THE USE OF ECONOMETRICS IN ANTITRUST ANALYSIS

John D. Harkrider
Axinn, Veltrop & Harkrider, LLP
Comments on this draft paper are welcome and should be sent to jdh@avhlaw.com.
Introduction

As private litigants, the Federal Trade Commission ("FTC"), Department of Justice ("DOJ") and courts turn to econometrics in antitrust analyses, it becomes increasingly important for antitrust practitioners to achieve a greater understanding of the terminology of, and assumptions underlying, econometrics. This is important for two reasons: (i) antitrust lawyers cannot effectively represent their clients if they are unable to defend against or affirmatively use econometric evidence; and (ii) the use of flawed econometric analyses, and the failure to detect such flaws by counsel, can lead to inefficient results in the merger review process, most notably the prevention of procompetitive transactions and the consummation of anticompetitive transactions.

Part One of this Article briefly introduces the reader to econometrics and then traces the growing reliance on econometric evidence in antitrust litigation and in the HSR merger review process.

Part Two of this Article discusses some practical evidentiary issues raised by the use of econometrics in both the courts and the federal agencies. In particular, we discuss the admissibility of econometric analysis under Daubert,¹ and the burden of proof assigned by courts in screening econometric evidence. We also discuss the lack of transparency in the use of econometrics by the federal agencies, caused largely by the use of confidential CID responses from competitors.

Part Three of this Article discusses three commonly used types of econometric methods used in antitrust analysis – Ordinary Least Squares ("OLS") regression, Antitrust Logit Model ("ALM"), and demand structure estimation – and highlights the key assumptions upon which such evidence relies for its probity.

Part Four of this Article will apply the concepts in the first two parts of this Article to WorldCom’s attempted acquisition of Sprint. In particular, we will see how substantive conclusions as to the merger’s effect on the price of long distance service can be determined by the econometrician’s choice of method and data. While this point may seem obvious, it further underscores the importance of using lawyers who can understand the importance of certain methodological choices.

Finally, Part Five of this Article suggests specific discovery requests and deposition questions that will help a party determine the reliability of their adversary’s econometric analysis. These questions will help determine (i) whether the data relied upon by the opposing expert are capable of supporting statistical inferences; and (ii) whether the basic statistical assumptions underlying the statistical model used by the opposing expert have been tested and met. These questions are designed to produce testimony that would allow the expert’s testimony to be challenged during a Daubert proceeding, potentially leading to its exclusion.

PART ONE

What is Econometrics?

Econometrics is the use of statistical measurement tools in the aid of economic analysis. In more sophisticated forms, econometric analysis may allow an advocate to, among other things, estimate a relationship between events, estimate post-merger price increases, or estimate demand elasticities. The principal values of econometrics are that it allows an advocate to control for other variables and to state conclusions with a specific level of certainty.

Most econometric techniques attempt to show a correlation between two or more variables. The explanatory variables are called the independent variables and the explained variable is called the dependent variable. Put in other words, the independent variables are those variables that theory leads you to believe would cause the dependent variable to fluctuate. For example, if you believed that industry concentration lead to higher profitability, industry concentration would be the independent variable (the explanatory variable) and profitability would be the dependent variable (the explained variable).

Two prominent problems are raised in simple regressions: first, the confusion of causal direction – that is, thinking that the cause is the effect, and vice versa; and second, the omission of an independent variable that truly caused the dependent variable to fluctuate, an omission that may be fatal if the omitted variable is correlated with an included dependent variable.

Dealing first with the confusion of causal direction, it is important to realize that correlation is distinct from causation; the fact that one variable is correlated with a second variable does not mean that there is a causal link between the two. Rather, the correlation is simply circumstantial confirmation of a hypothesized relationship. If the hypothesized relationship does not make theoretical sense, the existence of a correlation between the two variables is irrelevant.

One danger in inferring causation from statistical relationships is that the “cause” of an event may be mistaken for the “effect” of an event. For example, a correlation could exist between two variables such as “speed of rain” and “speed of windshield wipers” because I, as a driver, increase the speed of my windshield wipers as the speed of the rain increases. But what if I mistakenly believed that the “speed of the windshield wipers” caused the “speed of the rain” to increase? If I were to construct a regression with “speed of rain” as the explained variable and “speed of windshield wiper” as the explanatory variable, I would find a very high correlation between the independent and dependent variables. Thus, I would have circumstantial confirmation for my hypothesis that the speed of the wipers caused it to rain faster.

In this context it is important to recognize that the purpose of econometrics is to provide empirical support for pre-existing theoretical propositions. Using econometric results to
generate theory is quite dangerous -- especially in antitrust, where errors have tremendous social costs.

A second potential error is the omission of an independent (explanatory) variable. In general, there may be literally hundreds of causes of an event, and the inclusion of all of them is neither necessary nor prudent. For example, the price of potash may be correlated with hundreds of factors and a regression that included all of them is no more valid than one that includes the key factors influencing price. Frank Fisher notes, "Without some theory about which variables are likely to matter, throwing a great number of variables into the hopper is likely to lead to spurious results."2

However, the omission of a critical factor -- such as the privatization of a key industry player or the institution of anti-dumping proceedings -- may be fatal to a comparison of before and after pricing to prove the existence of a conspiracy to increase price.

Econometric evidence can be put to a variety of uses in antitrust analysis. In particular, econometric evidence can be used to support the positions that:

- Two products belong in the same (or different) market(s)
- A merger will lead to a quantified price increase taking into account claimed efficiencies
- New entrants are more (or less) profitable than existing firms
- Previous acquisitions in the relevant market did (or did not) lead to increased prices
- Previous acquisitions in the relevant market created efficiencies that were (or were not) passed on to consumers in the form of lower prices

Econometrics can significantly improve theoretical arguments supported only by anecdotal evidence.

First, econometric evidence allows an antitrust lawyer to quantify a relationship between two variables and, more importantly, provide some degree of certainty that the relationship exists. For example, the econometrician is able to say that he is ninety-five percent sure that a certain relationship did not occur by chance.

Second, econometric evidence allows the lawyer to control for other factors. For example, without an econometric analysis of a previous merger, the lawyer can only say that prices rose (or fell) after the merger, but is unable to control for other factors -- such as inflation, demographic changes, or other macroeconomic changes -- that might have caused prices to rise (or fall) after the prior merger.

Third, econometric evidence allows for aggregation, and thus avoids the danger of cherry-picking unrepresentative data points. For example, a lawyer may point to a single example of a successful or unsuccessful entrant and argue from this example that entry is or is not timely, likely or sufficient to deter supracooperative pricing. However, the lawyer relying upon an econometric analysis is able to analyze all attempts at entry over a time period and,

by controlling for other factors, should be able to draw more reliable conclusions about entry conditions than can be drawn from a single example.

For these reasons, a viable theoretical proposition supported by robust and replicable econometric results should be more persuasive than a theoretical proposition that is without econometric support.

