DAUBERT AND LITIGATION-DRIVEN ECONOMETRICS

Kevin A. Kordana* and Terrance O’Reilly**

PROFESSOR D.H. Kaye’s paper raises interesting and important questions. The central issue that his paper tackles is when expert testimony employing statistics, particularly econometrics, should get to the jury. Based on the case study he discusses, his answer appears to be “not if it’s not up to snuff.” Kaye nonetheless emphasizes that the answer also must not exclude too much. In other words, we should admit studies that are less than ideal but still “within the range of reasonable debate by experts.”

Kaye argues that the admissibility of evidence based on a statistical theory or technique should turn on whether the statistical method “has been subjected to sufficient study to establish its validity as applied to a class of problems that includes the one being investigated in the litigation” (i.e., the “major premise,” or overarching scientific theory). The question “[w]hether such a method is being applied properly to the problem at hand” (the “minor premise,” or case-specific claims) would generally be left to the jury.

While as a conceptual matter this approach is appealing, we are less optimistic than Kaye is about its prospects for providing trial courts with principled guidance about the admissibility of statistical evidence. He concedes that the line between method and specific application can sometimes be difficult to fix, but we are afraid that in the case of econometric evidence and social science statistics, drawing that line will usually be problematic. Of course, there will be extreme cases in which the expert cannot justify the use of one model rather than an apparently reasonable alternative, or where

* Professor of Law, University of Virginia.
** Sullivan & Cromwell.

2 Id. at 2013.
3 Id. at 1972.
4 Id. at 1973.
5 Id. at 1972.
6 Id. at 1973.
7 Id. at 1976 (arguing that the distinction was blurred in Daubert).
there are elementary errors in a model, but those are not cases in which courts require guidance. We will argue that Kaye's discussion of Conwood Co. v. United States Tobacco Co.\(^8\) illustrates the potential problems in distinguishing between major and minor premises in evidence employing econometrics. For example, many of his criticisms of the Leftwich damage estimates are not well-suited to the task of making a decision about admissibility—because they are of the sort that are routinely leveled in the pages of refereed economics journals against other refereed articles—and, though we may misunderstand Kaye, we do not take him to be making what would seem to be the counterintuitive argument that, in general, a relevant study published in a refereed (i.e. peer-reviewed) professional journal should not be prima facie admissible.

We suspect that the distinction between major and minor premises will, more often than not, prove to be extremely arduous to apply to evidence based on econometrics. In litigation, it often will not be practicable to find an off-the-rack theory that is specific enough to answer the question of how the relevant particular market operates. Suppose you have an antitrust case involving baby food. Perhaps no one has studied the baby food market recently or at least the relevant geographical market. There are conventional techniques for approaching the question, but a study of the market for baby food might plausibly be the subject of a doctoral dissertation because modeling a particular market, identifying and collecting suitable data to test hypotheses, and tailoring the econometrics to deal with the limitations and idiosyncrasies of the available data sets typically will require more than just routine application of general principles to an arbitrary product market. It is not even obvious that litigation involving barley prices could necessarily largely rely on the methodology of a recent definitive study of the market for, say, wheat,\(^9\) but certainly novel econometric questions—"scientific" questions—are likely to arise in studying a particular biotechnology, information technology, financial or labor market, even if superficially similar markets have recently been

---

\(^8\) No. 00–6267 (6th Cir. Jan. 4, 2000).

\(^9\) Structural differences between the wheat and barley markets could include the geographic dispersion of farms, sensitivity of the grains to parasites, transportation costs, the number of exchanges trading the commodity, etc.
the subject of refereed studies or if there is a generally accepted model for them. Perhaps in certain areas—wrongful death suits and cases alleging discrimination in employment or lending, for example—judicial precedent may come to sanction a certain class of statistical procedures across a range of different markets, but in principle the line between major and minor premises may not be any easier to draw in the case of labor or credit market studies. Moreover, the fact that the courts have become comfortable with particular procedures does not guarantee that they would be considered reliable by scientists.

