 F. Supp. 108 (N.D. Cal. 1980).

5. *In re Ampicillin Antitrust Litigation*, 88 F.R.D. 174 (D.C. Cir. 1983).

6. 499 F. Supp. 108 (N.D. Cal. 1980).

States Mail. They feed, moisten, and close the mail after it is impressed with postage by the meter.

Defendant Pitney Bowes is the largest manufacturer of mailing machines and postage meters in the world, accounting for roughly 90% of the mailing machines in use. Plaintiff Pacific Mailing Equipment (PME) is a small corporation that has been in business since 1967. PME acquires and reconditions used Pitney Bowes mailing machines. It sells, leases, rents, and services these reprocessed mailing machines in the greater Los Angeles area.

The postage meter is a sealed unit which, by law, must remain the property of the manufacturer and is rented to customers in conformity with regulations of the United States Postal Service. Postage meters and machines are manufactured by Friden and Singer as well as by Pitney Bowes. Each manufacturer's meters couple only with mailing machines it has manufactured; that is, only Pitney Bowes meters are compatible for use with Pitney Bowes mailing machines.

PME brought its action under section 2 of the Sherman Act alleging that Pitney Bowes monopolized trade in mailing machines by virtue of its dominant position in the market and a series of practices that prevented PME from achieving a competitive share of the total market.⁷ The most important such practices were: (1) Pitney Bowes' widespread use of trade-ins when selling new machines; (2) its heavy and increasing offering and use of leasing as an alternative to outright purchase of mailing machines;⁸ and (3) its practice (until after the case had been started) of mutilation and scrapping of virtually all⁹ of the used mailing machines it acquired via trade-in or lease termination.

The district judge decided the case in an innovative manner, in part because the litigation had earlier been subject to two mistrials and two changes of judges that had made a protracted litigation seem interminable. Judge Schwarzer persuaded the parties to agree to bifurcate the case and first let the judge issue findings of fact and conclusions of law on the questions of the relevant market and defendant's monopoly power therein. He defined the relevant market as consisting of the placement (sales, lease, or rental) of new and used mailing machines in the greater Los Angeles area.¹⁰

7. There were also charges of unfair competition (under a state statute) arising from such alleged practices as delays in selling repair parts, installing meters, and denigration of PME's machines by Pitney Bowes personnel. They were found unpersuasive by the jury and are neglected by the authors.

8. Leasing, as used in the industry, is distinguished from rental. A rental is a genuine hire of services. Leasing is a device to permit the purchaser to buy on time. At lease termination (typically three years), the lessee can acquire title to the machine by a nominal additional payment. Leasing here corresponds to what elsewhere is called hire-purchase.

9. Pitney Bowes retained some reacquired machines for demonstrators and loaners. It cannibalized certain older models for spare parts, and it itself reprocessed and resold a small number of the machines it acquired.

10. See *Pacific Mailing Equip. Corp.*, 499 F. Supp. at 111-13. From an economic/antitrust point of view perhaps the most interesting aspects of the case were eliminated by this ruling. Defendant argued that for a durable product such as mailing machines, the market should be a market for *services* of mailing machines, not for the *placement* of them. The judge rejected this argument, perhaps because he felt that it was unnecessary given the evidence on supply availability. Had the case been decided otherwise by the jury, this issue would have been interesting on appeal, raising novel issues of the application of economic theory to the law.

The only issues at the subsequent jury trial were whether the practices constituted monopolization, in the sense of allowing Pitney Bowes thereby to maintain monopoly power which had been lawfully acquired, and whether this monopoly caused injury cognizable under the antitrust laws. For the present discussion it is sufficient to focus solely on the trade-in and scrapping practices which formed the heart of plaintiff's case.¹¹

The central fact was that PME, year by year, sold a relatively small number of machines at a very high profit per machine.¹² Plainly, if PME (or someone like it) could make a substantial variable profit per machine, it could make much more money if it could expand its sales by a factor of ten or twenty.

The total size of the market, as Judge Schwarzer had defined it, was approximately 5000 machines per year, of which 90% were new placements and the rest used placements. PME was the largest seller of used machines, with Pitney Bowes itself second. Pitney Bowes' machines (by whomever sold) accounted for more than 90% of the total, with Friden and Singer sharing the balance.

Plaintiff's cognizable injury was based on its claim that but for the practices complained of, its share would have been much greater than the 5% actually achieved.¹³ In the background of the litigation (but very much in the forefront of the lawyers' thoughts) was the possible claim that a class of similar reconditioners, like PME, would have had much more than a 10% share of each of the regional markets in which they sold.

2. *Plaintiff's and Defendant's Theories.* The persistent combination of small volume but high incremental profit naturally raised the question of why PME and others did not seize the apparent money-making opportunity to expand their operations. High per unit profits on incremental sales normally induce profit-maximizing firms in any industry to expand their volume of output and thereby increase their total profits. This expansion, in turn, would be expected to lead to falling profit margins as the market became saturated and selling prices fell. In used and reprocessed mailing machines, however, high variable profits persisted over the whole decade. Competing explanations were advanced to account for this persistence of small volumes and high variable profits.

Plaintiff's Hypothesis: There was ample demand to support a large increase in sales, at current prices, if only an adequate supply of used mailing machines had

11. Although leasing practices occupied a great deal of trial time, the issue was effectively removed from consideration by the judge's (unusual) statement to the jury that although the described arrangements were complex, he regarded them as perfectly normal industrial arrangements.

12. The year 1975 may be taken as representative. During that year, PME sold ninety-five rebuilt or reconditioned mailing machines. For these ninety-five machines PME incurred an average acquisition cost of \$117, and spent \$179 for reprocessing, resulting in an average cost of \$296 per machine. It sold the average rebuilt machine for \$688, thus realizing an average "variable profit" (or "contribution to overhead") per machine of \$392. Considering the four years 1972-75, PME sold 333 rebuilt mailing machines whose average acquisition and reprocessing cost was about \$300 and average sales price was over \$700. There were no other substantial incremental costs. The variable profit was 58% of sales, or 138% of variable cost, a handsome markup.

13. Plaintiff's initial claim was that but for the defendant's restrictive practices, reprocessed sales would capture 50% of the total market. The judge ruled that this claim was mere assertion and could not be presented to the jury. See *Pacific Mailing Equip. Corp.*, 499 F. Supp. at 119-21.

been available for reprocessing. But no adequate supply existed. The high markup per machine was due to a supply shortage caused by the defendant's trade-in and scrapping practices. As a result, it was difficult or impossible for used equipment dealers to acquire used machines at a price at which they could profitably be reconditioned and sold. If supplies had been made available, sales would have expanded greatly.

Defendant's Hypothesis: There was an ample supply of used machines and reprocessing facilities to support a large increase in production of reconditioned machines at existing cost levels, but such an increase in supply could not be marketed at anything like the current prices. Demand was thin (in the business phrase) or "relatively inelastic" (in economists' terminology). To increase sales greatly would require either massive selling expenditures or substantially reduced price and profit margins. Given those demand conditions, it was more profitable for PME and others to maintain high margins on their small volume than to accept the small margins that would be required to expand sales greatly. For example, it is more profitable to sell ninety-five machines at a profit of about \$400 each (profit = \$38,000) than to sell 950 machines at a profit of \$35 (profit = \$33,250).

As a theoretical matter, either hypothesis could be correct. Either is consistent with the observed fact of small sales volume with a high incremental profit margin.

The plaintiff sought to establish a *prima facie* case simply in terms of the number of mailing machines physically destroyed, since every such machine was plainly thereafter unavailable for reprocessing.¹⁴ Plaintiff argued that but for the scrapping, it could have acquired those machines and then increased its sales. As a result it ought to receive damages based on an estimate of lost profits. Such a theory could quickly lead to a substantial measure of damages.¹⁵

The possibility that Pitney Bowes' trade-in practices and prices effectively prevented firms such as PME from acquiring machines which they could profitably resell was examined. It was found that the actual trade-in policies made sense if the trade-in was an inducement to the owner to buy a new machine, but not if the purpose was to maximize the acquisition of used machines.¹⁶ Pitney Bowes'

14. Witnesses testified to observing the physical destruction, with sledge hammers, by Pitney Bowes' branch managers of mailing machines. Indeed, Pitney Bowes' procedures guides gave explicit instructions as to how this was to be done to render the machines inoperable before they were sold to scrap dealers. (One of Pitney Bowes' attorneys referred to this line of testimony, which sometimes became flamboyant, as "the slaughter of the innocents.") Roughly 1000 machines per year were so scrapped in southern California branches, and another 16,000 were scrapped elsewhere in the United States. The source of the great bulk of these machines scrapped were trade-ins taken by Pitney Bowes for placements of new Pitney Bowes machines. A less important source was machines reacquired at lease-end that were not retained by the lessee.

15. For example, 1000 machines per year for ten years at \$400 per machine adds up to \$4 million. Trebled, this amount is \$12 million. From the defendant's point of view, such an amount would be but the beginning: there were, nationally, 50 to 100 other firms like PME. After the judge's ruling, *Pacific Mailing Equip. Corp.*, 499 F. Supp. at 119-21, the plaintiff abruptly shifted to asking damages only in the amount of \$50,000, a truly *de minimis* amount after years of litigation. One suspects the strategy was to try to get the jury to find a violation and award this small amount out of sympathy with a small plaintiff suing a corporate giant. The value of such an award, however, would be the potential collateral estoppel as to the issue of an antitrust violation in subsequent class action suits.

16. The nature of the practices tended to cast doubt on this possibility. Pitney Bowes' trade-ins are

trade-in values are (and were) regarded as low by its customers, a small fraction of the replacement cost. In fact, the evidence clearly showed that Pitney Bowes' regular trade-ins were not preemptive in dollar amount.¹⁷

Moreover, if PME and others were having difficulty obtaining used mailing machines, they could have employed commonly used techniques, such as advertising and calling on potential sellers, to persuade customers who did trade in to Pitney Bowes not to do so, but rather to sell to PME instead. The record shows that PME had not undertaken any systematic means of contacting owners concerning its willingness to buy their used machines before they were traded in. *PME's efforts and its advertising focused exclusively on selling, not buying.* Indeed, if the supply of used machines were truly the bottleneck, major expenditures of time and money would have been devoted to searching for and purchasing used machines from present users. All of these factors suggest that the traded-in machines (in the quantities available) were not, on average, worth their trade-in value as used machines. This was not because the trade-in cost was preemptive relative to existing resale prices; it was rather that there appears to have been an elastic supply of machines at the trade-in price, combined with an inelastic demand for such machines.¹⁸

Supposing that, as defendants claimed, there was no shortage of used machines. Does the high variable profit and low quantity sold of reprocessed machines imply output restriction and monopoly power in the secondary market? While that is, of course, a possible answer, it is neither necessary, nor do we believe it to be the fact. The reason is that the effective "supply curve of used machines to reprocess" must reflect both the supply-price of a used-machine once located by

offered as an inducement to potential customers to persuade them to trade up to a newer, better machine. Pitney Bowes' sales literature, especially when there are special promotions, establishes that amounts offered for trade-ins are discounts for trading up and are not offered in an effort to purchase used machines. The trade-in is available only if the customer in fact buys or leases a new machine. Pitney Bowes refuses to buy machines outright from owners. If Pitney Bowes were actually attempting to limit the supply available to PME and other dealers, it would presumably be willing to purchase machines outright and to allow as many machines to be traded in at a time as possible.

17. The amount of each particular trade-in depends upon the age and the most recent list price (MRLP) of the machine being traded in, not on the traded-in machine's condition. Mailing machines are highly durable, with an average useful life of 10-20 years, and many last even longer. For machines over about three years old, trade-ins were in the neighborhood of 10-15% of MRLP.

Every machine actually traded in to Pitney Bowes was potentially available to PME and other used equipment dealers before it was traded in, at a price equal to or slightly in excess of the traded-in price. Customers are willing to accept 10% of MRLP as a trade-in only if they cannot do better in the open market. PME's actual acquisition price for its machines was essentially the same as the lowest trade-in prices—about 10% of MRLP. This was true model by model. Had PME been willing to offer 15% of MRLP, it could have outbid Pitney Bowes' regular trade-in allowance on the vast majority of its trades. This would have increased PME's variable cost by about \$60 per machine, but left a variable profit of nearly \$350.

18. Some trade-ins by Pitney Bowes were at a substantially higher percentage of MRLP as part of campaigns to induce sales of particular machines. Special trade-in promotions are a common selling device for mailing machines, office equipment, and many other durable goods. Analysis of Pitney Bowes' (relatively infrequent) special promotions showed them to be equivalent to special discounts of from 5% to 15% of the sales price of the new machine. Such campaigns did sell new machines, but they never stopped the steady flow of trade-ins at the "regular" trade-in figure. Indeed, Pitney Bowes eventually decided to abandon such special promotions on the evidence that they merely accelerated trade-ins that were due to occur within a few months anyway.

the reprocessor, and the finding-cost of locating the owner and persuading him to sell the machine.

Suppose there is a very large (even infinite) number of old and reprocessable machines most of which have no economic value to their owners. A few of these are aggressively offered by owners for resale. Others might be called forth by an advertising campaign; still others must be ferreted out by purchasing agents; yet others require persuading the owner-user to abandon them in favor of a newer, better, more reliable machine. These conditions lead to an upward sloping supply curve.

Even if the supply-price paid for used machines, once "available" were constant at, say \$100 per machine, the supply-cost would rise from \$100 as the machines became increasingly difficult to locate. The apparent "high price, high margin, low quantity" equilibrium is all of those things only if one confusedly believes the supply curve to be perfectly elastic, rather than sharpening increasing in output.

This supply-adjustment is quite equivalent to the need to subtract selling expenses from the demand curve to have a net-demand curve for firms in monopolistic competition. It is often simpler to deal with unadjusted demand and supply-price curves and recognize the need for a substantial variable markup of product price over the out-of-pocket supply price to cover the other costs of finding and selling to either customers or suppliers.¹⁹

Without more, the defendants might have urged the jury to attribute the shortage of machines available for reprocessing to the failure of firms such as PME to become active acquirers or to seek to outbid Pitney Bowes for trade-ins. However, some direct evidence supporting this proposition was also available. In addition to outbidding Pitney Bowes for machines traded in, a potential rebuilder of used mailing machines could have obtained used machines from owners of the following types of machines: (1) Privately owned mailing machines that went out of service and were not taken in trade (or otherwise acquired) by Pitney Bowes; (2) Machines in service whose owners might be persuaded to sell or trade in to a used dealer rather than using the machine for an additional year. These owners could then use the money realized to apply to the purchases of new or better machines.²⁰

As to the first potential source, a large number of machines in fact left active service in southern California every year and were not acquired by Pitney Bowes. Defendant's economists attempted to estimate this number. (Because every machine in service has a rented meter, and because detailed records of meters in service must be kept by Pitney Bowes, such estimation could be done with reason-

19. The subsequent availability of the DAL, discussed *infra* p. 78, might have been expected to flatten the supply curve [but raise its intercept level]. A lesser margin would be required because the finding cost was reduced. This would be expected to lead to an expansion in used sales, as indeed occurred. A vast expansion, with low variable profits was surely prevented by the "thin" market on the demand side, and the need to spend increasing efforts on selling expenditures.

20. The age of machines in service is particularly revealing on this point. Roughly 50% of all mailing machines in service have been in their present location for five or more years, 25% for ten or more years and 10% for fifteen or more years. Such older machines were ripe for purchase and constituted a major potential source of supply to reprocessors.

able accuracy.) These estimates suggested nearly 1,500 machines in Los Angeles left service per year in addition to those that were reacquired. Thus for every machine traded in there was another (and more) that disappeared into a back room or junk yard. Some of them were the ultimate source of PME's supply, but many more were never acquired and presumably ended up as scrap metal.

As to the second source, there were two groups of machines in service that were potentially ripe for acquisition. One group consisted of older models no longer produced, of which there were about 10,000 still in service in the Los Angeles area in 1975. The second group consisted of old machines generally, whether or not they were still being produced, as determined by years in service. One must assume these machines were ripe for acquisition and reconditioning. For example, in 1976 in southern California there were at least 100 machines in service over five years old and forty over ten years old *for every machine that PME sold*.

To summarize, a relatively simple statistical compilation suggested a large potential source of supply of used machines. For example, for 1976 the (rounded) numbers for the southern California area were as follows:

Machines traded in to Pitney Bowes	1,000
Machines leaving service, not traded in	1,500
Machines in service over 10 years	8,000
Machines in service 5-9.9 years	9,000
<i>Potential Supply</i>	19,500

Even 1% of this potential supply would have exceeded the total of machines PME reprocessed. And 1.5% would have equaled the total of all used machine placements in southern California.²¹ In this context the scrapping of 1000 machines does not appear to be a significant reduction in the total supply, actual or potential.

3. *Resolution of the Issues.* Whether a statistical showing of this kind would have proven decisive to a jury is not known, because an additional body of data became available. Using *all* of the available evidence, the lawyers for the defendant were able to refute the supply shortage hypothesis. We will describe the nature of this additional evidence in a moment. First, since such a definitive resolution is not typically possible, it is worth considering whether the data of the kind discussed so far would have been sufficient to overcome plaintiff's prima facie case if not at the jury stage, then at some stage in the appellate process. The reason for its decisiveness, in the authors' view, is that once the nature of the competing hypotheses is *clearly posed*, data can be readily assembled and a definitive conclusion can be reached. It is the framing of the problem that provides the major statistical breakthrough to the analysis.

The approach described here—to clearly delineate alternatives in ways that are empirically testable—is not always the lawyer's instinctive approach. In the Pitney Bowes' litigation, the first reaction of the defendant's lawyers was to try to

21. The used-machine market was about 500 per year in southern California with PME and Pitney Bowes accounting for roughly half; another six smaller firms accounted for the rest. Only PME, however, had sales of reconditioned Pitney Bowes machines as its main source of income.

develop an affirmative defense for scrapping, based on an alternative motivation for mutilation: to protect the integrity of the postage meter by preventing someone from defrauding the postal service by tampering with the meter. Such a defense had serious holes in it;²² moreover, it was neither necessary nor sufficient. The defense was not necessary given the superabundance of used machines; it was not sufficient, given monopoly power, for there were surely ways to achieve this asserted objective that did not remove so many machines from service. There seems to be little doubt, given current antitrust interpretation, that *if* used machines were truly scarce, other means to protect the integrity of the meter would surely have been required of one with monopoly power.

Nevertheless, the decisive evidence arose because continuously since April 1, 1977, Pitney Bowes has made available to dealers virtually all used Pitney Bowes mailing machines it acquired, for prices that are not much in excess (on average) of 15% of the most recent list price of new machines. These machines, acquired by Pitney Bowes as trade-ins, at lease-end, or in other ways, are placed on a Dealer Availability List (DAL) for sale within thirty days to any authorized dealer.

During the following twelve months (April 1977-April 1978), Pitney Bowes placed over 20,000 used mailing machines on the DAL. A total of 682 were purchased by dealers. Over the thirty-six months for which data are now available (April 1977 to April 1980), 75,000 machines were placed on the DAL and only 2,427 were taken. The percentage of available machines taken in each twelve-month period varied from a high of 3.4% to a low of 2.8%. In each twelve-month period, over 96% of available machines were not purchased. This superfluity proved true model by model. This is revealing because such machines could have been sold at a great profit per machine if they could have been sold at anywhere near the prices charged by PME and others for rebuilt machines. Since the enormous profits were being left unclaimed, plaintiff's hypothesis was refuted.²³

The DAL experience also decisively dealt with the significance of Pitney Bowes' scrapping and mutilation practices. Over 95% of machines offered for sale on the DAL were eventually scrapped because there were no takers; that is, there was no (nonscrap) market for used machines in the quantity leaving service and being reacquired by Pitney Bowes. In short, disposal was an efficient response to excess supply. Disposal does not affect the supply or the price of the few used machines that do find some aftermarket. The reason is the unscrapped available machines at times for which we have data were sufficiently numerous to avoid any supply shortage.

22. For example, the scrapping did not protect against similar abuses for all of the other potential machines available to a would-be defrauder.

23. It was, of course, necessary to consider the question of whether the small number of purchases from the DAL resulted from prohibitive prices. The available data showed conclusively that they did not. Utilizing PME's reprocessing costs and assuming acquisition at DAL prices would have led to variable profits averaging over \$300 per machine if extra reprocessed machines could have been sold at the prices of machines sold by PME. That is, if demand was infinitely elastic and volume limited by a supply shortage prior to 1977, the DAL should have made a multiple expansion of sales of reprocessed machines possible. The "phantom profits" left unclaimed if plaintiff's hypothesis was fully correct amounted to roughly \$4 million per year nationally. This did not happen, and the unavoidable conclusion was that demand was not elastic and that it was not a supply shortage that limited volume of sales.

Given these data, the jury found for the defendant, despite the court's instruction to assume the existence of monopoly power, and the admitted use of the challenged practices of trade-ins and scrapping. Judgment was entered,²⁴ and costs were awarded to the defendant.

B. The Dangers of Simplicity

As illustrated by the mailing machines case, the presentation of some rather straightforward empirical calculations can be dispositive. In most cases the dangers associated with simplicity, either in the use of theory or in the application of statistical principles to the analysis of data, are substantial. Some direct examples of this fact are given in the discussion of the ampicillin case which follows.

1. *Introduction to the Ampicillin Litigation.* The plaintiffs in this case²⁵ were a substantial number of cities, counties, and states (CCS entities) who purchased ampicillin, a semisynthetic penicillin developed and patented by Beecham, a British drug firm. The defendant, Bristol-Myers (Bristol), received an exclusive license for manufacture and sale in the United States in 1959 from Beecham, which included the exclusive right of Bristol to sublicense in the United States. As a part of the agreement to grant the license, Beecham reserved the right to sell finished form ampicillin in the United States under its own trademark. This allowed Beecham to have a limited access to the U.S. ampicillin market when and if it chose to do so. However, Beecham did agree not to market bulk powder in the United States which would be available for finished production. This provided some assurance to Bristol-Myers that Beecham's role in the American market would be limited. Bristol did license some others to produce in finished form, but refused to allow its licensees to sublicense. Plaintiffs claimed these restrictions and licensing practices caused bulk to be unavailable to "generic houses" and others (wholesalers) who, it was alleged, would have entered the market and sold ampicillin at prices lower than those charged by Bristol, Beecham, Beecham's industrial customers, and Bristol's industrial customers. The exclusion of the competition from generic houses was alleged to have caused the plaintiffs to pay higher prices for ampicillin than they otherwise would have paid. Such conduct was said to have been an illegal exercise of monopoly power and thus violative of section 2 of the Sherman Act. Moreover, the licensees themselves were said to effect an unreasonable restraint of trade under section 1 of the Sherman Act.²⁶

The CCS entities all purchased ampicillin through a sealed-bid competitive bidding process. Over the relevant time period the winning bid prices tended to fall, a pattern consistent with both an increasing number of firms bidding and a steady decline in the cost of producing ampicillin. The central issue in the case was whether, but for the license agreement between Bristol and Beecham, the

24. *Pacific Mailing Equip. Corp. v. Pitney Bowes*, 499 F. Supp. 108 (N.D. Cal. 1980).

25. *In re Ampicillin Antitrust Litigation*, 88 F.R.D. 174 (D.C. Cir. 1983).