PART TWO

EVIDENTIARY ISSUES RAISED BY THE USE OF ECONOMETRICS

Econometric Evidence in the Courts

Courts recognize that “[a]s a general rule, a properly constructed regression analysis ‘can play a vital role in legal proceedings. Used properly, it is an accurate and reliable method of determining the relationships between two or more variables, and it can be a valuable tool for resolving factual disputes.” 3 “The caveat, of course, is that the study must be performed correctly to have any probative value.” 4

Although courts have used econometric evidence in antitrust cases for more than twenty years, 5 in recent years its use in litigated cases has expanded dramatically. One reason for this increase is the fact that the Department of Justice and Federal Trade Commission have increasingly relied upon econometric evidence in their own analysis, most notably in the estimation of post-merger price increases. Another reason for the increasing acceptance of econometric evidence in antitrust cases is the growing familiarity of judges with the econometric evidence used in non-antitrust contexts such as employment discrimination cases.

An illustration of the value of econometric evidence can be seen in a comparison of the treatment of two antitrust class actions – one using econometric evidence and the other not. In In re Beef Industry Antitrust Litigation, the court denied class certification under Rule 23(b)(3), finding that so many factors affected the price of beef that individual issues predominated over common issues. 6 In contrast, in In re Polypropylene Carpet Antitrust Litigation, the court rejected this very same argument, finding that the plaintiff’s econometrician “intends to control for the fluctuations identified by Defendants in a manner that employs evidence that is common to all members of the proposed class.” 7

Courts have relied on econometric evidence to determine the existence (or lack thereof) of market power, 8 the likelihood of unilateral effects, 9 the existence of antitrust injury, 10 and the

---

4 Id.
7 In re Polypropylene Carpet Antitrust Litig., 996 F. Supp. at 29.
amount of antitrust damages.\textsuperscript{11} Econometric evidence has even been used, along with other evidence, to demonstrate the existence of an antitrust conspiracy.\textsuperscript{12}

Courts agree that the admissibility of multiple regressions are determined under the test set forth in \textit{Daubert v. Merrell Dow Pharmaceuticals, Inc.}\textsuperscript{13} The \textit{Daubert} standard is “whether the reasoning or methodology underlying the testimony is scientifically valid and whether that reasoning or methodology properly can be applied to the facts in issue.”\textsuperscript{14}

Because some courts do not regard either economics or statistics as “science,” they have modified \textit{Daubert} to state that the court should determine whether “the proffered testimony is based upon valid economic, statistical or econometric reasoning that can properly be applied to the facts of [the] case.”\textsuperscript{15}

Courts have held generally that econometric evidence can satisfy \textit{Daubert} if performed properly.\textsuperscript{16} In doing so, the court is instructed to look at the “principles and methodology” and not at the resulting conclusions.\textsuperscript{17}

Although econometric evidence frequently passes the \textit{Daubert} test, it does not automatically do so. For example, in \textit{Blomkest Fertilizer, Inc. v. Potash Corporation of Saskatchewan}, the court held that the failure of the plaintiffs’ expert to include two key explanatory variables in his regression analysis violated \textit{Daubert}.\textsuperscript{18} Furthermore, courts have held that an expert’s opinion failed \textit{Daubert} where the expert relied upon assumptions because certain data was unavailable. In \textit{Johnson Electric North America, Inc. v. Mabuchi Motor America Corp.}, the court excluded the expert’s testimony under \textit{Daubert} because the expert—without data on whether the infringing product was actually sold in the United States—used the count of the infringing products marked with a UL stamp.\textsuperscript{19} The court held that “a product’s certification for export to the United States does not make export to the United States a certainty.”\textsuperscript{20} This is particularly important because virtually all models make assumptions—some heroic—regarding data that is otherwise unavailable.

\textsuperscript{10} See, e.g., \textit{In re Polypropylene Carpet Antitrust Litig.}, 996 F. Supp. at 22-29.
\textsuperscript{11} See, e.g., \textit{Colorado ex rel. Woodard v. Goodell Bros., Inc.}, 1987-1 Trade Cas. (CCH) ¶ 67,476 (D. Colo. 1987).
\textsuperscript{12} See \textit{Petruzzi’s IGA Supermarkets, Inc. v. Darby-Delaware Co.}, 998 F.2d 1224, 1241 (3d Cir. 1993).
\textsuperscript{13} However, in \textit{Petruzzi’s}, the Third Circuit held that while regression analysis probably would not be sufficient to defeat summary judgment by itself, regression evidence, coupled with evidence that tended to exclude the possibility of unilateral conduct, would be sufficient to defeat a motion for summary judgment. \textit{Id. See also Ohio ex rel. Montgomery v. Louis Trath Dairy, Inc.}, 925 F. Supp. 1247 (S.D. Ohio 1996).
\textsuperscript{14} \textit{Daubert}, 509 U.S. at 592-93.
\textsuperscript{15} \textit{Louis Trath Dairy}, 925 F. Supp. at 1252.
\textsuperscript{17} \textit{Daubert}, 509 U.S. at 595.
\textsuperscript{18} \textit{Blomkest Fertilizer, Inc. v. Potash Corporation of Saskatchewan}, 203 F.3d 1028, 1038 (8th Cir. 1999).
\textsuperscript{20} \textit{Id.}
Typically, an antitrust plaintiff will have the burden of production and persuasion on the admissibility of econometric evidence.\textsuperscript{21} Courts have differed, however, on whether the burden, once satisfied, then shifts to the defendant to show that the plaintiff's analysis is flawed.\textsuperscript{22}

In general, courts hold that the failure to include an independent variable in the equation relates to its probative value, not its admissibility.\textsuperscript{23} This generally is the correct result. However, the failure to include a key explanatory variable may make render the model inadmissible.\textsuperscript{24} Furthermore, the failure to include an independent variable that is correlated with other independent variables may bias the equation in a way that strips it of probative value, as we shall see in the next section.

In some cases, violations of the statistical assumptions set forth in the next section should result in the rejection of the econometric results. For example, regressions or other methods that result in biased estimates may be without any evidentiary value.\textsuperscript{25} This is especially true when the bias is not signed and corrected.\textsuperscript{26} In addition, regressions that result in large standard errors may also be without evidentiary value because the fact finder may be without sufficiently precise information to determine whether the coefficients are different from zero.

\textsuperscript{21} For an excellent theoretical discussion of the hidden social “costs” of allocating the evidentiary burdens of production and persuasion on statistical matters to one party as opposed to another see Daniel L. Rubinfeld, \textit{Econometrics in the Courtroom}, 85 Colum. L. Rev. 1048 (1985).

\textsuperscript{22} For example, some courts have held that once plaintiffs have satisfied their burden of production, the defendants have “a burden to produce tangible evidence, beyond mere speculation and inference, to support their claims that [the plaintiff’s] analysis reflects flawed econometric techniques.” \textit{In Re Polypropylene Antitrust Litig.}, 996 F. Supp. 18, 26-27 (N.D. Ga. 1997); \textit{see also} Estate of Hill v. ConAgra Poultry Co., 1997 U.S. Dist. LEXIS 13083, *18 (N.D. Ga. Aug. 25, 1997) (“[T]he burden of explaining how [the plaintiff’s] failure to satisfy the constant variance assumption impacts the regression analysis belongs to Defendants, as they are the parties raising the issue.”).

