Deposing an Econometrics Expert[1]

By Roy J. Epstein Ph.D.[2]

INTRODUCTION

Expert economic testimony using econometrics is often a vital element of antitrust litigation. Counsel needs to deal with the other side's econometrics expert and to recognize the implications of this strategy for their own side. To be effective, the deposition generally requires substantial preparation by the lawyer, who may have limited knowledge of econometrics, and close cooperation with his or her own expert in advance. This article focuses on helping an antitrust lawyer understand basic questions relevant to the reliability of econometric analysis to help structure the deposition.

Before dealing with economic analysis, the lawyer must first deal with procedural questions. Counsel from both sides need to agree on the scope of Rule 26 as it applies to items such as discoverability of drafts, computer runs, other expert work product, and expert communications with counsel. It is critical that all databases, computer programs, command scripts (e.g., MS Access commands) and spreadsheets relied on by the expert are turned over promptly to allow replication and verification. Suitable technical and IT staff should be available on both sides to iron out quickly the manifold problems that can arise in importing information from one computer system to another. It can be a great handicap if the other expert's materials are not immediately usable, especially when a clock is ticking down to the deposition date or the due date of a rebuttal report.

The lawyer also needs to identify goals for the deposition. One goal may be to establish a basis to move to limit or exclude the expert opinion. This basis could be insufficient background in econometrics or the particular type of analysis used. In addition, the expert may have failed to employ accepted statistical procedures used to screen out spurious and unreliable results.[3]

A second goal may be to identify precisely what case-specific knowledge the expert has and how the statistical analysis can be reconciled with other facts and qualitative information about the industry and conduct in question. This goes to the expert's credibility and reliability.

A third goal may be to confront the expert with reasonable alternative analyses that yield significantly different conclusions. For trial preparation, it would be important to discover how the expert intends to argue that the proffered opinions are more accurate and reliable than those based on alternative analyses.

A fourth goal may be to identify weaknesses in the analysis to facilitate settlement negotiations later on.

What is Econometrics?

Econometrics is a discipline that combines economic theory, data, and statistical methods to draw inferences from economic data. The most common type of econometric analysis is called "regression," which boils down to using the data to generate a formula to answer a question of interest. In legal settings, these questions involve determination of liability, impact, and damages. For example, one might use regression to determine if release of a pollutant lowered property values and, if so, by how much.

A regression predicts a "dependent" variable in terms of one or more "explanatory" variables. For example, an analysis of crop yield data might yield a formula:

$$\text{Yield} = 0.5 \times \text{Rain} + 2 \times \text{Fertilizer}$$

Here the coefficients of 0.5 and 2 "best fit" the rain and fertilizer data to predict the yield. In this case, the regression says that, all else equal, one can expect 0.5 more bushels per additional inch of rain and 2 more bushels per additional ton of fertilizer. Regression is widely used because in principle, given sufficient data, it allows one to quantify the individual effects of multiple explanatory variables when each one is a possible influence on the dependent variable.

There are two principal outputs from a regression analysis. First is the estimated value of each regression coefficient for the associated explanatory variable, as in the crop yield example above.

In addition, the regression procedure generates a test of the "statistical significance" of each coefficient. In any regression analysis, there is always a margin of error for each estimated coefficient known as the "standard error." The standard error arises for several reasons, e.g., using a limited amount of data to calibrate the model and the fact that a dependent variable is typically driven to some extent by randomness that cannot be explained. The standard error is used to calculate a "plus or minus" range around the estimated coefficient, called a confidence interval, to assess statistical significance.[4] If the confidence interval does *not* include the value zero, the coefficient is said to be statistically significant. On the other hand, the coefficient is not statistically significant when the interval includes zero. Lack of statistical significance means the effect of the associated explanatory variable is not statistically distinguishable from zero, regardless of the value actually calculated for the regression coefficient.

In the above example, suppose the coefficient of 0.5 on rain was statistically significant but that the coefficient of 2 on fertilizer was not significant. The results would be consistent with a claim that rain helps explain yield but that fertilizer has no effect (despite its larger coefficient). Suppose the calculated confidence interval for the fertilizer coefficient ranged from –1 to +5. The interval would indicate there is considerable uncertainty about the true magnitude of the fertilizer effect. Even though one may believe the effect is positive, and the confidence interval includes values as large as 5, the fact that the interval contains zero signifies the effect is not statistically significant. This procedure reflects a sound principle of conservatism in statistics to help guard against accepting potentially spurious "false positive" results.