26. After at least four years of pretrial activities, trial of the case began in May 1981, in the District Court of the District of Columbia. Near the end of plaintiff's case, a settlement was reached between the parties.

prices paid by plaintiffs for ampicillin would have fallen faster and thus would have been lower than they actually were. The "but for" issue will return at several points in the use of the ampicillin example; for the moment, however, a brief discussion of the product cycle of a drug may help to understand why costs fell and the number of bidders increased over time.

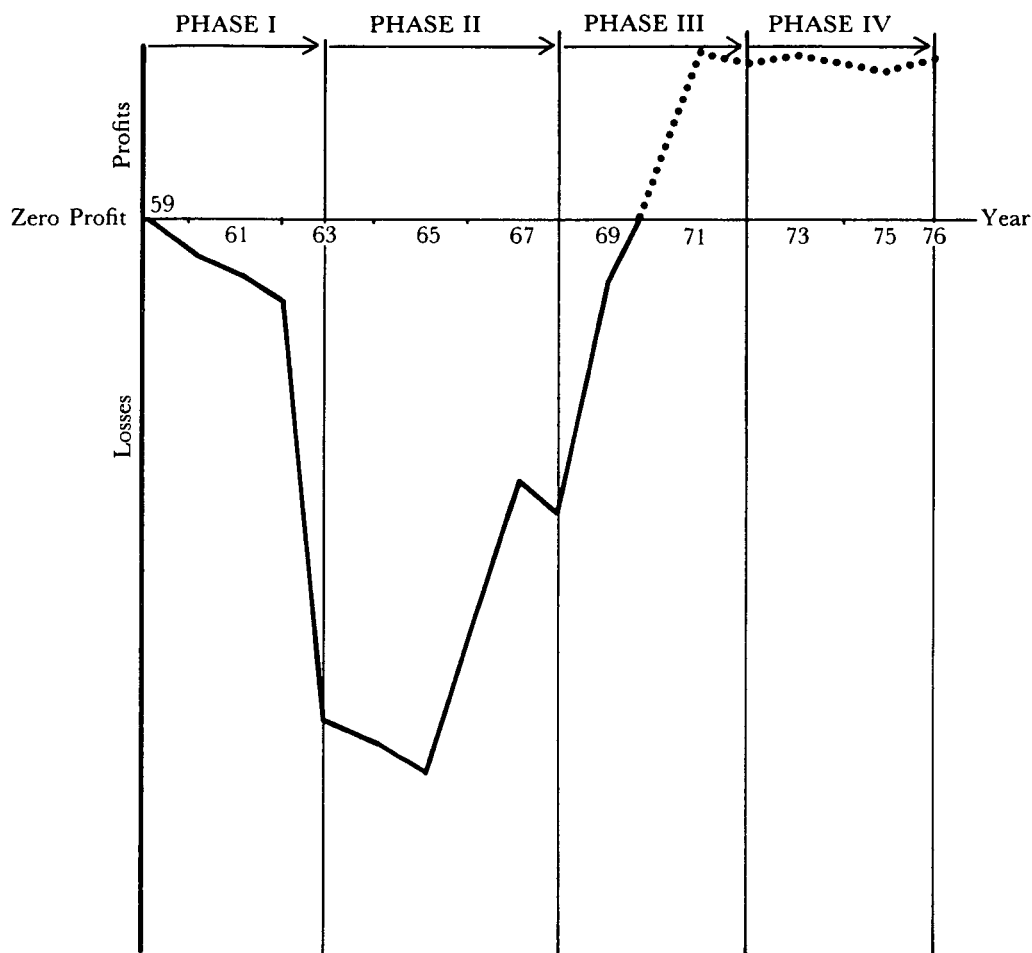
Most new and successful drugs go through a more or less common product cycle. Any new drug product must first be discovered, usually as a result of substantial research efforts. After chemical discovery of a possible new cure or preventive, the drug must be extensively tested for both effectiveness and safety, and permission must be received from government agencies to market the drug. Doctors and hospitals must then be introduced to the drug, persuaded of its potential benefits, and induced to prescribe it. Before a drug can be marketed, a variety of production problems have to be solved. It is one thing to develop an experimental drug in a laboratory; it is another to get the drug into large-scale production and to solve all the finishing, packaging, and selling problems, as well as the additional problem of bringing the product to the point where the revenues from its sale pay for the costs of development, manufacture, and marketing. Many drugs never make it beyond the initial stages. The drugs that do make it beyond this stage, having proven themselves both medically and commercially successful, begin to attract additional firms that want to get into the market either by a license from the patentholder or by infringing on the patent. New firms wish to enter the market in order to share in the future profits which exist after the development, testing, and acceptance of the drug in the market place.

Assuming that most drugs are not successful, motivation for innovation and development of drugs necessarily requires a strong profit potential for those drugs that ultimately prove successful. Trademark and patent laws are designed to meet this need by delaying immediate entry of competitors, thus allowing innovators and developers of drugs to obtain substantial profits when the drugs are successful. Although profits on a successful drug will seem large relative to the costs of developing that drug, when viewed in light of the costs of developing all drugs (both successful and unsuccessful) the profits seem substantially smaller.

Ampicillin was typical of many drugs in terms of the pattern of its product cycle. It was special only in that, in the end, it was a big success. In the ampicillin experience (see Figure 1), the first phase ran from early 1959 to late 1963 with this period devoted to inventing, testing, and acquiring the necessary certification from the FDA (the actual profit numbers are omitted since the changes in profits over the product cycle, not the magnitudes, are of concern). The product was first offered for sale in 1964. This phase was followed by a period of market development and production debugging (1964 to 1968); one marked by high but gradually falling costs per unit and by a relatively small number of firms producing the product. The third phase, 1969 to 1972, marked a period of market maturity during which ampicillin reached peak acceptance and costs per unit continued to fall. By that time, many other firms desired to acquire and sell ampicillin. During this period, patent restrictions became significantly binding and patent infringe-

ment began. Finally, Phase IV, from 1973 to the present, marks a period of a declining market as newer drugs begin to replace ampicillin in the market.

FIGURE 1
THE AMPICILLIN PRODUCT CYCLE, 1959-1976



The pattern of rising production and gradually falling costs in the ampicillin market from 1965 to 1975 thus is typical of most new drugs that prove successful. The question in this particular case was whether the patent agreement which initially excluded generic firms from entering the market would have caused prices to be higher than otherwise in a particular year. As a matter of theory, the answer could surely have been yes or no. It could be yes if output and sales had been artificially restricted and if generic firms would have increased supplies. It could be no if generic firms would have had neither the desire nor the ability to sell their product in those years even absent the patent restriction. The question of what would have happened but for the patent restriction is an empirical question, about which statistical techniques can provide useful information.

To utilize such techniques, it is first necessary to understand something about the drug industry and the markets in which the industry sells ampicillin. The ultimate users of ampicillin are individuals who purchase the drug after receiving a prescription from doctors or receive it in the course of hospital care. The main distributors or dispensers of ampicillin to ultimate customers are drug stores and hospitals. Drug stores purchase ampicillin in what is known as the wholesale/retail market. Part of this market is the brand-name market. In most states during the litigation period, if the doctor's prescription called for a specific brand of ampicillin, the pharmacist could not legally substitute another brand.

Another part of the wholesale/retail market is the generic market. If the prescription was written just for ampicillin, any brand—including unbranded products—could be used to fill it. Such prescriptions are called generic prescriptions. Generic prescriptions tend to become more important late in the product cycle as numerous firms demonstrate ability to produce a product of proven quality. It is important to note that branded as well as unbranded products are sold in the generic market. Private hospitals buy essentially the same way that drug stores do, except that large private hospitals often buy at auctions, using competitive bidding. Government hospitals, including the city, county, and state (CCS) hospitals that were the focus in this case, do most of their purchasing via sealed bids for contracts. These bid requests are necessarily generic bid requests and typically involve requests to supply all of the needs of a hospital for a period such as a year. While the bid requests can be filled by any acceptable brand, virtually all awards made for ampicillin end up being supplied by a brand-named ampicillin.

Some drugs—indeed most—are not licensed by the patentholder. In such situations generic prescriptions or bidding does not occur until late in the product cycle. Ampicillin was an example of a multisource drug from the very start, since both Bristol and Beecham produced and sold the drug. Moreover, Bristol early on licensed others to produce and sell in competition with it.

It is useful to think of the sellers of ampicillin in five groups: (1) Drug manufacturers who both make and sell ampicillin (this group includes Bristol, Beecham, Ayerst, Wyeth, and Squibb); (2) Other major drug manufacturers who make other drugs, but choose to buy their ampicillin in bulk quantities from the manufacturer; (3) Foreign manufacturers whose ampicillin production may find its way into the United States despite the fact that they do not hold patent licenses under U.S. patent law (these firms are called "infringers" and sometimes sell their product in bulk form to the next group); (4) Generic houses which buy ampicillin either in bulk or in bulk quantities of finished products and typically seek to resell it in unbranded form to the retail/wholesale market for filling generic prescriptions; and (5) Wholesalers who may take over the distribution from any seller of finished products by buying in large quantities and reselling it to drug stores in smaller quantities.

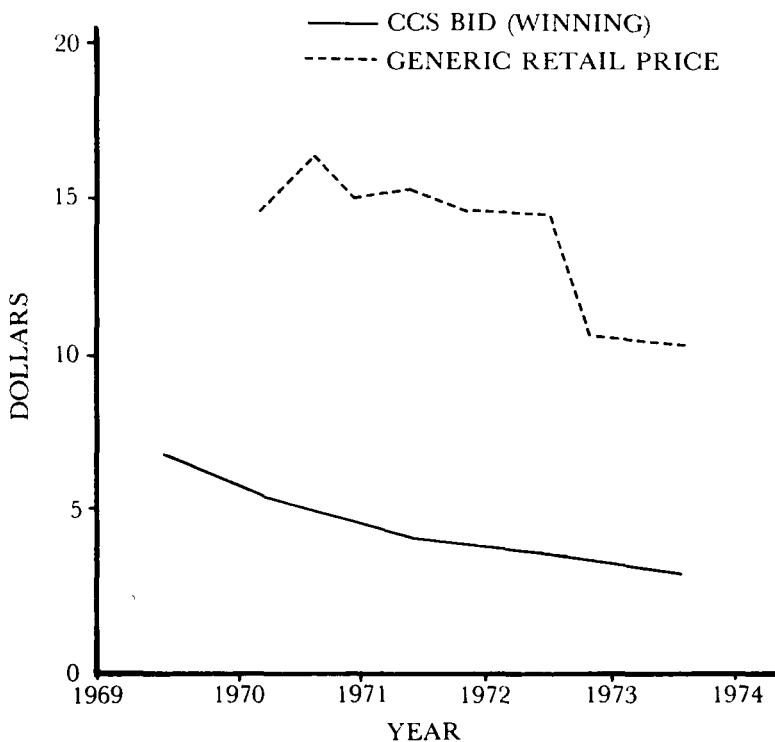
Ampicillin is sold in a number of forms—as capsules, as oral suspensions, and in injectable form. Each of these comes in many dosages and sizes. During the ampicillin litigation, however, most of the court's attention was focused, if only for reasons of cost and simplicity, on two of the most common dosages of capsules, bottles containing 100 or 500 capsules each with a 250 milligram dosage strength.

The quantitative analysis presented here is limited to the former, although the results were essentially the same for both groups.

2. *An Unsatisfactory Graphic Analysis of the Ampicillin Experience.* As illustrated by the mailing machines case, the use of basic economic theory and simple statistics can sometimes be dispositive in an antitrust case. Such is not always the case. There are dangers that the simple approach can bring with it. In this section, relying primarily on the ampicillin case, a few examples of apparently simple arguments that proved to be wrong are presented.

The plaintiffs' expert in the ampicillin case argued that one of the primary causes of the decline in the price of ampicillin in the CCS market was the growth of the number of generic firms selling ampicillin. His argument relied on two facts: (1) over time, as the number of generic houses increased, the retail price decreased; and (2) as generic retail prices decreased, so did the CCS bid price (see Figure 2). From this he inferred that increases in the number of generic firms caused the decrease in CCS bid prices. Crucial to his possibly "plausible" hypothesis was the seemingly strong visual correlation between the generic average retail price of ampicillin and the winning CCS bid price.

FIGURE 2



Such a correlation is suggestive, but it is surely not dispositive. What, for example, accounts for a link between the number of firms in the retail market and prices in the bid market? How were prices computed? Do they even accurately

reflect what they purport to show? Are there other omitted variables that affect the relationships?

The graphic argument made by plaintiffs' expert, although simple and appealing at first glance, suffers from disabilities in all of these respects. The major problem is the problem of "spurious correlation," in which two variables that are correlated are not necessarily causally related but instead move together due to other variables. Seen in the context of the ampicillin case, variables other than generic bidders and generic retail price provide a persuasive explanation of the CCS winning bid price, while the variables selected by the plaintiffs' expert do not. This analysis, however, involves the use of multiple regression analysis, and will be examined at length later in this article.

Even at a much less analytic level, however, there are major flaws in the apparently suggestive relationship. For example, the plaintiffs' expert obtained the generic average retail price by taking a simple average of the listed retail prices of firms selling ampicillin in generic form. This approach failed to take account of any discounts off retail price that might have been made. Nor was the weighting done by the amount of generic ampicillin actually sold. A high price at which no sales are made was given the same weight as a low price which moved large quantities. Moreover, the argument relating generic prices to CCS prices took as implicit the fact that what happens in the retail market is directly linked to what happens in the CCS market. But what is the link?

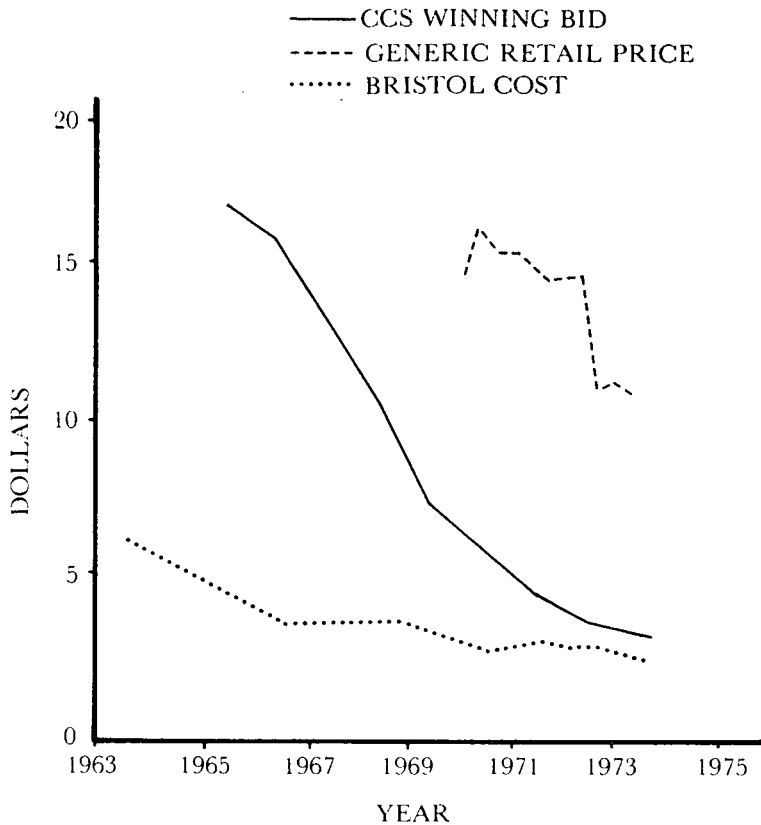
As suggested earlier, the CCS market is different from the retail market, both because of the method of selling and because of the relatively low profit margins involved. Such competition, which may have been stimulated by generic firms after 1970 in the *retail* market, did not necessarily have anything to do with the winning bid price in the CCS market. As Figure 3 shows, prices were falling in the CCS market over the period between 1963 and 1970 when there were no generic houses selling in any market. A measure of the cost of producing ampicillin is included here as a useful point of comparison for later discussion.

Did generic houses affect CCS bid prices after 1970? Two facts cast doubt on that proposition. First, generic houses did not even often bid in the CCS market. Of sixty-eight wholesalers and generic houses, thirty-nine never bid at all; fifty bid four or fewer times. In fact, of the sixty-eight generic houses and manufacturers who bid in the CCS market over the entire bidding period, only five bid twenty or more times.²⁷ Even this is a very small number when one compares it to the total number of bids made by a major manufacturer such as Bristol-Myers, which bid roughly 1000 times throughout the same period. Second, generic houses never won a bid in the CCS market.

But, ignoring all of these factors and assuming that generic houses exerted as strong an influence on prices in the CCS market as they did in the retail market, is the pattern of declining retail price and rising number of sellers indicative of the underlying relationship between the variables? The answer is that this is not nec-

27. The five "big" bidders were Reid-Provident, Zenith, Spence-McCord, Biocraft, and Sherry.

FIGURE 3



essarily so since there may be a third or fourth variable which is directly responsible for declining CCS market price. Retail price and number of sellers may be moving in the same direction but may not be causally related to CCS price. As the multiple regression analysis will make clear, this was the case. Thus, although graphs can be very helpful in describing relationships, they cannot settle the question of which variable is causally responsible for the movement of another variable.

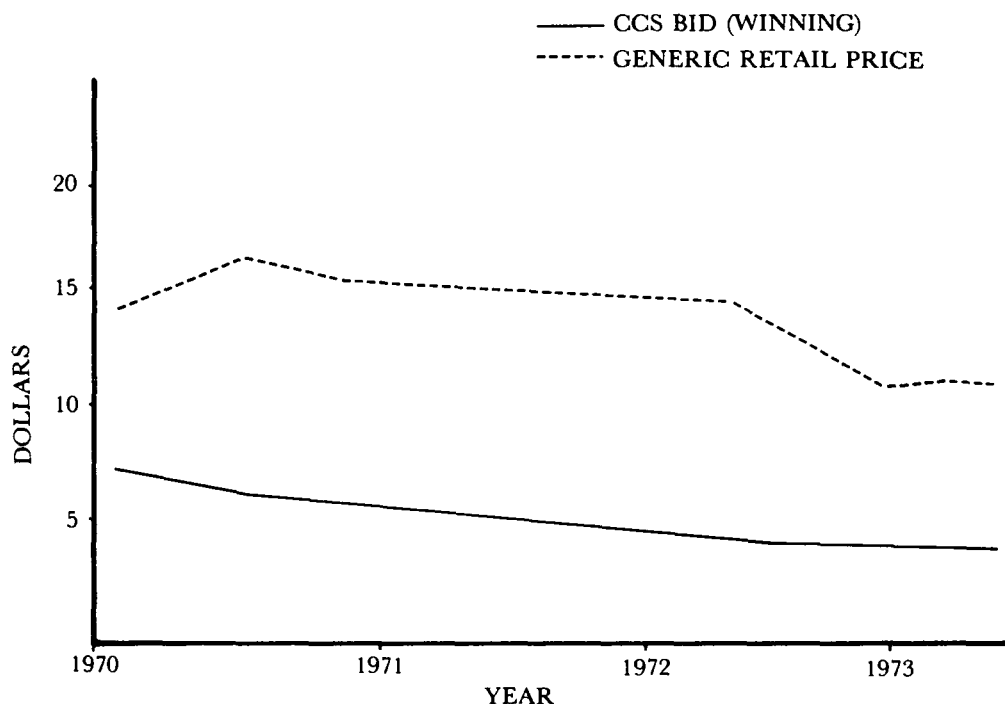
Before leaving this example, two recurrent problems with use of graphs in an adversary proceeding should be noted:

(a) *Choice of time period.* Figure 2 illustrates the relationship between retail price and CCS bid price for the period 1970 to 1974, a relatively short time in which the apparent relationship between the two variables was quite strong. When a graph of CCS bid price for a longer period of time is reexamined, as in Figure 3, it becomes apparent that the relationship between the two variables is not nearly as strong as one might have thought. Of course, it is hard to generalize for the period prior to 1970 since no retail generic firms were in the market; however, the impression of causation certainly does not jump to the eye in Figure 3. Thus, one must be careful when examining graphs to see that the period of time

for display has not been chosen to suggest a misleading picture of the overall pattern of the relationship between the variables.

(b) *Units of measurement of the variables and the scaling technique.* Compare Figure 2 with Figure 4. The two graphs plot the identical numbers, but a change in scale has flattened out the curves to the point where the changes seem much less interesting.

FIGURE 4

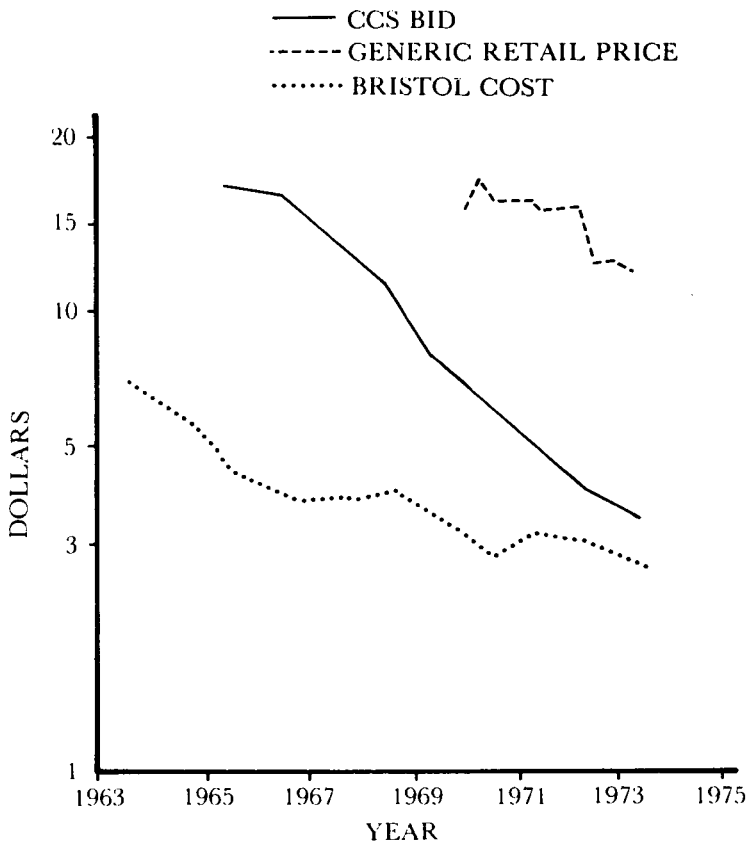


Next, consider a more subtle scaling issue by comparing Figure 3 and 5. Each plots the same data, but one uses an arithmetic scale, the other a ratio or percentage scale. They look quite different, and the apparent significance of generic retail prices is very different.

To use graphs effectively, it must first be determined what question is to be answered through the use of the graph. If, for example, the objective is to compare changes (decreases) in the retail price of generic drugs to the changes (decreases) in the price of the CCS drugs, a strong relationship between the two will occur when the percentage change in the retail generic price is approximately equal to the percentage change in the CCS price. The graphs in Figure 2 or Figure 3, however, focus on the *levels* of prices and costs.

A ratio or logarithmic scale, on the other hand, measures equal units of percentage changes, rather than equal absolute units. For example, on the ratio scale, a fall in price from \$20 to \$10 (a 50% decline) will have the same distance as a fall

FIGURE 5



in price from \$10 to \$5 (also a 50% decline). Notice on Figure 5 that a cursory examination of the graph yields a different impression than that which might be gleaned from an examination of Figure 3. In Figure 3 Bristol's direct cost and CCS bid price are apparently not closely related. In Figure 5, however, the relationship between Bristol's direct cost line and the CCS bid line seems to be much stronger than any relationship with generic retail price. The relationship between generic retail average price and the CCS bid line is still there, but it appears substantially weaker. If Figure 5 had been examined prior to Figure 2, it could have been concluded that declining costs provided the explanation for the decline in the CCS bid price. (As will be seen later, declining cost is in fact an important explanatory variable—but that cannot be proven by a graph.)

