ECONOMETRICS AND THE LAW:
A CASE STUDY IN THE PROOF OF ANTITRUST DAMAGES

BY

R. S. DAGGETT AND D. A. FREEDMAN

TECHNICAL REPORT NO. 23
SEPTEMBER 1983
REVISED
APRIL 1984

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY

Neyman-Kiefer volume
April, 1984

ECONOMETRICS AND THE LAW:
A CASE STUDY IN THE PROOF OF ANTITRUST DAMAGES

by

R. S. Daggett
Brobeck, Phleger, Harrison
One Market Plaza
San Francisco, California 94105

and

D. A. Freedman
Statistics Department
University of California
Berkeley, California 94720

Preface

An econometric model was used to prove damages in antitrust litigation.[1] The difficulties in such proof will be explored. For plaintiffs' counsel, an econometric analysis may seem quite attractive. Here is scientific, objective proof of market power and damages—backed by the prestige of mathematics, economics, and the computer. However, defense counsel need not despair. On closer examination, the econometric analysis may turn out to be no more than a series of unsupported assumptions even if these are expressed as formidable equations. Once articulated, the assumptions may prove to conflict with each other or with common sense. Also, the statistical calculations may turn out to be quite vulnerable to attack. Furthermore, in many econometric studies the data are of poor quality; this alone may vitiate the analysis.

---

[1] Enrico Farms and Diedrich Ranch v. H. J. Heinz et al. Daggett represented one defendant, and Freedman gave expert testimony on behalf of that defendant. This paper was adapted in part from materials prepared for the Federal Court Practice Program, Northern District of California, Lecture and Workshop Session on Expert Witnesses, February, 1982; and in part from Freedman's deposition testimony.

The validity of models, the reasonableness of their assumptions, their usefulness in forecasting, in policy analysis, or their probative value in the courtroom, must therefore be assessed very carefully on a case-by-case basis. No general rule can be given except this: mathematical symbols and computer printouts are not in themselves reliable indicators of scientific merit. All depends on where the equations come from, and what the computer programs do.

Econometric models are often used in damage litigation. The present case is used only as a concrete example, to focus the issues and make them clearer. In any endeavor involving fields as different as the law and economics, there is ample room for misunderstanding. The object of this paper is to identify the sort of technical assumptions that are involved in the legal use of econometrics, and the sort of questions that can be raised about those assumptions. For other views, see Finkelstein (1980), Fisher (1980), and the opinion of Higginbotham (1980) in Vuyanich vs. Republic National Bank 505 F Sup 224. We tried to make this paper accessible to a fairly wide group of readers, and hope that specialists will not be put off by the explanations which are unnecessary for them.

The Case

The litigation was a private antitrust action. The plaintiffs, who grow and sell tomatoes in the Fresno and Stockton regions, alleged that defendants Heinz and five other major canners conspired throughout the 1970's in violation of Section One of the Sherman Act to fix, stabilize and maintain at depressed levels prices the defendant canners paid plaintiff growers for tomatoes. Plaintiffs contended that as a result of the conspiracy, the prices plaintiffs received from defendants for tomatoes were lower than the prices they would have received in a free market, entitling plaintiffs to an award of treble damages and attorneys' fees. Defendants maintained that no such conspiracy existed, that each defendant canner determined the price it wanted to offer for tomatoes in the exercise of its own independent judgment based upon estimates made from year to year of market conditions, and that in fact tomato prices for the greater portion of the period in suit were fixed and controlled by growers themselves through the California Tomato Growers Association, which enjoys immunity from antitrust liability under the Capper-Volstead Act.

Plaintiffs proposed to call as an expert witness Richard E. Just, Professor of Agricultural and Resource Economics, University of California, Berkeley. This expert had developed an econometric model of the California tomato processing industry, which he used to prove the exercise of market power by the canners in the purchase of raw tomatoes, and to estimate what the free market prices for the tomatoes would have been, absent collusion by the canners. These prices were the basis for plaintiffs' proof of damages. A preliminary version of the model, and its data base, are described in a published monograph, Chern and Just (1978). The model itself is described in a journal article, Chern and Just (1980). Additional detail was given in deposition testimony.[1]

The defense engaged David A. Freedman, Professor of Statistics, University of California, Berkeley, to review the model and rebut it if possible. The present article will give some of the institutional background, and then summarize the model and the review. Some tentative conclusions will be drawn about the usefulness of econometric analysis in litigation.

The case was Enrico Farms and Diedrich Ranch v. H. J. Heinz et al., Nos. CV-75-206-EDP and CV-79-186-EDP (E.D.Cal). This case was called for trial in April, 1981 at Fresno, but the Court declared a mistrial upon grounds, among others, that a fair and impartial jury could not be obtained in the Fresno area. A new jury trial was set for Sacramento; before trial, all defendants except one settled. The trial was then rescheduled in Fresno; depositions were taken, after which plaintiffs dismissed the case voluntarily in March, 1982.

---

[1] In a civil case, before trial each party can examine the other's witnesses to discover their anticipated testimony; these proceedings are "depositions."

Background

Throughout the San Joaquin Valley in the Eastern District of California, canning tomatoes are harvested over a short season beginning in early July and ending about mid-October in each year. Canners, sometimes called processors, contact, purchase and pay for tomatoes for a variety of end-uses, e.g., canned tomatoes, catsup, canned tomato juice, tomato soup, spaghetti sauce, and tomato paste. During the harvest season the work goes on 24 hours a day, seven days a week. Tomatoes are picked up in the field by trucks and transported to processing plants. If the quantity of tomatoes delivered to a particular processor's plant on a given day exceeds the capacity of the plant, tomatoes may be delivered for processing to another canner's plant with open capacity at the time, and accounts are settled among the processors later on.

Until 1963, tomatoes in California were picked by hand by migrant workers. In 1963, mechanical harvesting equipment began to be used by commerical growers as the result of development of new tomato varieties suitable for harvesting by machine. Other and different varieties of tomatoes continue to be grown and picked by hand for sale as fresh tomatoes in stores. It is uncommon in the industry for a single grower to raise tomatoes for both the canning and fresh tomato market. Introduction of the mechanical harvester enabled growers to produce greater quantities of tomatoes per acre at lesser harvesting costs, but purchase of the equipment required growers to borrow money and incur substantial financing and other fixed costs. Canners purchasing tomatoes harvested by machine incurred significant changeover costs resulting from new requirements for handling tomatoes damaged by machine picking and the like. It is not likely that consumer demand for tomato products was affected or changed by the introduction of mechanical harvesting.

During the 1950's, some 57 canners purchased and processed California tomatoes, and by 1972 the number decreased to 28. Twenty-five of these were at first named as defendants in the action. Smaller canners were dismissed before trial, by plaintiffs voluntarily in some instances and by the Court in others, so that the case went to trial against remaining defendants who were the largest purchasers of processing tomatoes in the California market. The named plaintiffs were large growers of financial means, and each invested in mechanical harvesting equipment. In their complaints, plaintiffs alleged the litigation was a class action in which plaintiffs represented the claims and interests of all growers of California tomatoes. The Court denied motions to certify the litigation as a class action upon grounds, among others, that conflicts of interest among growers precluded plaintiffs from representing a class.

The purchaser market for California processing tomatoes includes, in addition to the defendant growers at trial and the smaller growers previously dismissed, several large member-owned cooperatives. In some instances cooperatives take delivery of tomatoes grown by members in return for nonmonetary scrip or points, and distribute funds to members from cash received from the harvest in proportion to the quantity supplied to the cooperative by the grower-member. In other instances, cooperatives buy tomatoes from members and other growers for cash like any other purchaser.

As with other fungible agricultural commodities, tomatoes available for sale to canners have tended to seek and find a prevailing market price during each growing year or season. The prevailing pattern of contract purchase of processing tomatoes by canners from growers is this: Large growers tend to sell their tomatoes to the same canner year after year, but not always. To preserve the soil, crops are rotated in the region from year to year among tomatoes, grains, alfalfa hay, sorghum, and, in the Fresno area, cotton and safflower. A grower does not grow tomatoes on the same land over a period of years; hence each grower may decide in a particular year to grow tomatoes or some other crop suitable to the region. After the harvest each year — at about Thanksgiving or Christmas — canners' field representatives discuss with growers how many acres of tomato production the canner is likely to want to purchase from the grower under contract and how many acres of tomatoes the grower is planning or willing to plant for the coming season. A few months later, usually in the January to March period, each canner informs each grower of the base price the canner proposes to pay for tomatoes harvested beginning in July. Sometimes, but not always, the canner will offer a premium price over the base price for purchase of late tomatoes harvested in September or October. This is done because the risk of bad weather makes growers more reluctant to plant tomatoes for late harvest than for maturity in earlier warm months. When agreement is reached on price among the canner and growers who have decided to sell tomatoes to the canner, written contracts are signed by the grower and canner under which the grower will sell the canner the tomatoes harvested from stated acreage at the contract price. The grower, having already prepared the soil, then plants his tomato crop. During some growing seasons, the contract price has been modified upward, on occasion retroactively. Generally, prevailing base prices have risen from about \$20 a ton in 1951 to over \$50 a ton in 1975, when the premium paid for late tomatoes was about \$5 a ton.

Advocates of "fair prices" for growers attempted during the 1960's to obtain, but did not obtain, a government marketing order which would have fixed the price canners paid growers for tomatoes by law. More recently, beginning in the mid-1960's, grower advocates formed the California Tomato Growers Association, called "CTGA," to present a united front to the canning industry in negotiating a price at which members would contract each year to sell their processing tomatoes. By about 1974, CTGA had assembled enough grower support to influence significantly, some say, and exercise control over, say others, the annual prevailing contract price for processing tomatoes in the region.

California Canners League, called "CCL," is a trade association formed many years ago to which most, but not all, of the defendant canners belonged. The purpose of CCL, and the nature of its activities, were in dispute. CCL and its members maintained that CCL exists and acts to deal with industry-wide problems of a technical nature unrelated to competitive behavior in the market. Plaintiffs' counsel contended that CCL has at least provided a fund of information concerning anticompetitive activities in the canning industry, and that CCL might be found by the jury to have provided some assistance in, or to have served as a vehicle for, illegal price fixing by canners of prices paid to growers for tomatoes.

Main points in the deposition testimony of plaintiffs' expert

An econometric model was used to show the existence of "price-leadership oligopsony" in the market for California canning tomatoes, on the basis of statistical tests of hypothesis.[1] The price leader may have been a single firm or a set of firms acting "as if" in collusion to set grower prices. The model, together with additional assumptions and calculations, was then used to estimate the "effective colluding share" of the market. This share turned out to be substantially larger than the market share of the largest single firm. The conclusion: there was a group of firms acting as if they were colluding to fix grower prices to maximize the joint profits of the colluders. Finally, the model was used to estimate what "free-market" prices would have been, absent this collusion. To establish some background for this argument, we present a brief review of two basic concepts in economics.