Other courts have suggested that the plaintiffs bear the burden of demonstrating “that the assumptions necessary for the [regression] model's use [have] been satisfied.” Sobel v. Yeshiva University, 566 F. Supp. 1166, 1182 (S.D.N.Y. 1983). \textit{See also} Carroll v. Sears, Roebuck & Co., 514 F. Supp. 788, 801 (W.D. La. 1981) (“[T]he burden is upon plaintiffs to establish a prima facie case by competent proof; this they cannot do by making a self-serving assumption which underlies their statistical presentation and for which no competent proof was adduced.”), \textit{aff’d in pertinent part}, 708 F.2d 183 (5th Cir. 1983); Penk. v. Oregon State Bd. of Higher Educ., 1985 U.S. Dist. LEXIS 22624, at *170 (D. Or. 1985) (“[A]s the party offering regression models as proof and as the party with the ultimate burden of persuasion, plaintiffs bear the burden... of demonstrating that the assumptions necessary for the model’s use have been satisfied.”).

\textsuperscript{22} \textit{See} Bazemore v. Friday, 478 U.S. 385, 400 (1986).

\textsuperscript{23} \textit{See}, e.g., Blomkest Fertilizer v. Potash Corp. of Saskatchewan, Inc., 203 F.3d 1028, 1038 (8th Cir. 2000) (affirming grant of summary judgement to defendants because, \textit{inter alia}, plaintiffs’ expert failed to account for two critical events in his regression analysis).

\textsuperscript{24} Because lawyers and courts might not understand the meaning of technical terms such as “bias,” it is not surprising that they frequently make mistakes. For example, in one case, a district court held that the regression results were entitled to some weight even though “these variables are biased” and “the bias has led the regression analysis models to generate” statistically significant results. \textit{Penk}, 1985 U.S. Dist. LEXIS 22625, at *136-37. This result is completely incorrect because no inference can be drawn from a biased coefficient where the bias is not corrected or signed.

\textsuperscript{25} Bias can be upward or downward. If it is upward, the reported coefficient is higher than it really is, and if it is downward, it is lower than it really is.
A related question is whether a defendant attacking the assumptions underlying the plaintiff’s statistical model must present an alternative model. Jonathan Baker argues that “criticism of an econometric or simulation methodology should be treated with skepticism absent a demonstration that a reasonable alternative leads to a substantially different result, where such an analysis is possible.”

Courts have followed this view in some instances. For example, one court held that “[w]here the defendant adduces no evidence of alternative methodologies or statistics, but merely criticizes those employed by the plaintiff’s expert, acceptance of the projections of plaintiff’s expert is appropriate, since they do have a rational basis.”

Other courts have disagreed, and have rejected “plaintiffs’ notion that these flaws [in a regression’s statistical assumptions] must be ignored unless the [defendant] comes forward with a regression model in which the flaws are corrected.”

Where the party proposing the model has the burden of production and persuasion on the issue the model supports, it would seem that criticisms that demonstrate that the proposing party’s model is flawed and not capable of generating statistical inferences should be all that is required of the party opposing the model. Nonetheless, even in these instances it is not sufficient simply to point out flaws in the regression analysis – the party opposing the model must demonstrate that the flaws bias the results.

The foregoing makes it important to elicit from the opposing expert at a deposition or at trial the conditions under which a biased or inefficient estimate is entitled to no evidentiary weight. This concession should reduce the danger that a court will underestimate the impact of violating econometric assumptions.

**Econometric Evidence at the Antitrust Agencies**

The Department of Justice and Federal Trade Commission frequently use econometrics in their antitrust analysis. Their use of econometrics has taken the form of modeling post-merger price increases, product and geographic market definition, and evaluating the existence of post-merger efficiencies.

---

30 *See, e.g.*, EEOC v. Gen. Tel. Co. of Northwest, 885 F.2d 575 (9th Cir. 1989). The Ninth Circuit panel in that case distinguished the court’s earlier decision affirming the district court’s decision in *Penk*, which had discounted the plaintiff’s statistical evidence due to flaws in its analysis, despite the fact that the defendant had not shown that the flaws affected the outcome. The flaws in *Penk* were “so central to [the issue at hand] that the defendants, merely by pointing out such omissions, could defeat any inference of discrimination.” *Gen. Tel. Co. of Northwest*, 885 F.2d at 582.
For example, econometric analysis was quite important in the DOJ’s decision to bring suit to enjoin MCI Worldcom’s acquisition of Sprint – especially in the long distance markets. Similarly, Jonathan Baker stated that it was unlikely that the FTC would have brought the Staples case had the theory suggested by the documents not been confirmed by the FTC’s econometric analysis.\textsuperscript{31}

There is, of course, a danger in using econometric analysis in the HSR review process. Econometric analysis is a time-consuming and painstaking process that, if performed under intense time constraints, may lead to incorrect results. Unfortunately, intense time constraints are the hallmark of agency review, where the agency is facing a clock ticking down to the end of the waiting period, and the parties are often facing a clock ticking down to the merger agreement “drop dead” date.

Of course, one might argue that antitrust analysis is, by its very nature, a time-consuming and painstaking process and that econometric analysis does not present a unique problem. The key difference between substantive antitrust analysis and statistical analysis is that substantive analysis is more or less transparent. That is, the facts that the parties are relying upon are apparent to all, and are subject to analysis by the other party.

Econometric analysis often is not transparent. The government, for example, may be relying upon a data set compiled from competitor CID responses and may be unwilling or unable to share that data with the parties. Moreover, the statistical methods used by econometricians may not be fully understood by the lawyers in the case, who are, after all, not schooled in econometrics. As a result, the lawyers may be unable to elucidate the methodological problems in the government’s analysis. While one might argue that the government will be required to disclose its data in litigation, the mere threat of litigation is often sufficient to deter otherwise benign or even procompetitive transactions.\textsuperscript{32}

Indeed, Daniel Rubinfeld argues that “the expanded use of multiple regression techniques is accompanied by the possibility of their misuse” and that “substantial normative questions about their use remain unanswered.”\textsuperscript{33} Without full disclosure of the data sets and open discovery concerning the methods used by both the government expert and the party’s experts, these normative questions are, if anything, more pointed in an agency setting.

\textsuperscript{31} Baker, supra note 27.

\textsuperscript{32} Comparing the use of econometrics in an academic setting with the use of econometrics in an agency setting highlights some of the above-mentioned concerns. In an academic setting, publication of the data set relied upon by the author is often a pre-condition to publication in a journal. Despite the absence of intense time constraints on the initial research, analysis of the data set by other academics often reveals methodological problems in the author’s analysis that repudiate the published results. Given the time constraints raised by antitrust review at the agency level and the fact that the decision makers often are not schooled in econometrics, it is difficult to believe that enforcement errors at the agency level are less likely than errors at the academic level. When one considers the social costs of stopping a merger that is otherwise efficient or letting a merger go forward that is otherwise inefficient this is a very serious problem.

\textsuperscript{33} Rubinfeld, supra note 21, at 1094-95.
Given this, a strong argument can be made that neither the government nor private parties should be permitted to use econometric analysis unless their data sets and methods, including alternative specifications, are subject to discovery and cross-examination.

Jonathan Baker argues as much when he states that interested parties submitting econometric studies must present the raw data, describe how it was collected and cleaned, describe whether the data was transformed, describe which observations were included (or dropped), and describe what sorts of variables were included. At a minimum, the agency should be required to do the same.