In a price-fixing case, the most important coefficient in the model often is the coefficient on a so-called "dummy" explanatory variable that purports to measure the effect of an alleged conspiracy on prices. Such a variable might be coded as "1" for observations pertaining to the conspiracy and "0" for other observations.[5] For example, suppose data are available from 1990 to 2007 and a price-fixing conspiracy is alleged from 2001 to 2004. The dummy variable would be set equal to 1 for the four conspiracy years and 0 for the other time periods. The estimated coefficient on this variable is often interpreted as the magnitude of the conspiracy effect on prices, and statistical significance of this coefficient is often interpreted as empirical proof of economic impact from the conduct identified in the liability claim.

This is fairly straightforward. But regression analysis has many pitfalls that can easily result in false or misleading coefficients, regardless of statistical significance or lack thereof. Indeed, in antitrust litigation it is quite common for one side to find large and statistically significant coefficient(s) to show impact and/or damages while the other side finds small and/or statistically insignificant effects associated with the same allegations.

Returning to the crop yield example, suppose the model omitted a relevant factor such as temperature. If the data points with high fertilizer application also had a higher incidence of crop freeze, the effect of fertilizer could be masked by temperature and the fertilizer coefficient could be biased downward. A more appropriate model in this case should include an additional explanatory variable to control for temperature. Estimating this expanded regression model could show a positive and significant coefficient for fertilizer, as well as a negative coefficient for the effect of freezing temperature.

This example illustrates a fundamental issue about using regression analysis to draw conclusions about causal relationships. The estimated coefficients and the results of significance tests can be highly sensitive to which variables are included in the model, as well as other considerations such as quality and appropriateness of the data and the method used to calculate the regression coefficients.[6]

But this is not really a defect of regression. Instead, it reflects the inherent difficulty of empirical validation of alternative theories, especially in a field like economics where controlled laboratory experiments are seldom feasible. A regression model can be viewed as a valuable way to organize the available information and to help establish what assumptions go into reaching a particular conclusion. Ideally, this process should promote transparency of an analysis.

In a controversy such as a lawsuit where regression is used, there will usually be competing models leading to different conclusions yet with each side able to claim their results are plausible. The focus for the lawyer is how to establish that the other side's model is not credible or reliable, regardless of initial plausibility. A model in an antitrust case is built up from economic theory, data, statistical assumptions, and estimation methods. There are established procedures to help assess the quality of a model in these dimensions. For the deposition, the lawyer needs to become familiar with some technical vocabulary and basic statistical concepts that are grounded ultimately in common-sense. Indeed, many of the key questions to probe weaknesses in the opposing expert's econometric work are quite down to earth and require minimal use of mathematics.

Assessing the Conceptual Basis for the Model

Before getting into the details of the expert's statistical work, the lawyer should use the deposition to explore the justification for the model. For example, the author developed an econometric model successfully used by MIT to defend a case of alleged price fixing with Ivy League universities in financial aid.[7] This model used a competitive benchmark based on other schools, including public universities and liberal arts colleges as well as other private universities. It was fair game to ask why this benchmark was appropriate and the model survived this challenge.[8]

Weak or arbitrary assumptions can undermine a model, regardless of the statistical results. For example, expert economic testimony has been excluded because the empirical model was based upon an economic theory the Court found was unsupported by the facts of the case[9] or because the Court found the model was too simplistic, given the actual conditions in the industry.[10]

Generally, if there is uncertainty about how the model relates back to underlying allegations, case facts, or economic principles, the lawyer should ask for explanation. For example, how are prices negotiated for particular customers? How does the presence or absence of excess industry capacity relate to pricing? Does the model assume the identical conspiracy was in operation during the entire claimed damages period? What determines the bounds of the claimed damages period?

Every model is necessarily a simplification and will not be able to account for every fact. Nonetheless, at some point a model may ignore too many industry details to serve as a credible explanation of observed prices and other possible indicators of an antitrust violation. To determine whether this threshold has been reached, however, it should not be enough to make a laundry list of possible criticisms. Criticisms should be put to the test by suitably revising the model at issue to see if the conclusions are sufficiently robust to withstand such changes.

Assessing Statistical Reliability of an Econometric Model

A central task of the deposition is to probe how the expert will argue that the results presented are reliable. As will be discussed, many assumptions go into constructing an econometric model. The results can be very sensitive to the assumptions made. The deposition must probe the key assumptions driving the results and elicit how the expert justifies them.