This example should serve to warn that while graphic analysis may be suggestive, it requires caution. It is possible with graphs both to mislead and to suggest an answer to the wrong question. As a general rule, while a graphic analysis may disprove some hypotheses (that prices were constant, for example), it can never definitively decide questions of causal impact. Yet, causation is central to most legal questions. Here, more sophisticated techniques can help.

III

MULTIPLE REGRESSION

A. An Introduction to Multiple Regression

There are a number of situations in which the use of simple statistics and/or the use of economic theory will not suffice in an antitrust case. The reasons may vary from case to case, but they almost always relate to the possibility of spurious correlation. Spurious correlation arises when two variables move in the same direction and are apparently closely related to each other. In fact this relationship is purely a chance relationship with no particular import or meaning. For example, the fact that death rates are high in Florida (due to the high mean age of the population) where the sun shines a great deal of the time suggests nothing about the casual effect of bright sunlight on mortality. To avoid the problem of spurious correlation, explicit account must be taken of the systematic relationship between a variable being studied in the case and the other variables that are asserted to affect the one in question. A good illustration of this point can be taken from the ampicillin case described earlier.

The plaintiffs' expert testified that the price of ampicillin sold in the CCS market from 1965 to 1972 was "too high." The reasoning and the assumptions underlying this assertion might be made explicit, as follows:

- (1) There is a negative association (or correlation)²⁸ between the price of ampicillin in the retail market and the number of generic firms in the market.
- (2) This correlation is explained by the entrance of generic firms, which increase competition and thus cause the price of ampicillin to fall.
- (3) What was true about the retail market would also have been true for the CCS market.
- (4) Since generic firms caused a lower price when they did enter, they would have caused a lower price earlier in time had their entrance into the market not been barred.
- (5) The patent licensing agreement between Beecham and Bristol-Myers served as a barrier to entry for the generic firms in the CCS market, and thus delayed their entry.
- (6) Therefore, the patent agreement delayed price decreases in the CCS market.

Several of the assumptions underlying this argument can be tested empirically, but for the moment it will suffice to focus on one issue. Does the correlation between the number of firms and the price of ampicillin explain anything about causality? The answer could well be yes, since a series of events can be imagined in which the entrance of new firms would lead to a lower price. Such a causal link might arise if: (1) new firms may win bids, thus having a direct effect on price; and (2) manufacturers currently in the market lower their bids to fight off the competi-

28. Two variables are positively correlated when "on average" values above the mean of one variable tend to be associated with values above the mean of the second variable, and vice versa. Negative correlation occurs when relatively high values for one variable are associated with relatively low values for the second. The concept of correlation used here is "simple correlation" since only two variables are involved. "Multiple correlation" involves three or more variables.

tion from the new firms, thus lowering prices even though the new entrants win no bids.

Of course, the mere possibility of such a causal link is not itself evidence. There is an important reason for this. The facts may show that neither factor occurred: if no new firms win bids, *and* manufacturers do not respond to the new firms in choosing their bidding strategies, then any correlation must be spurious. (This empirical question will be discussed later.)

How might such a spurious relationship occur? The price decline in the ampicillin market (which correlates with the increase in number of sellers) may be due to a third (or fourth) factor not previously mentioned, such as the decline in the cost of producing the product. Assume for the moment that cost is the true explanatory factor, while the number of firms is not. Given that cost declined, price would have declined whether the number of firms increased, decreased, or stayed the same. As it happened, the number of firms did increase. It could then be asserted that the relationship between the number of firms and the price of ampicillin involves spurious correlation, not because the correlation is not there, but because the causal link behind the correlation is invalid.

Statistics involving the analysis of one or two variables are usually insufficient to settle issues involving spurious correlation, although they may raise them. If each of two different variables is correlated with some third variable that each of them helps to explain (the "dependent" variable), it is possible that either both or neither are causally related. Multiple regression, on the other hand, is generally well suited to this purpose. Multiple regression applied to a well-developed theory of the case allows one to choose among alternative hypotheses and to sort out those correlations which are plainly spurious from those which are not. It does so by including as explanatory variables not only those variables of direct relevance to the legal issue, but also those which must be controlled for, lest a false conclusion be reached. Before applying this concept to the ampicillin example, a brief overview of the multiple regression technique might be helpful.

B. The Technique of Multiple Regression²⁹

1. *The Choice of Variables.* Multiple regression involves the statistical analysis of the relationship among three or more variables. The variable to be explained is called the "dependent" variable. Other variables, the "explanatory" variables or "independent" variables, are chosen because they are thought to aid in the explanation of the dependent variable. Some of the explanatory variables are chosen because they are hypothesized to be the causal variables in one or another of the alternative theories of the case. Cost and number of generic bidders are two such explanatory variables in the ampicillin case. However, other variables which represent theoretically plausible additional influences on the dependent variable

29. The following discussion is necessarily brief. For a more complete description of multiple regression, see Fisher, *Multiple Regression in Legal Proceedings*, 80 COLUM. LAW REV. 702 (1980). See also *Vuyanich v. Republic Nat'l Bank*, 24 FEP Cases 128 (1980). For a textbook treatment of applied econometrics there are a number of sources. One which the authors find particularly congenial is R. PINDYCK & D. RUBINFELD, *ECONOMETRIC METHODS AND ECONOMIC FORECASTS* (2d ed. 1981).

should also be included, even if they are not the direct focus of the case. The econometrician's initial concern is with providing the opportunity for any variable thought to be potentially relevant to have its actual influence examined. In the ampicillin case, the number of manufacturing firms bidding in the CCS market would clearly be one sensible variable for inclusion in the model on the ground that competition among manufacturers is widely believed to affect price. Other variables of probable initial interest would be the number of industrial customers in the market and the size of the drug order. The first would be related to competition, the second to the cost of distribution.

Not all possible variables which might have some influence on the dependent variable can be included. There are literally thousands of possible variables. Some that might otherwise be helpful cannot be measured, while others are simply not known or understood. Nevertheless, an attempt should be made to identify all known or hypothesized and measurable major explanatory variables. Excluded variables should be either variables shown to be not important or variables that one is content to treat as minor. The summary effect of these excluded variables is included in the multiple regression model in the random disturbance or "error" term. If, in the end, the unexplained portion of the relationship seems unacceptably high, one may seek to discover whether there is some (previously undetected) major variable missing.

Why is it so important to include all major explanatory variables whether or not they are directly relevant to the case? The answer is that in many cases, one explanatory variable (for example, number of manufacturing firms) may be related to (or correlated with) the explanatory variable of interest (for example, number of generic firms). Failure to account for the number of manufacturers may lead to a misleading estimate of the effect of the number of generic firms on the dependent variable. Thus, one may credit to one included variable an effect that is really caused by the excluded variable.³⁰ For example, in the ampicillin case, the estimate made of the effect of generic firms on price may be a poor one unless one controls for cost. Technically, it can be said that the omission of explanatory variables which are correlated with the one of interest leads to biased and inconsistent estimation. Essentially this means that any results obtained are likely to be misleading.

What about omitted variables that are not correlated with other variables? Does their omission bias the analysis as well? The answer is no. Assuming *arguendo*, that: (1) the multiple regression analysis is concerned solely with the effect of the number of generic firms on price and (2) cost is explicitly accounted for in the analysis but a variable affecting the distribution cost of the drug (for example, a measure of the volume or size of the drug order) is inadvertently omitted, the

30. For example, suppose that both the number of manufacturers and the number of generic firms increase year by year, but that the number of manufacturers is the causal variable that explains the movement in CCS price. If the number of manufacturers as a variable is omitted from consideration and remains in the error term, one would mistakenly assume or conclude that the number of generic firms is a primary variable which explains movement in price. In effect, the number of generic firms variable is getting credit not only for its own direct effect on price, but also for the indirect effect of the number of manufacturers variable.

measure of the effect of generic firms on price will be essentially unchanged or unbiased, provided that distribution costs are unrelated to the number of generic firms. This may be a reasonable expectation if distribution costs have varied very little over time, and the number of generic firms has increased steadily. Of course, a measure of the effect that distribution costs themselves might have on price will not be obtainable, but an estimate of the effect of generic firms will be acceptable.

While the results in such a circumstance are not biased, there is still a potentially serious problem. If the effect of the omitted variables is substantial, that is, if there are many omitted variables that themselves do much to explain the changes or variation in price, the statistical model itself will be unreliable. What happens, in effect, is that the random noise generated by the rather large error term is likely to swamp or mask the signal given by the measured effect of the number of generic firms on price.³¹

What does this mean for attorneys who wish to make use of multiple regression studies? There is an important distinction to be made between bias and “explanatory power” or reliability. Bias caused by the omission of important variables that are related to included variables of interest is a serious problem. Since reliance on regression results assumes the absence of a significant systematic relationship between the error term and the explanatory variables in the model, the presence of such a relationship introduces bias. By either overstating or understating the magnitude of an effect of one variable on another, one can misinterpret the results of the analysis.

However, even if the estimates of the included explanatory variables’ influence are obtained in an unbiased manner, problems still exist since the unbiased result will only be correct on the average. In other words, any estimate of an effect obtained will just as likely be an underestimate as an overestimate. But, how accurate the measure of an effect will be also involves the issue of reliability. In general, the more complete the explained relationship between the included explanatory variables and the dependent variable, the more reliable the results. Thus, the list of variables unaccounted for goes to the possible reliability of the results, for the explanatory variable or variables under study may be only a trivial part of the total.

2. *Estimation of a Multiple Regression Model.* Once a decision about the variables to appear in the model has been made, the model can be fully specified and then estimated. Starting with a very simplified example in which the variable that will be explained is price, denoted P , while (unrealistically) the only possible explanatory variable is the number of generic houses bidding, denoted N , the specification of the model requires not only an assumption about the variable or variables to be included, but also an assumption about the “functional form” in which the variables enter. In the interests of simplicity, the following linear form will be

31. In technical terms, it can be stated that the larger the variance of the disturbance term, the more unreliable is an estimate of the model.

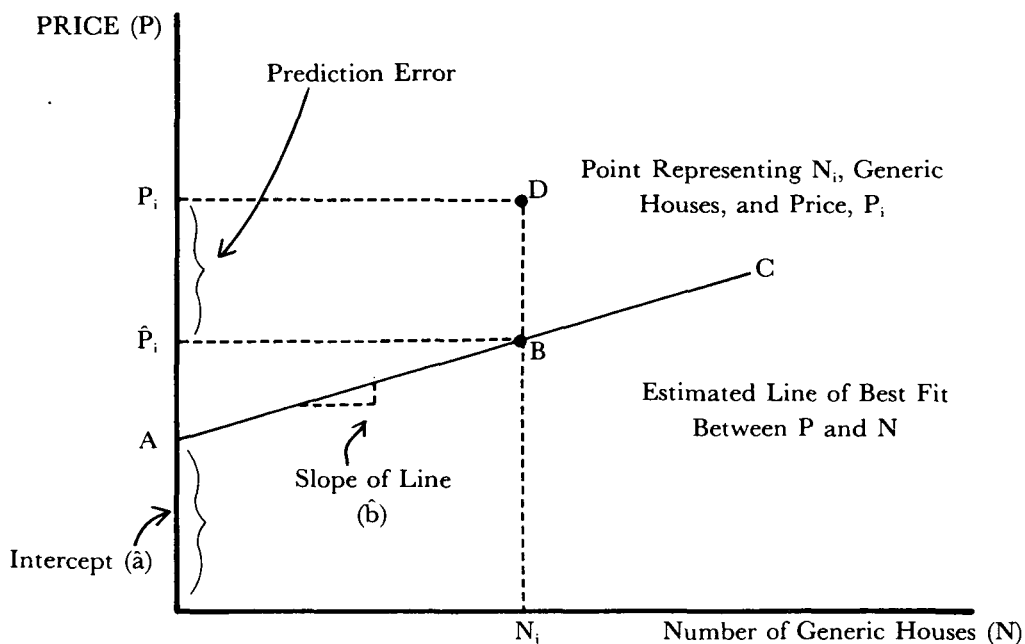
assumed: $P = a + bN + e$.³²

In the equation set out above, a and b are the "parameters" to be determined from the data, while e represents the random disturbance term. Were there no disturbance present, the relationship between P and N is specified to be a straight line that describes the systematic relationship between the two variables. Since not all of the actual observations fall on the line, the variable e is required to account for omitted factors. In this simple two-variable example, a represents the intercept of a straight line while b represents its slope.³³

To measure the specific relationship between number of generic firms, N , and price, P , numerical estimates of the parameters a and b must be obtained. Econometricians have developed a relatively simple and straightforward means of determining these parameter estimates. This approach is to analyze all of the information about the variables N and P in order to determine those estimates which yield the "best fit" between the data points and the straight line which is determined by the estimated parameters.

To see how the "best fit" is determined, we might denote the parameter estimates in Figure 6 by \hat{a} and \hat{b} . Then the "predicted value" of the equation is $\hat{P} = \hat{a} + \hat{b}N$, where \hat{a} is the intercept of the line, and \hat{b} is the slope. The variable \hat{P}

FIGURE 6



32. Variables such as the number of manufacturing firms and costs are omitted solely to simplify this example, not because they are unimportant variables.

33. The intercept represents the value of the variable P when the explanatory variable N takes on the value zero. The slope represents the change in P associated with a unit change in N . If the relationship between P and N is a positive one, that is, if P increases as N increases, the slope will be positive. On the other hand, if P falls when N increases, the slope will be negative.

represents our best guess for P , given information about N . In terms of the graphical relationship, it represents the point on the straight line associated with the value of N that is chosen. For example, in Figure 6, line ABC represents the estimated line which best fits the available data relating P to N . Only one actual data point, D , is shown on the graph. Unfortunately, point D , representing price P_i and number of firms N_i , does not lie on the line. Had this information been unavailable, the best guess for the price associated with N_i would be \hat{P}_i , as given by point B .

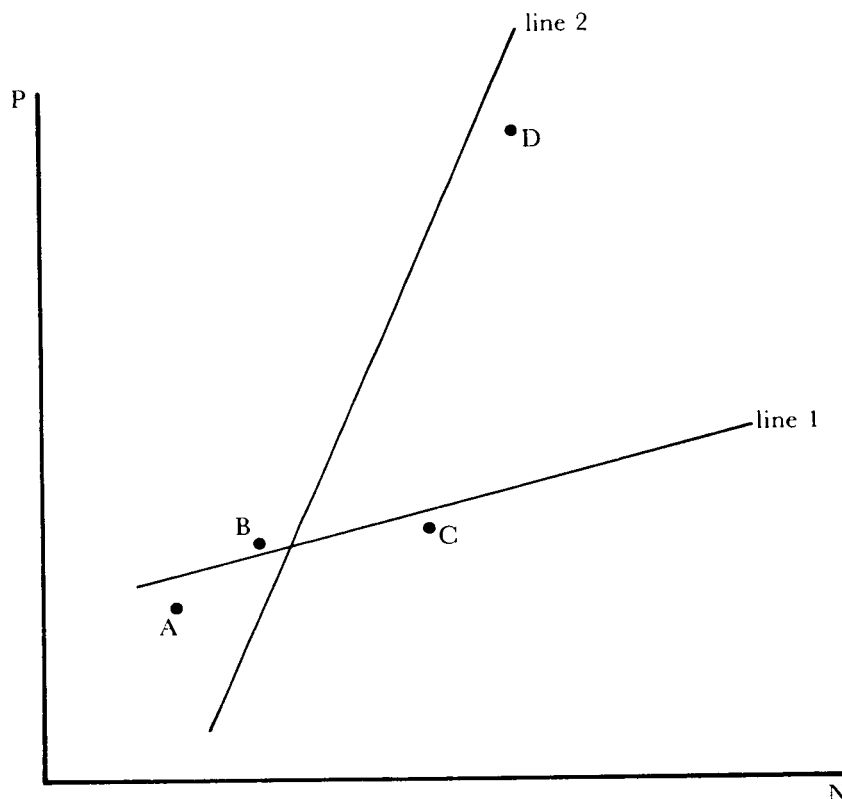
If all points lie on the straight line, each predicted value of price (sometimes called fitted value) will equal the actual price. However, when more than two data points are given, it is possible and even likely that some points associated with predicted values will not lie on the straight line. The mistake, prediction error, or deviation which results is given by $P_i - \hat{P}_i$. For example, in Figure 6, the deviation associated with predicted price \hat{P}_i is given by line segment DB .

At this point it is necessary to define what makes one estimated equation better than another. Plainly, if one equation leads to smaller prediction errors than another, it is the better method. But what is meant by smaller prediction errors, and what if there is more than one technique to choose from? On the basis of simplicity and statistical advantage, most econometricians choose the best fitting equation by determining parameter estimates which minimize the sum of the squared deviations between the predicted P and the actual P . The sum is taken over all observations on the variables which are available. Least-squares regression, as this technique is called, is by far the most popular form of estimation available. It receives almost universal usage and is therefore the obvious technique for legal applications. In fact, the least-squares technique is appropriate in most cases. It has substantial advantages over most other alternative estimation techniques.³⁴

Least-squares regression does have one disadvantage, however, that is worth some consideration. The least-squares regression line can be very sensitive to extreme data points. This sensitivity can most easily be seen with reference to Figure 7. Assume initially there are only three data points, A , B , and C , relating information about the variable N to the price P . The least-squares line describing the best relationship between points A , B , and C is represented by line 1. Point D is called an *outlier* because it represents a point that lies far from the regression line that fits the remaining points. When a new best-fitting least-squares line is reestimated to include point D , line 2 is obtained. Clearly, Figure 7 shows that the presence of the outlier point D has a dominant effect on the slope and intercept of the least-squares line. It is because least-squares attempts to minimize the sum of

34. First, least-squares yields a unique regression line. Second, least-squares is relatively inexpensive from a computational point of view. Third, it has a number of technical statistical properties that make it very easy for one to test the validity of the results obtained. Other techniques can be devised which have some of these advantages and perhaps others, but the alternatives are more costly and substantially more complex. From the point of view of litigation it has the advantage of being the almost universally used method.

FIGURE 7



squared deviations that this sensitivity of the line to individual points can sometimes be substantial.

What makes the problem even more difficult is that the impact of an outlier may not be readily seen if deviations are viewed from the final regression line. The reason is that the influence of point *D* on line 2 is so substantial that its deviation from the regression line is not necessarily larger than the deviation of any of the remaining points from the regression line. While not as popular as least-squares, alternative estimation techniques which are less sensitive to outliers are available.³⁵

What does all this mean for legal statistical work? Because of the sensitivity of least-squares to outliers, analysts should at a minimum test to see whether the results are sensitive to individual data points. One relatively simple way of doing this is to reestimate the least-squares regression line, dropping one or more data points that are of concern, to see whether any of the regression results are decisively changed. If there is little change, outliers should be of no concern. If there is

35. Although the number of methods is substantial and they vary along a number of dimensions, the primary difference between these techniques and least-squares lies in the weight given to deviations. Techniques which put less emphasis on extreme data points will be less sensitive to those data points when the regression lines are determined. These techniques are often described as robust estimation techniques. They are discussed in D. BELSLEY, E. KUH & R. WELSCH, *REGRESSION DIAGNOSTICS: IDENTIFYING INFLUENTIAL DATA AND SOURCES OF COLLINEARITY* (1980).

a substantial effect or change, then the outlier must be closely examined to see whether the data point is in fact reliable. If the data point is unreliable because of an error in transmission or collection of data, it clearly should be removed. Otherwise, outliers provide important information about unusual events and should be retained in the analysis. In that case the sensitivity of the results to specific data points must simply be reported.

3. *Interpreting the Regression Results.* To get a clear idea of how regression results are interpreted, it will now be useful to expand the above example to consider the possibility of a second explanatory variable—cost, C . When two or more explanatory variables are included in the model, the technical estimation becomes somewhat more complex, but the general arguments made previously still apply. To understand how to interpret regression results with either one or more explanatory variables, consider the following example:

$$(1) P = a + bN + e_1$$

$$(2) P = a + bN + cC + e_2.$$

The estimated regression “lines” are as follows:

$$(3) \hat{P} = 6.38 - .81N$$

$$(4) \hat{P} = -12.4 + .04N + 3.84C.$$

The first two-variable regression describes the “best-fitting” straight line relating price to the number of generic bidders. The intercept 6.38 indicates that, if there were zero generic bidders, the estimated price would be \$6.38. The slope coefficient indicates that on average for each additional generic bidder in the market, the winning bid price is expected to fall by 81¢. In technical terms, the slope coefficient explains the change in price with respect to N , that is, the change in price resulting from a one unit increase in the number of bidders.

The importance of including all systematically important variables in the model is clearly suggested by the change in the regression results after the cost variable is added to the model. (Compare equations (3) and (4).) Note that the intercept is now negative. Taken literally, the intercept warns that if there were no generic bidders and zero costs of production, the price of the winning bid would be $-\$12.40$. This interpretation is clearly farfetched and helps to provide a general warning about the use of regression models. Recall that the model was specified to be a linear one. Whatever the specification, linear or not, any results obtained are likely to be reliable only in the range of values of the variables for which substantial information exists. Since none of the available data were generated in zero cost situations, the assumption of linearity must come into question if an attempt is made to predict the price so far outside the data range.³⁶ As a general rule the intercept term, implying zero values of *all* included explanatory variables, is so far out of the range of data that it has no significance for policy conclusions. Thus, lawyers ought to be very suspicious about any predictions or policy conclusions which rely on the estimated intercept of a regression line.

36. The cost coefficient has related interpretation problems. See *infra* text following note 37.

Continuing with an interpretation of the second regression equation, it should be noted that the coefficient of the N variable has changed substantially, from $-.81$ to $+.04$. Apparently an extra generic firm, instead of leading to a price decrease of 81¢ , leads to a rise of 4¢ . The inclusion of cost, C , in the regression makes sense, however, since economic theory dictates that changes in costs will be reflected in price. Thus, the regression results in (4) ought to be more reliable than those in (3). Why did the coefficient change so substantially? The answer gets back to the heart of the multiple regression analysis. The coefficient $.04$ measures the effect of an additional generic bidder on price, *holding the effect of cost constant*.³⁷ By including cost explicitly in the model, changes in cost can be controlled for when the effect of N on P is examined. Since cost is an important determinant of price and (as it turns out) cost and number of bidders are inversely related (costs fell over the same time interval as the number of bidders increased), the failure to include cost as a variable biases the coefficient in the first regression. To put the matter somewhat differently, cost in equation (3) has been inadvertently placed in the error term. Unfortunately, as a disturbance term it violates an important assumption of least-squares, the assumption that the disturbance is systematically unrelated to the explanatory variables. Here cost is inversely related to N , so that the coefficient of N in equation (3) is biased downwards. Thus, failure to accurately control for changes in cost would lead to the false conclusion that N and P are directly related. *While they are negatively correlated, that correlation is a spurious one.* Once the proper controls are accounted for, there is no systematic negative relationship between N and P .

The importance of including all variables which are importantly correlated with the explanatory variable of interest can hardly be overstated. So long as this is done and the disturbance and the explanatory variables remain uncorrelated, the least-squares parameter estimates have two very useful properties: they will be unbiased and consistent. The first means they will be right on average. Thus, if an entirely new sample were chosen and the model reestimated, the new parameter estimate for the coefficient of N would be as likely to be higher than the current estimate as it would be lower. Least-squares will also be consistent, a property which states that as the sample size gets larger, the probability that the parameter estimate will equal the true slope parameter gets larger and approaches one.