Indeed, it is illuminating to contrast the use of econometrics in academia with the use of econometrics in the HSR process. In academia, economists conduct econometric analysis generally indifferent to the direction (positive or negative) of the statistical relationship (other than hoping that there is a statistically significant one), generally without time pressures, almost always with complete transparency in methods and data, and within a rigorous peer review process where other qualified academics review and referee the submitted work. In the HSR process, econometrics is performed (at least in the private sector) by economists who have clients who are hardly indifferent to the substantive outcome, under extreme time pressure, with less than ideal transparency, and not only without peer review, but also where the referee is a lawyer without econometric background. Of course, econometrics in the HSR process does have its advantages, namely that tremendous resources are spent to collect and analyze data. However, the process is less than ideal and one would expect that significant errors would be made, just as significant errors are made in the academic world despite the safeguard of peer review.

PART THREE

Introduction to Basic Econometric Methods and Models

Three common forms of econometric analysis are ordinary least squares ("OLS") regression, the antitrust logit model ("ALM"), and the estimation of demand equations from empirical evidence. We will discuss those methods, and the assumptions necessary to generate valid results under each of them, in this section.

OLS Regression

The Method

An OLS regression seeks to determine the relationship between two or more variables – the explained or dependent variable and the explanatory or independent variable(s). To illustrate how the method works, we will use a model that hypothesizes a relationship between gross margin (the dependent variable) and HHI (the independent variable). This model is grossly simplified for illustrative purposes and is not suggested as a method to determine whether a merger will have a price effect.

34 Baker, supra note 27.
In the graph below, the X axis represents HHI and the Y axis represents Margin.

The OLS regression fits a line that “best” expresses the relationship between the dependent and independent variable. The points along the fitted line are the “predicted” values of the dependent variable for a given value of the independent variable.

Returning to the example above, the OLS regression predicts values of margin for a given level of industry concentration. The predicted values are represented by triangles and the fitted line connects these triangles. Assuming that our results are statistically significant, this model suggests that there is a positive correlation between Margin and HHI, which is consistent with our hypothesis.

---

35 The method by which OLS determines the “best” estimate is to fit a line that minimizes the sum of the squared “errors,” where the error (or residual) is the distance between the line (where the points “should be”) and a data point (where the point “is”).
The relationship between the dependent and independent variables is expressed as follows:

The Dependent Variable \( (Y) \) = An Intercept \( (A) \) – which is the value of the Dependent Variable when all of the Independent Variables are equal to zero) plus The First Independent Variable \( (X_1) \) plus The Second Independent Variable \( (X_2) \) plus The Third Independent Variable \( (X_3) \) ... plus an Error Term \( (E) \) – which includes, among other things, other independent variables that were not specified in the model).

When the data is fed into the statistical package, the regression program will report coefficients for each variable as well as the intercept. The coefficient or Beta \( (B) \) indicates the marginal effect of an additional unit of the dependent variable, controlling all other variables. What this means is that for each additional increment of an independent variable, we would expect the dependent variable to increase by \( B \).

Putting all of this together:

\[
Y = A + B_1X_1 + B_2X_2 + \ldots + B_kX_k + E
\]

Consider the following example. The FTC has long taken a position that supermarkets do not compete with superstores such as Wal-Mart. One way to answer the question of whether Wal-Mart and supermarkets compete is to construct a regression where the dependent variable would be the profitability of some bundle of groceries sold by a supermarket and the independent variable would be the number of Wal-Marts within a specific radius. The theory is that Wal-Mart directly competes with supermarkets and, therefore, the presence of Wal-Mart should lower the store’s profitability on certain items.

Of course, there may be many reasons why supermarkets earn higher or lower margins on their products. For example, the number of other supermarkets, labor conditions, demographic conditions, the cost structure of the competitors in the market may differ. These other factors are also “independent” variables and may be included in the model.

Let us consider a simplified regression:

\[
\text{Gross Margin} = \text{Intercept} + B_1 \times \text{Number of Wal-Marts} + B_2 \times \text{Number of Supermarkets} + B_3 \times \text{Median Household Income} + B_4 \times \text{Unionization} + E
\]

The Bs or Betas would represent the marginal effect that each variable has on the price of products, holding the other variables constant. \( E \) would represent, among other things, omitted variables that also affect the profitability of groceries.

In order to run this regression, an econometrician would collect data on the profit margins on a bundle of groceries sold by supermarket competitors in the market, and also collect data on the number of Wal-Marts in the market, the number of competitors in the market, the median
household income of the store’s customers, and whether the store was unionized. Then this information would be fed into a computer and the output would be analyzed.

Let us say that the program reported the following results:

\[ \text{Gross Margin} = .30 \times (\text{Intercept}) + -.01 \times \text{Wal-Mart} + -.02 \times \text{Supermarket Competitors} + .005 \times \text{Median Household Income} + -.02 \times \text{Unionization} + E. \]

Assuming that each of these coefficients is statistically significant, these results suggest that the gross margin of the supermarket is 1% lower for each Wal-Mart in the market, 2% lower for each competing supermarket, .5% higher for each unit increase of median household income, and 2% lower if the store is unionized.

**Statistical Significance vs. Substantive Significance**

The coefficient or Beta on an independent variable explains what happens to the dependent variable when there is a one-unit change in the independent variable. For example, the coefficient on Wal-Mart is negative 1%. This suggests that increasing the number of Wal-Marts in the market by one will decrease the supermarket’s margin by 1%, after controlling for the other independent variables. If the gross margin is 35% if there is no Wal-Mart, the gross margin should be 34% with one Wal-Mart, 33% with two Wal-Marts, etc.

The regression output reports a standard error on a coefficient. The standard error tells us about the variability in our estimate of the coefficient. A coefficient may be large and positive, but its estimate may have such a high degree of variability that we have little confidence in that estimate. The T statistic is a measurement of whether the size of the standard error affects our confidence in our results. In large samples, a coefficient is only statistically significant if it has a T statistic that is at least 1.96. In such cases, we would be ninety-five percent confident that the results we are seeing did not occur simply by chance.

Importantly, statistical significance is not the same thing as substantive significance. A regression may suggest that the price of a product decreases by .0003 cents for each additional competitor in the market. The regression may indicate that we are very confident of this relationship. However, a merger that reduces the number of competitors by one may have a negligible effect on the price of the product, and this price increase must be balanced against the cost of merger enforcement and the efficiencies claimed by the parties.\(^{36}\)

**Assumptions**

OLS regressions depend upon certain assumptions known as the Gauss-Markov assumptions. If these assumptions are met, the equations used to generate the results are valid and the

\(^{36}\) A related concept concerns the adjusted R-squared statistic, which indicates to us how much of the dependent variable, in this case price, is determined by the independent variables. In general, the adjusted R-squared is not a meaningful statistic because we are not concerned about how much of price is affected by the model, but whether there is a statistically significant relationship between price and a particular independent variable, a relationship that is revealed simply by looking at the coefficient and its standard error.
estimates generated by the OLS regression are the best linear unbiased estimates available ("BLUE"). If these assumptions are not met, the equations used to generate the results may be invalid, and, if they are, the results generated by the regression may be without any evidentiary value.

We will discuss three of the assumptions here.\textsuperscript{37}

1. The error term has a constant variance ("homoscedasticity").
2. Error terms in different observations are not correlated (no "autocorrelation").
3. The error term is not correlated with any independent variable.

These assumptions will be discussed separately with particular attention both to the statistical tests that are used to ensure that these assumptions are valid and to real world instances in which these assumptions may not be met.