Unfortunately, showing that a particular regression coefficient is statistically significant (or not) is not sufficient to prove reliability. This follows from what was said above about how significance can depend on the data, what variables are included in the model, and the method of estimation.

An assessment of reliability often is best carried out by "stress testing" the model to see if reasonable changes in approach lead to materially different conclusions. The basic suite of stress tests should focus on three main areas: the data, the choice of explanatory variables, and the estimation technique. If major sensitivities are found, the focus should then turn to in-depth scrutiny of the justification for the competing model specifications. This may be achieved through additional statistical testing or consideration of other factual information or implications of economic theory.

Some examples of stress testing in regard to each of the three areas—data, variables, and estimation technique—may be helpful here. The author was recently involved in a price fixing case where the plaintiff claimed a large and statistically significant overcharge for an industrial commodity, based on a dummy variable for the alleged conspiracy period. The prices analyzed were supposed to be for first-quality product. Yet the source data included sales at very low prices to brokers known to specialize in purchasing off-quality and waste. Moreover, these sales were made during the "non-conspiracy" benchmark period, which could skew the model. Re-estimating the model but omitting the questionable data resulted in the overcharge coefficient dropping by more than 30%.

As another example, the same model included a price index for houses as an explanatory variable although economic theory and case facts provided at best a tenuous justification. Re-estimating the plaintiff's model but omitting this variable resulted in the overcharge coefficient dropping by more than 50%.

This case also illustrated the importance of a more technical econometric issue. An econometric model uses the explanatory variables to predict each observation for the dependent variable but there will always be prediction error—no model is perfect. These errors are known as the residuals. A valid regression model requires certain properties in the residuals. In particular, when using time-series data the residuals should not display "autocorrelation," which may be evidenced by long sequences of positive residuals followed by long sequences of negative residuals.[11] Standard statistical tests exist to check for autocorrelation. It is well known that autocorrelation causes tests of statistical significance to be biased and also causes coefficients to be estimated with less precision.[12] In the plaintiff's model, there was evidence of a very high degree of autocorrelation. Employing a standard estimation method to correct for this problem resulted in the overcharge coefficient falling nearly to zero and losing statistical significance.[13]

The deposition must probe the extent to which the expert stress tested the results. An expert should be vulnerable if stress testing was superficial and the results are sensitive to reasonable changes in assumptions.

After the other expert has turned over all relevant materials, the first task is to ensure the analysis can be replicated. Then attention can shift to exploring the effect of different assumptions, and mapping out sensitivities. Replication can be surprisingly time-consuming and the exploration stage can require a significant amount of professional time for a complete and thorough analysis. A month is by no means a lot of time for this, especially in a complex case where the opposing expert may have spent upwards of a year developing a model. The lawyer would be well advised in a high-stakes case to seek more time than this, if possible.

An Initial Examination of the Model

An initial line of questioning should cover the model reported by the expert. It is important to probe the data used in the model to make sure it is accurate and relevant. The importance of this issue has grown in recent years with the increasing use of defendant sales invoice databases in antitrust modeling. For example, the transactions data may include returns, lump sum rebates, and other price adjustments that the opposing expert did not properly account for.

In addition, the regression results should be examined for plausibility in terms of algebraic sign and magnitude.[14] For example, suppose in a model to explain price that one of the explanatory variables was the cost of raw materials. A negative coefficient should be a warning that something is amiss because higher costs are expected to lead to higher prices, not lower prices. Alternatively, suppose the cost coefficient was positive but very small. The expert should explain why a large increase in costs should have only a minor effect on price.

In a model with many explanatory variables, there is a particular risk of a "kitchen sink" approach in which variables are added (or dropped) until a desired result is obtained. The problem is that, as a practical matter, the amount of data available to work with is limited. In a textbook world, each new version of a model should be estimated with a fresh set of data. In reality, the same data are often used repeatedly in what can easily be dozens of different estimation attempts.[15] This "data reuse" can greatly distort significance tests because it can amount to the modeling equivalent of gerrymandering.[16] This consideration further highlights the importance of a clear a priori economic or statistical basis for each variable in the model.

The lawyer should also ask about what the expert does not report. In particular, it is likely that the expert estimated additional models that were not reported. It is appropriate to ask how many models were estimated and how they differed in formulation. For example, data could be aggregated in different ways (across time, across firms), alternative time periods may have been used, other explanatory variables could have been used, and different statistical estimation procedures could have been employed.