Given the severe danger which can arise if a "relevant" variable is omitted from the multiple regression study, why not include all possible variables that might conceivably be relevant to the study? There are two good answers to this question. First, there is a trade-off between bias and reliability. While including additional variables does not bias any of the parameter estimates of the already included variables, it does tend to make all of the information about the existing estimates less reliable. As more variables are added to the model, the existing sample information is directed in part to obtaining slope estimates for the coefficients of the new variables. For a fixed amount of available data, less information is available with which the remaining slope coefficients can accurately be deter-

37. This coefficient is statistically insignificantly different from zero. See *infra* p. 98.

mined. Of course, in certain situations an alternative of collecting more data may exist. In this case, the argument about sample information would become substantially more complex.

Second, some of the added variables may be highly correlated with one or more of the included variables.³⁸ For example, the number of wholesalers might also have an effect on price. However, if the number of wholesalers and the number of generic bidders are highly correlated, it will be impossible to obtain accurate information about the effect of a change in one variable on price with the other variable being held constant. This problem, called multicollinearity, is not a debilitating problem—that is, the multiple regression model still functions appropriately. It does, however, cause problems in attempting to obtain the most reliable estimates possible.³⁹

C. Determining the Reliability of the Regression Model

Least-square regression provides not only parameter estimates that indicate the direction and magnitude of the effect of a change in the explanatory variable on the dependent variable, but it also estimates of the reliability of the parameter estimates and measures of the overall goodness-of-fit of the regression model. Each of the last two factors is considered in turn. To illustrate these measures of reliability and goodness-of-fit, the multiple regression results described earlier are repeated, this time including a set of additional notations that prove to be useful statistics as well:

$$\begin{array}{lll}
 (4) \quad \hat{P} = -12.4 & +.04N & +3.84C \\
 & (.60) & (.08) & (.12) \\
 t = -20.5 & t = .44 & t = 31.7 \\
 R^2 = .94
 \end{array}$$

1. *Standard Errors of the Coefficients and t-Statistics.* The standard error of the coefficient (given in the parentheses just below the coefficient) provides a measure of the reliability of the coefficient. Of course, the coefficient is measured exactly since it is calculated from the available sample information. But the concern here is whether the coefficient provides an accurate measure of the effect of one variable on another. Unfortunately, it may not. Focusing for the moment on the effect of the variable N (number of generic firms) on P (price), it is believed that there is a number represented by b , the slope coefficient, which indicates the true relationship between N and P . According to plaintiff's theory, this number is negative and substantial, while according to defendant's, it should be zero. If one had a substantial amount of information, all of the appropriate controls, all variables measured correctly, and the model specified correctly, it might then be possible to

38. The additional variables must be included, of course, if they have an effect on the variable to be explained since their omission will cause bias.

39. In general, the solutions to the multicollinearity problem involve more data or more information about the regression model. Absent either of these, one must simply interpret regression results in light of any multicollinearity that is present. Since multicollinearity does not create bias in any regression coefficient or statistical test, this presents no special problem. For a further discussion of the problem of multicollinearity and its solution, see R. PINDYCK & D. RUBINFELD, *supra* note 29.

expect the estimated coefficient to exactly measure the effect of N on P . Nevertheless, given the limitations involved with the model as estimated in equation (4), one must expect that the measured coefficient may deviate from the "true" coefficient. In all probability, the measured coefficient will change every time the model is estimated with a different sample. For example, if as time goes by the sample size is increased by including new data points for new years, new slope estimates for the parameter b result. One would expect that the more data one has, the more accurate will be the resulting estimate.

The standard error is obtained as follows. If one varied the sample a large number of times, for each sample one would obtain a new estimate of b . These estimates would vary around the true parameter, but would not necessarily equal it each time. The standard error measures the variation of these parameter estimates about the true parameter b . The result of a greater variance in parameter estimates from sample to sample will be a larger standard of error, and less reliable regression results.⁴⁰ Very small standard errors imply results that are likely to be very reliable.

To be more specific about how reliable the results are, one can utilize basic statistical methods to obtain estimates of how confident one can be about our results. For relatively large samples (often thirty or more data points will be sufficient), the probability that the estimated parameter lies within an interval of two standard errors around the true parameter will be about .95, or 95%. For example, if the estimated parameter b was zero, and our imperfect knowledge about b was summarized by a standard error of .08, one could state that the probability is .95 that the true parameter will be in the interval from $-.16$ to $+.16$, which we call a 95% confidence interval. Since one can never be sure of what the true parameter is, one uses statements of the kind just made to describe the results. Thus, in the ampicillin example, one takes the estimated parameter derived from the least-squares regression, .04, plus or minus two standard errors, .08, and uses that to form the interval $-.12$ to $+.20$. Were one to guess what the true parameter is, one would put odds at 19 to 1 (95% to 5%) that it was in the $-.12$ to $.20$ interval.

Let us now view the example of the .04 coefficient on the variable N in light of the issues in the ampicillin case. When examining any coefficient of a least-squares regression, it is important to look for three items: (1) the sign of the coefficient; (2) the magnitude of the coefficient (is the variable important or unimportant?); and (3) the significance of the coefficient (is the coefficient significantly different from zero?). In the case of the coefficient on the variable N , the coefficient is positive. It is hard to rationalize a positive sign here since, while one can imagine generic firms having a downward or no effect on price, it is difficult to see why their presence should lead to higher prices. Thus, to attorneys, the plus .04 value may seem troubling in that it appears "illogical." The result is a bit less troubling when one realizes that the magnitude of the coefficient is relatively small. A .04 coefficient implies a 4¢ effect of an additional firm on price, a relatively small effect in light of the fact that the average price of drugs in this period was roughly \$5.00. The

40. Actually, the standard error of the coefficient is equal to the square root of the variance of the estimated coefficient about its true value.

problem becomes even less worrisome when one considers the significance of the coefficient. The standard error of .08 suggests that if the true value were zero, one might readily get a parameter of plus or minus .08. Since this range includes zero, as well as negative numbers, it is not necessary to argue for a positive true value for the slope parameter.

Why not argue that the effect of generic firms on price is negative, since negative numbers are included in the $(-.12 \text{ to } .20)$ range? While that possibility cannot be ruled out with certainty, the most likely value is .04, and the likelihood that one would have gotten this sample if the true parameter were negative is substantially less than 50%. In fact, it is no more than 31%.

Now, consider the coefficient on cost. The standard error of .12 is very small relative to the coefficient 3.84. The range of two standard deviations about the coefficient in this case is the range from 3.60 to 4.08. Clearly, this range does not include the value zero. Plainly, the effect of cost is both significantly different from zero and positive. One can be quite confident that cost does not have a zero effect on price. However, the exact measure of the effect of cost on price is still somewhat uncertain. In this case the error possibility relates solely to amount, not to influence or direction, and its approximate magnitude is not in doubt.

A common alternative way of looking at statistical significance involves the use of the t -statistic. The t -statistic is simply the ratio of the estimated coefficient to the standard error of the coefficient. Thus, the t of .44 reported in the regression results in equation (4) is roughly equal to .04 divided by .08. (The unrounded numbers .037 divided by .083 were used in computing.) When describing 95% confidence intervals, the rule is that the true parameter lies in a band of two standard errors on either side of the estimated parameter. In its most frequent usage, the t -statistic provides a test of the hypothesis that the true parameter is zero. A t of 2.0 (for relatively large samples) implies (as before) that the probability that the parameter was generated by a situation in which the true coefficient is zero is equal to 5%. This hypothesis will be supported if the 95% confidence interval includes zero, but it will be rejected if zero is not in the interval. Notice that the intermediate case—where one is on the margin between supporting and rejecting the hypothesis of a zero coefficient—occurs when the estimated coefficient (in absolute value) is equal to twice its standard error. If, for example, the estimated coefficient were .04 but the standard error were .02, the 95% confidence interval would range from 0.0 to 0.8. The t -statistic in this case would be identically equal to 2.0.

A t -statistic of 2.57 is associated with a 99% confidence interval that includes a (larger) band of 2.57 standard deviations on either side of the estimated coefficient, and it indicates significance at the more demanding 1% level. (With a coefficient of .04 and a standard error of .02, for example, the confidence interval would range from approximately $-.01$ to .09, and the zero coefficient hypothesis could not be rejected.)

Either 5% or 1% might be regarded as the statistical equivalents of “beyond a reasonable doubt.” At the 5% level, a coefficient is labeled significant even though it is not significant one time in twenty—at the 1% level, one time in 100. Loosely

speaking, because it is less than 2.0 (i.e. .44) the t -statistic of .44 on the coefficient of the variable N indicates that the probability that the slope is zero or less is substantially higher than 5%. It cannot, therefore, be assumed that it has been shown to be significant, that is, significantly different from zero. By way of contrast, the t -statistic of over 31 for the coefficient of the cost variable C indicates that the coefficient on cost is significantly different from zero. Therefore, the probability of the true slope being zero is very, very small.

When dealing with the burden that the evidence be beyond a reasonable doubt, the 5% and 1% tests may be appropriate. Whether such tests are appropriate in legal proceedings when a preponderance of the evidence is the issue is problematic. A 50% significance level might seem more reasonable on the surface, but when that proposition is examined more thoroughly, there is a problem.⁴¹ No statistical test actually indicates a probability that the coefficient is zero. Thus, to say that the probability is 50% that a variable's true coefficient is zero would be an incorrect statement. The coefficient is either zero (with probability 1) or it is not. The question of how one might decide on the appropriate test when a preponderance of the evidence is at issue is presented later in this article.

Therefore, the attorney is to be warned to be very careful about the use of sophisticated statistical tests. The real question is whether .04 marks a measurable effect of the variable N on price. Rather than choose any particular significance level, a better procedure might be to state that the best estimate is .04 and that the probability of getting this sample result when the true parameter is negative is only .31. The court and the jury are then left with the problem of evaluating the importance of the statistical results, rather than leaving the decision entirely in the hands of the expert.

2. *R-squared.* R -squared (R^2) is a statistic that measures the percentage of the variation in the dependent variable that is accounted for by all of the explanatory variables. Thus, it provides a measure of the overall goodness-of-fit of the multiple regression equation. Its value ranges from 0 to 1, with an R^2 of zero meaning that none of the explanatory variables contribute to explaining the variation of the dependent variable around its mean. An R^2 of one, on the other hand, means that the explanatory variables explain the variation in the dependent variable perfectly. (In the two-variable model, this would mean that the data points all lie on the regression line.) Thus, the R^2 of .94 states that number of generic bidders and cost taken together account for 94% of the variation in the price variable.

What level of R^2 leads to a conclusion of reliability? The problem in interpreting levels of R^2 , a serious one, is that the magnitude of the statistic depends on the characteristics of the data series being studied: whether the data vary over time or over individuals, whether aggregated or disaggregated, and so on. The reason for a low R^2 in cross-sectional studies is that such studies involve attempts to explain differences in individual behavior. These individual differences are

41. See Fisher, *supra* note 29.

likely due to a large number of factors, many of which cannot be measured. As a result, one cannot hope to explain most of the variation in individual behavior. In time series studies, on the other hand, one is explaining the movement of aggregates over time. Since most aggregate time series have substantial rates of growth in common, it will not be difficult to "explain" one time series with another time series simply because both are moving together.

What level of R^2 is satisfactory for legal purposes depends very much on the nature of the problem being studied. In addition, the value of R^2 can be quite sensitive to the number of explanatory variables that are included in the multiple regression analysis. In general, not much weight should be placed on the mere numerical value of this particular statistic. An expert must be asked to state whether the result is unusually high or low for studies of a comparable nature.

Note as well that a high R^2 does not by itself prove that the variables actually included in the model are the appropriate ones. The underlying economic theory is a vital ingredient here. Consider, for example, the argument that the number of generic bidders should not affect winning bid price since generic firms never win bids and are not competitive even when they lose. This argument suggests that a better choice of variable would be the number of all firms bidding in the market, call it *NBIDS*. When the multiple regression was re-run using *NBIDS* in place of *N*, the following results were obtained:

$$\begin{array}{rcll}
 (5) & \hat{P} & = & -12.0 & +.06NBIDS & +3.79C \\
 & & & (.79) & (.04) & (.13) \\
 & t & = & 15.2 & 1.5 & 29.1. \\
 & R^2 & = & .94 & &
 \end{array}$$

The coefficient of *NBIDS* has the "wrong" sign since one would expect more bidders to have a negative impact on price. However, the result is insignificant at the 5% level (since the *t*-statistic of 1.5 is less than 2.0 in magnitude). Thus, this model is preferable to the previous one, yet, the R^2 statistics are the same.⁴² Actually, they are not identical—this R^2 is slightly higher if one looks at the fourth decimal place. The main point should be clear, however. A comparison of R^2 statistics cannot show which model is better—the underlying theory can and should.

To sum up the analysis to this point, three simple estimating equations have been examined:

$$(3) \hat{P} = 6.38 - .81N$$

$$(4) \hat{P} = -12.4 + .04N + 3.84C$$

$$(5) \hat{P} = -12.0 + .06NBIDS + 3.79C.$$

While the number of generic houses is a possible choice for an explanatory variable, a better measure of competition is the total number of firms bidding in the market at any one point in time. In either case, the cost of production does much more to explain price than does any measure of the number of firms that has been used so far. The ampicillin case is later analyzed to see what other explanatory variables might help to explain price.

42. It is preferable both in that the result conforms to the theoretical expectation and that it includes important variables, namely bidders other than generic houses.

D. The Technique of Econometric Forecasting

Frequently in antitrust cases, the question arises as to what the world would have been like had a certain event not taken place. This determination can have relevance for the question of whether there has been some impact from a given event as well as the problem of measuring damages. For example, in a price-fixing case one can ask what the price of a product would have been had a certain event associated with the price-fixing agreement not occurred. Whether the answer is that prices would be no different than they are or that they would be different, the evidence goes to the question of impact. If prices would have been lower, the evidence suggests impact. If one can predict how much lower they would have been for particular transactions or particular classes of customers, the data can help one to develop a numerical estimate of the amount of damages.

In general, a forecast is a prediction about the values of the dependent variable, given information about the explanatory variables. This article describes several important types of forecasts in the context of a time series model, but forecasting techniques can have wide applicability in the cross section context as well. Most econometricians perform *ex ante* forecasts in which they seek to predict values of the dependent variable beyond the time period in which the model has been estimated. (If the values of the explanatory variables are known, the forecast is unconditional; if they must be predicted as well, the forecast is conditional.) More useful in antitrust cases are *ex post* forecasts, in which econometricians predict what would have been but for some allegedly illegal action. An *ex post* forecast has a forecast period such that all values of the dependent and explanatory variables are known. Thus, *ex post* forecasts can be checked against existing data and provide a direct means of evaluation.

As an example, reconsider the ampicillin price regression, equation (4): $\hat{P} = -12.4 + .04N + 3.84C$. One can predict price when $N = 3$ and $C = 5$ to be $-12.4 + (.04)3 + 3.84(5) = \6.92 . This will be an *ex post* forecast for a period when N equalled three and C equalled five.

Given such a forecast, can one tell how reliable it is? The answer is yes for both *ex ante* and *ex post* forecasts, and the answer lies in a statistic called the standard error of forecast. To understand how the standard error of forecast is determined, reconsider for the moment the regression line estimated previously. R-squared was examined as a measure of goodness-of-fit, but one could also have examined the standard error of the regression, SE , which was equal to \$.90. The standard error of the regression is a measure of how far, on average, actual prices deviate from the prices predicted by the multiple regression equation.⁴³ A zero standard error occurs whenever $R^2 = 1$, that is, when all of the data points lie on the regression line. Other things being equal, the larger the standard error of regression, the poorer the fit of the data to the model. However, unlike R^2 , the value of the SE is not measured in percentage terms but rather in the same units as the dependent

43. Actually, SE is calculated as the square root of the mean sum of squared deviations. Squared deviations are used to rule out the possibility that negative and positive deviations will cancel each other out. While other methods also have this attribute, the SE has statistical properties that make it particularly valuable.

variable. To see whether the SE is large or small, it is usual for one to compare it in magnitude to the mean of the dependent variable. This comparison provides a measure of the relative size of the standard error, a more meaningful statistic than its absolute size. In the example, the mean price was equal to \$5.87 so that the ratio of SE (\$.90) to mean price is about 15.4%. A reasonable rule of thumb is that a model fits well if the ratio is 0-15%. A ratio of greater than 20%, whatever the t -statistic, suggests a questionable model.

Now the standard error of forecast can be defined. Assume that one is about to forecast price, given particular values for each of the explanatory variables. The standard error of forecast measures the deviation between the forecast that one might make and the actual price. Thus, it describes how far on average the forecasted prices will deviate from actual prices.⁴⁴

The standard error (SE) and the standard error of forecast (SEF) measure similar concepts and for most purposes can be used interchangeably. However, there is a difference between the two. The standard error measures the errors that are made within a sample in which the explanatory variables are known with certainty. The standard error of forecast measures the error that might be made as the data describing the explanatory variables change over time or over units of study. In general, since forecasts involve data which may be very different from the sample data used to determine the standard error, the standard error of forecast can be substantially larger than the standard error. The farther one is forecasting from the sample information, the larger the standard error of forecast relative to the standard error.

In general, the SEF can be used to determine how reliable a given forecast is. Thus in the example of equation (4) if the sample size is sufficiently large, the probability that the forecast will deviate more than two standard errors from the actual value will be less than .05. Thus, for explanatory variables valued near their means, the probability is roughly 95% that the actual price will be within the interval $\$6.92 \pm \1.80 , or from \$5.02 to \$8.72.

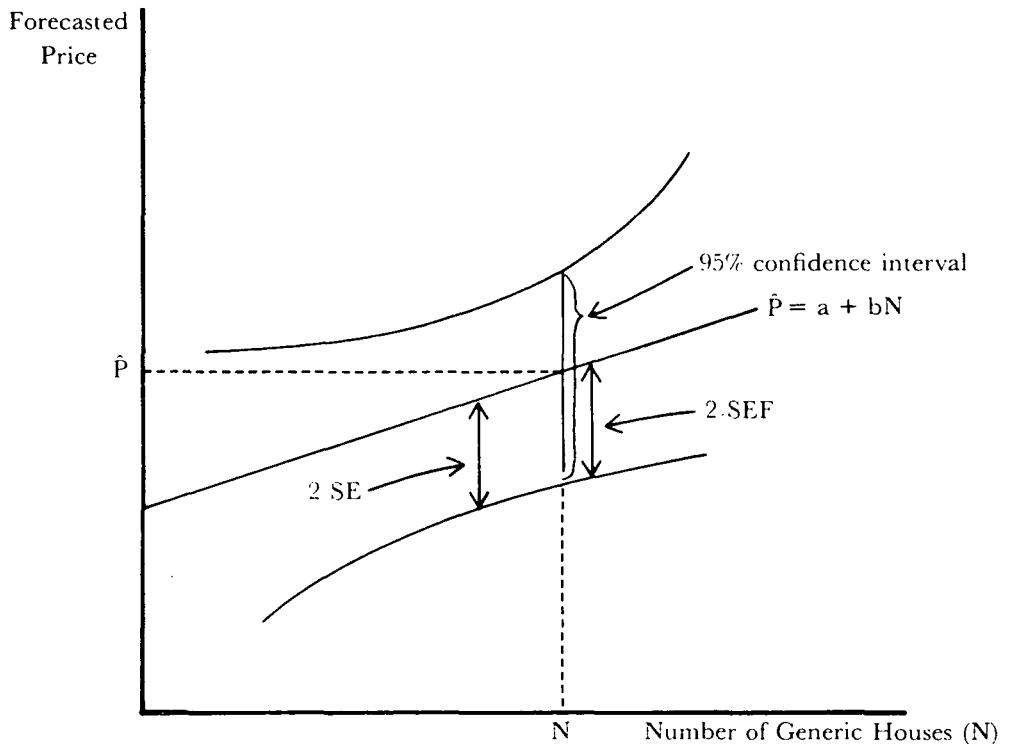
For many purposes the forecasted value may provide the only useful empirical information about liability or damage, but the reliability of the estimate can vary substantially. Thus, an interval of two standard errors of forecast about the predicted value provides a useful summary measure. There is a danger with the SEF . It is calculated on the assumption that the model includes the correct set of explanatory variables. If the choice of variables is wrong, the forecast error is likely to be larger. Thus, the SE reliability measure should if anything be viewed as providing an overly optimistic picture about the forecast.

The difference between SEF and SE can be seen in Figure 8. The standard error of the regression measures deviations within the sample period. However, the SEF is more general since it calculates deviations within or without the sample period. If the model is not correctly specified, deviations are likely to increase once one makes forecasts based on values for the explanatory variables which are far

44. Actually, the standard error of forecast measures the square root of the variance of the forecast error.

from those values available within the sample. The difference between SEF and SE increases with the difference between the forecasted predicted values of the explanatory variables and the predicted value calculated at the sample mean. The 95% confidence interval created by the measurement of two standard errors of forecast about the regression line is shown in Figure 8. In the Figure the distance $2SE$ is measured at the point representing the means \bar{N} and \bar{P} . SEF is slightly greater than SE in the diagram.

FIGURE 8



IV

DEALING WITH VIOLATION—THE USE OF MULTIPLE REGRESSION

The multiple regression technique just described has a number of potentially valuable uses in antitrust litigation. This section treats the question of whether any cognizable effect has occurred; the impact/damage issue will be discussed in section V. First, the ampicillin example discussed in section III will be used to show how hypothesis testing in a multiple regression framework can lead to useful tests for violation. Second, the plywood case will be used to illustrate how econometric forecasting tools can also be instructive.

A. Hypothesis Testing—The Ampicillin Case

The issue in the ampicillin case, it will be recalled, was whether the pattern of price decline in the CCS market would have occurred earlier and to a larger extent

had generic houses not been excluded by patent from entering the market. Such an issue is typical of many antitrust issues in that it poses a set of competing hypotheses to be evaluated. The plaintiffs' hypothesis was that the presence of generic houses in the market resulted in a lower price; the defendant's hypothesis was that the presence of generic houses had no effect on price. The two hypotheses are clearly mutually exclusive. The hypothesis of no effect is a clear, well-defined hypothesis, suggesting no violation and (of course) no impact or damages. Violation is obviously a matter of law, not economics or econometrics. For the purposes of this article, "no effect" will be equivalent to "no violation," while some effect is at least consistent with a cognizable violation. If there is no effect, there is obviously no impact or damages. If there is an effect, there may or may not be cognizable impact on any particular plaintiff, or class thereof. The hypothesis that there is some effect suggests violation, but it leaves the question of impact on a particular plaintiff or amount of damages to be determined. It is also a somewhat vague hypothesis. Does it mean 1¢ or 10¢ or \$1.00 when it speaks of "some effect?"