1. Homoscedasticity.\textsuperscript{38}

Homoscedasticity means that we assume that each error term for each observation has the same variance. Remember, the "error term" is the difference between the value of the dependent variable predicted by the model and the actual value of the dependent variable.

This assumption is frequently violated in econometric analysis. For example, consider a model that measures the profitability of firms and assume that the population we are examining includes both large and small firms. It may be that large firms exhibit greater variability in profit because of product diversification and greater research and development expenditures; thus we would expect that the variance would not be the same for larger firms as for smaller firms.

A model that violates the assumption of homoscedasticity is said to be heteroscedastic. A model that exhibits heteroscedasticity results in unbiased point coefficients but biased standard errors. This means that the coefficients are correct, but the standard error is larger or smaller than the model predicts. Consequently, a model may suggest a statistically significant relationship between variables when none in reality exists.

2. Uncorrelated Errors.\textsuperscript{39}

\textsuperscript{37} Four additional assumption are necessary for the equation to be BLUE: first that the functional form is linear in its coefficients, second that the error terms have a mean of zero, and third that the independent variables are linearly independent (the independent variables are not perfectly multicollinear and can be identified). The third assumption is commonly known as the identification condition. It means that the independent variables cannot be perfect linear combinations of other independent variables because, if they are, it is difficult to determine with precision what effect any single variable has on the dependent variable. Because this assumption is most frequently violated when independent variables are multicollinear, we will discuss it in the "data problems" section of this article.

\textsuperscript{38} This assumption may be expressed as follows: \( \text{Var}[e_i]=\sigma^2 \) for all \( i \).

\textsuperscript{39} This assumption may be expressed as follows: \( \text{Cov} [e_i, e_j]=0 \) if \( i \neq j \).
Uncorrelated error terms assume that the error terms for different observations are not correlated. That is, it assumes that errors for one observation are not correlated with errors for another observation. Autocorrelation is common in time-series data.

Autocorrelation may occur when omitted explanatory variables are correlated. For example, when omitted explanatory variables are correlated over time, then errors in one time period are correlated with errors in another time period. This would result in temporal autocorrelation and frequently occurs in time-series data. For example, consider a model that measures real investment as a function of real GNP and real interest rate. A plot of the residuals in this example would likely reveal that misses in one period (investment is lower than expected in one year) are likely correlated to misses in the next period (investment again is lower than expected in the next year). This may occur because an omitted variable such as the level of inflation in one year may be correlated with the level of inflation in the next year. This would suggest the presence of temporal autocorrelation.

The effect of autocorrelation is that the least squares estimator is inefficient.\textsuperscript{40} This means that the point estimate is correct (the B) but the t-tests on the point estimate may not be correct. In effect, if autocorrelation is not corrected, there may be instances in which the fact finder does not have sufficient confidence in the results to draw conclusions from the results of the model.

3. The Error Term Is Not Correlated with the Independent Variables\textsuperscript{41}

The error term represents the difference between the model’s predicted value and its actual value. In general, this error term includes both measurement error and independent variables that are omitted from the model. The error term is, by definition, correlated with the dependent variable. However, if the error term is correlated with the independent variables then the resulting estimates are biased and, in some cases, statistical inferences may not be drawn from the model.

For example, consider a model that predicts profitability as a function of market share, firm size, industry concentration, and raw material costs and consider that the equation demonstrates a statistically significant relationship between the firm’s market share and firm profitability, after controlling for firm size and costs. Clearly there are omitted variables in this model, such as marginal costs. There is a high probability that raw material costs are correlated with firm size as the firm takes advantage of its size to obtain purchasing efficiencies. In such an instance, the error term likely is correlated with the independent variable and the coefficients reported in the equation may be biased.

Whether the assumption that the error term is not correlated with the independent variables is true or not can be readily detected using the Hausman specification test.

\textsuperscript{40} Autocorrelation, like heteroscedasticity, can be diagnosed and corrected. The most common statistical test for autocorrelation is the Durbin-Watson test, which enables the researcher to determine whether it is possible to reject the hypothesis that autocorrelation is present. In many instances, however, the Durbin-Watson test is inconclusive, meaning that the researcher is unable to reject this hypothesis.

\textsuperscript{41} This assumption may be expressed as $E[x, \epsilon]=0$ for all i and j.
The ALM is an econometric model frequently used to predict consumer welfare losses from mergers involving heterogeneous products.\textsuperscript{42} In the ALM, demand is modeled as logit, a method where the parameters are estimated using maximum likelihood.

As an initial matter, it is important to recognize that while the ALM is used in mergers where market shares are believed to give a misleading picture of competitive effects, it depends upon market shares in order to quantify post-merger price effects. This is so because the size of the cross elasticities are determined by the product’s market share.

It is also important to realize that the ALM assumes that there is no new entry or product repositioning. This assumption can be quite unrealistic in mergers between sellers of differentiated products where entry may be accomplished quickly by product repositioning.

The basic logit demand model is stated as follows:

\[
U_{ij} = a_j - Bp_j + e_{ij}\]

\(U_{ij}\) = the indirect utility of consumer \(i\) associated with choice of product \(j\).
\(a_j\) = the perceived quality differences among products.
\(p_j\) = price for product \(j\).
\(e_{ij}\) = individual-specific components of utility.

The ALM predicts price increases using the following inputs:

- \(S\) which represents market shares;
- \(P\) which represents price;
- \(B\) which represents elasticity of demand, controlling for substitutability of the goods of the merger parties (the inside goods);
- \(E\) which represents aggregate elasticity of demand, controlling for substitutability of inside goods and an outside good.

\(S\) and \(P\) are directly observed. \(B\) and \(E\) are estimated.

The ALM is based upon a number of assumptions that are critical to the validity of its results, the most controversial of which is that competing goods have identical cross-elasticities of demand.\textsuperscript{43}


\textsuperscript{43} Other assumptions include that \(E_{ij}\) is not heteroscedastic and that individual-specific components of utility are not correlated with price. Logit models, like OLS models, assume that the error terms are not heteroscedastic. In an OLS regression, heteroscedastic error terms result in unbiased and consistent coefficients, but inefficient standard errors. This means, in effect, that our estimates are correct, but we have a wide confidence interval. In a logit model, heteroscedastic error terms result in inconsistent estimates. This means that our estimates are not usable because in large samples (the real world) our estimates of a coefficient do not converge to its true value.
Logit models assume identical cross-elasticities of demand with respect to any given product.\textsuperscript{44} This is referred to as the assumption of Independence of Irrelevant Alternatives ("IIA").\textsuperscript{45} In the context of the ALM, the IIA assumption implies that consumers switch to other products in proportion to the relative market share of these competing products.\textsuperscript{46} For example, assume a merger between Shredded Wheat and Grape Nuts where the relative market shares in the ready to eat (RTE) market are:

- Shredded Wheat: 5%
- Grape Nuts: 10%
- Granola: 15%
- Captain Crunch: 20%
- Lucky Charms: 20%
- Fruit Loops: 30%

The IIA assumption would suggest that if the merged entity were to increase price, consumers would substitute Fruit Loops, Lucky Charms and Captain Crunch before Granola. In this case, it is difficult to believe that the IIA assumption would be met. Indeed, it seems very odd to assume identical cross elasticities of demand in markets characterized by differentiated competition. Hausman notes: "Since the size of the post-merger unilateral effect depends crucially on the pattern of demand substitution between the products of the merging firms, restricting the demand substitution patterns would seem to defeat the purpose of performing the unilateral effects analysis."\textsuperscript{47}

Fortunately, the IIA assumption can be tested empirically through a test developed by Hausman and McFadden. Predictably, this test reveals a very high level of rejection of the IIA assumption.\textsuperscript{48}

Other models may be used to estimate post-merger price increases that do not rely upon the IIA assumption. In some instances, these models estimate price increases that are very


\textsuperscript{45} Werden, supra note 42 at 378. One attempt to solve the IIA problem is to use a nested logit demand system where products that have different attributes that matter to consumers (e.g., in the case of cars, size, speed, degree of luxury) are placed into different nests. However, the nested logit model still makes the assumption of identical cross elasticities of demand applies only to products within the same nest, and this is a problem because consumers consider heterogenous factors when making decisions upon what car to buy within a class of products (e.g., high performance luxury cars vs. mid-level performance luxury cars). \textit{Id.} at 379-80.