Supply and demand

Two basic concepts in economics are *supply* and *demand*.[2] These are terms of art, whose meaning differs somewhat from ordinary usage. "Supply," for example, refers not to a single quantity of some commodity, but to the relationship between the price and the total quantity producers will offer to the market for sale at that price. Such a relationship is represented by the *supply curve* in Figure 1 below. Price is shown on the horizontal axis, quantity on the vertical; for each price, the curve shows the total quantity that would be offered for sale in the market at that price. Notice

---

[1] An "oligopoly" is a small group of firms controlling the sales of a commodity, an "oligopsony" is a group controlling purchases. A "price leader" sets the price, taking into account the interests of the other firms in the group.

[2] A standard "principles of economics" text is Samuelson (1980). A more advanced reference is Mansfield (1982). A rigorous treatment of competitive markets is in Debreu (1959). A standard text on oligopoly theory is Scherer (1980).

that the curve slopes up: other things being equal, as price increases so does the quantity offered for sale.

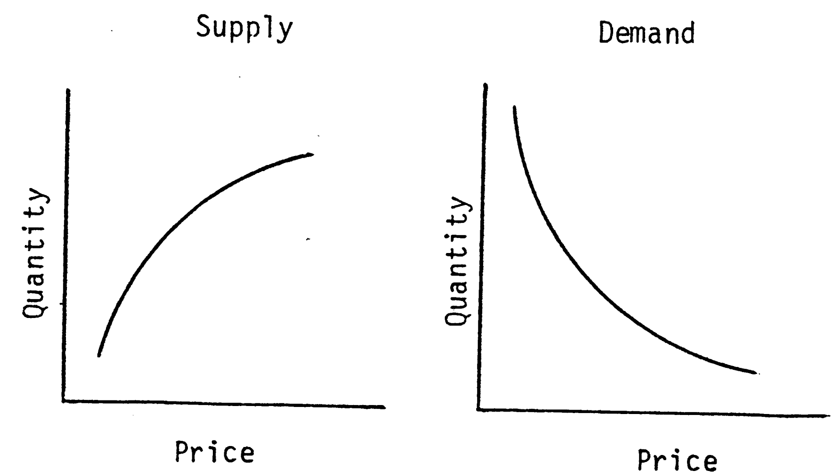

Fig. 1. Supply and Demand

"Demand" is similar. The *demand curve* in Figure 1 shows, for each price, the total quantity of the commodity that would be demanded at that price by buyers in the market. This curve slopes down; other things being equal, demand goes down as price goes up. As shown in Figure 2, the market price and the quantity transacted are considered to be determined by the intersection of the supply curve and the demand curve. At the intersection of the two curves, the market clears: supply equals demand. At a higher price, economists argue, the quantity sellers offer will exceed the quantity buyers want: this excess supply should drive the price down. Similarly, at a lower price, the quantity buyers want will exceed the quantity sellers offer: the excess demand should drive the price up.

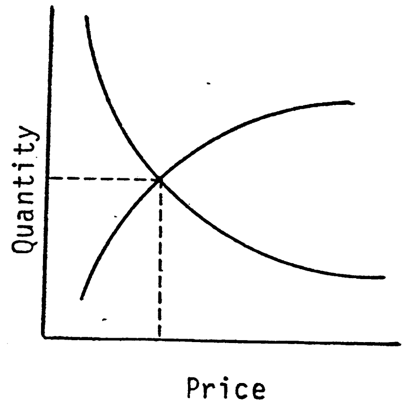

Fig. 2. Market clearing

The supply and demand curves express important behavioral assumptions: that there is a stable relationship between the quantity supplied and price; likewise for demand. This relationship is assumed, as economists say, *ceteris paribus*; other things being equal. That is, there are factors other than the price of the commodity which affect its supply: these are called *the determinants of supply*. For tomatoes, the price of fertilizer, the price of alternative crops, and agricultural labor rates may all be important determinants of supply. Likewise, there are *determinants of demand*, such as consumer income. In Figure 1, the determinants of supply and demand are held constant.

Of course, the curves are in a sense hypothetical constructs, for they represent quantities and prices in transactions which are not observed: only the market-clearing prices and quantities are observable. Furthermore, the curves in Figure 1 are so far qualitative rather than quantitative: no numbers appear on the axes. On the other hand, as will now be explained, by imposing additional assumptions, economists can make the curves quantitative and estimate them from actual transactions data.

In order to proceed, a crucial assumption must be made, as to the algebraic form of the curves in Figure 1. Economists refer to the choice of algebraic form as a *specification*. They may specify linear forms:

(1a) Supply $Q = a_0 + a_1 P$

(1b) Demand $Q = b_0 - b_1 P$

Or, they may specify log linear forms:

(2a) Supply $\log Q = a_0 + a_1 \log P$

(2b) Demand $\log Q = b_0 - b_1 \log P$

Still other choices are possible, and present economic theory is not sufficiently well developed to dictate which is the right choice in any given situation.

In these equations, Q represents quantity and P, price; a_1 and b_1, for example, are parameters, or constants of nature which govern the market. Usually the numerical values of the parameters are unknown, and must be estimated from data.

Plaintiffs' expert specified the linear form (1) for California canning tomatoes,[1] and we focus on that by way of example. It is customary to think of a succession of *market periods*, taken as calendar years in the tomato example. These are represented by the subscript t for "time." The primary data are the price P_t which obtained in year t, and the quantity Q_t transacted in that year. For the tomatoes, t ran from 1951 through 1975. Rewriting the equations,

---

[1] In Chern and Just (1980); in Chern and Just (1978), however, the specification was log linear.

(3a) Supply $Q_t = a_0 + a_1 p_t$

(3b) Demand $Q_t = b_0 - b_1 p_t$

We turn now to the determinants of supply and demand. First, take supply. The determinants of supply are usually considered to come in through the intercept a_0 of equation (3a). For example, suppose the only important determinant of tomato supply is the agricultural wage rate. Let W_t be this wage rate in year t . An economist may specify that

(4a) $$a_0 = \alpha_0 - \alpha_1 W_t$$

As W changes, the supply curve (3a) shifts up and down parallel to itself; the slope a_1 does not change. Likewise, consumer income may be the important determinant of demand for tomatoes, and in the framework of (3b) an economist may assume that

(4b) $$b_0 = \beta_0 + \beta_1 I_t$$

Here, I_t denotes income in period t.

Again, economic theory does not dictate specific choices for the major determinants of supply and demand. Even given these choices, equations like (4a) and (4b) are by no means logically inevitable. The right equations, for example, might turn out to be

(5a) $$a_0 = \alpha_0 - \alpha_1 W_t^2$$

(5b) $$b_0 = \beta_0 + \beta_1 \sqrt{I_t}$$

Still other choices are possible.

It is usually considered that many random phenomena will influence supply and demand; for example, the supply of an agricultural commodity will be influenced by weather. Likewise, many variables will have some small, indirect influence on supply and demand; for example, fluctuations in the stock market may have a marginal effect on consumer decisions to spend or save. The impact of random phenomena, and of omitted variables, is usually summarized by adding a "random disturbance term" to the equations: in effect, shifting the supply and demand curves up and down. These terms may be denoted by ε_t or δ_t. Our linear supply and demand equations now take the form

(6a) Supply $Q_t = \alpha_0 - \alpha_1 W_t + a_1 P_t + \varepsilon_t$

(6b) Demand $Q_t = \beta_0 + \beta_1 I_t - b_1 P_t + \delta_t$

At this point, it is convenient to introduce still another distinction, between *exogenous* and *endogenous* variables. In the equation above, W and I would be taken as exogenous: determined outside the system. However, Q and P are endogenous, determined within the system. Thus, the two simultaneous linear equations can be solved for the two endogenous variables Q and P in terms of W, I, ε and δ. A pair of equations like (6a-b) constitute an *econometric model*.

The equations above have six parameters: α_0, α_1, a_1, β_0, β_1, b_1. To estimate these from actual transactions data, statistical assumptions must be made about ε and δ. This completes the *specification* of the model, i.e.,

— Selection of the major determinants of supply and demand.

— Decisions as to the algebraic form of the equation.

— Assumptions about the statistical behavior of the disturbance terms, sufficient to enable the estimation of the parameters from the data.

The statistical assumptions might be that $(\varepsilon_t, \delta_t)$ are independent and identically distributed from year to year, but have some fixed, arbitrary covariance matrix within a year. To state this more vividly, imagine a box of tickets. Each ticket shows two numbers; the first is an ε, and the second a δ. The average of ε's is zero, and likewise for the δ's. For each year t, draw a ticket at random with replacement from the box; the first number is ε_t in (6a); the second is δ_t in (6b). The box stays the same from year to year, whatever the values of the "exogenous" variables I and W. Technically, exogenous variables are assumed to be statistically independent of the disturbances. Then the pair of simultaneous equations (6a-b) can be solved for Q_t and P_t, in terms of the exogenous variables and disturbances. To estimate the present system, at least one additional exogenous variable, or *instrument*, would be needed.

How can the parameters of a model be estimated from data? This is quite a complicated topic, but the idea is straightforward. Suppose that over time, the demand curve stays put while the supply curve shifts up and down. The market-clearing prices and quantities will then trace out the demand curve (see Figure 3). Conversely, if the supply curve stays put while the demand curve shifts up and down, the demand curve can be traced out. In general, both curves will be shifting simultaneously, and statistical estimation techniques must be used. Since Q and P are obtained by solving (6a-b), both involve ε and δ. This simultaneity is a crucial technical feature of econometric models: since P is correlated with ε and δ, ordinary least squares is inconsistent. To overcome this difficulty, econometricians have developed a technique called "two-stage least squares;" a reference is Theil (1971); or see Appendix C.

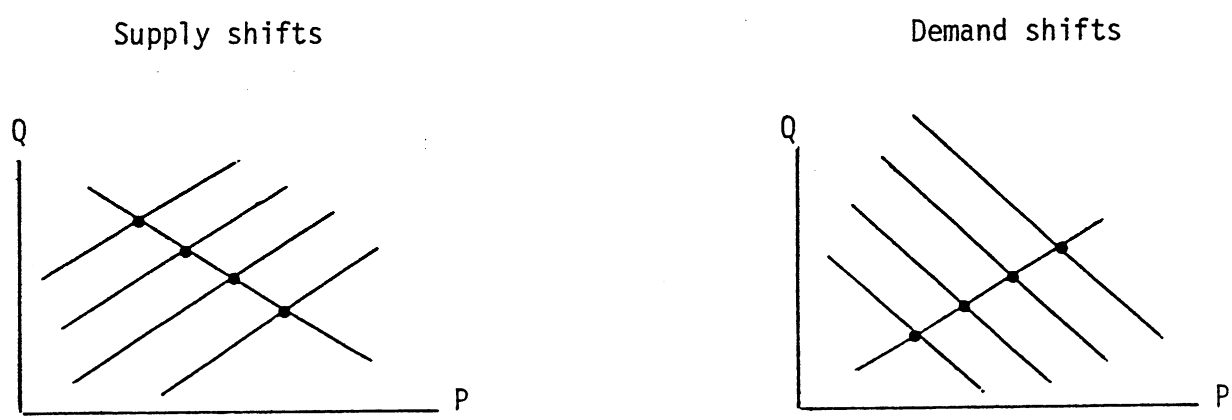

Fig. 3. Tracing out the demand and curves.