The choice between these hypotheses is one that can be handled in the context of classical statistics and econometrics. The null or initial hypothesis is set to be the well-defined hypothesis of no effect. This hypothesis is then tested against the alternative hypothesis that there is some effect. A null hypothesis of this type can be tested directly using a *t*-test within a multiple regression framework. If the null hypothesis is rejected at the appropriate level of significance, such as 1%, 5%, or 10%, the conclusion is that the data support the plaintiffs' hypothesis. However, a failure to reject the null hypothesis supports the defense. It will often be difficult to reject a null hypothesis, a result which seems reasonable only in cases in which the plaintiff bears the burden of proof.⁴⁵

The initial step in analyzing whether there is an effect in any particular case is to define the list of variables to be studied. It is then useful to classify the explanatory variables according to the underlying reasons why they might explain the dependent variable. One usual distinction, for example, is between variables that affect price through the supply relationship and those that affect price through the demand relationship. Once the variables have been listed and classified, verbal hypotheses can be translated into specific, statistically testable null hypotheses. Usually, the null hypothesis or null hypotheses will involve statements about individual variables or about classes of variables. For example, a null hypothesis might be: "the average cost of production has no effect on price." Or, the null hypothesis might be: "all supply-related variables have no effect on price." Once the null hypothesis has been specified, one chooses the appropriate form for the multiple regression and estimates the model using an appropriate data set. The estimated model is then used to perform the test of hypotheses described previously. The result should indicate whether or not there has been an effect.

In the ampicillin case, the first step is to use an underlying understanding of

45. What makes this point especially difficult is that the likelihood of rejection of a null hypothesis varies with the size of the sample being studied. A large sample makes rejection more likely than a small sample. As a result, when considering tests for effect, one must also consider whether such an effect is or is not substantial. This article returns to the problem in the section on impact and damages.

economics and of the institutions of the ampicillin market to model the factors expected to determine the price of ampicillin in the CCS market. Specifically, assume that the price of a bottle containing 100 250-milligram capsules of ampicillin is determined by four types of variables:

(1) *Cost Variables* (summarized as *COST*). Cost and price are expected to be positively correlated, that is, higher costs are expected to lead to higher prices in the market.

- a. C = a cost index based on the average cost of production of ampicillin [by Bristol-Myers].⁴⁶
- b. QD = a dummy variable to reflect the higher costs associated with selling ampicillin in low volume. It will be equal to one if the bid involves less than 100 bottles, zero otherwise. Note that the larger numerical value consists of *small* orders.⁴⁷

(2) *Measures of Competition* (*COMP*). This reflects the competition for sales by other firms in the market.

- a. $NMAN$ = number of manufacturing firms that bid.
- b. $NBRCUS$ = number of Bristol-Myers customers that bid. (These companies purchased bulk from Bristol and resold the final product to the CCS entities. None of these companies were generic houses.)
- c. $NBECUS$ = number of Beecham customers that bid.
- d. NW = number of wholesalers (firms that bought bulk ampicillin from nonmanufacturers).

(3) *Time Adjustment Variables* (*TIME*). Each of these time variables is a dummy variable which allows for differences in price which occur over the various years listed and is accounted for by the model.

- a. $T67T068$ = 1 in 1967 or 1968; 0 otherwise.
- b. $T69T070$ = 1 in 1969 or 1970; 0 otherwise.
- c. $T71T072$ = 1 in 1971 or 1972; 0 otherwise.
- d. $T73T077$ = 1 from 1973 to 1977; 0 otherwise.

These variables are introduced to account for cost differences that were not reflected in the crude cost index cited previously and for other structural changes in the ampicillin market that might have occurred.

(4) *Generic House Competition Variable*.

- a. N = number of generic houses that bid.

The model which explains the determination of the price of ampicillin can then be written as follows:

$$P = a + bN + cCOST + dCOMP + fTIME + e,$$

where e is a random disturbance term. Note the eclectic nature of this formulation. It is possible for any or all of the variables to play a role. A variable that does play a role will have a coefficient which is significantly different from zero. Variables that do not play a role are likely to have coefficients which are "insignificant," that is, coefficients that are reasonably close to zero.

The null hypothesis of no effect with respect to generic house competition is the

46. It is possible, of course, that had generics been allowed to manufacture ampicillin they would have found ways to do so at lower production cost. The analysis here assumes that the production costs of Bristol-Myers are less than or equal to those of any potential competitor.

47. A dummy variable is a variable which takes only two values, usually a zero and a one. A dummy variable in a multiple regression allows for a different intercept associated with observations in each of the two distinct zero-one categories.

hypothesis that $b = 0$, while the alternative hypothesis is that b is not equal to zero. Because the plaintiffs' hypothesis is that more firms bidding would cause price to fall, a more accurate formulation of the alternative hypothesis is that b is less than zero, i.e. negative.

To test the null hypothesis, data for nine states and eighty bidding situations were used to estimate the multiple regression model described above. The statistical results are presented in Table 1A.

At the start, let us examine the regression results to see whether they appear sensible. First, the cost index coefficient is positive and significant, indicating that cost changes are an important predictor of price changes. The fact that cost has the highest t -statistic suggests it may be the most important predictor of price.

Likewise, the coefficient of the quantity discount *QD* variable is positive and significant at the 5% level, indicating that bids on very small volume sales are generally higher than average.

Of the competition variables other than the number of generic houses at issue, *NMAN* and *NBRCUS* are negative, as expected; however, neither is statistically significant. (The insignificance here is partly due to the multicollinearity which

TABLE 1

Variable	Table 1A		Table 1B	
	Coefficient	t-Statistic	Coefficient	t-Statistic
N	.06	.67	.01	.12
C	3.45	14.03	-2.76	-2.67
C ²			.54	6.05
QD	1.46	3.66	1.00	3.14
NMAN	-.12	-1.12	-.07	-1.02
NBRCUS	-.27	-1.90	-.04	-.32
NBECUS	.23	.80	.02	.07
NW	.11	.84	.02	.20
T69T070	-.79	-1.94	-1.14	-3.33
T71T072	-.22	-.42	-1.59	-3.46
T73T077	-.12	-.31	-1.71	-3.53
CONSTANT	-9.73	-6.03	7.78	2.55
	R ² =.96	SE=.76	R ² =.97	SE=.64

arises because *NMAN* and *NBRCUS* are relatively highly correlated. If *NMAN* were dropped from the sample, for example, *NBRCUS* would become significant.) Thus, to the extent that competition determines price, the competition comes from manufacturers and from customers of the defendant in the case, Bristol-Myers.

Now consider the null hypothesis that generic houses had no effect on price. The coefficient of *N* is .06, suggesting that, other things being equal, an additional generic house bidding adds 6¢ to the average winning bid price. However, the 6¢ coefficient is highly insignificant, that is, not statistically different than zero (the t -

statistic is only .67).⁴⁸ In statistical terms, an appropriate way to look at the ampicillin results is to state that the null hypothesis that generic houses have no effect on price cannot be rejected. Moreover, given the positive sign of N_j , these data provide no support whatever for the view that generic houses might have a negative impact on price in the CCS market.

In addition to the signs and the significance of the coefficients of the regression results in Table 1A, it is also important and helpful to examine the magnitudes of the coefficients. In particular, one should ask whether the magnitudes make sense in light of the problem being studied. As an example, consider the coefficient on the cost variable, 3.45. Taken literally, this coefficient suggests that an additional \$1.00 of cost will lead to an increase in price of \$3.45. On its face this is clearly a surprising result, and it suggests that there may be a problem with the regression results.)⁴⁹

A sense of what the problem might be can be gleaned by examining the coefficients of each of the time dummy variables. The value -.79 on the dummy *T69T070* implies that from 1967-68 to 1969-70, price fell by 79¢ more on average than the change indicated by other variables. Price fell somewhat less in 1971-72 and 1973-77. A reexamination of the cost data from figure 3 is helpful here. Costs of production were generally falling throughout the period of analysis, and, in particular, costs were higher in 1967-68 than at any later period of time. Thus, the pattern of the coefficients on the time variables may well be picking up some of the effect of the decline in costs over time. If costs had been declining continuously at the same rate each year, the cost variable in the equation would have a negative coefficient reflecting this fact, and the time dummies would show no effect. Therefore, since a cost term is already included in the regression, the relationship between cost and price appears to be more complex, with the effect of cost on price varying with the level of cost, rather than being constant. Of course, one cannot be sure that cost is the only misspecified variable. It seems likely that other variables related to the product cycle of the drug are also important and omitted from the model. At this point, however, the product cycle effects alone do not seem to tell the whole story, since the product cycle effect would suggest a pattern of time dummy coefficients increasing, rather than decreasing, in magnitude as knowledge of the drug diffuses throughout the medical industry.

Failure to take account of the complex nature of the effect of cost on price provides a good explanation for the difficulty in interpreting the cost coefficient. Accordingly, the regression model has been reestimated with an additional variable, C^2 , the square of the cost variable, as shown in Table 1B. In this formula-

48. In fact, it should not be surprising to see a coefficient that is not exactly equal to zero. Nonzero coefficients for insignificant variables can be caused by many different reasons. The model may be slightly misspecified, variables may be measured with error, or the functional form may be slightly different than what is most appropriate. Expecting a coefficient exactly equal to zero is like expecting 50 heads and 50 tails to arise when a coin is tossed 100 times. The fact that 52 heads and 48 tails result in a given experiment does not obviously imply that the coin tosser was cheating or that the coin was unbalanced.

49. In a competitive market, price would be equal to marginal cost, so that one would expect at most a \$1.00 increase in price. Even if the market were not competitive, there would be no reason to believe that the cost increase would in itself allow firms to increase their market power and raise price more than \$1.00.

tion, the marginal effect of an increase in cost on price is not constant. Since the coefficient of C^2 is positive, the effect of cost on price increases at higher costs. Note also that the time dummies remain negative and significant, thus reflecting either a cost specification that is not yet satisfactory or the omission of other relevant variables from the equation. Nevertheless, the regression makes the interpretation of the effect of cost on price more reasonable. For example, in the 1967-68 period, when costs were relatively high, a \$1.00 increase in cost had an effect on price that was just under \$1.00. (Figure 5 shows that cost during this period was approximately 3.4 to 3.5.) Subsequently, the effect of cost on price decreased. Here, however, the pattern of time coefficients is consistent with the product cycle argument. The omitted relationship between prices in the CCS market and other related markets, including the acceptability of generic products over time, suggests prices declined at a greater rate than costs.

Finally, notice that the equation used in Table 1B alters the magnitudes of the other coefficients in the regression model. It is most interesting to note that the coefficient on the generic house variable is now .01 and once again insignificant. In terms of the question of violation, the conclusion is unchanged and, if anything, strengthened.

The question of why generic houses would have no impact will be considered with the question of impact and damages in section V. For the moment, however, it is worth reflecting a bit on the limitations of such an hypothesis test. It was stated that the hypothesis of no violation could not be rejected. In statistics, as in life, little is certain. The failure to reject the null hypothesis at the 5% level of significance leaves open the possibility that any of a number of mistakes were made.

The implications of such a failure to reject are complex but important, if one is to fully appreciate the usefulness of multiple regression analysis for hypothesis testing. Suppose (contrary to fact here) that the null hypothesis was rejected, and the alternative hypothesis, that the number of generic houses affects price, was accepted. The statistical test used ($t \geq 2.0$) implies that one would have incorrectly decided there was an effect, when in fact there was no effect, about 5% of the time.⁵⁰

However, the statistical test tells much less about the issue at hand here since the null hypothesis has not been rejected. How often would the null hypothesis not be rejected when an effect had in fact occurred? The answer is harder to spell out because the alternative hypothesis has been specified in such vague terms. The answer would be different depending on whether the alternative hypothesis stated that an additional generic house would lower the price by 50¢ or more, or that an additional generic house would have \$1 or greater impact.

To begin with, it is certainly possible that if a violation did occur and if price had actually been trivially lower, for example 1¢, the sample might have shown a

50. One might argue that an alternative statistical test would have been appropriate here since one could not seriously consider the alternative of a positive effect of the number of generic houses on price. A "one-tailed" test would test the hypothesis of a zero coefficient against the alternative that the coefficient is negative.

coefficient of $+.06$. However, at greater effects involving a lowering of price by 5¢, 10¢, 20¢, or 40¢, the likelihood of obtaining a coefficient of $.06$ falls. Some specific calculations can show how the probability of mistakenly failing to find an effect depends upon one's specific notion of what such an effect might involve. How often might a coefficient of $+.06$ or larger occur if the "true" effect were negative? It depends on how negative the effect is, as Table 2 shows. The larger the true negative effect, the less likely that one would get a $.06$ result. Thus, while the small positive coefficient shows that a negative price effect is not likely, it cannot rule out the possibility of a small effect. The $.06$ coefficient does, however, make even a 15¢ negative effect extremely improbable.

Where does all this lead? In a suit in which the burden of proof is on the plaintiff, rejection of the null hypothesis (which embodies the defense view of no effect) provides strong support for the plaintiff's case. Even then, a final conclusion about the case is still not certain because rejection of any null

TABLE 2

Price effect of additional generic house	Probability of getting coefficient of N of $+.06$ or more
.00	.25
-.05	.12
-.10	.04
-.15	.01
-.20	.00

hypothesis depends on the level of significance chosen (a somewhat arbitrary number), as well as the accuracy of the regression model, which in turn is a function of the amount of data available.

Failure to reject the null hypothesis provides support for the defense. However, failure to reject the null hypothesis does not necessarily rule out the possibility that plaintiff's case is correct. Nothing is ever certain, but some things are highly improbable. Statistical analysis can indicate the probabilities.

To summarize:

- (1) It is easier to show that an alleged effect is not proved than to show that it is proved.
- (2) Any such proof of effect requires a *specific* hypothesis about how much effect is expected.
- (3) Given that the burden of proof is generally on the plaintiff, plaintiff's specific hypothesis must prevail over the null hypothesis for plaintiff's case to be valid. Of course, one could argue for a burden of proof on the defendant since one might begin with the presumption that more competition will lead to a decline in price.

Of course, it is difficult to suggest specific hypotheses without reference to individual cases. One sensible approach might be to select as the magnitude of violation what is deemed to be a *de minimis* impact. Anything that shows such an effect

or a more substantial one will support plaintiff's case. In the ampicillin example, one might decide that a 5% impact, roughly 25¢, is *de minimis*.⁵¹ Then plaintiff would be required to show such impact in order to prove his case. If a 5% impact were the assumption, one could conclude quite definitively that there was no violation in the ampicillin case. Another approach would be to calculate that one could be quite sure (with a probability of 95%) that the conclusion of no violation was correct if one were to define the violation to be an impact of $\$X$ or greater. X turns out to be 11¢ in the ampicillin equation.

Finally, a note of warning to the reader: all of the analyses described turn on the equations being correctly specified. Any judgment reached on the basis of quantitative information of this kind will be dependent upon the adequacy of the model being utilized. Thus, favorable results on their own do not relieve the defendant and its experts from explaining why the model and the results that follow from the model make sense.

B. The Plywood Case

In some antitrust cases, it is difficult to specify hypotheses in a form that can be readily tested in the manner just described. In some of these situations, the use of econometric forecasting techniques provides a valuable alternative to the hypothesis testing approach. On both the issue of existence of effect and the question of damages, the econometric forecasting technique can be used, but it can be abused as well. This fact can be illustrated using as an example the plywood antitrust litigation.⁵²

Before the forecasting method is described, it is essential to develop some background concerning the plywood case. Parts of this case are also interesting as another example of situations (as in section II) when the direct application of microeconomic theory ought to have been sufficient in evaluating some of the plaintiff's theories.

1. *Background.* The cases involving plywood pricing that have come to be known either as *Boise Cascade* or as *Plywood Antitrust Litigation* offer some interesting (and disturbing) insights into the use of economic and econometric analysis in antitrust litigation. The central question was whether use of a uniform and superficially artificial method of quoting prices for southern pine plywood violated the law. There were, in fact, two different cases: the earlier FTC case⁵³ in which the major plywood manufacturers were charged with violation of the Federal Trade Commission Act § 5,⁵⁴ and the private class action cases⁵⁵ where the complaint

51. Of course a determination of what is *de minimis* must depend not only on the price effect but also on the impact in terms of dollars of damages. Therefore, the question of how to decide what is a substantial impact remains unanswered.

52. *In re Plywood Antitrust Litigation*, 655 F.2d 627 (5th Cir. 1981), *cert. granted*, 456 U.S. 971 (1982).

53. *Boise Cascade Corp. v. FTC*, 637 F.2d 573 (9th Cir. 1980) (reversing the order of the FTC given in *Boise Cascade Corp.*, 91 F.T.C. 1 (1978)).

54. 15 U.S.C. § 45 (1982).

55. See *In re Plywood Antitrust Litigation*, 655 F.2d 627 (5th Cir. 1981), *cert. granted*, 456 U.S. 971 (1982).

concerned Section 1 of the Sherman Act, which is the immediate concern of this article. Precisely the same practices were at issue in both cases, and in the end precisely the same empirical questions came to be dominant. This discussion first illustrates how the use of economic theory should have helped to reject some of the hypotheses in the case. The plywood case is then used to show the use of econometric techniques dealing with the additional hypotheses of the plaintiff.

The undisputed facts were as follows:

1. Before 1964 softwood plywood was manufactured primarily from Douglas fir, found only on the west coast, and particularly in the Pacific Northwest. Since plywood is a heavy commodity and is characteristically shipped long distances by rail, transportation costs represent a significant part of the total cost of delivered plywood. Freight rates for plywood are zone rates, and initially all producing plants were in the same west coast originating freight district. Thus, quoting freight rates to buyers anywhere in the United States (outside of the west coast itself) involved what was known as Portland (or "west coast") freight. West coast freight meant the rail freight from Portland, Oregon to the buyer's freight zone, with one minor modification. Prices quoted to buyers were delivered prices stated in the form of mill prices to which would be added the appropriate amount of west coast freight. The modification was that all sellers and buyers had come to use a standard set of (assumed) weights, known as "association weights" for plywood of specified thickness.⁵⁶ These standardized weights, rather than actual weights, were the bases for buyers' freight payments. Because plywood of a given ply, thickness, and surface quality is sold by (and valued for) its area, differences in weight do not affect its value to buyers. However, plywood does vary in weight due to differences in moisture content. Since freight charges vary according to actual weight, the effect of using standardized weights is to give manufacturers an incentive to control the moisture content (which they can do by the nature of the drying operations engaged in) and also to give buyers a firm delivered price quotation independent of the actual weight of the particular plywood they will ultimately receive.

2. Competition among west coast plywood manufacturers led to weekly, daily, even hourly fluctuations of quoted prices for plywood. The quoted prices were highly variable over time, but the west coast freight changed infrequently—only when the ICC approved new freight rates.

3. Although all manufacturers negotiated and quoted prices in the same manner, there was no uniformity in the "west coast mill price" quoted by plywood plants to a given buyer at a given time. Typically, buyers phoned several sellers and solicited bids; often they negotiated for yet lower bids. The freight component to any buyer was, however, constant, since the buyer's location alone determined the freight rate, and the weight was always the uniformly used association weights.

It is important to note for what followed that the FTC charged no violation of

56. "Association weights" appear to have had their origin in the National Recovery Act, and they were used by the government in World War II price controls as well. By the 1960's, the use of association weights was habitual. Of course, this avoids the question of why association rates were utilized in the first place.

the law based on these three facts. In essence, the unconstrained variability of mill prices led to an unconstrained variability of delivered prices. Two price reporting services (*Crows* and *Random Lengths*) published weekly or semiweekly listings of prices of standard types of plywood. These services, like the market participants, quoted a west coast mill price. A buyer knowing his own freight district and the association weights could translate this nominal "price" into a delivered price. If, for example, his west coast freight for a bundle of 3/8 inch plywood was \$30, a quote of \$85 meant a delivered price of \$115, and a quote of \$82 meant a delivered price of \$112.

One feature of this arrangement was to occasion concern in the litigation. Over time, *actual* weights for fir plywood had decreased (due to improved techniques for drying plywood), but the use of original association weights continued unabated. On average this led to "underweights"—since the actual freight incurred and paid by sellers was (on average) a few percent below the freight charged to buyers. The amounts of these expected underweights were known to all sellers and buyers and were a source of net mill returns to sellers.

4. In 1963, engineers for Georgia Pacific, the leading west coast producer of plywood, perfected a technology for manufacturing softwood plywood from southern pine; the plywood so produced was to prove functionally identical to western fir plywood. This new technology had two dramatic effects on the supply of plywood. First, it made vast stands of softwood timber in the South—particularly in Louisiana, Arkansas, Georgia, and North Carolina—available for plywood production. Second, it made it possible to produce plywood much closer to key consumption areas of the South, Southeast, Midwest, and East Coast and thus to reduce shipping distances. Given high freight costs, this locational advantage made production in the south potentially profitable, even though, at the outset, mill costs were substantially higher than on the west coast.⁵⁷ The first southern plant in Fordyce, Arkansas came into production in 1964, and its success triggered the growth of the southern plywood industry.

5. The growth of the southern industry, both in number of plants and in production, is shown in Table 3. This expansion was the result of both entry of new independent producers and the opening of southern plants by the established western producers, Georgia Pacific, Boise-Cascade, Weyerhaeuser, and others.

6. Southern plywood, naturally enough, found its initial market in the areas surrounding its own plants where its locational advantage was greatest. While some southern plywood was shipped by rail, most was delivered by truck, since relatively short distances were involved.

7. Southern plywood was priced originally to compete with the established west coast plywood. In fact, for the first couple of years, it was priced several dollars below its west coast competition to persuade buyers to try the new product in preference to the known one.⁵⁸ However, after a relatively short period, buyers

57. Southern pine trees were smaller in diameter than Douglas fir trees. Their moisture content also necessitated higher gluing and drying costs.

58. For a short period there were fears that the southern plywood would literally come unglued.

TABLE 3
SOUTHERN PLYWOOD PRODUCTION AND PLANTS

Year	Plants	Production ^a
1963	0	0
1964	3	80.0
1965	12	501.7
1966	23	1239.5
1967	28	1879.1
1968	33	2472.4
1969	34	3525.4
1970	40	4610.0
1971	49	5705.1
1972	52	6613.8
1973	53	6853.6
1974	55	6425.2

^a Production is measured in millions of board-feet 3/8" basis.

came to accept southern plywood as a perfect substitute for western softwood plywood and shopped between plywood sources purely on the basis of delivered price.

8. Southern producers quoted prices to buyers in a form that made them directly comparable to west coast prices. This involved quoting an adjusted west coast price in the form of an index price to which one would add west coast freight to calculate the delivered price.

9. As long as southern and western plywood were sold in direct competition with one another, use of the index method of pricing was both convenient to buyers and wholly understandable. As the production costs of southern plywood were brought down toward the levels of western plywood, the freight advantage of southern mills gave them an opportunity to profitably undersell western mills in large sections of the country. Suppose a western mill was selling in New Orleans for \$81 + \$19 freight = \$100 delivered (for a mill realization of perhaps \$83, given \$2 of underweights). A Louisiana mill with only \$5 freight to New Orleans and production costs of even \$85 could very profitably sell at a delivered price of less than \$98, the west coast zero profit price. It could, for example, sell at \$95 by quoting an index of \$76 and make a \$5 profit. Such extraordinary profit opportunities would be expected to lead to opening of more and more southern plants, as indeed happened—thirty new mills opened between 1965 and 1968 and another twenty-two between 1968 and 1974. As this expansion occurred, western plywood was gradually squeezed out of an ever larger area. For example, the key competition for areas such as New Orleans, Miami, and Atlanta came to be among southern bidders. Notwithstanding this development, the west coast index method of price quotations was retained even in areas where southern buyers looked exclusively to southern sellers.⁵⁹

59. There was one mill in Moncure, North Carolina, that did actually quote F.O.B. mill prices instead

2. *Underlying Issues.* Before considering the major issue in this case, it seems worthwhile to consider a less important point: whether uniform use by sellers of upwardly biased association weights led to an overcharge, an element of “phantom” freight charged but not incurred. It is easy to show that in an otherwise competitive market this would not occur. Assume, *arguendo*, that there was an illegal agreement to use association weights. Economic theory establishes that such an agreement could not affect delivered price if there was truly competitive bidding. The reason is that any competitive mill seeking a certain markup over its out-of-pocket costs in determining its bid would consider both its direct mill excess of revenues over cost and its phantom freight from underweights. Competition among sellers would squeeze the total mill return to the competitive level. Whatever that level might be, it could be achieved by any combination of direct mill return and phantom freight due to underweights. If, for example, all sellers agreed to adjust downward association weights with the effect of lowering charged freight by \$2 on a given shipment (but leaving actual freight unchanged), they would have to raise quoted mill price by \$2 to achieve the same delivered price and the same total mill return. Thus, underweights allow competitive firms to lower mill prices.