\textsuperscript{48} \textit{Id.}
different (sometimes higher and sometimes lower) than the price increases predicted by the ALM.\footnote{\textit{Id.} at 336. In the case discussed in this article, the logit model assumptions of identical cross elasticities were not consistent with real world data.}

**Demand Equation Estimation**

An alternative method, which is preferable to the ALM where data is available, is to directly estimate a two-stage demand system, using available data regarding real world purchases by consumers. One stage would relate to the demand for the product, while the other would relate to competition among brands. By using both stages together, own and cross-price elasticities for each brand can be estimated. Having estimated own and cross elasticities, using a Nash-Bertrand price-setting model for differentiated products, post-merger price effects can be estimated.

There are numerous advantages to using real world data to estimate price effects in mergers among differentiated products. The principal advantage is that doing so avoids the IIA assumption discussed in the above section. The premise behind unilateral effects is that a significant number of consumers regard product X and Y as next best substitutes. Thus, one would expect consumers of X to switch to Y in a proportion higher than Y’s market share and vice-versa.

If real world data reveals that consumers of the two products switch in a proportion higher than their market share would suggest, the ALM would underestimate the post-merger price effects. Conversely, if real world data reveals that consumers of the two products switch in a proportion lower than their market share would suggest, the ALM would overestimate the post-merger price effects.

This analysis can only be performed if time series data on the price and quantity of purchases of competing goods is available. As a result, it is not the sort of data a company would be able to collect by itself. Sources for time series data would, therefore, be scanner data, survey data, or competitor data. Because this data is often expensive or not available, it is frequently quite difficult to perform this analysis. These issues will be discussed below. However, if the data issues can be overcome, this analysis is superior to the ALM alternative.

**Data Issues**

The critical input in econometric analysis is data. There is a saying in statistical analysis – GIGO – or “garbage in garbage out.” What this means is that even the most scientific statistical analysis cannot glean reliable inferences from bad data. Moreover, without the correct data type, it may be impossible to use the proper econometric method.

**Data Types**

For econometric methods, the best type of data measures price and quantity over time. It is critical that the data include information on the merging firms and all other competitors. This
may be a problem not only in the context of a government investigation where private parties lack subpoena power, but also in private litigation, as courts may be unwilling to burden a third party with responding to a detailed information request for price and quantity information from a direct competitor.

As discussed above, one type of time series data is point of sale scanner data that collects data regarding store purchases. One limitation with this data is that it typically is available only for retail goods. Another issue regards its reliability. For example, EarthGrains, a multi-billion dollar bakery products company, attempted to use scanner data from grocery stores to transform its distribution processes. Scanner data from many retailers, however, proved to be of variable quality, limiting its usefulness. Another issue is that it may be important to augment scanner data files to monitor advertising, promotional activities, and competitors’ actions. Finally, despite the sheer volume of price and quantity information, scanner data typically lack information pertaining to socio-demographic profiles of consumers. In sum, we should resist the temptation to use data simply because it is available and should rigorously examine the quality and completeness of the underlying data.

Another type of time series data is panel data, which follows the purchases of a group of consumers over time. There are many advantages to this type of data, most notably that one is able to examine the same individuals over time. But the problem with panel data is that individuals often are selected on some non-random basis – for example, they self-select by agreeing to send their bills to a survey firm for $50/month. This may bias the results if the consumers differ from most people on a dimension that is related to the dependent variable – for example, if people who are likely to send in their bills for $50 are also likely to switch long distance carriers for $50 payments.

It is important to understand that the Government is almost always at a substantial advantage over private litigants regarding data. Consider an example where the parties use the ALM because they do not have access to real world data. Assume that the government has access to real world data and not only uses a two-stage demand equation but also tests – and rejects – the assumptions made by the parties in their econometric analysis. If this occurs, no amount of advocacy will convince the Government that the parties are right. Similarly, consider an example where the parties use survey data from a third party source. Again assume that the Government has access to the “universe of real world data” and again not only estimates elasticities directly but also determines that the parties’ survey data is not random. Again, no amount of advocacy will convince the Government that the parties are correct.

In both cases, the parties have wasted substantial time and money for a result that may be rejected out of hand by the Government. As discussed earlier in this Article, this result may be unfair as the parties are not only fighting a losing battle, but also are unable to test the reliability of the Government’s econometric analysis.

Data Problems

In addition to the issues discussed above, there are four types of data problems typically encountered in econometric analysis.

1. **Multicollinearity**: Two or more independent variables are too highly correlated to permit precise analysis of their individual effect.

2. **Improper Aggregation**: The data contains two or more subsets of observations and these different subsets are subject to different competitive dynamics – and thus, if disaggregated, would have statistically different coefficients.

3. **Missing Observations**: The data includes missing observations that are either dropped from the data set or generated through some statistical method.

4. **Influential Observations**: The data includes “influential observations” that have undue influence on the regression results.

Efforts to solve these four problems must be undertaken carefully, with consideration of the effect of the solution on the viability of the results. In the extreme case, efforts to solve these problems may destroy the viability of the results generated by the data; in less extreme cases, efforts to solve these problems may considerably undermine the reliability of the statistical results.

1. **Multicollinearity**

The dependent variables in multiple regressions are almost always correlated with each other to some extent, and generally this does not present a problem. If two variables are highly correlated with each other, however, they are said to be multicollinear.

Although multicollinearity does not itself present a violation of the Gauss-Marcov theory, multicollinearity may have several important consequences:

- The equations will not be robust in that small changes in the data may cause large changes in the coefficient estimates.

- Individual coefficients will have large standard errors (we are not sure that either coefficient is different from zero) even though together, the coefficients will be highly significant (we will be sure that both coefficients are different from zero).

- The coefficients will have the wrong sign (e.g., the presence of another competitor causes the price to be lower than expected, rather than higher than expected) or an implausible magnitude (e.g., the post-merger price increase is much higher than it should be).