It should be noted that the specification of a model involves making a series of assumptions. These may be more or less reasonable. However, in its present state of development, the science of economics does not permit any exact, objective way of determining the major determinants of supply and demand, nor the algebraic form of the equation; nor is there usually any way to deduce the requisite statistical assumptions about the disturbance terms from generally accepted theory. And there are no generally accepted statistical procedures for testing the validity of an econometric model; some of the assumptions can indeed be examined, but only by introducing still other assumptions.

One powerful empirical test of a model is often available: use it to make predictions about the future, and see how accurate these turn out to be. Such tests are seldom made, and when made give conflicting results: see [9] or [13] and the references cited there, especially Christ (1975) and Zarnowitz (1979). The ability of a model to predict the future must be distinguished from its ability to reproduce the past. Given the flexibility described above in specifying a model, it is often possible by trial and error to get equations which reproduce the past very well indeed; such equations may be quite unsatisfactory when used to predict the future and may not capture the essential features of the market in question. For a more technical discussion, see [2] and [10].

The Model and the Proof of Market Power

Coming back to our case, plaintiffs' expert used an econometric model to conclude that the market behavior he examined was consistent with collusion among canners in buying processing tomatoes and inconsistent with competitive behavior in that market. Summarizing the previous discussions, the econometric model includes as hypothetical constructs a supply curve and a demand curve for processing tomatoes in the area studied. The supply curve represents the quantity of tomatoes which will be sold at a given price; the demand curve represents the quantity of tomatoes which will be purchased at a given price. The supply curve and demand curve thus represent relationships between quantities and prices. They are hypothetical constructs in the sense that they represent quantities of tomatoes which would be supplied or demanded at prices which have not been observed in actual transactions. The curves are estimated by statistical methods, however, from data in actual transactions.

In a market consisting of numerous tomato growers in competition with each other, introduction of machine harvesting must cause a shift in the supply curve in two respects, according to plaintiffs. First, at any given price the supply of tomatoes will be larger than it was before machine harvesting. Second, the tomato supply will be less responsive to any given change in price after the advent of machine harvesting than it was before. The explanation is that investment by growers in harvesting equipment made the grower less free to choose to grow crops other than tomatoes on his land while the harvesting machinery remained unused. In short, the grower is "locked in." This shift in the supply curve is illustrated in Figure 4.

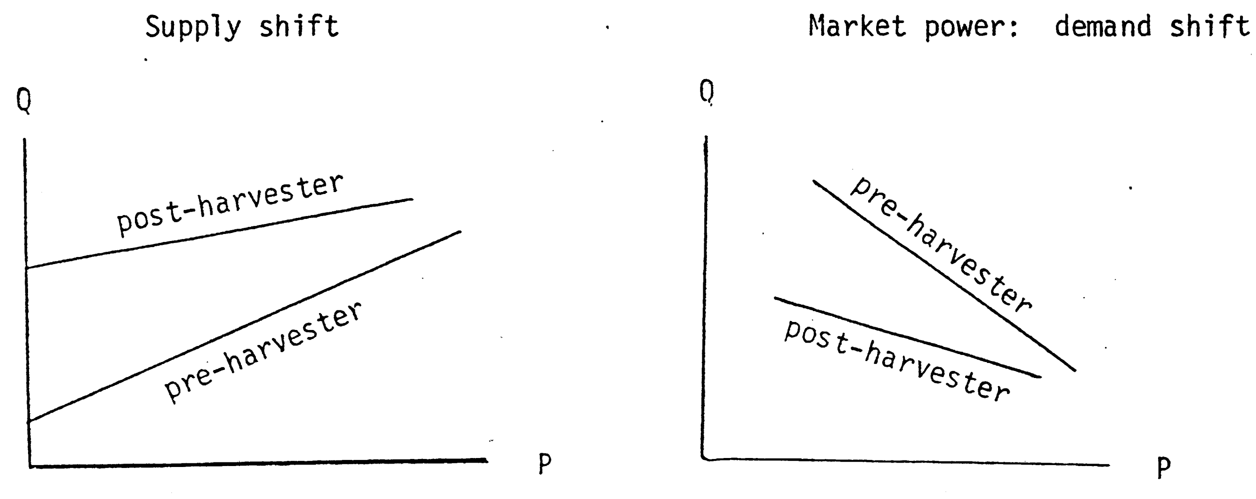

Fig. 4. The harvester shifts the supply curve. If the canners exercise market power, the perceived demand curve will shift when the supply curve does.

If the canners compete in buying tomatoes, a shift in the supply curve will not cause any shift in the demand curve. The reason is that demand for canning tomatoes is a function of consumer demand for the tomato end-products and is not determined by supply. On the other hand, if tomato prices are determined or tend to be determined by canners acting in collusion, after the introduction of machine harvesting the *perceived demand curve*, i.e., the demand curve facing the growers, will shift in two respects. First, at any given price, the demand for tomatoes will be less. Second, tomato demand will be less responsive to changes in prices than before machine harvesting (see Figure 4). Likewise, tomato demand will become less responsive to changes in the determinants of demand, such as income.

Perceived demand is a nonstandard construct. In brief, with a linear supply curve for canning tomatoes, and a linear demand curve by consumers for tomato endproducts, and canners exercising market power, there exists a linear "perceived demand curve," whose intersection with the supply curve gives the transacted price and quantity for canning tomatoes. The slope of this curve depends on the slope of consumer demand curve and on the slope of the supply curve; the intercept depends on the foregoing and on the intercept of the consumer demand curve; the intercept of the supply curve does not come in. See Appendix A for details.

From the data in Chern and Just (1978), plaintiffs estimated supply and demand curves for the years before and after introduction of machine harvesting, omitting the transition period 1964-66 when hand-picking and machine-harvesting were both used. Post-machine harvesting, the supply curve shifted to show supply was less responsive to changes in price, consistent with free competition among growers. On the other hand, the demand curve shifted to show that demand was less responsive to changes in price after machine harvesting than before, consistent with the existence of a "colluding share" of the canner market and inconsistent with tomato price behavior in a market having canner-purchasers acting in free competition with each other. Likewise for the determinant of demand, income. See Tables 1 and 2 on pp. 596-7 of Chern and Just (1980).

Weaknesses in the Argument

There are three major reasons for concluding that plaintiffs' statistical argument failed:

- The logic of the model is open to serious question.
- The statistical calculations are wrong.
- Even taking the empirical results at face value, they admit several plausible interpretations other than price-leadership oligopsony.

These reasons will now be considered in more detail.

With respect to the logic of the model, many assumptions are made without any empirical support. For example:

1) Competition among the growers.

2) Supply curve for processing tomatoes linear in price.

3) Demand curve for tomato end-products linear in price.

4) Constant unit processing costs.

5) Perfect knowledge by the colluding processors, of the supply and demand curves, and processing costs.

Assumption #1 is important, because if any significant number of the growers exercise market power, there is no supply curve to estimate: a monopoly or oligopoly does not have a supply curve.[1] Assumptions #2-3-4 are also critical, because they are needed to derive the "perceived demand curve," without which construct the model is vacuous. Assumption #5 is needed to do the profit maximization on behalf of the colluding processors.

Plaintiffs' expert admitted in deposition testimony to having made no study supporting these assumptions: indeed, he made no study of individual processors, and no study of demand at the consumer level. No arguments were given in support of the linearity assumptions, except that any smooth function can be approximated by a linear one: a function differentiable at a point can be approximated near that point by a line. In short, for small changes, linear supply and demand curves may be good enough. But the relevance of such observations to the issues at hand is not so clear.

---

[1] This may seem like a paradoxical assertion, but it is standard doctrine. The reason is that a monopolist sets price to maximize profit, and in the calculation uses the slope of the consumer demand curve. Hence, there is no stable relationship between the price set and the quantity offered. All depends on the demand curve. See p.285 of Mansfield (1982).

Plaintiffs used linear supply and demand curves over a 25-year period, during which California tomato production increased threefold, while grower prices doubled and consumer income went up by a factor of four.[1] These are not small changes.

What does it mean to say that e.g., the consumer demand curve is linear in price? Just this: that a 10¢ increase in the price of a can of tomatoes has the same impact on consumer demand in 1951 as in 1975, whether the can costs a quarter or a dollar, whether the consumer is making $5,000 a year or $15,000. A very strong assumption.

Two additional assumptions are needed to develop plaintiffs' theory of price-leadership oligopsony:

6) Segmented market for tomato end-products.

7) Fungible product within each segment.

Assumption #6 is easiest to explain in terms of a hypothetical example: that there are "red-label" and "blue-label" canned tomatoes, virtually identical but differentiated by the label and perhaps by advertising. The competitive canners, such as cooperatives, sell the "blue-label" product; the colluding canners sell the "red-label" product. The colluding canners fix the grower price to maximize joint profits; the competitive canners take the grower price as given.

No evidence is presented to support assumption #6, which is a nonstandard way to develop price-leader theory.[2] This unorthodox approach

---

[1] We are following plaintiffs here in using current rather than constant dollars, an issue to be discussed later.

[2] For the standard theory, see Mansfield (1982, pp. 349-50) or Scherer (1980).

is necessitated by assumption #4 above, of constant unit costs. It is also worth noting that plaintiffs treated the two markets independently, admitting no cross-elasticity of demand. To use the hypothetical example given above, consumer demand for red-label canned tomatoes depends only on the price of the product; it does not depend at all on the price of the blue-label product: again, this is against standard doctrine.

Assumption #7 is unrealistic: consumers do not drink a glass of chilled tomato paste before dinner; neither do they have a bowl of hot, nourishing tomato juice for lunch. On occasion, plaintiffs suggested that while canned tomatoes, juice, puree, catsup and paste may not be interchangeable, their quantities may be measured in terms of the weight of raw tomatoes needed for their manufacture. However, the cost of a tomato end-product does not depend solely on the weight of its raw tomato content. (With a No. 303 can of tomatoes, for example, the can costs more than the raw tomatoes.) In sum, this suggestion of plaintiffs' violates the idea that in a given market, a commodity has only one price. Specifically, the price of a ton-equivalent of the "commodity" will depend on how that quantity is distributed amongst the end-products.

We turn now to another set of issues. Plaintiffs began with an "analytical model" of the market for processing tomatoes. This considers only one geographical market for the tomatoes, and does not bring in determinants of supply and demand. A linear supply curve is assumed, of the form

$$Q = a_0 + a_1 P$$

Plaintiffs then introduced an econometric model which considers eight county markets for the tomatoes, and does bring in determinants of supply and demand. Thus, the post-harvest supply equation is assumed to be:

$$Q_{ct} = a_c + a_1 P_{ct} + a_2 Y_{ct} + \varepsilon_{ct}$$

In county c and year t,

Q_{ct} is the quantity of canning tomatoes sold in thousands of tons

P_{ct} is the price of canning tomatoes, in dollars per ton

Y_{ct} is the average yield over the previous three years, in tons per acre; an "expectations" variable

ε_{ct} is the disturbance term

a_c is a county-specific intercept

a_1 and a_2 are parameters, the same for all counties

Likewise for the demand equations.