Precisely the same issue concerns “phantom freight” existing due to differences between actual freight and west coast freight. However, because the index price quotation practice is the very heart of this litigation, some elaboration will prove useful. Consider, for example, the viewpoint of a buyer of plywood in New Orleans who is faced with price quotes from west coast mills of \$83 and \$81 to which it would add west coast freight of, say, \$19—for delivered prices of \$102 and \$100. Imagine that the buyer is seeking a telephone price quotation from each of two southern plywood mills, one in Fordyce, Arkansas, and the other in Ruston, Louisiana. One way it might do so would be to ask for *delivered* price quotes from the southern mills which it could then compare with the implicit prices of \$102 and \$100 in the west coast quotes. A second way would be to ask for the *mill* price at each of the southern mills. To compare these to each other, and to the west coast mills, the buyer would need to determine the point-to-point actual freight between the particular southern mill and New Orleans. A third way would be to receive west coast-equivalent quotations. These would be directly comparable to the west coast bids of \$83 and \$81. For example, if Fordyce quoted \$80 and Ruston \$78 as west coast-equivalent prices, these imply delivered prices of \$99 and \$97. The lowest index price is the lowest delivered price.

Presumably in a competitive market sellers would be willing to quote by whichever of these methods buyers asked for. Either the first or the third method permits accurate instantaneous comparison among all the alternative sources of supply. The second does not. In practice the third method—the west coast, or

of index prices. Its F.O.B. mill prices were of course higher than the index prices of nearby mills, but its delivered prices were comparable.

"index" basis—was adopted from the start and was more or less ubiquitous until this litigation.⁶⁰

The apparent anomaly of southern mills quoting prices to southern buyers in terms of western prices plus west coast freight triggered the FTC's lawsuit. Surely if the plywood industry had started in the South, it would have been unnatural for producers competing among themselves to quote prices in terms of a distant point. (Would Detroit car manufacturers quote prices in California with freight from Yokasuka? Conceivably they could have if the Japanese industry had been there first.) The survival of the index method of quoting plywood prices is capable of competing interpretations:

- a) It was an historical conventional practice that was a convenient and harmless form of quoting delivered prices even to those buyers for whom west coast plywood was no longer a realistic alternative source of supply; it continued for three reasons: (i) inertia; (ii) because it was found to be the most convenient form of price quotation for those buyers at the boundary of the predominant southern zone (for example, in Chicago) for which both west coast and southern plywood continued to be viable alternatives;⁶¹ and (iii) it permitted buyers and sellers to have a ready market reference point. If *Crows* reported that last week's closing (index) prices were in the range of \$80, buyers in any location could regard a \$90 (index) quote as reflecting a sharp jump in market prices, and they would have the opportunity to decide whether this increase in prices reflected a sharp change in the overall supply situation or was simply a high quote by a seller not much interested in making the sale.
- b) It was a device by which southern sellers, although competing *inter se*, sought to preserve part of their freight advantage by restricting competition so as not to erode phantom freight. In modern oligopolistic terminology, the use of this price quotation device may have been a form of quasi-agreement intended to limit competition to certain dimensions and thus lead to less price competition than would occur with another form of price quotation.

3. *Plaintiffs' Theories.* With this background about the plywood case, it is now possible to consider in a more serious way several of plaintiff's hypotheses and theories concerning the plywood industry. This analysis must begin with the basing point theory of the FTC and then proceed to a more complex theory of violation which underlies the private case.

a. *The basing point theory of the FTC.* The initial FTC complaint charged the plywood industry with using the index system of pricing as a means of administering a classical single basing point scheme to fix prices. Under a single basing point scheme, buyers will be charged a single price at a base point, usually a mill, plus identical transportation charges (regardless of the actual point of the shipment or of the freight costs) to the buyer's plant. As a result a buyer faces the same delivered price (mill price plus transportation) for the good he wishes to purchase

60. Of course, after a while the typical southern index price was different from the western index price. The reporting services, by 1967, were reporting them separately.

61. As long as both western and southern plywood have natural areas of comparative advantage, there is always some boundary where the delivered cost of western and southern plywood will be equal.

independent of the location of the source of product he buys. The consequence is that price competition among sellers is effectively nonexistent.⁶²

Superficially, the use of a west coast price plus rail freight from the west coast looks to be of the single basing point form. It was practiced by sellers with mills far removed from the west coast who were characteristically shipping (often by truck) to buyers in Pennsylvania, Ohio, and Florida, at actual freight cost well below the "west coast freight" that is part of the delivered price.

Use of basing points to eliminate price competition rests on the ability to assure that at any time the price quoted to any buyer is determined solely by the buyer's location and is therefore identical among all bidding sellers. To better understand what is and is not a true basing system, compare the following two pricing schemes each *arguendo* adopted by *all* sellers:

- | | | |
|----------------------------------|--|--|
| I. Price delivered to buyer x | = Fixed mill price from a specified mill | + Freight from specified mill to buyer x (fixed freight amount) |
| II. Price delivered to buyer x | = Variable price quotation | + Freight from specified mill to buyer x (fixed freight amount). |

In I, *both* elements of the delivered price are fixed—every potential seller following this formula will arrive at an identical bid; in II, one element is fixed, and the other element is variable. Although method I, if followed by all sellers, constitutes uniform price fixing, it is apparent that method II need not provide a disincentive to competitive pricing. If the variable element of price is truly unbounded, there is no necessary limit placed on the delivered price by virtue of fixing the "freight" element. Thus while I is plainly price fixing, II is not.⁶³

This is not to say that method II could not have an effect on price. For example, it might be that the fixed freight element places an effective minimum on the delivered price (that is, if "west coast freight" to some location is \$30, and no one ever quotes index prices below zero, this places a \$30 floor on delivered prices). If the west coast freight element were large enough, this might conceivably constitute an effective limit on price cutting. Moreover, because quoted mill prices are easy to verify, limitation of competition to one easily verified dimension of pricing may make secret discounting harder and thus eliminate one source of oligopolistic competition.

Although the use of the west coast index method of price quotation was not "just like a basing point," it was theoretically possible that it led to stable, uniform prices of the kind that are characteristic of basing point schemes. There is little doubt that the FTC thought it had found a simple basing point case. This possibility was quickly dispelled, however, by direct testimony and by very simple statistical compilations. Individual buyers testified to the variety of prices offered and to their practice of never buying without calling three to five sellers. Sellers

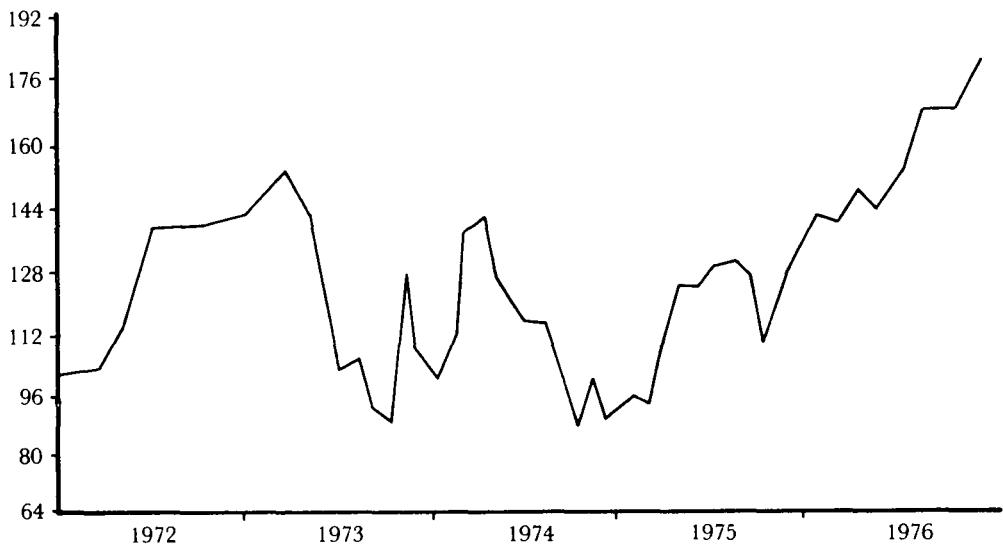
62. For a further discussion of basing point systems, see *FTC v. Cement Institute*, 333 U.S. 683 (1948); *Triangle Conduit & Cable Co. v. FTC*, 168 F.2d 175 (7th Cir. 1948), *aff'd by an equally divided Court sub nom Clayton Mark & Co. v. FTC*, 336 U.S. 956 (1949).

63. If this is not clear, consider whether the standard use of, for example, a 10¢ deposit on bottles in any way limits the competitive prices charged for soft drinks.

described the organization of their selling rooms—dozens of salesmen manning a battery of phones, surrounded by blackboards which listed available supplies of plywood of different sizes and grades at different mills. Changes on the blackboards led them to vary price quotations, often hour by hour, and certainly day by day.

The possibility that the overall effect was to lead to substantial price stability over time (of the sort characteristically found in basing points) was dispelled in part by superficial examination of reported monthly prices (Figure 9). In 1973, for example, prices varied from \$93 to \$155 and seemed to exhibit sharp and virtually continuous seasonal and cyclical fluctuations, despite the use of the index method of price quotation.

FIGURE 9
INDEX PRICE OF 1/2" STANDARD EXTERIOR PLYWOOD
(SOURCE: RANDOM LENGTHS)



Price uniformity and price stability are key indicia of price fixing via basing points, and those indicia were not present in pricing of softwood plywood, either west coast or southern. By the time of trial before an FTC administrative law judge, the basing point theory of violation had been quietly dropped.

b. More sophisticated theories of violation. The possibility existed that while plywood prices were nonuniform and highly volatile they nevertheless exhibited a different pattern of fluctuations and fluctuated around a higher average level as a result of the price method employed. The FTC did not so claim; rather it rested upon the assertion that the artificial method of price quotation constituted an unfair method of competition, whether or not it affected prices. The Ninth Circuit

disagreed and refused to enforce the FTC order since it had failed to show either agreement or effect.⁶⁴

In the civil suit, private plaintiffs necessarily asserted an effect on the level of price. Prices, they said, while volatile, were artificially high at all times, albeit responsive to market forces. Plaintiffs argued that the use of west coast index pricing led to overcharges due to the presence of both underweights and phantom freight.⁶⁵ They claimed that whatever the competitive price path for a buyer at a given location had been, the effect of this method of pricing would cause the delivered price paid to include phantom freight (or equivalently the mill realization would equal the quoted mill price plus phantom freight).

Although the plaintiffs' assertion about the competitive price path proved to have sufficient forensic appeal to be accepted by a jury and a district judge,⁶⁶ it is, nevertheless, theoretically unsatisfying because it fails to indicate how such a result could come about and, if it did, how it could be maintained. Suppose phantom freight is \$6 from a seller's mill to some location—that is, its actual freight is \$6 less than the west coast freight to that location. *This is purely a matter of the structure of freight rates.* Plaintiffs' theory was that any index bid by a seller made to a buyer at that location—whether of \$130 or of \$60—was necessarily inflated by the same \$6. If a seller bids, for example, \$79 in an attempt to outbid a price quote of \$80, it is argued that the seller collects \$6 more than its minimum acceptable bid. If the seller then discovers that a rival quotes \$78, and it cuts to \$77, it is still allegedly collecting \$6 more than its minimum acceptable bid. Even if the seller reduced its bid to zero, it would still be collecting phantom freight of \$6.⁶⁷

Consider a cartel that agrees to charge "the competitive price + \$6." Unless one can independently define the level of a competitive bid, it is impossible to police an agreement to charge "the competitive price + \$6" or to prove or refute the allegation that the actual price is supracompetitive by the alleged fixed mark up of \$6.

If this theoretical argument is not persuasive, can statistical analysis help? Analysis of data relating to profits can be used to attempt to evaluate the assertion of supracompetitive pricing, for consistent overcharges by southern producers would assuredly lead to persistent excess profits. In the actual litigation, neither party felt obliged to undertake such an analysis. The plaintiffs believed the jury would be less troubled than an economic theorist by the logical difficulties of their theory; the defendants believed the logic would prove persuasive, if not to a jury, at least on appeal. The defendants appear to have been wrong.

It is also of some note that plaintiffs were not required to undertake the empirical study that alone could provide scientifically persuasive evidence to support it.

64. *Boise Cascade Corp. v. FTC*, 637 F.2d 573 (9th Cir. 1980).

65. Southern plywood, being heavier than western plywood, did not exhibit the clear pattern of underweights shown by using association weights for fir plywood. Plaintiff's damage theory was based on underweights for west coast shipments and the difference between west coast and southern freight rates for southern plywood.

66. The jury decision was upheld on appeal and certiorari was granted by the U.S. Supreme Court. The case, however, was settled before the Supreme Court could hear it. *In re Plywood Antitrust Litigation*, 655 F.2d 627 (5th Cir. 1981), *cert. granted*, 456 U.S. 971 (1982).

67. The beauty of this theory for computing damages is that it makes overcharges purely a function of quantity purchased and location, not of price charged.

The judicial process allowed an unsupported assertion to replace a weighing of competing possibilities. Of course it is not just the plaintiffs who were to blame in this process. The defendants, who had access to much of the relevant data, chose not to present all of it. This could be indicative of the nature of the defendants' case, although it more likely has to do with the competing views about the appropriate strategies involved in the defense of such an antitrust case. Once again, it seems readily apparent that a clearer presentation of the theory in the case and an appropriate empirical test might have led to a different outcome.

C. *Testing for Violation—The Use of Forecasting*

One of the primary issues in the plywood case was whether the growth and development of the southern plywood industry, in coordination with an allegedly conspiratorial method of quoting delivered prices, served to raise plywood prices in the South higher than they would have been had the market been competitive. One approach to settling this issue uses statistical analysis to explain the movement of plywood prices during a period in which the conspiracy was alleged to be in effect. This model might then be used to predict what hypothetical conspiratorial prices would have been in a period in which no conspiracy was alleged and therefore when prices were agreed by all parties to be competitive. A comparison of the predicted conspiratorial prices and the actual competitive prices can provide a test for violation as well as a measure of damages if violation has occurred. If predicted prices are substantially higher during the nonconspiratorial period than actual prices after controlling for relevant variables, the analysis provides support for the theory that the conspiracy raised prices. However, if there is little or no systematic excess of predicted prices over actual prices, the defense position is supported.⁶⁸

This technique is illustrated by applying it in the plywood case. The price of ½" standard exterior southern pine plywood is the dependent variable, and the explanatory variables are those that economic theory suggests are related to demand (housing starts in the region, income) or supply (stumpage price, forest sales, a production capacity measure, log costs, drying costs, and glue costs).⁶⁹ In terms of the plywood case, there is no reason to believe that a cartel (if it existed) could have controlled the explanatory variables that have been listed. Housing starts and income are clearly outside the control of such a cartel, and no suggestion has been made that input prices and costs of production were controlled by the cartel. In such a situation, it is desirable to see if demand and supply variables can

68. There is of course the question of how many years such an experiment must be carried out. A reasonable period is necessary since historical factors may serve to maintain higher price levels for a while.

69. In specifying any model, one must consider whether the question of simultaneity arises. In a simultaneous model one would account for the fact that price and quantity are jointly determined, and neither can be said to have an independent effect on the other variable. The equation used here was a "reduced form" equation, rather than a "structural equation," in that quantity did not appear on the right-hand side as an explanatory variable. For most purposes, reduced form equations are preferable for forecasting purposes, although strictly speaking the estimates obtained for the parameters may not be the most efficient (accurate). Structural equations can be used if properly estimated, but the use of simultaneous equation techniques complicates the analysis and will be avoided here. For further details on this issue, see Fisher, *supra* note 29, at 726.

account for most of the movement in the southern plywood price. If so, the results will at least suggest that a cartel is not needed to account for the actual pattern of price. This argument should be contrasted to the uranium case to be discussed in section V. In that case, the cartel was able to affect some of the determining supply and demand variables. As a result, the analysis becomes substantially more complex.

TABLE 4
PLYWOOD REGRESSION RESULTS

Dependent Variable = SPLYCROW (Mean=121)
Time Period = May 1964 - February 1977

Variable ^a	Coefficient	t-Statistic	Mean
CONSTANT	-85.10	-2.80	
NEHST	.40	1.24	19.7
NCHST	.19	.93	29.9
SHST	.37	1.91	53.0
WHST	-.20	-.67	30.1
GNP	.09	1.95	1018.3
FIRSTUMP	.46	6.07	97.5
FORSALES	-.01	-2.84	1817.4
SOPINEL	-.22	-.51	69.3
SLOGNET	-6.52	-3.69	31.1
SDRYING	40.78	3.95	6.8
SGLUE	-6.01	-2.40	7.1
$R^2 = .84$		SE=13.16	

- ^aSPLYCROW = index price of southern plywood (1/2" standard)
 NEHST = housing starts in northeast region (thousands)
 NCHST = housing starts in northcentral region (thousands)
 SHST = housing starts in southern region (thousands)
 WHST = housing starts in western region (thousands)
 GNP = Gross National Product (billions)
 FIRSTUMP = Douglas fir stumpage price (\$/million board feet)
 FORSALES = Volume of public timber sold in Northwest (million board feet)
 SOPINEL = Southern pine stumpage price (\$/million board feet)
 SLOGNET = measure of cost of southern pine logs at Georgia-Pacific plant
 SDRYING = resume of drying cost in the South (Georgia Pacific plants)
 SGLUE = measure of glue cost in the South (Georgia Pacific plants)

The regression results obtained for the plywood case using least-squares are presented in detail in Table 4.⁷⁰ The econometric results suggests that the model fits quite well, with an R^2 of slightly higher than .80.⁷¹ In addition, many of the individual variables are statistically different from zero at the 5% level of significance.⁷² A useful way to examine goodness-of-fit is to graph the relationship

70. The equation used here involved serially correlated disturbances, but corrections were not made. A more nearly correct technique yielded essentially similar results.

71. R^2 tends to be rather high with time-series data, so a value of .80 is not unusually high.

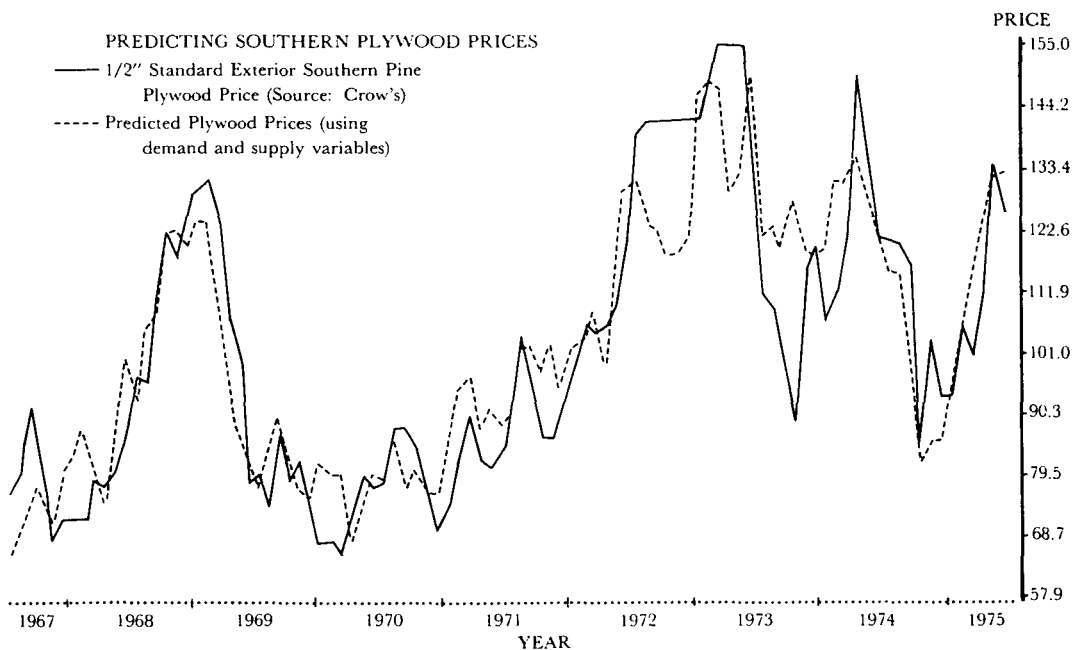
72. A number of the coefficients of the variables had the "wrong" signs, but such a phenomenon is

between the actual plywood prices and those prices that would be predicted from the multiple regression model. In Figure 10, the actual prices are represented by a solid line, and the predicted prices by a dotted line. With the exception of the period running from late 1972 to early 1973, the predicted prices move quite closely with the data. The ability to predict or "track" fluctuations in the data—turning points—is a good test of a regression model. Models that predict only growth paths for variables without explaining turning points are poor forecasting models.⁷³

Why does the model predict so poorly in the 1972-73 period? The answer is probably that during this period the government controlled prices, thereby restricting the ability of the market to equate demand and supply. The fact that the supply/demand model works badly in a period when supply and demand were not free to work is encouraging.

What may be surprising about the results is that they show predicted prices below actual prices during the period of price control. The analysis suggests that the government erred in setting prices by setting the price of southern plywood too high.⁷⁴ This error may have occurred in part because of the price controllers' con-

FIGURE 10



typical when there is substantial multicollinearity. For example, all cost variables in the model were positively related to price, but it was impossible to accurately separate out individual effects.

73. There are, of course, data limitations in predicting turning points. For example, if quarterly data were used, it would be impossible to predict fluctuations or turning points for prices on a weekly basis.

74. Price controls provided only for a ceiling, not a floor. It appears that the control price provided a focus of tacit agreement for companies to raise prices above the competitive level, and buyers were lulled into thinking they should pay these prices.

fusion about the difference between western and southern freight.⁷⁵

Can the results of the regression model be used to test for an antitrust violation? Not directly. The problem that remains is that the regression model can be used to explain the price behavior of a cartel just as well as that of a competitive firm or firms. It does not take much economic theory to realize that a cartel of price-setting firms would find it appropriate to vary price, depending upon conditions of supply as well as demand. For example, an increase in demand for plywood caused by an increase in housing starts will allow the cartel members to raise price, just as it would lead to higher prices through competitive market forces. Of course, the cartel's price will be higher than the competitive price, but that is a question involving the level of price, not price variation.

It is critical to realize that regression models are not particularly well suited to explaining levels of variables; rather, they explain variation about the mean. Imagine, for example, two price series to be explained, both having the same pattern of variation, but the monopolist's price always being 50% higher than the other. A regression model would explain both price variables equally well, with the same R^2 and same prediction of turning points. The only difference would be reflected in the intercepts, due to the level differences. Regression lines always fit the data on average—the intercept term is calculated in such a way to assure that. Thus, the fact that the average of predicted prices equals the average of the true prices does not prove anything.