---

51 Even if multicollinearity is high, the OLS estimators are still BLUE, that is, best linear unbiased estimators. In fact, multicollinearity does not violate the other Gauss-Marcov assumptions. Thus, the only effect of multicollinearity is to make it hard to get coefficient estimates with small standard errors, which is essentially the same problem presented by having a limited number of observations.
Multicollinearity is easy to identify. However, it is not as easy to fix. For example, one solution is to drop one of the multicollinear variables. However, the effect of omitting a statistically significant variable that is, by definition, correlated with another dependent variable is to violate the Gauss-Markov assumption that the error terms are not correlated with the independent variables.\(^5^2\)

2. Improper Aggregation

Data can always be disaggregated into component parts. For example, data on the minute price of a long distance call can be disaggregated into consumer minutes and business minutes, and within consumer minutes, minutes from customers on basic plans and minutes from customers on high usage plans; data on the price of products in a nationwide chain can be disaggregated into prices at stores in different geographical regions; data on the price of newspaper advertisements can be disaggregated into the price of classified advertisements and the price of display advertisements.

Using aggregate data in a single econometric model can be very dangerous when there are different competitive dynamics for different subsets of the data. For example, Hausman and Leonard argue that the FTC pooled all the Staples stores into a single model even though a Chow test revealed that the coefficients on the variables of interest were different depending upon the region of the country.\(^5^3\)

It may not be enough, however, simply to point out that a single model using pooled data is incorrect. Rather, an econometrician should construct an alternative model using disaggregated data and review the results, because, in some cases, the results of the unpoled data contradict the results of the pooled data.

3. Missing Data

A frequent problem encountered in analyzing survey data is the problem of missing observations. For example, a client conducts a survey of customers to determine whether and to what extent customers believe there will be a post-merger price increase, but not all the customers surveyed answer the survey and not all survey respondents answer all questions. Another frequent problem is when the client collects data on pricing in the ordinary course of business but certain data entries are missing.

One solution to the problem of missing data is to drop observations with missing data. However, if the fact that data is missing is correlated with the dependent variable, this solution is extremely dangerous and can destroy the viability of the statistical results. For example, in the case of the survey of customers, let us say that the economist used the survey data in a regression where the dependent variable was the level of price increase (for example, one percent, two percent . . . ) and the independent variables were a vector of client

\(^{5^2}\) Greene, supra note 43, at 270.

characteristics such as size, industry and geographic location. If clients who believe that there will be a large post-merger price increase are more likely to answer survey questions on their reaction to the merger than those clients who do not believe that there will be a price increase, then simply performing a regression on the complete data set -- without attempting to correct for self-selection bias -- may result in regression results that are biased and without statistical validity.

Because data sets are rarely clean, analysts frequently drop or generate missing variables so that they can perform statistical analysis. Because the fact that the data set was cleansed is not apparent even to a retained expert, dubious statistical practices can often escape detection.

4. Influential Observations

In regressions with small numbers of data points, it is often the case that certain data points have an undue influence on the regression. Look at the following regressions and graphs to see how this might occur.
This regression plots price as a function of industry concentration. In the first regression there was one concentrated (HHI 1500) market with an average price well above the other industry points. The regression suggests that there is a positive relationship between the industry concentration and price. When we delete that observation we see that there is a slightly negative relationship between price and industry concentration. Thus it appears that a single observation may have been driving the econometric results.

An observation that has an unexplained and inordinate power is known as an outlier, and there are a number of statistical tests that can be used to identify outliers and control for their effect. For example, we might find out that the market in which the high concentration occurred experienced a labor strike on an essential input that drove up the cost to all competitors in the market. While we should not drop influential observations simply because they are outliers, we might identify additional control variables that might explain the apparent relationship between the dependent and the independent variable.

The Special Problem Presented by Accounting Data

Many of the data problems discussed above are illustrated by the use of accounting data. Accounting rates of return in the form of gross or net profitability are frequently used by advocates as evidence of market power. For example, a plaintiff may claim that a defendant’s net profit is higher in markets where the defendant has a high market share than in markets where the defendant does not. Conversely, the absence of a correlation between net profits and market share may be used as evidence of the lack of market power.

The use of accounting data has superficial appeal. It is collected by all companies in the ordinary course of business, and because it is relied upon by shareholders and submitted to the government in the form of tax returns, it has an air of reliability.

The problem with using accounting profits to derive inferences on the existence of market power is that accounting profits are not economic profits. The economic rate of return on an investment is equal to the discount rate that equates the present value of its expected net revenue stream to its initial outlay. Accounting profits are a valid measure of market power only when accounting profits have some relationship to economic profits. However, because of depreciation schedules, the timing of investments, the mix of different types of

---

54 See Frank Fisher and John J. McGowan, *On the Misuse of Accounting Rates of Return to Infer Monopoly Profits*, AM. ECON. REV., Mar. 1983, at 82. For this reason, courts have rejected the use of accounting profits to demonstrate the market power. See, e.g., Blue Cross & Blue Shield United of Wisconsin v. Marshfield Clinic, 65 F.3d 1406, 1412 (7th Cir. 1995) (“[M]easured rates of return reflect accounting conventions more than they do real profits (or losses), as an economist would understand these terms.”). In addition, monopoly power is not associated with a high rate of return. *Id.* (“[T]here is not even a good economic theory that associates monopoly power with a high rate of return. Firms compete to become and to remain monopolists, and the process of competition erodes their profits. Conversely, competitive firms may be highly profitable merely by virtue of having low costs as a result of superior efficiency, yet not sufficiently lower costs than all other competitors to enable the firm to take over its market and become a monopolist.”)
investments in a reporting entity, and the rate of growth of a firm, "it should be obvious that only by the merest happenstance will the accounting rate of return on a given investment . . . be equal to the economic rate of return."^{55}

These various factors raise many of the issues discussed above. For example, the mix of different types of investments in a reporting entity raises the issue of improper aggregation. Consequently, a model that uses accounting profits for its dependent and/or independent variable(s) may not generate valid results even if all of the statistical assumptions made above are met.

PART FOUR

WorldCom’s Proposed Acquisition of Sprint

WorldCom’s failed acquisition of Sprint provides an interesting case to examine many of the data and method issues discussed above. Although the author was retained by the Department of Justice to assist in its econometric analysis in that matter, the discussion in this section will be confined to information found in the public record – most notably in the public comments filed with the FCC. In addition, the views expressed in this section – and in this Article as a whole — are strictly the author’s and do not necessarily reflect the views of the Department of Justice.

WorldCom and Sprint are two of the three largest long distance carriers in the United States. According to the Complaint filed by the Department of Justice, more than 80% of residential lines in the United States are presubscribed to one of the Big 3 as their long distance carrier, with approximately 53% of residential lines subscribing to AT&T, 19% of residential lines subscribing to WorldCom and approximately 8% subscribing to Sprint. The remaining 20% was split among a number of other carriers, none of which had more than 3%.

The Complaint further alleged that a disproportionate number of mass market consumers who leave WorldCom for a new long distance provider switch to Sprint, and further that a disproportionate number of mass market consumers who leave Sprint for a new long distance provider switch to WorldCom. In other words, their diversion ratios were not proportionate to their market shares.

The competitive effects of the merger were analyzed by three different groups – the DOJ, the parties, and third parties – each of whom had access to different types of data, which, in turn, drove the econometric method that they were able to use.