The econometric model is at the level of raw tomatoes: the supply considered is by the growers; demand is by the canners. The following aspects of the econometric model must be considered:

1) The supply equation includes a lagged three-period moving average of yields as an "expectations variable," and the corresponding standard deviation as a "risk variable" in the pre-harvester period.

2) The pre-harvester period supply equation includes the price of only one alternative crop, viz., sorghum; post-harvester, no alternative crop prices are included.

3) The demand equation includes an end-product price index.

4) The analytical model applies to a single, integrated market for raw tomatoes; the econometric model applies to eight county markets.

Points #1 and 2 may be minor, but they are indicative of the arbitrariness in plaintiffs' approach. There is little justification for the choice of the expectations and risk variables. For example, why not use a four-period moving average, or the mean absolute deviation? There is also little justification for including sorghum rather than, say alfalfa or sugar-beets. To quote (Chern and Just, 1978, p. 19)

> For example, more than 100 crops are grown in San Joaquin County, and it is impossible to single out one or two crops as the most common alternative crops to processing tomatoes.

In his deposition, plaintiffs' expert argued at times that on the one hand the omitted crop prices are collinear with the sorghum price, so there is no point in putting them into the equation; and that on the other hand, including such prices gave rise to estimated coefficients with signs known à priori to be incorrect. These two points stand in contradiction to one another, and both seem wrong. First, the correlation between e.g. the price of sugar beets and the price of sorghum in the pre-harvester period (1951-63) is only about 0.32, so these two crop prices moved nearly independently of each other. Second, if sugar beets belong in the model, but putting them in gives incorrect signs, plaintiffs' expert concluded he should drop the beets; rethinking the model seems a better alternative.

Point #3 is a bit more serious. The development of the analytical model is aimed in part at eliminating the price P^d for end-products from the perceived demand curve for raw tomatoes: Chern and Just (1980), pp. 591-92); or see Appendix A. All that belongs in the demand equation are the "determinants of demand," i.e., the variables which affect consumer demand, other than price. There is only one such variable in the model: income. Putting product price back into the demand equation represents a theoretical inconsistency, and makes quite a difference in the results, as shown in Appendix B.

Point #4 is another major difficulty. The analytical model shows how processor demand for raw tomatoes is driven by consumer demand for tomato end-products. This consumer demand is national in scope; indeed, the major processors sell into a national distribution network. Plaintiffs should have indicated how this national consumer demand is to be parcelled out amont the eight counties, and failed to do so.

To see the tension between the one-market analytical model and the eight-market econometric model, consider this basic consequence of the existence of price-leadership oligopsony in the one-market analytical model for canning tomatoes. After the introduction of the machine harvester (Chern and Just 1980, p. 590):

> ... the perceived demand (the estimated econometric demand) would be shifted downward ...

That is to say, other things being equal, after the machine harvester comes in, at any given price the canners will demand fewer tomatoes from the growers. Now consider the estimated demand curve for canning tomatoes in Yolo county, according to the eight-county econometric model. (In 1975, Yolo was the largest producer among the eight counties considered in the econometric model.) In round numbers, the demand equations are:

pre-harvester (1951-63) $Q = -8 + .7\,I + 132\,R - 14\,P$

post-harvester (1967-75) $Q = 696 + .14\,I + 46\,R - 3.3\,P$

where

Q = tomato demand (thousands of tons)

P = grower price (dollars per ton)

R = product price index (dollars)

I = U.S. disposable personal income (billions of dollars)

To implement the idea of "other things being equal," choose in round numbers values for I and R typical for the whole period 1951-75 as follows: $I = 500$ and $R = 5$. The equations become

pre-harvester (1951-63) $\quad Q = 1002 - 14\ P$

post-harvester (1967-75) $\quad Q = 996 - 3.3\ P$

Demand in the analytical model.

Estimated demand in the econometric model; Yolo county.

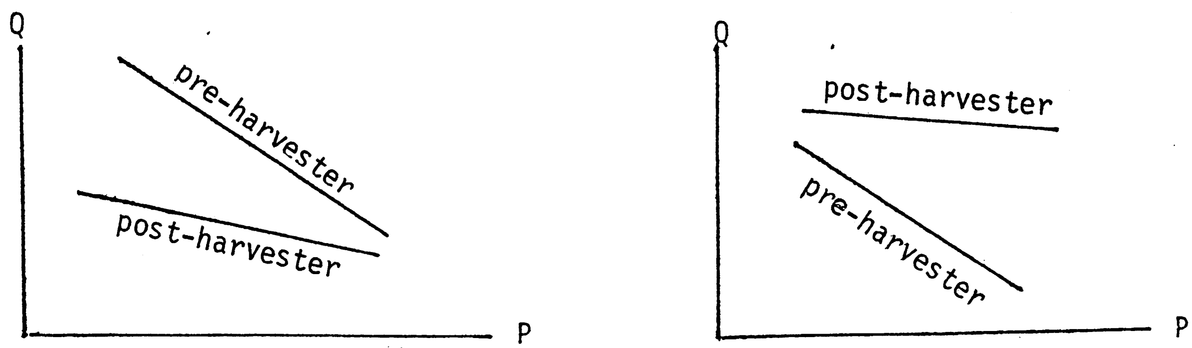

Fig. 5. The analytical model vs. the econometric model: the demand for canning tomatoes.

Historically, P ranges from about \$20 to \$60 a ton. Over this entire range, the post-harvester demand by processors for raw tomatoes exceeds the pre-harvester demand, at any given price. See Figure 5. In other words, one of the most basic conclusions of the single-market analytical model is contradicted by the findings of the eight-county econometric model. Only two resolutions of the difficulty seem possible: either the models are wrong, or no market power is being exercised by the canners. This completes our discussion of the first main reason for rejecting plaintiffs' argument: the logic of the model is open to serious question.

We turn now to the second reason: the statistical calculations are wrong. Plaintiffs' argument depends upon differences which are "statistically significant" between the estimates of the demand curves made separately for the periods before and after machine harvesting, by the t-test. For this it is necessary to compute standard errors. The formulas used by plaintiffs to compute standard errors are founded on assumptions shown to be false by a mere statement of them, for example, that weather patterns in neighboring counties are unrelated. (For quantitative estimates of the impact of such assumptions, see Appendix C.) Further, in contexts like the present one where the amount of data is limited relative to the number of parameters being estimated, the parameter estimates may be seriously biased; plaintiffs' statistical procedures do not take this bias into account (see Appendix C). Apart from such considerations, plaintiffs' admission that in order to avoid implausible results many variations of the econometric model were attempted before settling on the final version is more than sufficient to invalidate the statistical computations: see [2] and [10].

We turn now to the third main reason for rejecting plaintiffs' statistical argument. Even taking the empirical findings at face value, they admit of several plausible interpretations other than price-leadership oligopsony:

1) Inflation, which plaintiff did not take into account. If a dollar is worth less in 1975 than in 1951, a dollar increase in the price of a ton of tomatoes will matter less. Adjusting for inflation makes quite a difference in the empirical results: see Appendix B.

2) Plaintiffs excluded Fresno and Kern county tomato production from the eight-county econometric model; in 1961, these counties were not significant producers, but by 1975 they contributed 25% of California processing tomato production (Brandt et al., 1978). A very short version of the statistical argument offered by plaintiffs' expert is as follows: assuming competition, he can figure out how many tomatoes the canners ought to have wanted to buy, post-harvester, in the eight counties; he sees they bought less; he concludes they did not compete as buyers. The alternative: the canners bought the missing tomatoes in Fresno or Kern or for that matter, anywhere else. Extending this line of thought, a canner who commits himself to buying the bulk of his tomatoes in Fresno may well become less sensitive to variations in price or consumer demand when buying tomatoes in plaintiffs' eight counties; this would explain the change in coefficients reported by plaintiffs. So would an increase in processing costs arising from the introduction of the harvester.

3) Awareness or belief on the part of canners that growers, after 1963, were not only enabled to supply increased quantities of tomatoes to the market by their investment in harvesting machines but were impelled to do so. In short, seeing that the growers were "locked in" to tomato production, the canners bargained harder on price.

Plaintiffs' expert excluded the period 1964-66 as transitional. To rebut the sort of argument now made, in terms of alternative explanations, he conducted a "sensitivity analysis," moving the excluded period back and forth, and finding the most significant differences when this excluded period is close to the original one. Plainly, this argument does not meet point #3, which attributes the change in perceived demand to the introduction of the harvester. Nor does it meet point #2, for Fresno and Kern only became significant shippers of canning tomatoes at the time of the introduction of the harvester. It does not even cope with point #1 especially well, since the rate of inflation accelerated substantially, by coincidence, during the period of the introduction of the harvester. For example, as measured by the CPI consumer prices only increased by about 20% from 1951 to 1963; but they increased by about 60% from 1967 to 1975. Likewise, fuel prices and fertilizer prices increased rapidly even in real terms from 1973 through 1975, due to the Arab oil embargo.

Free Market Prices

To prove damages, plaintiffs used the model to estimate the size of the "effective colluding share," and to estimate what the free-market prices would have been, absent collusion by the canners. Even granting the validity of the model, these subsequent calculations are seriously flawed.[1]

To estimate the model, plaintiffs' expert disaggregated to the eight counties. To estimate free-market prices, he reaggregated back to the state level. However, there is nothing to stop us from using plaintiffs' equations to compute the "free market" prices in 1975 for Yolo and Sacramento counties, after the introduction of the harvester, assuming the "effective colluding share" was 80%. The exogenous variables are set in round

---

[1]The methodology used is not explained in Chern and Just (1978, 1980), but it is described in desposition testimony, summarized in Appendix A below. Plaintiffs' expert first estimated the effective colluding share, and then used the estimate as a datum in computing the prices; his estimate changed from time to time during the deposition, but 80% seems to be a representative figure.

numbers at typical values for 1975 as follows: I = 1000, R = 10, Y = 25. Large differences across county lines in the model free-market prices are almost inevitable, due to the county-specific intercepts. For details of our computation, see Appendices A-B.

The results are shown in Table 1 below. Comment may be superfluous, but Sacramento is adjacent to Yolo. Differences in hypothetical prices across the county line of $50 or more a ton are implausible. The actual difference in 1975 was only 25¢ a ton.

TABLE 1. The model's free market prices in two counties in 1975.

	Case 1	Case 2
Yolo	$126 per ton	$173 per ton
Sacramento	$ 74 per ton	$ 87 per ton

Case 1: the canners colluded in buying the tomatoes, but competed in selling end-products.[1]

Case 2: the canners colluded both in buying and in selling.

[1]The table is computed using plaintiffs' formulae, which include a parameter to differentiate between the two cases. However, it may be questioned whether plaintiffs have a coherent theory for case 1. On plaintiffs' assumptions, it can be argued that if the canners compete in sales, and fix the price of raw tomatoes, they can only fix it at what the free-market price would have been.