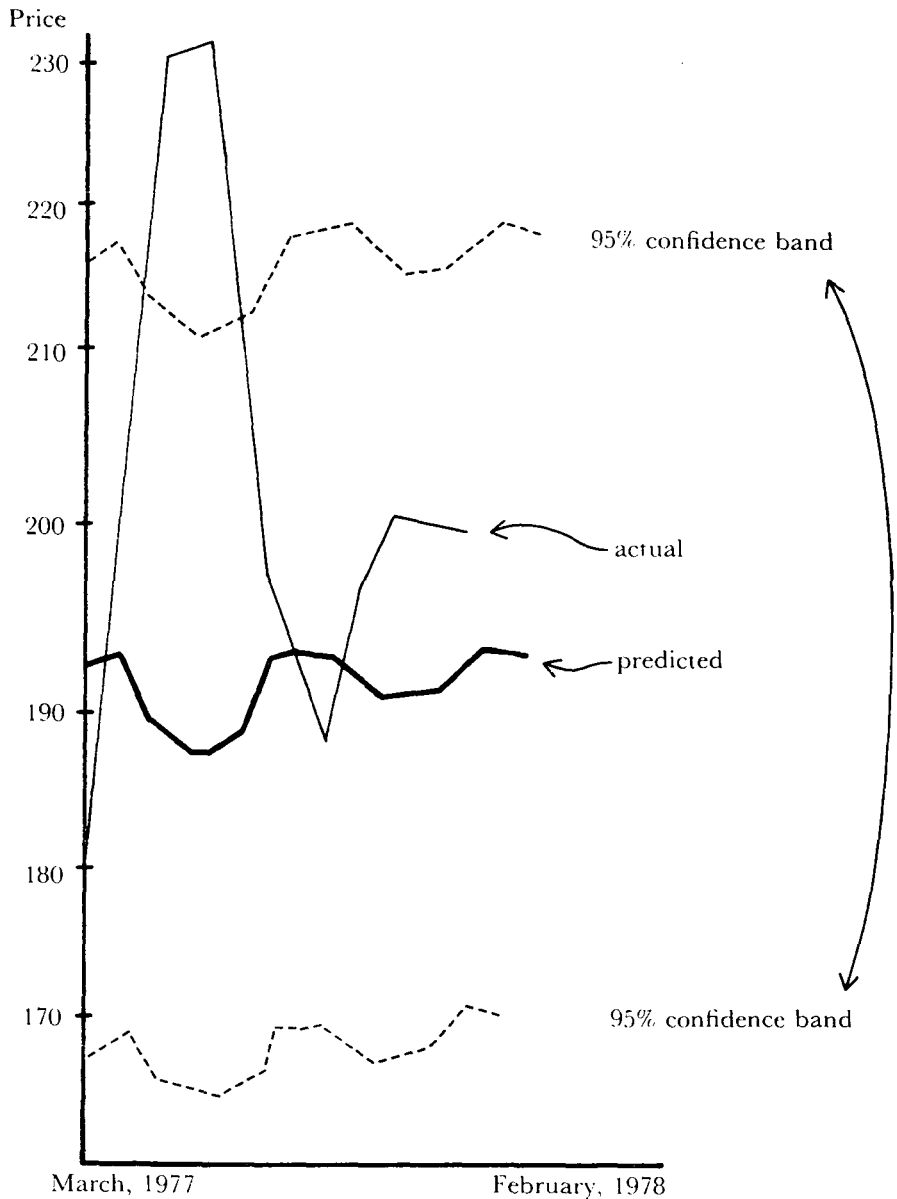
The regression model can overcome this limitation if the forecasting technique described previously is used. The southern plywood model was estimated through the month of March 1977. March 1977 represents the time when the "price conspiracy," if it occurred, would in any event have broken up because at that time the defendant plywood companies altered their form of pricing to eliminate the alleged west coast freight add-on. The estimated coefficients were then used to forecast what prices would have been from March 1977 through early 1978 had the pricing behavior prior to March 1977 remained in effect. The results of comparing the forecasted prices to actual prices are shown in Figure 11.⁷⁶ Had there been a conspiracy prior to March 1977, predicted prices should be substantially higher than actual prices. Had there been no effective conspiracy, there should be no systematic relationship between the levels of predicted and actual prices. The results suggest no conspiracy. While predicted prices were higher than actual prices some of the time, they were lower at other times. There is no clear pattern and certainly no distinct decline in price during the post-March 1977 period.⁷⁷

75. In the western plywood market, predicted prices were approximately at the price control level, suggesting that controls had little or no impact.

76. The forecasts involved corrections for the presence of serial correlation and thus are somewhat more sophisticated than implied. For a discussion of this issue, see R. PINDYCK & D. RUBINFELD, *supra* note 29.

77. There is a related but somewhat different econometric approach to distinguishing cartel prices from competitive prices. In it, one would estimate the econometric model for all time periods for which there was data, both conspiratorial and nonconspiratorial. However, this estimating model would include one additional variable, a dummy variable equal to 1 when the cartel was in effect, and 0 otherwise. The coefficient on the cartel dummy would then indicate whether the average price during the cartel's existence

FIGURE 11



Even this test is not a fully satisfactory one, however. As described earlier, forecast reliability needs to be taken into account. In the regression model, the

was significantly different and, in particular, higher than the average price during the competitive period. Thus, one could translate the forecasting approach into a hypothesis testing approach. However, there is an important limiting assumption made here. That assumption is that the overall behavior of the regression model, that is, the behavior of supply and demand, can be modeled in precisely the same way during both the conspiratorial and nonconspiratorial periods. This may be a reasonable assumption in some cases, but it should not be made without substantial thought. In contrast, the described forecasting approach requires no such assumption and is therefore more general than the hypothesis testing approach.

standard error of forecast was approximately 12.0. (It is actually somewhat higher than the standard error of the regression, but not substantially.) When two standard errors of forecast are drawn on either side of the predicted values in Figure 11, the 95% confidence band is so large that almost anything is possible. Thus, while the forecasting test adds no support to the conspiracy hypothesis, one cannot say with certainty on the basis of this test that a conspiracy could not have occurred. A better test would be provided by a better model or a forecast for a longer period of time.⁷⁸

Given the difficulties of distinguishing cartel-determined plywood prices and competitive plywood prices through the use of multiple regression, some complementary approaches are worth considering seriously. These rely to a greater extent on the basic economic theory of noncompetitive markets and to a lesser degree on multiple regression analysis. If the plywood companies effectively set prices higher than competitive levels, they must have restricted output to do so. Lower output and higher prices will lead, other things being equal, to higher than competitive profit levels, and will induce new firms to enter to take advantage of such supracompetitive profits.

There are two related approaches of analysis suggested here. One focuses on comparing rates of return in the plywood industry in the South to rates that would be expected in a competitive industry. Were southern plywood profits higher than competitive industry profits? The rate of return analysis involves a substantial set of calculations with a number of necessary implicit assumptions. Once the calculations are done, one has an *ex post* comparison of rates of return in southern plywood with competitive industry rates of return. It must still be decided whether any differences in these rates of return are due to the cartel or to unaccounted-for factors.

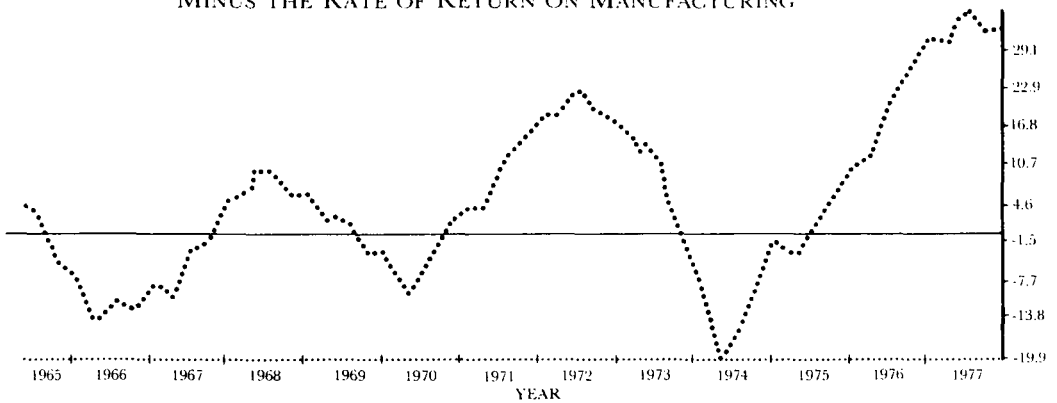
The second approach focuses on entry. Was the assumed cartel able to limit entry to some extent by delaying the construction of some southern plywood plants and not building others? Indeed, it must still be decided how rapidly firms would have entered had the market been competitive versus entry under regulation by a cartel. One must then decide exactly how to test to distinguish between these alternative patterns of entry.

The current analysis cannot fully resolve either the issue of rate of return or that of entry since both issues involve rather complex econometric factors of time lags, cyclical demand, the measurement of rates of return, and so on. However, the following brief description of the plywood data in relationship to the rate of return issue suggests why an econometric approach is likely to be useful. Figure 12 shows the rate of return on southern plywood plants relative to the rate of return on all manufacturing.⁷⁹ Initially, rates of return were quite low relative to all manufacturing, a typical result when a new industry is in its early development stage. Subsequently, rates of return showed a cyclical pattern over the 1966-76 period, with returns being higher than manufacturing returns for a portion of the

78. With a long forecast period one can estimate a new model using only the "postconspiracy period" and see whether it is different from the "preconspiracy model."

79. Data sources are shown in Table 4.

FIGURE 12
SOUTHERN PROFIT RATE ADJUSTED FOR CAPITAL COSTS
MINUS THE RATE OF RETURN ON MANUFACTURING



period and lower for the remaining portion. The important and most obvious result is that rates of return were not systematically higher in the southern plywood industry than in all manufacturing. Were it not for some of the technical difficulties in creating the rate of return series, this evidence would be very persuasive, if not conclusive.⁸⁰

V

DEALING WITH IMPACT AND DAMAGES

This article has previously discussed how the use of statistics and econometrics can be very helpful in deciding whether a violation has occurred. This section discusses the issue of whether a violation, if it has occurred, has caused impact and, if it has, whether there are any damages. This section begins with a consideration of the appropriateness of using econometrics to answer the impact/damage question.

The econometric approach to measuring damages must overcome certain innate difficulties: (1) the need for a theory to explain which variable or variables should be used; (2) the need to collect the relevant data; and (3) the choice of an appropriate estimation technique and functional form. Once these obstacles have been overcome, an econometric approach can be extremely useful. In the most propitious situation the data analyzed cover the entire period during which violation has occurred. In such a situation, if the model is appropriate, one can obtain a direct measure of the total damages involved. Econometric techniques, however, can never be applied mindlessly, and on occasion they may provide a good answer

80. The rate-of-return series comes from individual firm data and includes adjustments to account for the economic depreciation (not the accounting depreciation) of the firm's plant and equipment. These calculations involve a substantial number of technical assumptions and therefore are likely to be more subject to attack than an analysis which relies more completely on published data. As a second test, deviations in rates of return from the average were checked to see whether they could be explained by demand/supply variables. They were, with the resulting $R^2 = .92$.

to the wrong question. Section A illustrates some of the dangers of the econometric method with reference to the international uranium cartel. Section B considers how econometric techniques might be used to estimate damages with respect to the ampicillin case. In many cases, however, the question of damages requires obtaining an answer to a "but-for" question for which a direct econometric calculation is not possible. There the second subsection shows how the technique of economic simulation can provide an alternative approach to getting at the impact and damage question within the context of the ampicillin case. In the ampicillin case the simulation and the econometric exercises suggest absence of impact, but the techniques apply equally well whether or not there is impact.

The final two subsections consider cases in which information is not available about the entire damage period, or if available, is not necessary to estimate impact and damages. The third subsection demonstrates the use of samples of statistical information about damages to obtain total damage estimates when the data do not permit a comprehensive economic analysis. Finally, the last subsection discusses some of the limitations of anecdotal evidence, as well as the appropriateness of choosing samples randomly.

A. *Uranium*

Suppose that, for some period of time, it is established that a producers' cartel did exist and did attempt to raise prices, and that prices in fact rose. There is little question that an antitrust violation has occurred.⁸¹ The questions for a private suit against a member of the cartel by purchasers during the period of cartel operations are whether the cartel caused any part of the price rise, and if so, how much. The economic-statistical problem is whether it is possible to measure the impact of the cartel on price.

For this type of case, superficially at least, the application of multiple regression would seem natural. Many things in a competitive market would be expected to affect price, some affecting demand and some affecting supply. While such forces would affect price in any market, competitive or cartelized, they would not necessarily do so in the same way, or to the same extent. Why not identify each important supply and demand variable and attempt to explain price variations in terms of them? Why not then add a variable for cartel activity and see if it adds to the explanation, and if so, by how much? The coefficient of the cartel variable might be taken as the measure of the cartel's effect on price after the influence of the other supply/demand variables has been measured.

This approach might be successful either if there was a long enough precartel period to establish the competitive impact of the so-called supply/demand variables, or if the other explanatory variables are known to be independent of the cartel activities and vice versa. If such independence is not the case, a multiple regression analysis might reasonably account for price changes, but it would not tell anything about the causal impact of the cartel. To see why, it is only necessary

81. Of course, there are any number of potentially exempt agreements, but it is assumed that none of them apply to this discussion.

to realize that while a cartel can (1) raise price, restrict output, and gain profits in the context of given "supply" and "demand," it can also (2) cause shifts in demand and in supply that lead to price increases, and (3) take advantage of shifts in demand or supply that occur outside of the activities of the cartel. Such "indirect" effects of a cartel make statistical determination of prices but for the cartel difficult. Such indirect effects are especially likely to occur if demand and/or supply are heavily affected by expectations, for the cartel can deliberately influence those expectations by its actions. The international uranium cartel that operated at least between 1971 and 1975 provides a case in point.

The uranium industry developed far in advance of a private commercial market for its product. Uranium was, of course, needed for the atomic bomb and for a long period was developed for and sold exclusively to the Atomic Energy Commission (AEC). When the commercial market for uranium for projected nuclear power plants came into existence, the supplies of known uranium were larger than short term needs; there was a large oversupply of enriched uranium in U.S. government stockpiles, and all aspects of the sale, enrichment, and storage of uranium in the United States were subject to massive U.S. regulations. There was, moreover, an embargo against the use (but not the enrichment) of uranium of foreign origin in the United States, where the greatest demand for uranium was expected. Uranium prices had once been supported at artificially high prices by AEC purchases in order to stimulate development, but after the AEC stopped buying, prices became heavily depressed. Even at \$6 a pound, uranium was in abundant supply and uranium producers were arguably selling uranium at below, perhaps well below, its replacement cost.

Sometime early in the 1970's, discussions arose among the major uranium producers regarding the formation of a "club" that would seek to restrict outputs, allocate markets, and enforce minimum prices. This club included producers from Canada, South Africa, France, and the United States, the largest existing uranium producers, and Australia, the sleeping giant in terms of its undeveloped reserves. Rather uniquely, in the history of covert cartels, detailed minutes of many of the club's meetings, starting in Paris in 1972 and continuing until 1974, have come to light. The revelations detail who was present, what the agenda was, and what was decided. This was a classic cartel that first sought to raise prices into the \$8-\$10 range and later sought to refuse to enter into fixed price contracts at all.

Use for fuel in nuclear plants is the only major source of demand for uranium. The expected boom in nuclear power plant construction and authorization materialized far more slowly than previously anticipated, owing largely to safety and environmental matters. Nevertheless, between 1972 and 1975, uranium prices exploded from the \$6 range to \$45 a pound.

While superficially the price behavior correlated with cartel activity to the point that some connection might seem obvious, the uranium case defendants (especially subsidiaries of Gulf) when sued by major American customers (especially Westinghouse, Exxon, and TVA) claimed no impact.⁸² Conceding the exist-

82. See *In re Uranium Antitrust Litigation*, 552 F. Supp. 518 (N.D. Ill. 1982).

ence of the cartel, they claimed it had had no effect on price. The cartel's modest efforts to overcome a depressed price, they claimed, were quickly overtaken by events which led price to rise farther and faster than the cartel's members' fondest hopes. These events, the argument continued, created a genuine fear of a uranium shortage and that led to exploding prices.

Among these events were: (1) the OPEC cartel and the exploding price of oil, which generated rises in prices of all competitive fuels, as well as instilling an almost irrational fear of a worldwide energy shortage; (2) numerous changes in U.S. government policy concerning uranium enrichment; (3) election of a labor government in Australia which then withdrew the undeveloped, but potentially cheap, uranium reserves of that country from the world market for several years; (4) Westinghouse Electric Company's revelation that it had sold more uranium for future delivery than it firmly possessed and was not going to honor its delivery contracts; and (5) decisions of the Canadian government related to the uranium cartel that greatly restricted exports to the United States and forbade exports at prices less than those agreed to by the cartel.

Although the details are not of concern here, a regression analysis explaining price changes in terms of these events would suffice to "explain" much of the rise in uranium prices without any reference to a producers' cartel. Because these events and the cartel meetings covered the same period, each was therefore closely correlated with the price rise. Had the events been independent of one another and the cartel, the relative importance of the cartel might have been sorted out by multiple regression. But many of these events were critically interdependent. If the oil embargo created fear among online nuclear plants of an energy shortage, the cartel members were in a position to convert that fear to virtual panic by a moratorium on commitments to sell uranium. When middleman Westinghouse defaulted, it did not thereby increase real demands or decrease real supplies of uranium. The latter were always in excess of needs at modest prices. However, when the newly uncovered commitments for uranium were not replaced by commitments from cartel members, panic developed and bid prices for uranium commitments shot up with no takers.

The uranium cartel could, and arguably did, affect price by restricting output to a specified quantity and enjoy the higher price that the limited output would command. It could also affect the level of demand at each price and decrease the elasticity of demand by the actions that increased buyers' willingness to purchase uranium. Notice the word "willingness." Demand reflects perceptions of need as well as requirements and thus has a subjective element affected by expectations. For example, if most people believe that something they use will be in short supply next week, they may rush to buy it this week, and if necessary, pay premium prices to fill their anticipated needs. Moreover, their actions may bring about the result they feared.

The inadequacy of attempting a dichotomy between cartel induced and supply-demand induced causes of price increases now becomes clear. The changes in price just described are all due to changes or shifts in supply and demand. Yet,

each of these shifts of supply or demand can be caused in the first instance by the cartel.

Moreover, if a cartel is affecting supply (or demand) simultaneously with genuinely independent (exogenous) events over which it has no control, how should one attribute the resulting cumulative effect? An example of this complicating interaction is shown in as follows. Suppose that a decrease in supply leads to an increase in price from \$10 to \$20. Suppose half of the decrease in supply is due to exogenous events and half to the activities of a cartel. Either half alone might have no effect on price with demand, absorbing an excess supply at any price above \$10. Together they create a shortage and cause a rise in price to \$20. The joint effect is greater than the sum of the separate effects. The cartel members might be expected to claim that but for the exogenous event, their restriction had no effect on price. Alternatively, but for the cartel, the exogenous supply decrease had no effect on price. It seems clear that when comparing events with and without the cartel, the cartel activity is the marginal activity, and it is thus the basic cause of the price increase in the example.

The facts in the uranium market generated competing hypotheses about what, if any, effect on price was due to the cartel. Neither economic theory nor straightforward econometric analysis was adequate to resolve the question.⁸³ Techniques that work some of the time do not work always.

Perhaps the major point of this discussion is that additive models of behavior, which are implicit in relatively simple regression models, can be dangerous if misapplied. Regression analysis depends critically on independence of explanatory variables from each other, so that the effects of one explanatory variable can be added to another. When numerous events are occurring simultaneously and inter-dependently, the interaction of the individual variables may be more important than their individual effects. There are other, more complex ways of shedding light on such events. Perhaps if the uranium cases had not been settled, the adversary process would have shown that here too economic analyses had a role to play.

B. *Estimating "But-For" Prices—Ampicillin*

In the ampicillin case, the question of how to test whether the presence of generic firms bidding in the CCS market affects price has been answered. Implicit in that analysis was some empirical information about the impact of generic bids. In this section, the discussion broadens to ask the question of what prices would have been but for the patent limitation on the sale of bulk ampicillin. The ensuing impact/damages analysis requires an analysis of a series of three questions, all of which must be answered yes if damages are to be present. The questions are:

- (1) If there had been no contract limitation, would generic houses have obtained bulk ampicillin?

83. This statement does not imply, however, that an appropriately specified econometric model could not be useful here. The specification would need to take into explicit account the fact that some of the explanatory variables relating to supply and demand can themselves be explained by cartel behavior. Thus, a single regression equation would have to be replaced by a complex system of equations, whose estimation and interpretation would be more difficult than the single equation case.

- (2) If the generic houses had obtained bulk ampicillin, would they have bid in the CCS market?
- (3) If the generic houses had been in the CCS market would their bids have lowered the purchase price of ampicillin?

Economic analysis helps to address each question.

1. *Would Generic Houses Have Obtained Bulk Ampicillin?* First note that the costs of producing and marketing ampicillin vary by firm and by quantity produced. Costs of production of ampicillin fell sharply between 1963 and 1973 in part because larger scale production led to more efficient processes, and in part because technology improved. In addition, costs vary by firm—evidence strongly suggests that Beecham's costs were higher than those of Bristol-Myers. If generic houses could obtain bulk ampicillin without the contract between Beecham and Bristol, Bristol would have been the obvious producer to supply it. If Bristol (or Beecham for that matter) had excess capacity, how would they have used it? In particular, would it have made economic sense to sell the material to generic houses?

One use for the excess ampicillin would be to sell it in finished form under a brand name, in either the retail or CCS markets. Such self marketing was a major use by manufacturers of their production. Another would be to sell it to an industrial customer who could market it more effectively than the manufacturer, perhaps because the industrial customer had an underutilized sales staff. Or, such an industrial customer may have had a patented product of its own to offer in return, thus permitting the ampicillin manufacturer to diversify. A third use would be to sell it to a generic house for resale. However, a sale to a generic house is not likely to have had any appeal to a manufacturer such as Beecham because anything the generic house could do, other manufacturers could do as well or better. Such generic houses had no staffs of experts to market the product nor the marketing know-how to promote the product. Neither did they have other products to exchange for ampicillin in a barter transaction; thus, they had no comparative advantage as middlemen.

2. *Would Generic Houses Have Bid in the CCS Market?* But suppose generic houses had obtained bulk from Beecham. They would have had to sell at a markup in order to make a profit. However, analysis of costs and of winning bids suggests that the only opportunity for Beecham itself to sell profitably in the CCS market was in the years 1965 to 1969. During that period, both Beecham and Ayerst (its American distributor) did win bids in the CCS market. During that time, however, most sales by these companies were in the retail market, in which prices for both brand name products and generic products were higher than in the CCS market. Thus, to the extent that there might be damage (were bulk obtained), it would be limited to the 1965-69 period.

Nevertheless, consider the profit-maximizing motives of the generic house that buys bulk, either from Beecham or from a foreign manufacturer. Direct evidence shows that when generics did sell ampicillin they sold it in the generic retail market, not the CCS market, since prices and profit margins were higher. No

generic house (to the authors' knowledge) ever won a bid in the CCS market. It seems likely, therefore, that had generic houses obtained bulk, they would not have sold the resulting ampicillin in the CCS market.