The point of such questions is to assess to what extent the expert's own work may have generated results that contradict the reported model. The expert can then be asked for the basis for assuming the reported model is more reliable. Needless to say, this part of the deposition is much easier if work papers on the other models are available.

Stress Tests

Ideally, the lawyer's own expert will have subjected the model to a range of stress tests before the deposition so the results can be used to press the examination of sensitivity and reliability further. Since at this point the lawyer has probably put his or her own expert to work, these tests will not be described in detail here. Instead, the following list can be considered as a core set of stress tests that can be carried out by any capable econometrician.[17] The lawyer's expert can explain them as necessary.

Basic Stress Tests to Evaluate Econometric Models

Goal: Assess sensitivity of regression coefficients with respect to:
Autocorrelation (in conventional and panel models)
Dropping variables
Adding new variables
Adding/Removing fixed effects in panel models
Including lags
Pooling assumptions (in models with panel data)
Cross-sectional error correlation (in models with panel data)
Potential outliers and other questionable data
Level of temporal aggregation (with time series data)
Alternative functional forms
Weighting
Definition of competitive benchmark (time periods and/or geographic regions)

It is something of a mystery why so few of these tests are automated in existing commercial econometric software. By helping to identify relevant sensitivities, the stress tests facilitate more focused investigation of the key assumptions in a model.[18] The benefits include speeding up model evaluation and reducing cost. Moreover, enhancing the software along these lines has considerable potential to improve quality control in econometric work and screen out spurious results. The stress tests would greatly augment conventional regression diagnostic information such as standard errors and the Durbin-Watson statistic (a basic test for autocorrelation) and focus attention on additional aspects of model specification. The author has developed proprietary software to implement the tests on this list and it is to be hoped that other researchers will also take an interest in this area.

Analyses for Class Certification

There is an emerging trend for courts to require actual econometric work to address merits at the class certification stage.[19] The key issue typically involves empirically assessing whether impact from accused conduct can be proven for each class member using common proof. This contrasts with many older cases where plaintiffs broadly asserted that common proof of impact was possible and described general damages models without being required to prove validity of the approach for the case at hand.

Antitrust defendants presumably have an interest in developing evidence at the class certification stage that challenges both the claim that impact for all class members can be proven using common evidence and the reliability of a class-wide damages calculation.

Defendant sales databases are valuable for investigating commonality of price effects due to the great level of detail they provide. Did all customers face a similar trend in prices? Were all customers affected by the alleged conduct? Did some customers purchase products under fixed-price contracts or other contract terms that would have insulated them from the alleged conduct? At the deposition, questions like these go to the congruence of the expert's statistical analysis with the overall facts of the case.

Empirical analysis of class certification raises additional deposition questions, particularly for the defendant expert. A defendant econometric model that attacks commonality by, say, demonstrating systematic differences in pricing for different combinations of purchasers, products, and suppliers can potentially be stood on its head. The implication of such a model may be to show impact and to suggest a credible calculation of class-wide damages by identifying a tractable number of relatively homogeneous categories of products or purchasers.

Summary

Most econometric models are sensitive to one or more assumptions regarding the data, choice of explanatory variables, and the estimation procedure. Claimed findings are not reliable unless the model that generated them is stress tested and sensitivities are identified and explained. Moreover, an econometric model should be judged in relation to the overall facts of the case.

The deposition must address the expert's overall qualifications for carrying out the analysis, understanding of the relevant industry facts, and basis for specific modeling assumptions. This process requires substantial preparation by the lawyer to be most useful. The payoff is enhanced ability to prepare motions to exclude and limit scope of testimony, develop your own expert's analysis, and facilitate efficient settlement negotiations.

---

[1] This article is based on a presentation made by the author as a panelist on an ABA Antitrust Section Brownbag Teleconference on February 24, 2009. He thanks his fellow panelists Dan Gustafson, Esq. and Asim Varma, Esq., and the moderator Bruce Snapp Ph.D., for their excellent presentations and helpful discussions. He also thanks the editors, John Harkrider, Esq., Fred Knapp, Stan Young, and participants in the 2008 Fall North American Stata Users Group meeting for valuable comments.

[2] The author is an expert economic consultant in litigation matters and adjunct professor of finance at Boston College Carroll School of Management. He can be reached by email at rje@royepstein.com.