APPENDIX A

The Analytical Model

The object of this appendix is to give a concise statement of plaintiffs' analytical model for price-leadership oligopsony, to define perceived demand, and to indicate the methodology used in computing free-market prices. We follow Chern and Just (1980). The econometric model will be considered in Appendix B.

The supply equation for growers is

$$\text{(A1)} \qquad Q = a_0 + a_1 P_g$$

where Q is quantity (1000's of tons) and P_g is the grower price (dollars per ton).

There is a dominant firm or set of colluding firms, who sell into one segment of a market, and face the demand equation

$$\text{(A2)} \qquad Q = b_0 - b_1 P_r$$

where P_r is the price of the end-product, assumed fungible and measured in the same units as supply, viz., raw tomato content. These firms all have the same constant unit processing cost, θ. They set the grower price, P_g.

Next, there is a fringe of competitive firms, who take the grower price P_g as given, sell into another segment of the market, facing the demand equation

$$\text{(A3)} \qquad Q' = c_0 - c_1 P_r'$$

These competitive firms all have constant unit processing costs θ', so $P_r' = P_g + \theta'$. They will sell

$$\text{(A4)} \qquad Q' = (c_0 - c_1\theta') - c_1 P_g$$

and are allowed to buy this much from the growers.

The dominant share now faces the supply relation

(A5) $$Q = a_0 + a_1 P_g - Q' = \alpha_0 + \alpha_1 P_g$$

where

(A6) $$\alpha_0 = a_0 - c_0 + c_1\theta' \quad \text{and} \quad \alpha_1 = a_1 + c_1$$

Take the case where the colluding firms exercise market power in selling the end-product as well as buying the raw tomatoes. They set P_r and P_g to maximize joint profits

(A7) $$Q(P_r - P_g - \theta)$$

where Q is defined by (A5). They must also clear the market for canning tomatoes, so

(A8) $$\alpha_0 + \alpha_1 P_g = b_0 - b_1 P_r = Q$$

The optimal P_g can be shown to satisfy the equation

(A9) $$a_0 + a_1 P_g = d_0 - d_1 P_g$$

where

(A10) $$d_0 = (c_0 - c_1\theta') + \frac{\alpha_1}{2\alpha_1 + b_1}(b_0 - b_1\theta)$$

(A11) $$d_1 = c_1 + \frac{\alpha_1}{2\alpha_1 + b_1} b_1$$

The right-hand side of (A9) is the "perceived demand curve."

It is assumed that b_1, c_1, θ and θ' are constant over the period 1951-75; that $c_0 > c_1\theta'$ and $b_0 > b_1\theta$. If a_1, the slope of the supply curve, diminishes due to the introduction of the harvester, then $\alpha_1 = a_1 + c_1$ will diminish too, and so will d_1. Likewise for d_0. This is plaintiffs' argument relating the shift in post-harvester perceived demand to the exercise of market power in the canners.

We turn now to the model's free-market prices, following the deposition testimony of plaintiffs' expert. He assumes

$$\text{(A12)} \qquad \frac{b_0}{b_0 + c_0} = \frac{b_1}{b_1 + c_1} = \eta$$

No rationale was given for this strong assumption. The idea is that if the product price were P in both segments of the market, the colluding firms would have the market share

$$\frac{b_0 - b_1 P}{(b_0 - b_1 P) + (c_0 - c_1 P)} = \eta$$

Since the model is set up to allow different product prices in the two markets, it seems peculiar to interpret η as a market share. However, he calls η the "effective colluding share."

Let $\phi = (1-\eta)/\eta$, so $c_0 = \phi b_0$ and $c_1 = \phi b_1$. To avoid additional complications, we follow plaintiffs and set $\theta = \theta' = 0$; in essence, this redefines the intercepts b_0 and c_0. Recall (A6). Substitution into (A10-11) gives

(A13)
$$d_0 = \phi b_0 + \frac{b_0(a_1 + \phi b_1)}{2(a_1 + \phi b_1) + b_1}$$

(A14)
$$d_1 = \phi b_1 + \frac{b_1(a_1 + \phi b_1)}{2(a_1 + \phi b_1) + b_1}$$

Plaintiffs consider (A14) separately for the pre-harvester and post-harvester periods. The two demand price coefficients (d_1-before, d_1-after) and supply price coefficients (a_1-before, a_1-after) are estimated in the econometric model. This gives two equations in two unknowns, viz., ϕ and b_1: it is assumed that b_1 is constant over the entire period. Thus ϕ can be estimated, and then η; also b_1 is estimated. Now (A13) can be used to determine b_0 from d_0, the latter being estimated in the econometric model. In a competitive market with zero processing costs, the free-market price $P_r = P_g = P$ can now be determined by solving (A8):

(A15)
$$P = \frac{b_0 - \alpha_0}{b_1 + \alpha_1} = \frac{(1 + \phi)b_0 - a_0}{(1 + \phi)b_1 + a_1}$$

The empirical results will be presented after the econometric model is discussed in Appendix B.

APPENDIX B

The Econometric Model

The object in this appendix is to state briefly the econometric model in Chern and Just (1980), and show how the results depend on the specification. The free-market prices will also be computed.

The pre-harvester supply equation in the econometric model is

$$Q_{ct} = a_c + a_1 P_{ct} + a_2 Y_{ct} + a_3 D_{ct} + a_4 F_t + a_5 G_{t-1} + a_6 W_{ct} + \varepsilon_{ct} \tag{B1}$$

The post-harvester supply equation is

$$Q_{ct} = a_c + a_1 P_{ct} + a_2 Y_{ct} + \varepsilon_{ct} \tag{B2}$$

The demand equation for both periods is

$$Q_{ct} = d_c + d_1 P_{ct} + d_2 I_t + d_3 R_t + \delta_{ct} \tag{B3}$$

In county c and year t:

- Q_{ct} is the quantity transacted of processing tomatoes in 1000's of tons
- P_{ct} is the grower price in dollars per ton
- Y_{ct} is the average yield over the preceding three years in tons per acre
- D_{ct} is the SD of those three yields, in tons per acre
- W_{ct} is the agricultural wage rate, in dollars per ton
- F_t is the price of fertilizer, in dollars

G_t is the price of grain sorghum, in dollars per ton

I_t is the total U.S. personal disposable income, in billions of dollars

R_t is an index of tomato product prices in dollars

All money variables are in nominal dollars; a_c and d_c are county-specific intercepts; $a_1,\ldots,a_6$ and d_1,d_2,d_3 are parameters, the same across counties; G_t, I_t and R_t are the same for all counties.

ε_{ct} and δ_{ct} are stochastic disturbance terms, assumed to have mean 0, and to be independent and identically distributed across c and t, given the variables other than quantity and price. For each c and t, the pair $(\varepsilon_{ct}, \delta_{ct})$ has a fixed arbitrary 2 x 2 covariance matrix.

To explain more vividly the stochastic assumptions made by plaintiff, imagine a box of tickets; each ticket shows 2 numbers, the first being an ε and the second a δ; the ε's average out to zero, and so do the δ's. Focus on e.g., the pre-harvester period. For each year t in that period and county c, draw a ticket at random with replacement from the box; the first number on the ticket is ε_{ct} in (B1) and the second is δ_{ct} in (B3). The same box is used for each year, and each county, irrespective of the values of the exogenous variables (Y, D, F, A, W, I and R). Then equations (B1) and (B3) are solved for the two endogenous variables Q_{ct} and P_{ct} in terms of the exogenous variables and the disturbances.

In the present case, the stochastic assumption seem quite unrealistic. For example, these assumptions imply that each county shows the same random variation, big counties and small alike, in good years and bad; also, after adjusting for the exogenous variables, the remaining variations of supply and demand in one county are assumed to be uncorrelated with those in any other county.

TABLE B1. Plaintiffs' parameter estimates.

	Pre-harvester 1951-63		Post-harvester 1967-75	
	Supply	Demand	Supply	Demand
P	11.7	-14.2	3.9	-3.3
Y	9.9	--	15.4	--
D	-6.5	--	--	--
F	-2.1	--	--	--
G_{t-1}	-91	--	--	--
W	-335	--	--	--
I	--	.7	--	.14
R	--	132	--	46
County Intercepts				
San Joaquin	871	217	261	453
Yolo	625	-8	497	696
Solano	316	-319	-148	55
Sutter	295	-331	-147	67
Sacramento	378	-260	-294	-104
Stanislaus	300	-349	-307	-103
Santa Clara	274	-381	-325	-103
San Benito	227	-406	-354	-99

Coming back to the data-processing, we attempted to verify plaintiffs' estimates, and succeeded for the pre-harvester period. For the post-harvester period, as it turns out, there seems to have been one key-punch error made by plaintiffs in transferring data from Chern and Just (1978) to the computer. In 1973 in San Joaquin, Chern and Just (1978) give the price of grain sorghum as 2.87 dollars per ton; plaintiffs' computer data set (printed out as one of the deposition exhibits) has 1.43. When this error is corrected, the post-harvester demand price coefficient changes from -3.3 to -1.7, showing how sensitive the model must be to small errors in the data.[1]

We attempted to respecify the model by dropping the product price index R. In the pre-harvester equations, both price coefficients came in with the wrong sign: demand increases with price, and supply decreases. Post-harvester, demand and supply both increase with price. We also attempted to respecify the model, keeping R but deflating all monetary variables to 1967 dollars, using the CPI. Pre-harvester, the signs are right. Post-harvester, however, demand increases with price. In short, two attempts to correct misspecifications caused the model to produce inconsistent results. We did not attempt to adjust for population increase, which must also have a strong impact on demand.

As part of the data-processing, we fit a very simple dynamic model of our own:

(B4) $$Q_{ct} = a + b\, Q_{ct-1} + \zeta_{ct}$$

The estimates: $a = 14$ and $b = 1$. This simple model has only two parameters, it takes no advantage of data on price, income, yield, etc., etc. And it runs right through the transition from hand harvesting to machine harvesting. But it fits the data just about as well as plaintiffs' model. Our equation explains just over 90% of the variance; their equations explain between 87% and 94%.

We return now to the free-market prices, Case 2 in Table 1 above: $\eta = .8$ corresponds to $\phi = .25$; the post-harvester version of (A14)

[1]The corrected model is better for plaintiffs. The computer work described in this section was done by Dr. Thomas Permutt in SAS on the IBM 4341 at the University of California, Berkeley.

entails $b_1 = 5.78.$[1] These quantities are the same for all counties, by assumption. Then (A13) can be used to compute county-specific b_0's: the post-harvester econometric results from Table B1 give the left side of (A13):

$$d_0 = \text{county demand intercept} + .14\ I + 46\ R$$

Now (A15) can be used to get the county-specific free-market price P, with a_0 from the post-harvester results of Table B1:

$$a_0 = \text{county supply intercept} + 15.4\ Y$$

Case 1 is similar, and proceeds from the assumption that the dominant firms set the product price equal to marginal cost for the entire coalition, i.e., taking into account price reaction by growers to increased demand. Plaintiffs described this as the case where the canners in the coalition compete with each other in sales — a description whose merit is not apparent.