For example, Jerry Hausman, on behalf of SBC, used a two-stage demand estimation equation as discussed above. SBC, of course, did not have access to competitor data, and therefore was required to rely upon third party data provided by a survey company named PNR. Using this data, Hausman estimated demand elasticities for WorldCom and Sprint. Based upon these demand elasticities, Hausman argued that WorldCom customers would

^{55} ld.
experience a post-merger price increase of 5.4% and Sprint customers would experience a post-merger price increase of 8.9%.\textsuperscript{56}

The parties in their FCC filings, however, criticized Hausman’s use of PNR data, arguing that it omitted critical price information such as discounts paid by long distance competitors. If true, this is important because it could mean that the data could not correctly identify price changes. The parties also argued in their FCC filings that Hausman failed to show that the point estimates for the demand elasticities were statistically significant. Finally, the parties argued that Hausman’s econometric method reported marginal cost estimates for AT&T that were unrealistic. Furthermore, the parties relied upon data from another data provider, Paragren, to argue that the customers of WorldCom and Sprint regard emerging carriers as very good substitutes.

An alternative approach to direct estimation of demand equations was presented by Greg Werden in a paper written well before WorldCom and Sprint decided to merge. Utilizing the ALM, Werden argued that only mergers involving AT&T would be expected to lessen welfare significantly.\textsuperscript{57} The reason the ALM would predict a different result than direct demand equation is that the ALM assumes that all when a firm increases price, consumers switch to others in proportion to the relative shares of those products. This is the IIA assumption discussed above. However, as alleged in the Complaint, this assumption was incorrect. As a result, the ALM would underestimate the price effect.

The point of this exercise is not to demonstrate that one side or the other was correct, but rather to demonstrate the importance of data and method in driving conclusions. The ALM predicted that a merger between Sprint and WorldCom would not create a price increase. The direct estimation of demand from survey data predicted that there would be a price increase. This conclusion was subject to criticism because it relied upon an invalid data source. In sum, only by using a high quality data set and an appropriate method could one correctly analyze the effects of the merger, something the Department of Justice, because of its unique ability to collect data during the HSR process, was in the best position to do.

\textbf{PART FIVE}

\textbf{Deposing Experts}

Discovery is a critical weapon in attacking econometric results. Interrogatories and document demands should seek sufficient information (\textit{e.g.}, data, models, and methods) from which your expert can replicate the opposing expert’s substantive results and, in so doing, identify the data manipulation and statistical assumptions necessary to generate those results.


\textsuperscript{57} See Werden and Froeb, \textit{supra} note 46, at 407.
In addition, if possible, an attorney should depose the econometrician who conducted the statistical analysis and the analyst who collected and transformed the underlying data. These depositions should aim to do four things:

1. Determine the expert’s qualifications as an econometrician.
2. Identify potential data problems.
3. Identify violations of statistical assumptions.
4. Determine whether there exist prior inconsistent econometric results.

Before an attorney deposes an opposing expert, the opposing side should have produced:

1. All work papers
2. All original and final data sets
3. All current and prior models
4. All tests run to determine whether there were violations of statistical assumptions

Much of this information can be obtained by demanding production of the statistical program’s log file. Virtually all econometric programs have “log” files that contemporaneously save statistical results. “Log” files will identify alternative specifications, will identify the results of regression diagnostics and will allow your expert to determine whether the data was tinkered with in order to generate results.

One problem that may arise is that the opposing party has hired a non-testimonial statistician to perform the econometric analysis. Thus the opposing side will likely argue that discovery into the non-testimonial expert’s analysis is not permitted under Rule 26(b)(4)(B). However, to the extent that the testimonial expert is unable to provide sufficient information on the models used by the non-testimonial expert, discovery should be permitted.

Failure to make adequate pretrial disclosure of computer methodology and data that the expert would rely upon at trial may constitute reversible error.

The following sets forth some useful deposition questions. These are suggestions and clearly are not all-encompassing. The appropriate questions in a given situation should, of course, be developed after discussion with your expert.

1. **Determine the expert’s qualifications as an econometrician.**

   a. Identify all econometric courses the expert has taken.

---

58 See Pearl Brewing Co. v. Jos. Schlitz Brewing Co., 415 F. Supp. 1122, 1140 (denying discovery of alternative models from non-testimonial expert because exceptional circumstances had not been demonstrated).
59 Id. at 1138-39 (permitting discovery of printouts of econometric analysis and permitting deposition of the statisticians because testimonial expert had only superficial understanding of the statistical analysis performed by the non-testimonial experts). But see Bartley v. Isuzu Motors Ltd., 151 F.R.D. 659 (D. Colo. 1993) (permitting discovery of alternative analysis).
60 See, e.g., Shu-Tao Lin v. McDonnell Douglas Corp., 574 F. Supp. 1407, 1412-13 (S.D.N.Y. 1983) (granting motion to set aside verdict because of, inter alia, failure to make adequate pretrial disclosure of computer methodology and data that expert would rely upon at trial), rev’d on other grounds, 742 F.2d 45 (2d Cir. 1984).
b. Determine whether econometric methods were a major, first minor, second minor or cognate in the expert’s Ph.D. program.

c. Identify whether the expert took an oral examination in econometrics and whether the expert passed.

d. Identify all papers the expert has published in econometrics, especially for journals such as *Econometrica*.

e. Determine whether the expert has taught any classes on econometric techniques.

f. Determine whether the expert believes s/he is qualified as an expert on econometrics.

g. Ask basic questions testing the expert’s knowledge. For example:

(i) What are the Gauss-Marcov assumptions? Express them in matrix form and in plain english.

(ii) Express b hat in matrix form?

(iii) What are the symptoms of multicollinearity?

(iv) What is the consequence of dropping a multicollinear variable in an OLS regression?

(v) What is the consequence of heteroscedasticity in an OLS regression?

(vi) What is the IIA assumption in logit?

(vii) What is the consequence of heteroscedasticity in logit?

2. *Identifying Data Problems.*

a. Identify all of the dependent and independent variables.

b. What was the data source for each variable? Client? Public Source? Survey? Sample?

c. How was the data collected?

d. If a survey, were people paid to participate?

e. If a survey, how were participants selected?

f. Were there any problems in obtaining information?
g. Did certain parts of the data have to be estimated? How?
h. If certain data was unavailable, was a proxy for that data used? And what was it?
i. Did certain variables have missing observation?
j. Were observations with missing values deleted?
k. Were missing values estimated through an econometric technique? If so, what was the technique?
l. Were any of the variables transformed and why (Logs, square roots, etc.)?
m. Were influential observations identified? With what technique? Were they deleted?

3. Identifying Violations of Statistical Assumptions.
a. Have the expert identify all relevant assumptions for the statistical method being used.
b. Did the expert test for violations of each statistical assumption?
c. Did the expert test for heteroscedasticity? What test was used? What was the result?
d. Did the expert test for linearity? What test was used? What was the result?
e. Did the expert test for autocorrelation? What test was used? What was the result?
f. Did the expert test that the independent variables were not correlated with the error term? What test was used? What was the result?
g. What is the consequence of a violation of each statistical assumption?
h. What does it mean to say that an estimator is biased? Should the court give any weight to biased estimators, and in what circumstances?
i. What does it mean to say that an estimator is inconsistent? Should the court give any weight to inconsistent estimators, and, if so, in what circumstances?

4. Identify Prior Inconsistent Econometric Results.
a. What was the expert’s first theoretical hypothesis?
b. How were prior theoretical hypotheses specified?
c. Have the expert describe the results of each alternative specification.
d. Did any of the alternative specifications lead to inconsistent results?

e. Did any of the alternative specifications lead to inconclusive results?

f. Why were alternative specifications rejected?