[3] The problem of "spurious" regression has been recognized for many years. *See, e.g.*, C.W.J. Granger & P. Newbold, *Spurious Regressions in Econometrics*, 2 J. OF ECONOMETRIC OF ECONOMICS 111-20 (1974), and E. Leamer, *Let's Take the Con Out of Econometrics*, 73 AM. ECON. REV. 31-43 (1983).

[4] Because statistics deals with probabilities, there is always a chance of a "false positive," i.e., that the confidence interval will not include zero even though in reality there is no connection between the dependent variable and a particular explanatory variable. The conventional confidence interval is intended to limit the risk of a false positive to no higher than 5%.

[5] This approach uses the non-conspiracy period as a competitive benchmark. When data before the start of the alleged conspiracy and after the end are used, the approach in effect treats the average of the "before and after" prices as the benchmark, taking account of the effect of the other explanatory variables. In other settings, the benchmark could be defined using observations from a different geographic area or for a different product instead of a different time period.

[6] Additional aspects of model specification that can affect the outcome of a significance test are discussed later in this article.

[7] D. Carlton, G. Bamberger, & R. Epstein, *Anti-Trust and Higher Education: Was There a Conspiracy to Restrict Financial Aid?*, 26 RAND J. OF ECON. 131-47 (1995).

[8] The model for financial aid included variables to control for different school characteristics such as public vs. private and research university vs. liberal arts college.

[9] Concord Boat Corp. v. Brunswick Corp., 207 F.3d 1039 (8th Cir. 2000). The court wrote that "the Cournot model to construct a hypothetical market…was not grounded in the economic reality of the

stern drive engine market, for it ignored inconvenient evidence." Moreover, the model "did not separate lawful from unlawful conduct."

[10] American Booksellers Ass'n, Inc. v. Barnes & Noble, Inc., 135 F. Supp. 2d 1031 (N.D. Cal. 2001). Among other criticisms, the court noted that the model at issue "uses arbitrary retention ratios to determine the extent to which plaintiffs' and defendants' sales are affected by price changes" and that the expert "arbitrarily set the retention ratios at 0.5, without attempting to determine what those ratios actually were in the real world."

[11] To illustrate autocorrelation, suppose the dependent variable is 100 observations of monthly price data. The model yields a predicted price for each month and the residual for each observation is the difference between the actual price and the predicted price. Suppose the first 50 residuals have negative algebraic sign (i.e., the model overpredicted) and the last 50 residuals have positive algebraic sign (i.e., underprediction). These residuals would be highly autocorrelated and such a pattern indicates the model is making a systematic error that should be further investigated. On the other had, suppose the sign of the residual changed randomly from one observation to the next. There would be roughly equal occurrences of four possibilities—positive followed by positive, positive followed by negative, negative followed by positive, and negative followed by negative. A valid statistical model for time-series data should display this sort of randomness in the residuals.

[12] See any introductory econometrics textbook for a discussion of the necessary properties of the residuals, e.g., Damodar Gujarati, BASIC ECONOMETRICS (5th ed. 2008).

[13] Such a dramatic change in the model result may indicate that autocorrelation was not the only problem.

[14] It is worth reviewing relevant academic studies and other models, including the expert's own prior published work, to see if the model in the report is consistent with these other sources.

[15] A good econometric practice, when sufficient data are available, is to keep some of the observations in reserve to use as a validation sample. That is, a model is selected and estimated using part of the data and then re-estimated with the validation sample to see if similar results are obtained.

[16] Technically, this problem is known as data mining and is closely related to the statistical problem of "multiple comparisons." For an amusing but also disturbing discussion in a somewhat different context see P.C. Austin et al., *Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health*, 59 J. OF CLINICAL EPIDEMIOLOGY 964-69 (2006). The study showed how uncritical reuse of the data led to "statistically significant" findings such as people born under the sign of Leo having a significantly higher probability of hospitalization due to gastrointestinal hemorrhage compared to the rest of the population.

[17] This list of stress tests by no means exhausts the types of analyses one might carry out. It is intended as a set of procedures that are reasonably general and applicable. Each model may also present fact-specific and idiosyncratic features that are difficult to anticipate or describe in advance.

[18] An econometrician will be able to complement the stress tests with appropriate "specification tests" to ascertain whether the differences in results are statistically significant.

[19] *See, e.g.,* In re Hydrogen Peroxide Antitrust Litig., 552 F.3d 305 (3d Cir. 2008).