[1]Pre-harvester, $b_1 = 26.18$, casting some doubt on the constancy of b_1, over the period 1951-1975.

APPENDIX C

Two-Stage Least Squares and the Bootstrap

The object of this section is to present a brief account of the estimation procedures used by plaintiffs, known as two-stage least squares; and then some details on the bootstrap, a statistical procedure used to test plaintiffs' calculations.[1] We begin with the more basic generalized least squares. Consider the regression model

(C1) $$\underset{n\times 1}{Y} = \underset{n\times p}{X}\;\underset{p\times 1}{\beta} + \underset{n\times 1}{\epsilon}, \quad E(\epsilon) = 0, \quad \operatorname{cov} \epsilon = \Sigma$$

For historical reasons, X is called "the design matrix." With Σ known, the generalized least squares (gls) estimate is

(C2) $$\hat{\beta}_{gls} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$$

As is standard,

(C3) $$E(\hat{\beta}_{gls}) = \beta$$

(C4) $$\operatorname{cov}(\hat{\beta}_{gls}) = (X^T\Sigma^{-1}X)^{-1}$$

When Σ is unknown, statisticians often use (C2) and (C4) with Σ replaced by some estimate $\hat{\Sigma}$. Iterative procedures may be used, as follows. Let $\hat{\beta}^{(0)}$ be some initial estimate for β, typically from a preliminary ordinary least squares (ols) fit. There are residuals $\hat{e}^{(0)} = Y - X\hat{\beta}^{(0)}$. Suppose the procedure has been defined through stage k, with residuals

$$\hat{e}^{(k)} = Y - X\hat{\beta}^{(k)}_{gls}$$

[1]A standard reference on two-stage least squares is Theil (1971). For a more extended discussion of the bootstrap, see [11] and [12]. This section is joint work with Dr. Stephen Peters, Center for Computational Research in Economics and Management Science, MIT.

Let $\hat{\Sigma}_k$ be an estimator for Σ, based on $\hat{e}^{(k)}$: an example will be given below assuming a block diagonal structure for Σ. Then

$$\text{(C5)} \qquad \hat{\beta}_{gls}^{(k+1)} = (X^T \hat{\Sigma}_k^{-1} X)^{-1} X^T \hat{\Sigma}_k^{-1} Y$$

This procedure can be continued for a fixed number of steps, or until $\hat{\beta}_{gls}^{(k)}$ settles down. A convexity argument shows that $\hat{\beta}_{gls}^{(k)}$ converges to the maximum likelihood estimate for β, assuming ϵ is independent of X and multivariate gaussian with mean 0.

The covariance matrix for $\hat{\beta}_{gls}^{(k+1)}$ is usually estimated from (C4), with $\hat{\Sigma}_k$ put in for Σ:

$$\text{(C6)} \qquad \widehat{cov}^{(k+1)} = (X^T \hat{\Sigma}_k^{-1} X)^{-1}$$

This may be legitimate, asymptotically. In finite-sample situations, all depends on whether $\hat{\Sigma}_k$ is a good estimate for Σ or not. If $\hat{\Sigma}_k$ is a poor estimate for Σ, the standard errors estimated from (C6) may prove to be unduly optimistic: an example is given in [11]. Unfortunately, approximate *gls* estimators are often used when there is too little data to offer any hope of estimating Σ with reasonable accuracy. In such circumstances, the bootstrap is a useful diagnostic, and in many cases it gives a more realistic estimate of the standard errors.

To ease notation, $\hat{\beta}_{gls}^{(k)}$ will be referred to as the (gls,k)-estimator. The paper [11] only considers the $(gls,1)$-estimator, which in many situations has full asymptotic efficiency. In some examples, further iteration seems to make the coefficient estimates better, but also exaggerates the optimism of the standard error estimates. In other examples, iteration actually makes the coefficient estimators worse.

In econometric work, it is usual to constrain β to fall in some linear space Λ; these are the "identifying restrictions." Typically, many elements of β are constrained to vanish, and some are constrained to equal each other. The constraints are often incorporated by re-expressing the model in terms of linearly independent parameters: this involves linear manipulations of the columns of the design matrix X. A more elegant solution is to make unconstrained estimates, as indicated above; and then to project the unconstrained estimator $\hat{\beta}_u = \hat{\beta}^{(k)}_{gls}$, say, into the constraint space Λ. However, the projection must take into account the covariance structure of $\hat{\beta}_u$, i.e., the constrained estimator $\hat{\beta}_c$ is the element of Λ minimizing the distance

(C7) $$(\hat{\beta}_u - \hat{\beta}_c)^T \hat{W}^{-1} (\hat{\beta}_u - \hat{\beta}_c)$$

where $\hat{W}$ is the estimated covariance matrix of $\hat{\beta}_u$. Thus, the constrained gls estimator is found by two applications of unconstrained gls.

Changing the meaning of n, p and X, consider next an econometric model of the form

(C8) $$\underset{1\times q}{Y_i} = \underset{1\times q}{Y_i}\ \underset{q\times q}{A} + \underset{1\times p}{X_i}\ \underset{p\times p}{B} + \underset{1\times q}{\epsilon_i}$$

In this equation, A and B are coefficient matrices of unknown parameters, to be estimated from the data, subject to identifying restrictions; Y_i is the vector of endogenous variables at data point i; X_i is the vector of exogenous variables; and ϵ_i is the vector of disturbances. This is a system of q equations in q endogenous variables; there are p exogenous variables. The matrix $I - A$ is assumed invertible. To be more specific about the stochastic assumptions:

(C9a) The ϵ_{i} are independent and identically distributed

(C9b) $E(\epsilon_{i}) = 0$ and $\operatorname{cov} \epsilon_{i} = V$ for all i

(C9c) The X's are independent of the ϵ's

Here, V is a $q \times q$ positive definite matrix. As is conventional, we normalize A so $A_{jj} = 0$ for all j. Write Y_{ij} for the j^{th} component of the row vector Y_{i}. Then the j^{th} equation in the system explains Y_{ij} in terms of the other endogenous variables and the exogenous variables:

$$Y_{ij} = Y_{i} A_{j} + X_{i} B_{j} + \epsilon_{ij}$$

where A_{j} is the j^{th} column of A and B_{j} is the j^{th} column of B.

In e.g. the pre-harvester tomato model, i corresponds to the pairs ct of counties and years, so there are $n = 8 \times 13 = 104$ data points; $q = 2$, because there are 2 endogenous variables, price and quantity; and $p = 15$ because there are 15 exogenous variables, viz.,

8 county dummy variables

yield	Y
dispersion	D
fertilizer	F
grower price	G
wage rate	W
income	I
retail price	R

Coming now to two-stage least squares, by conditioning on the exogenous X's, we may suppose them constant: see (C9c). Multiply (C8) by $X_{i.}^T$ and sum:

$$\text{(C10)} \qquad \underset{p\times q}{R} = \underset{p\times q}{R}\ \underset{q\times q}{A} + \underset{p\times p}{S}\ \underset{p\times q}{B} + \underset{p\times q}{\Delta}$$

where

$$\text{(D11)} \qquad R = \sum_{i=1}^{n} X_{i.}^T Y_{i.} \qquad S = \sum_{i=1}^{n} X_{i.}^T X_{i.} \qquad \Delta = \sum_{i=1}^{n} X_{i.}^T \epsilon_{i.}$$

Notice that the j^{th} column of (C10) corresponds to the j^{th} equation in (C8).

In applications, $[A, B]$ is constrained to fall in some linear space Λ of dimension at most pq. Then A and B can be estimated from (C10) by some variant of least squares. Notice that S is constant (non-random), since X is. It is conventional to treat R on the right side of (C10) as constant. This may be legitimate asymptotically, but is false in any finite sample. In fact, R is not only random but also correlated with Δ, and this is the source of "small-sample bias" in 2SLS estimators. In the tomato model, this small-sample bias is serious, as will be seen below.

The matrix of errors Δ on the right hand side of (C10) does have some covariance structure, so generalized least squares is the procedure of choice. To make contact with the standard format of (C1), we stack the columns in (C10): column #1 on top of column #2,...,on top of column #q. In the stack, information corresponding to the first equation in model (C8) comes first, information about the last equation comes last.

The parameter vector β in (C1) is obtained by stacking A and B: column #1 of A, followed by column #1 of B,..., followed by column #q of A, followed by column #q of B. The design matrix is obtained by writing R and S down the diagonal, and padding with zeroes.

At this point, the design matrix is highly singular, having dimension $pq \times (pq + q^2)$. Usually the elements of β known *a priori* to vanish are suppressed, and the design matrix is adjusted accordingly by deleting the corresponding columns. An alternative approach is to use generalized inverses. In any event, the left hand side Y vector in (C1) consists of the stacked R matrix; the error vector is the stacked Δ matrix. The full system of equations (C10) is layed out in stacked form as follows, with $M_{\underline{j}}$ being the j^{th} column of any matrix M:

$$\text{(C12)} \qquad \begin{bmatrix} R_1 \\ \vdots \\ R_q \end{bmatrix} = \begin{bmatrix} R & S & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & R & S & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & R & S \end{bmatrix} \begin{bmatrix} A_1 \\ B_1 \\ \vdots \\ A_q \\ B_q \end{bmatrix} + \begin{bmatrix} \Delta_1 \\ \vdots \\ \Delta_q \end{bmatrix}$$

By (C9), the covariance matrix of the error vector (the stacked Δ matrix) is the Kronecker product

$$\text{(C13)} \qquad \Sigma = V \otimes S = \begin{bmatrix} V_{11}S & V_{12}S & \dots & V_{1q}S \\ V_{21}S & V_{22}S & \dots & V_{2q}S \\ \vdots & \vdots & & \vdots \\ V_{q1}S & V_{q2}S & \dots & V_{qq}S \end{bmatrix}$$

Pretending R is constant, the system (C10) should be estimated by generalized least squares, relative to $V \otimes S$. Of course, V is unknown and must be estimated from the data. Coming now to two-stage least squares, consider each column of (C10) in isolation. Take column j, for instance:

(C14) $$\underset{p\times 1}{R_{\cdot j}} = \underset{p\times q}{R}\ \underset{q\times 1}{A_{\cdot j}} + \underset{p\times p}{S}\ \underset{p\times 1}{B_{\cdot j}} + \underset{p\times 1}{\Delta_{\cdot j}}$$

Now $\operatorname{cov}\Delta_{\cdot j} = V_{jj}S$; the constant of proportionality V_{jj} is immaterial, so (C14) can be estimated by generalized least squares, treating R on the right as a constant. Constraints specific to the j^{th} equation would be imposed, but not the cross-equation constraints. In the tomato model, the only constraints set appropriate coefficients to zero; e.g., in the pre-harvester supply equation, income I does not come in; that coefficient is set to zero.