3. *Would Bids by Generic Firms Have Led to Lower Prices?* There are a number of reasons why one would expect manufacturing firms to sell ampicillin in the CCS market. First, the purpose might be promotional. With brand name ampicillin available for hospital prescription, doctors will become familiar with the product and presumably prescribe it for nonhospital use. Second, the sales can be viewed as a part of a general promotion of the drug ampicillin vis-a-vis other ampicillin substitutes. Third, the sales in the CCS market might be economically advantageous to a price-discriminating drug company. With sales of retail prescriptions at a relatively high price and sales to hospitals at a lower price, the latter could sufficiently increase the volume of sales so that the firm could take advantage of scale economies in the production of the drug.

With manufacturing firms selling in the CCS markets, the question of whether bids by generic houses would have affected winning bid price becomes a crucial one. To analyze this issue, assume *arguendo* that generic houses would have bid earlier in the CCS market. Would those bids have had any impact on price, and if so, what would have been the size of the impact? Despite the assumption about the presence of generic houses, the answer to this important question is essentially an empirical one that cannot be settled on the basis of pure economic theory. To decide the impact issue, one must first take into account the fact that there are two ways in which the presence of generic houses might affect price. First, generic houses might actually win sealed bids, at prices lower than what the winning bid would have been without the generics. Second, even if generics never won a bid, the bids of others in the market might have been lower in response to the presence of generics. If the bids of manufacturers are lowered to counter the presence of generics, for example, one would argue that impact has occurred despite the fact that a generic house never won a bid.

How can such an empirical question or questions be decided with reference to a period in which no generics were in the market? The answer lies in the use of either of two quantitative techniques—multiple regression analysis or simulation. For either to be applied to this question, one must first assume that the behavior of firms in the ampicillin market after 1972 can be used to shed light on what the behavior of those firms would have been pre-1972. Costs and prices need not be the same, but the bidding strategies of firms of each type must be the same. There is no reason to suspect otherwise in the ampicillin case. Notice that this assumption involves two parts: (a) the assumption that the bidding strategy of generic houses would have been the same before 1972 as after it; and (b) the assumption that the strategy of existing firms in response to the presence of generic houses after 1972 would have been the same before 1972, had generic houses been bidding.

This discussion then treats the question of impact in two stages, using each of these assumptions along the way. First, multiple regression analysis can be used to determine whether individual firms in the market select their sealed bids in part in response to the presence of generic houses. Such an analysis can be done with

respect to a number of firms, but the analysis for Bristol-Myers is typical. The question is whether the number of generic houses bidding had any impact on the bid price selected by Bristol-Myers. Because Bristol's "bid function" may depend upon other factors such as cost and other types of bidders, a statistical analysis is necessary to control for these other variables. Therefore, a multiple regression as described in section IV, subsection A is used. In that section, the objective was to test the hypothesis that generic houses had no effect on winning bids. The hypothesis was tested by comparing the coefficient on the variable N (number of generic houses) to its standard error. Here the concern lies with the coefficient of the variable N in a regression in which the dependent variable P = Bristol's bid price.

The multiple regression results are summarized in Table 5. Since the emphasis lies solely with the coefficient of N , the sign, rather than the magnitude, of the other coefficients is reported. The coefficient on the N variable turns out to be .05,

TABLE 5

Variable	Sign of Coefficient	Significant at the 5% Level
CONSTANT	—	Yes
NMAN	—	No
NBRCUS	—	Yes
NBECUS	+	No
NW	+	No
N	+(.05)	No (SE. = .09)
C	+	Yes
QD	+	Yes
T69T070	—	Yes
T71T072	—	No
T73T077	+	No
	N=79	R ² = .96

which is not significantly different from zero. This result suggests literally that the addition of a generic house leads to higher bid prices by Bristol, but a more realistic interpretation is that the addition of one or more generic house bidders has *no* impact on Bristol's bidding (or the bidding of other manufacturers or industrial customers). This result is not particular to Bristol. Essentially the same results were obtained for all manufacturers of ampicillin.

By thus answering one of the two original questions, one simplifies the analysis of the second question. Bristol's bids and the bids of other manufacturers were unaffected by the presence of generic houses. The only impact that generics could have had on price was a direct one—by winning a bid. Generic houses actually won no bids after 1972. Nevertheless, to answer the question of whether they would have won bids prior to 1972, one utilizes a simulation technique.

Roughly speaking, the simulation process is utilized to mimic what the CCS ampicillin market would have been like had generic houses been present to bid. Detailed information describing the bids of all types of firms, manufacturers, industrial customers, wholesalers, and generic houses were entered into a com-

puter. The bid descriptions were divided into each of four time periods, 1965-68, 1969-70, 1971-72, and 1973-77. Within each time period, bids were adjusted for cost to put them all on a comparable basis. Thus, for each type of firm the computer stored a set of possible bids, based on actual bidding experiences. In addition, the computer was given detailed information about the number and combination of nongeneric firms that bid in different types of bidding situations.

To simulate what the effect of generic houses would have been in 1965-68 (when in fact there were none of them), the following procedure was used. First, the computer selected at random a typical bidding situation—for example, one in which five manufacturers, three industrial customers, and two wholesalers were bidding. For each potential bidder, the computer then selected a characteristic bid for each of the firms in the market. The bid was chosen at random from the set of all bids that firms of that type had made throughout the period of study.⁸⁴ The computer then examined all the bids made and determined the winner and the winning bid. If this experiment had been performed only once, one could not be certain that the result would be typical of what would happen in different bidding situations. Fortunately, however, modern computers can handle a simulation problem of this sort at very low cost. Therefore, each simulation experiment was repeated one thousand times at a cost of about 20¢. The mean winning bid in each period was then calculated, as were the number of wins by type of firm.

Would generic houses have an impact on price? To complete the evaluation of this question, a simulation experiment just like the previous one was run, but with the addition of generic houses to the bidding process.⁸⁵ This simulation experiment was repeated an additional one thousand times to see (a) whether there was any change in the mean winning price, and (b) whether generic houses ever won a bid. The results are summarized in Table 6. In the period 1965-68 (the one primarily at issue in the case), generic houses never bid in practice and never won a bid in our simulation.

It was also asked what would have happened if more wholesalers entered the market. Wholesalers were difficult to distinguish from generic houses in the analysis and appeared to be in relatively similar cost situations. After two additional wholesalers were added to the simulation, it was found that in one thousand tries, one wholesaler won a bid. How did this happen, given the relatively high costs associated with wholesaling and the high bids that wholesalers typically make? The answer lies in the fact that in randomly drawing bidding situations, the computer selected one in which no other firms were bidding. Thus, the win by a wholesaler reflects a possible, but relatively implausible, situation which to the authors' knowledge has had no real world counterpart. That single wholesaler's bid had virtually no impact on price. The average impact of an increase of a few cents in price, occurring only once in a thousand times, is effectively zero.

84. The bids were adjusted for differences over time due to changes in cost and price levels.

85. Since it has been shown that generic houses do not affect the bids of others, the addition of generic houses is an easy process, but it would still be manageable had other firms' bid functions been more complex.

TABLE 6
SUMMARY OF THE SIMULATION EXERCISE

Period	Probability of a win (in 1965-68, by wholesaler; other periods, by generic house)	Mean winning price for all winning bids ^a (dollars)
1965-68		
Original situation	.000	11.00
Two wholesalers added	.001	11.03
1969-70		
Original situation	.000	5.21
Two generic houses added	.000	5.20
1971-72		
Original situation	.000	3.90
Two generic houses added	.000	3.90
1973-77		
Original situation	.000	3.55
Two generic houses added	.002	3.54

^a In all cases, the difference between mean winning prices are insignificant when comparing the original situation and the situation in which additional bidders are added.

Essentially the same result was obtained for later time periods. In the simulation exercise for 1969-70, no wholesalers or others won any bids, and when two additional generic houses were added, no bids were won, even out of a thousand tries. For the period 1971-72, the result was the same. The effect of generic houses was felt slightly after 1972. Two generic houses won bids in the regular one thousand bidding situations. However, this increase was so tiny that it had no measurable impact on price.

C. *Use of Samples*

What happens if the econometric analysis of the whole data set is not feasible? A second approach that can be quite useful is to take a sample of events and estimate the damage appropriate to those events, and then to generalize about total damages. In fact, if the number of possible incidents is very large, it may be practically necessary to rely on such a sample. And, as long as the sample is a "random" sample, the analysis presents little problem. A random sample is not a haphazard or biased selection, but rather one chosen independently of any of the characteristics of the case, especially those that might be related to the damages.

A careful statistical analysis of a random sample can show how reliable the results are, that is, how liable to error the damage measure is likely to be. An example of a nonrandom biased sample would be one in which only incidents generating high damages were sampled. Obviously, the use of such a sample would lead to an estimate of total damages in the case that was "biased upward," that is, too large. Similarly, a biased downward estimate would result if a sample excluded the high damage cases. The problem with haphazard samples is that they may be biased upward or downward, but to a degree that cannot be known

by studying the sample. Plaintiffs will seek upwardly biased samples, defendants downwardly biased ones. As a result, sampling in the context of an adversarial proceeding often takes on a rather unscientific, self-serving character.⁸⁶

Perhaps for this reason, there is a widespread sense among some attorneys that a sample is not nearly as adequate from an evidentiary point of view as are the total population data. On statistical grounds, however, an unbiased sample may be as good as, or even better than,⁸⁷ the entire population. If the sample is sufficiently large to provide accurate information about damages, then there is no need to obtain information about the entire population.

Consider the case of *United States v. Nissan Motor Corp. in U.S.A.*⁸⁸ The government charged that Nissan, the Japanese maker of Datsun cars, was guilty of resale price maintenance in sales in the United States. In addition to the case brought by the Justice Department, private plaintiffs also brought class actions on behalf of purchasers of Datsun cars.⁸⁹ The government case ended in a consent decree. Suppose, for the sake of this discussion, that there was a violation of the law, that is, that Nissan did systematically attempt to impose minimum retail prices on its dealers in ways that violated section 1 of the Sherman Act. Private plaintiffs claimed that this retail price maintenance (RPM) had the effect of reducing discounts below what they would otherwise have been and thus led to overpayments and damages to all buyers. An important legal question in the private cases was whether there was a sustainable class which could serve as the basis of a class action suit.⁹⁰ For such a class of buyers to be sustained, evidence was needed that there was *some* impact on all buyers of Datsuns within the class.

Consider the question of impact if the proposed class consisted of "all buyers of Datsun vehicles in the United States," and also the question of impact, state by state. It is surely conceivable that the violation had led to uniformly higher prices in some states (for example, California and Massachusetts) but not in others (for example, Florida).⁹¹ In that event, a national class might not be sustainable, but class actions in California and Massachusetts might survive. Of course, within any state, a similar set of questions might be asked: did *all* purchasers in California suffer a reduction of discounts, or did discounts vary according to, say, type of vehicle, so that the retail price maintenance scheme led to overcharges in some cases and not in others?

The private plaintiffs' *Nissan* case was fought on the question of class certification. In principle, every transaction during the damage period could be studied

86. Even nonrandom samples can be useful in certain contexts. For example, if one were attempting to find out whether there was a certain minimal or maximal effect and one knew the direction of bias in the sample, a nonrandom sample could be used.

87. Errors in measurement may render enumeration of population characteristics subject to errors that are larger than possible sampling errors.

88. 1973-1 Trade Cas. (CCH) ¶ 74,333 (N.D. Cal. Feb. 26, 1973) (consent decree).

89. See *In re Nissan Antitrust Litigation*, 577 F.2d 910 (5th Cir. 1978), *cert. denied*, 439 U.S. 1072 (1979).

90. See *P.D.Q. Inc. v. Nissan Motor Corp. in U.S.A.*, 61 F.R.D. 372 (S.D. Fla. 1973); see also *In re Nissan Motor Corp. Antitrust Litigation*, 552 F.2d 1088 (5th Cir. 1977).

91. Competitive pressures may have differed state by state, as might efforts by Nissan to enforce its RPM on dealers.

because every sale involved an invoice which showed both the actual price paid and sufficient detail to permit calculation of the RPM price that (it is assumed) Datsun urged its dealers to charge.

Defendants seeking to prevent certification of the class under Federal Rule 23 attempted to collect and analyze all invoices from selected dealers in fourteen states over six years.⁹² This proved an incredibly time-consuming and costly exercise. It involved over 40,000 individual transactions which were collected and tabulated by a major accounting firm. The collection of data alone cost nearly \$1 million. Whether it was all worth it is not entirely resolvable. Nissan won the case in the end, arguably in part because its data collection revealed such heterogeneity in prices across states and across models as to undermine the notion of uniformity of impact that was vital to sustaining the class action.

The question we address here is whether so extensive an enumeration was necessary for any purpose. When the time came for detailed analysis by Nissan's economists of the data collected, the decision was made not to analyze the whole data base, or population, but only a 12% random sample of it, simply to save on the costs of computation. The fact that such a sample may prove wholly adequate can be seen by looking at Table 7, which compares information taken from the sample with information from the population. The first thing to note is that the average dollar discount given on Datsuns is very similar whether one looks at the sample or at the population. The difference between the means is statistically and substantively insignificant—about \$0.75 per car. The average percentage discounts were nearly identical, approximately 2.4% for both the sample and the population. More important from the point of view of the case was that these discounts varied substantially, with a standard deviation of roughly 5.7%. Discounts from recommended prices were far from uniform. Further, and more importantly, the sample showed that for a substantial number of transactions, no discount was given, while other sales involved large discounts. Thus the sample of evidence was sufficient to show decisively the absence of uniform impact required to maintain the class action.

TABLE 7
SUMMARY DATA
POPULATION AND SAMPLE

	Mean Dollar Discount	Mean Percentage Discount	Standard Deviation of Dollar Discount	Standard Deviation of Percent Discount	Number of Obs.
Population	58.60	2.39	158.85	5.70	40,640
SAMPLE	59.35	2.47	160.95	5.77	4,866

92. This collection was, of course, itself a sample since there were more states and more years that might have been included.

Since the sample of 4886 transactions showed this absence of uniform impact decisively, would it not have been sufficient in the first instance to collect data on only 5000 transactions instead of 40,000, and thus save substantially on cost? From a statistical point of view the answer is plainly yes. Whether legally the sample would have satisfied lawyers, judges, or jurors is an open question. If the use of statistical techniques is to be available in cases where other than large firms are involved, it is plainly important to find ways to make scientific sampling an acceptable form of evidence; otherwise, cost considerations will invite use of less adequate anecdotal data.

Before leaving the data collected in the *Nissan* case, it is worth noting that analysis of the sample data revealed a great deal about what *did* determine price variables in Datsuns. The pattern of discounts responded substantially to the supply and demand variations over time and across regions of the country. Some of this evidence is illustrated in Table 8. This table shows that the dollar discounts varied substantially by car, with a negative discount—that is, a premium—attached to the very popular Datsun 280Z. Likewise, dollar discounts varied by year, generally falling as the popularity of Datsuns rose over time. In addition, they varied substantially by state. Note that all of the results here involve the use of simple statistics and samples. Regression analysis could have been used to further explain the factors describing how discounts vary. In the *Nissan* case, however, such analysis was not necessary to resolve the issue in question.

D. Anecdotal Evidence

Now consider a third and highly common form of damage estimation—the use of anecdotes which are (implicitly) taken to be typical. In a common example, plaintiff argues for damages by identifying one or more incidents in which significant damages are clearly exhibited. Such incidents are used to provide estimates of total damages by multiplying average damages for the selected incident with the total number of possible incidents throughout the damage period. It should be quite clear that such an approach has some fatal defects. The primary reason is that the sampling of incidents is not random. Plaintiff has the incentive to choose incidents in which damages are large, both because the proof of violation is likely to be easier in such cases and because evidence about damages is easier to obtain when damages are high. Thus, the adversary process invites biased analysis. Anecdotes are haphazard and not random, while science needs a random sample of the distribution of all incidents to obtain accurate damage estimates.⁹³ Because of

93. One unusual example of the use of anecdotes occurred in the ampicillin case. *In re Ampicillin Antitrust Litigation*, 88 F.R.D. 174 (D.C. Cir. 1983). Early in the case, plaintiffs argued that impact and damages could be determined on the basis of a single incident—the so-called Biocraft incident. Here a bid by a generic house (although subsequently disqualified) led to a rebidding of the contract. The winning bid was lower than the previous winning bid. Thus, the argument ran, the presence of the generic house led to a lower price. The implicit estimate of damages using this single plaintiff-chosen incident was very substantial when that percentage was applied to all transactions. Plaintiffs, however, later dropped the incident from consideration when evidence became available which suggested that the incident itself had been contrived by plaintiffs' attorneys as a device to establish impact and damages.

such defects, anecdotal-based measures of damages must be treated with considerable skepticism.

TABLE 8
DISCOUNT DATA BY CATEGORY

(holding other variables constant through multiple regression)

Variable	SAMPLE	
	Dollars	Percent Discount
<i>Datsun Model Type</i>		
311	125.0	3.60
280Z	-308.9	-8.69
Pickup	102.5	3.73
Standard	119.0	4.34
<i>Year</i>		
1967	178.2	7.87
1968	100.8	4.57
1969	117.4	4.91
1970	95.0	3.75
1971	30.9	1.33
1972	49.7	2.03
<i>State</i>		
Fla.	73.2	3.08
N.Y.	73.2	3.08
Calif.	31.0	1.29
Ariz.	111.8	4.54
Ore.	105.7	4.21
Tex.	85.8	3.57
Ark., Kan., Okla.	117.8	4.64
Mass.	25.1	1.17
Pa.	72.7	3.13
Ohio	87.6	3.42
Ill.	33.3	1.71
Minn.	34.3	2.00

VI

SOME TENTATIVE CONCLUSIONS

This article has sought to illustrate the value of quantitative methods in anti-trust analysis. Quantitative methods can be extremely valuable in studying a wide range of empirical questions involving antitrust and can prove helpful in the judicial disposition of cases. When properly utilized, the benefits can be substantial. There is, however, a serious possibility of misuse. This discussion has pointed out some of the dangers involved with quantitative methods. These may be regarded as "costs" of use of the methods. Under a balanced treatment of benefits and costs, the benefits are sufficiently great to warrant incurring, or avoiding, the costs.

A detailed recapitulation of the technical aspects of the article seems unneces-

sary. Rather, this discussion concludes with some possible implications for the legal process. First, we are urging a new view of quantitative methods and of the economic expert, to displace the "traditional view." To oversimplify a bit, the traditional view of experts with or without quantitative methods, is in conventional adversary terms: a battle between competing sets of experts. The experts for the plaintiff and the experts for the defendant each testify on certain issues; the jury or judge evaluates the credibility of each witness, relying in part perhaps on the validity of the empirical evidence, but equally on the expert's demeanor and credentials. Of course, the expert's opinion is going to be taken as valid evidence only to the extent that it is reasonably based and consistent with the underlying data. Nevertheless, in this traditional view the expert himself becomes tied directly to the evidence. Greater emphasis can be placed on the believability of the witness than on the technical quality of the empirical evidence presented. A good witness, or one who appeals to the factfinder's prejudices, can prevail with shoddy data or shoddy methods. This result is not surprising when the careful analysis of data is necessarily complex and sophisticated. The complexity of economic analysis makes short, simple, clear explanation difficult. At the same time, if presenting the data honestly, the witness will admit a dozen places where errors could occur. In contrast, a simple graph with a firmly asserted conclusion, even if spurious, can be quickly grasped and readily accepted.

This traditional view, which places great weight on the character and background of the witness along with the underlying data, is not satisfactory. What is required is a scientific or professional standard as to what is adequate evidence in antitrust cases. The standards need to be based less on the credentials of the expert, although credentials are not irrelevant, and more on the technical quality of the data and analysis being presented. As one example, too much is known to allow lawyers to denigrate data merely because they are based on samples or are "subject to error."

Of course, there are serious difficulties in presenting technical analyses in antitrust litigation. If the trial is by jury, there are obvious problems of comprehension of the material. Sadly, these problems are not necessarily less in the typical bench trial.⁹⁴ This article does not present a panacea to resolve this problem, but urges reform if the legal process is to have the benefit of modern economic analysis. Changes in the process ought to encourage the courts to distinguish serious empirical research and theory from misleading, even shoddy, use of quantitative methods. Current law is perverse: serious studies are detailed and technical and in part for that reason difficult to enter as evidence. Once entered, they are often easy to attack, in part because of their complexity, in part because of the necessary assumptions that must be made to make the analysis possible. It ought to be possible to make the experts' data and findings, rather than his asserted conclusions or himself, the evidence in the case.

How may the law approach this problem? First, it may be advisable in certain cases for the court to impanel an external expert or a panel of experts to advise it

94. For a discussion of the advantages and disadvantages of nonjury trials, see Lempert, *Civil Juries and Complex Cases: Let's Not Rush to Judgment*, 80 MICH. L. REV. 68 (1981).

on the professional nature of the technical material. The advantage of such a procedure is that an "unbiased" observer can advise on detailed matters that may not be easily handled through the normal adversary process.⁹⁵ There are, of course, substantial difficulties here because it would be easy for the external expert to become simply a second judge in the case.⁹⁶ But the difficulties must be balanced against the plain deficiencies of the traditional role, and there is certainly room for greater use of court-appointed experts than is now common.⁹⁷

Second, serious consideration ought to be given to allowing appellate review of technical evidence. Of course, one cannot be certain without further discussion and analysis whether this solution would work better than the usual "competition of the experts" at the district level. Nevertheless, it is worth pursuing. Briefs by parties as well as by amici on the scientific adequacy of techniques might be allowed.⁹⁸ If so, it may even make sense in certain kinds of cases to have outside experts available (as temporary clerks, perhaps) to the appellate court, to advise, at least in a limited way, on the validity of the technical empirical evidence entered by experts.⁹⁹

In short, for technical quantitative material of the kind becoming common in major antitrust cases, the trial court may simply not be the best place for the final determination of facts. It may well make sense to allow questions of fact to be tried at both the trial and appellate levels, and possibly remanded for further study. This does not seem so large a step: if a court can remand (or reverse) a case on the ground that market definition was neglected, it should be able to do so on the ground that no scientifically acceptable evidence of market definition was introduced.

95. For additional discussion in support of the broader use of masters in courtroom proceedings, see Kaufman, *Masters in the Federal Courts: Rule 53*, 58 COLUM. L. REV. 452 (1958); Comment, *An Adjudicative Role for Federal Magistrates in Civil Cases*, 40 U. CHI. L. REV. 584 (1973).

96. There are a number of constitutional as well as policy issues that suggest that the role for masters and experts ought to be limited. See, e.g., *Ira S. Bushey & Sons v. W.E. Hedger Transp.*, 167 F.2d 9, 17 (2d Cir.), cert. denied, 335 U.S. 816 (1948); see also Note, *Masters and Magistrates in the Federal Courts*, 88 HARV. L. REV. 779 (1975). The issue presented is that the use of magistrates can conceivably conflict with article III, section 1 of the Constitution which suggests that judges with permanent tenure in the court system should not have to worry about legislative or executive coercion.

97. The Constitution does allow judges broad discretion in using experts in the fact finding process. See Silberman, *Masters and Magistrates Part Two: The American Analogue*, 50 N.Y.U. L. REV. 1297 (1975). See also Lempert, *supra* note 94.

98. In a sense the trial courts are able to exercise some appellate review because they can pass on the issue of whether expert opinion should have been allowed in the first place.

99. The notion of trying fact at the appellate level runs counter to current American legal process. There are exceptions, of course, as when the Supreme Court has original jurisdiction in a federal case. In addition, magistrates and masters have been used in district courts to review the determinations of administrative agencies.