Let $\hat{A}$ and $\hat{B}$ denote the two-stage least squares estimates. Let

(C15) $$\hat{\epsilon}_{i\cdot} = Y_{i\cdot} - Y_{i\cdot}\hat{A} - X_{i\cdot}\hat{B}$$

(C16) $$\hat{V} = \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_{i\cdot}^{T}\hat{\epsilon}_{i\cdot}$$

The covariance matrix of $[\hat{A},\hat{B}]$ can be estimated from (C6), with $\hat{V}$ used to estimate V in (C13) and get $\hat{\Sigma}$. It is conventional, for the purpose of estimating $\operatorname{cov}[\hat{A},\hat{B}]$ only, to inflate $\hat{V}_{jj}$ by $n/n - r$, where r is the number of variables actually coming into the j^{th} equation. Call this estimated covariance matrix $\widehat{\operatorname{cov}}$.

Turn now to the bootstrap. Let $\hat{A}$ and $\hat{B}$ be the two-stage least squares estimates for A and B in (C8). It will be assumed that $I-\hat{A}$ is invertible. The residuals are defined by (C15). These are estimates for the true disturbances in the model (C8). Let μ be the empirical distribution of the residuals, assigning mass $1/n$ to each of $\hat{\epsilon}_1,\ldots,\hat{\epsilon}_n$.

Consider next a model like (C8), but where all the ingredients are known:

- Set the coefficients at $\hat{A}$ and $\hat{B}$ respectively.
- Make the disturbance terms independent, with common distribution μ.

The exogenous X's are kept fixed. Using this simulation model, pseudo-data can be generated for $i = 1,\ldots,n$. These pseudo-data will be denoted by stars:

(C17) $$Y_i^* = (X_i\hat{B} + \epsilon_i^*)(I - \hat{A})^{-1}$$

where the ϵ^*'s are independent with the common distribution μ.

Now pretend the pseudo-data (C17) come from a model like (C8), with unknown coefficient matrices. Since there are county-specific intercepts, assumption (C9a) is satisfied. Estimate these coefficients by two-stage least squares from the pseudo-data; denote the estimates by $\hat{A}^*$ and $\hat{B}^*$. Also compute from the starred data the estimated covariance matrix $\widehat{\text{cov}}^*$, using (C15-16) applied to the starred data.

The distribution of the pseudo-errors

$$\hat{A}^* - \hat{A}, \quad \hat{B}^* - \hat{B}$$

can be computed, and used to approximate the distribution of the real errors

$$\hat{A} - A, \quad \hat{B} - B$$

This approximation is the bootstrap. It is emphasized that the calculation assumes the validity of the model (C8). The distribution of the pseudo-errors can be computed, e.g., by Monte Carlo, simply repeating the procedure some number of times and seeing what happens. The distribution of the pseudo-errors is of interest only as an approximation to the distribution of the real errors.

The next object is to test plaintiffs' statistical calculations using the bootstrap methodology. Four sets of tables are presented; each of the four tables has in turn four parts:

i) pre-harvester supply

ii) pre-harvester demand

iii) post-harvester supply

iv) post-harvester demand

The tables correspond to different sets of assumptions about the stochastic disturbance terms ε and δ in plaintiffs' model, as specified in equations (B1-3). In all cases, it is assumed that the 16-vectors

$$(\varepsilon_{ct}, \delta_{ct}: \; c \text{ varies over 8 counties})$$

are independent and identically distributed in time t, with mean zero. Furthermore, the pre- and post-harvester disturbance terms are assumed independent.

Case 1. The 16-vectors of disturbances have an arbitrary 16×16 covariance matrix across equations and counties.

Case 2. The disturbance terms are independent across equations; within an equation, the 8-vectors of disturbances have an arbitrary 8×8 covariance matrix across counties.

Case 3. The pairs $(\varepsilon_{ct}, \delta_{ct})$ are independent and identically distributed, as c and t vary; but have an arbitrary 2×2 covariance matrix across equations.

Case 4. The disturbance terms are independent across equations; within an equation, they are independent and identically distributed.

A priori, Case 1 is the most plausible. Implicitly, Case 3 is assumed in Chern and Just (1980); and the discussion of the bootstrap presented above used Case 3. Case 4 is more special than Case 3, so plaintiffs' analysis must cover this case too. To apply the bootstrap in the other cases, it is only necessary to modify appropriately the error distribution for the ε^*.

The main conclusions from the bootstrap analysis are as follows:

- In Cases 2 and 4, there is serious "small-sample bias" in the coefficient estimates. This bias distorts the statistics in plaintiffs' favor. The bias can be seen by comparing the "parameter" column 1 with the "mean" column 3 for e.g. post-harvester demand price in Tables C2 and C4.

- In Cases 1 and 2, the conventional asymptotic formulae for standard errors are off by as much as 40%, in either direction, presumably due to the covariance across counties and to the difference in variances between counties, which are ignored in plaintiffs' analysis.[1] This can be seen by comparing the "SD" column 4 with the "RMS Nominal SE" column 5 in Tables C1 and C2, especially for e.g. pre-harvester supply price and post-harvester demand price. A downward distortion of the SE's helps the plaintiffs; an upward distortion hurts.

Columns 1 and 2 in the tables show the results obtained by fitting plaintiffs' model to plaintiffs' data, with one data error corrected (see Appendix B). Pre-harvester, the results agree quite closely with those reported in Table B1; the small differences are probably the result of round-off error, since the work was done on two different computers using

[1] In deposition, plaintiffs' expert made an ad hoc adjustment in certain exhibits for these factors. This adjustment is highly suspect, for the reasons given in [11].

two different software packages. Post-harvester, the results differ appreciably from Table B1, as discussed in Appendix B.

In the tables, column 1 shows the parameter estimates; column 2, the standard errors from the conventional asymptotic formulae. The columns are the same in all four tables. Columns 3-4-5 give the results of a "bootstrap" simulation experiment. In this experiment, plaintiffs' model (B1-3) is taken as true, with parameter values as given in column 1 of the tables. The disturbances terms obey the conditions laid out above, e.g., Table C1 corresponds to Case 1. On the computer, the disturbance terms are generated by resampling the residuals from the fit reported in column 1. This is done 100 times, to generate 100 simulated data sets, according to plaintiffs' theory as expressed in equations (B1-3). The exogenous variables are held fixed, and the equations are solved for the endogenous price and quantity variables. For each simulated data set, we use plaintiffs' fitting procedure to "estimate" the parameters in the simulation model, and to compute the conventional approximations to the standard errors. The mean of the 100 parameter estimates $[\hat{A}^*, \hat{B}^*]$ is shown in the "mean" column 3; the standard deviation, in the "SD" column 4. The root-mean-square of the 100 conventional standard errors is shown in the "RMS Nominal SE" column 5. This is the square root of the mean of the diagonal entries in $\widehat{\text{cov}}^*$ above.

In Table C4, for example, the post-harvester demand price coefficient is set in the simulation model as -1.68; however, in the 100 simulated data sets, the mean of the estimates done by plaintiffs' method is 0.32, so there is an upward bias of 1.68 + 0.32 = 2. Plaintiffs' method for estimating parameters is strongly biased.

Likewise, in Table C1, the standard deviation of the 100 estimates for the pre-harvester supply price is 6.58; this is a true measure of the variability in the estimates, in the simulation model. However, the root-mean-square of the 100 conventional estimates of standard error is only 4.01. Plaintiffs' method for approximating the standard error is biased downward by

$$1 - \frac{4.01}{6.58} \approx 39\%$$

Table C1. 100 Bootstrap Replications Resampling 16-Vectors

	Pre-harvester supply				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	11.6	4.11	12.5	6.58	4.01
Y	9.69	5.09	10.1	6.11	4.77
D	- 6.62	11.2	- 6.85	14.4	10.4
F	- 2.10	2.28	- 1.90	3.59	2.03
G	- 91.9	45.3	- 91.9	73.5	42.2
W	- 328.	154.	- 334.	229.	162.
San Joa	871.	335.	832.	449.	272.
Yolo	625.	336.	586.	450.	273.
Solano	316.	336.	274.	453.	274.
Sutter	295.	336.	252.	454.	273.
Sacto	378.	335.	336.	454.	272.
Stanis	300.	332.	260.	447.	270.
Santa C	273.	337.	231.	455.	274.
San Ben	227.	338.	183.	452.	275.

	Pre-harvester demand				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	- 14.3	5.71	- 13.7	4.52	4.46
I	.700	.115	.717	.129	.118
R	132.	32.1	129.	28.4	25.2
San Joa	217.	62.2	212.	74.3	61.4
Yolo	- 8.36	62.1	- 11.2	75.2	61.3
Solano	- 319.	62.2	- 325.	72.1	61.4
Sutter	- 332.	62.0	- 339.	72.3	61.2
Sacto	- 260.	62.2	- 268.	70.8	61.4
Stanis	- 349.	61.9	- 355.	72.4	61.2
Santa C	- 381.	62.4	- 387.	73.0	61.6
San Ben	- 406.	62.3	- 412.	73.4	61.4

Table C1. 100 Bootstrap Replications Resampling 16-Vectors (cont'd)

	Post-harvester supply				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	4.10	1.48	3.82	1.04	1.13
Y	14.9	8.08	14.1	6.51	6.88
San Joa	263.	161.	282.	153.	145.
Yolo	500.	165.	538.	145.	149.
Solano	- 145.	167.	- 116.	141.	150.
Sutter	- 144.	172.	- 113.	146.	154.
Sacto	- 291.	160.	- 266.	142.	144.
Stanis	- 304.	164.	- 277.	146.	148.
Santa C	- 321.	174.	- 295.	152.	156.
San Ben	- 349.	190.	- 319.	164.	170.

	Post-harvester demand				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	- 1.68	3.98	- 1.46	2.21	3.28
I	.156	.146	.155	.115	.142
R	35.2	28.5	32.7	18.7	24.4
San Joa	445.	80.2	444.	77.5	79.5
Yolo	689.	80.0	707.	83.5	79.7
Solano	48.4	80.1	58.1	56.9	79.8
Sutter	60.4	80.2	70.7	60.1	79.8
Sacto	- 111.	80.1	- 105.	62.4	79.7
Stanis	- 111.	80.7	- 104.	59.0	79.5
Santa C	- 111.	80.6	- 105.	60.1	79.3
San Ben	- 107.	80.5	- 99.5	58.2	79.6

Table C2. 100 Bootstrap Replications Resampling 8-Vectors

	Pre-harvester supply				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	11.6	4.11	11.0	5.27	3.77
Y	9.69	5.09	9.42	7.00	4.80
D	- 6.62	11.2	- 7.42	13.2	10.5
F	- 2.10	2.28	- 2.14	3.26	2.00
G	- 91.9	45.3	- 89.7	69.2	42.4
W	- 328.	154.	- 325.	184.	155.
San Joa	871.	335.	890.	483.	280.
Yolo	625.	336.	644.	480.	280.
Solano	316.	336.	334.	478.	280.
Sutter	295.	336.	312.	479.	280.
Sacto	378.	335.	395.	477.	279.
Stanis	300.	332.	318.	471.	276.
Santa C	273.	337.	289.	479.	281.
San Ben	227.	338.	244.	479.	282.

	Pre-harvester demand				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	- 14.3	5.71	- 11.8	3.59	3.83
I	.700	.115	.658	.108	.108
R	132.	32.1	120.	22.0	22.6
San Joa	217.	62.2	226.	63.1	59.2
Yolo	- 8.36	62.1	- .747	64.0	59.3
Solano	- 319.	62.2	- 312.	60.5	58.7
Sutter	- 332.	62.0	- 325.	60.4	58.4
Sacto	- 260.	62.2	- 254.	59.9	58.7
Stanis	- 349.	61.9	- 340.	61.1	58.5
Santa C	- 381.	62.4	- 374.	61.7	58.8
San Ben	- 406	62.3	- 399.	61.1	58.6

Table C2. 100 Bootstrap Replications Resampling 8-Vectors (cont'd)

	Post-harvester supply				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	4.10	1.48	3.34	.765	.920
Y	14.9	8.08	17.5	6.71	6.06
San Joa	263.	161.	227.	148.	130.
Yolo	500.	165.	484.	170.	135.
Solano	- 145.	167.	- 172.	153.	135.
Sutter	- 144.	172.	- 172.	160.	139.
Sacto	- 291.	160.	- 322.	143.	129.
Stanis	- 304.	164.	- 334.	149.	133.
Santa C	- 321.	174.	- 354.	154.	140.
San Ben	- 349.	190.	- 386.	171.	153.

	Post-harvester demand				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	- 1.68	3.98	.223	1.57	2.01
I	.156	.146	.156	.0949	.115
R	35.2	28.5	23.4	14.6	18.5
San Joa	445.	80.2	438.	57.8	67.6
Yolo	689.	80.0	713.	76.1	74.5
Solano	48.4	80.1	57.9	50.0	66.6
Sutter	60.4	80.2	70.0	52.0	67.0
Sacto	- 111.	80.1	- 110.	53.7	66.3
Stanis	- 111.	80.7	- 108.	49.7	65.9
Santa C	- 111.	80.6	- 108.	51.0	65.9
San Ben	- 107.	80.5	- 102.	50.8	65.8

Table C3. 100 Bootstrap Replications Resampling 2-Vectors

	Pre-harvester supply				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	11.6	4.11	12.2	3.84	3.60
Y	9.69	5.09	10.3	4.66	4.70
D	- 6.62	11.2	- 6.75	10.6	10.4
F	- 2.10	2.28	- 2.08	1.84	1.95
G	- 91.9	45.3	- 89.1	40.9	41.8
W	- 328.	154.	- 331.	167.	156.
San Joa	871.	335.	844.	250.	264.
Yolo	625.	336.	596.	250.	265.
Solano	316.	336.	288.	248.	265.
Sutter	295.	336.	269.	249.	264.
Sacto	378.	335.	354.	248.	264.
Stanis	300.	332.	272.	246.	262.
Santa C	273.	337.	246.	251.	266.
San Ben	227.	338.	198.	251.	267.

	Pre-harvester demand				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	- 14.3	5.71	- 13.9	4.77	4.51
I	.700	.115	.686	.104	.112
R	132.	32.1	130.	27.9	25.2
San Joa	217.	62.2	221.	58.8	58.5
Yolo	- 8.36	62.1	- 5.42	57.4	58.4
Solano	- 319.	62.2	- 316.	55.0	58.5
Sutter	- 332.	62.0	- 328.	58.5	58.4
Sacto	- 260.	62.2	- 256.	58.1	58.6
Stanis	- 349.	61.9	- 347.	59.3	58.3
Santa C	- 381.	62.4	- 379.	58.7	58.7
San Ben	- 406.	62.3	- 402.	59.3	58.5

Table C3. 100 Bootstrap Replications Resampling 2-Vectors (cont'd)

	Post-harvester supply				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	4.10	1.48	3.92	1.20	1.17
Y	14.9	8.08	14.5	6.27	7.17
San Joa	263.	161.	269.	131.	150.
Yolo	500.	165.	516.	131.	154.
Solano	- 145.	167.	- 126.	130.	156.
Sutter	- 144.	172.	- 135.	129.	160.
Sacto	- 291.	160.	- 275.	134.	150.
Stanis	- 304.	164.	- 285.	134.	153.
Santa C	- 321.	174.	- 308.	143.	162.
San Ben	- 349.	190.	- 332.	151.	177.

	Post-harvester demand				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	- 1.68	3.98	- 1.78	3.52	3.94
I	.156	.146	.150	.112	.133
R	35.2	28.5	35.2	23.4	27.7
San Joa	445.	80.2	447.	67.3	78.1
Yolo	689.	80.0	698.	66.9	78.5
Solano	48.4	80.1	60.4	69.9	78.3
Sutter	60.4	80.2	62.4	62.5	77.9
Sacto	- 111.	80.1	- 104.	76.4	78.8
Stanis	- 111.	80.7	- 98.7	73.6	77.8
Santa C	- 111.	80.6	- 104.	70.3	77.9
San Ben	- 107.	80.5	- 98.3	65.2	77.9

Table C4. 100 Bootstrap Replications Resampling 1-Vectors

	Pre-harvester supply				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	11.6	4.11	11.2	3.06	3.69
Y	9.69	5.09	9.13	4.79	4.77
D	- 6.62	11.2	- 5.67	9.43	10.6
F	- 2.10	2.28	- 2.20	1.90	2.01
G	- 91.9	45.3	- 98.8	39.7	42.2
W	- 328.	154.	- 343.	166.	162.
San Joa	871.	335.	922.	254.	266.
Yolo	625.	336.	676.	252.	267.
Solano	316.	336.	370.	253.	267.
Sutter	295.	336.	347.	249.	266.
Sacto	378.	335.	428.	251.	266.
Stanis	300.	332.	348.	249.	264.
Santa C	273.	337.	324.	257.	268.
San Ben	227.	338.	277.	253.	269.

	Pre-harvester demand				
	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	- 14.3	5.71	- 11.9	3.75	4.70
I	.700	.115	.676	.107	.121
R	132.	32.1	119.	22.4	26.8
San Joa	217.	62.2	221.	50.1	59.2
Yolo	- 8.36	62.1	- 2.35	53.6	59.3
Solano	- 319.	62.2	- 312.	54.0	59.6
Sutter	- 332.	62.0	- 325.	50.6	59.0
Sacto	- 260.	62.2	- 253.	51.0	59.5
Stanis	- 349.	61.9	- 343.	48.7	58.3
Santa C	- 381.	62.4	- 377.	51.5	59.8
San Ben	- 406.	62.3	- 398.	52.7	59.7

Table C4. 100 Bootstrap Replications Resampling 1-Vectors (cont'd)

Post-harvester supply

	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	4.10	1.48	3.35	.879	1.02
Y	14.9	8.08	16.9	5.20	6.48
San Joa	263.	161.	245.	115.	136.
Yolo	500.	165.	484.	110.	139.
Solano	- 145.	167.	- 162.	110.	141.
Sutter	- 144.	172.	- 166.	117.	145.
Sacto	- 291.	160.	- 315.	108.	135.
Stanis	- 304.	164.	- 313.	109.	139.
Santa C	- 321.	174.	- 345.	114.	146.
San Ben	- 349.	190.	- 372.	126.	160.

Post-harvester demand

	The Fit		The Bootstrap		
	Parameter	Nominal SE	Mean	SD	RMS Nominal SE
P	- 1.68	3.98	.320	1.51	2.02
I	.156	.146	.159	.0928	.120
R	35.2	28.5	22.1	12.2	15.3
San Joa	445.	80.2	454.	62.5	66.9
Yolo	689.	80.0	699.	59.1	66.2
Solano	48.4	80.1	51.9	66.8	69.1
Sutter	60.4	80.2	68.1	53.3	65.9
Sacto	- 111.	80.1	- 104.	59.7	66.5
Stanis	- 111.	80.7	- 102.	56.7	66.0
Santa C	- 111.	80.6	- 102.	53.1	67.9
San Ben	- 107.	80.5	- 98.3	60.1	67.0

REFERENCES

[1] J.A.Brandt, B.C.French, E.V.Jesse (1978). Economic performance of the processing tomato industry. Bulletin 1888, Division of Agricultural Sciences, University of California at Davis.

[2] L.Breiman and D.A.Freedman (1983). How many variables should be entered in a regression equation? Journal of the American Statistical Association, Vol. 78, No. 381, pp.131-136.

[3] W.S.Chern and R.E.Just (1978). Econometric analysis of supply response and demand for processing tomatoes in California. Giannini Foundation Monograph No. 37, University of California at Berkeley.

[4] W.S.Chern and R.E.Just (1980). Tomatoes, technology, and oligopsony. The Bell Journal of Economics, Vol. 11, pp.584-602.

[5] C.Christ (1975). Judging the performance of econometric models of the U.S. economy. International Economic Review, Vol. 16, pp.54-74.

[6] G.Debreu (1969). Theory of value. Monograph 17, Cowles Foundation for Research in Economics at Yale University. (Wiley, New York).

[7] M. Finkelstein (1980). The judicial reception of multiple regression studies in race and sex discrimination cases. Columbia Law Review, Vol. 80, pp. 737-754.

[8] F. Fisher (1980). Multiple regression in legal proceedings. Columbia Law Review, Vol. 80, pp. 702-736.

[9] D.A.Freedman (1981). Some pitfalls in large econometric models: a case study. Journal of Business, Vol. 54, pp.479-500.

[10] D.A.Freedman (1983). A note on screening regression equations. American Statistician, Vol. 37, pp. 152-155.

[11] D.A.Freedman and S.Peters (1984). Bootstrapping a regression equation: some empirical results. Journal of the American Statistical Association, Vol. 79, pp. 97-106.

[12] D.A.Freedman and S.Peters (1983). Bootstrapping an econometric model: some empirical results. Journal of Business and Economic Statistics, Vol. 2, pp. 150-158.

[13] D.A.Freedman, T.Rothenberg, R.Sutch (1983). On energy policy models. Journal of Business and Economic Statistics, Vol. 1, No. 1, pp. 24-36.

[14] P.Higginbotham (1980). Opinion in Vuyanich vs. Republic National Bank of Dallas. 505 F Sup 224.

[15] E.Mansfield (1982). Microeconomics, 4th edition. (Norton, New York).

[16] P.A.Samuelson (1980). Economics, 11th edition. (McGraw-Hill, New York).

[17] H.Theil (1971). Principles of Econometrics. (Wiley, New York).

[18] F.M. Scherer (1980). Industrial Market Structure and Economic Performance, 2nd edition. (Rand-McNally, Chicago).

[19] V.Zarnowitz (1979). An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level. Journal of Business, Vol. 52, pp. 1-34.

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

TELEPHONE: (510) 642-2781
CABLE: UNIVCALB
FAX: (510) 642-7892

UNIVERSITY OF CALIFORNIA
DEPARTMENT OF STATISTICS
367 EVANS HALL # 3860
BERKELEY, CALIFORNIA 94720 - 3860

Date: 9/8/96

To: Eric Lund

FAX# (612) 470-4700

From: Pat Hardy
University of California
Department of Statistics
367 Evans Hall # 3860
Berkeley, California 94720-3860
FAX # (510)642-7892

Memo: As you requested Tech Report #23.

No. of pages sent (include this page in the total): 65