If we are right about the frequent unavailability of well-tested off-the-rack theory in the context of litigation, then much litigation-inspired econometrics will include largely new and previously untested "major premises," which might exclude even evidence that is as good as any single new study could be. This might well exclude more evidence than Kaye appears to favor. To analyze this, we turn to Kaye's criticisms of the Leftwich damages study for the Conwood plaintiffs, which Kaye intends to be instructive in deciding whether econometric studies should be admissible. Kaye identifies a number of "flaws" in the analysis, and for the sake of argument we assume all were presented to the trial court.

Apparently Leftwich, the plaintiff's expert, argued that the statistical evidence showed that (1) the smaller Conwood's market shares in 1990 in a particular market, the smaller Conwood's growth in that market over the period 1990-97 (when, per the plaintiff, the defendant was engaging in anticompetitive behavior), (2) this relationship between growth and market share did not hold over the 1984-90 period (a period during which the defendant was not engaging in anticompetitive conduct), and (3) the new relationship from 1990-97 demonstrated that Conwood's

---

10 This is not to deny that litigation-driven econometrics may be result-driven. Cf. Robert Thornton & John Ward, The Economist in Tort Litigation, 13 J. Econ. Persp. 101, 106 (Spring 1999) (discussing reports of "tremendous pressure placed on the economist" to deliver favorable estimates).

11 Kaye, supra note 1, at 1987-88.

12 While this makes for a useful exercise, it appears that some of the objections were waived by Conwood. See id. at 1993 n.283. Some objections do appear to have been made, however, at least with respect to damages issues. Id. at 2006 n.312.

13 Id. at 1992–93.

14 Id. at 2000–01.
growth was inhibited by defendant's anticompetitive behavior over the later, but not earlier, period, because anticompetitive behavior would be more effective the lower the defendant's market share.

Based on Kaye's description of Leftwich's analysis, it may be quibbling to worry about the grounds for rejecting Leftwich's damage estimates. Perhaps it is unnecessary to worry about finding the right statistical technique to test Leftwich's theory, since the underlying model itself does not seem to be drawn from an established theory of market behavior. Even if it were an accepted economic theory of something, it would not seem to be reliably applied to the question of the defendant's antitrust liability. Still, it is not unusual for an academic economic model of a market, as well as the statistical techniques for testing it, to be adapted creatively to the data available. If the legal test is the one that Kaye seems to advocate—with admissibility turning on whether a model is a variation of an accepted model or is "invented for use" in litigation—then prospective experts can be counted on to emphasize the inspiration for their approach as enthusiastically as academics enumerate the originality of their research. Judges, however, may find it difficult to reliably draw the line between a suitable adaptation of an accepted social science model and an entirely new one. The distinction between a new model and a variation on a model may have only legal significance. For scientific purposes, the distinction may not be significant (even if similar distinctions are drawn in decisions regarding granting tenure to university faculty, perhaps not the ideal model of the scientific method), in which case scientists' opinions on the question would not be particularly helpful. In principle, a judge who understood the degree to which a model resembled or varied from other widely used models could make the distinction, but that is a lot to expect from nonspecialist judges. Kaye's agenda for implementing Daubert does not seem to resolve

---

15 We have not reviewed Leftwich's testimony or reports.
17 Kaye, supra note 1, at 2007.
the dilemma Justice Stephen Breyer recognized: "[J]udges are not scientists and do not have the scientific training that can facilitate the making of such decisions." 18

Kaye does critique Leftwich's economic theory, 19 but Kaye's theme is statistics and econometrics, so he goes on to discuss estimation. 20 On that score, Kaye's criticisms include 21 potentially omitted variables 22 and failure to employ a proper hypothesis test. 23 He also expresses some disappointment about goodness of fit 24 and a failure to check for heteroskedasticity. 25 Some of his criticisms he would have a jury weigh—whether sensitivity to "outliers" should have been probed, for example; 26 others would be left for the judge to consider in determining admissibility. 27

It is not clear that Kaye's criticisms of Leftwich's testimony, whether more properly characterized as modeling critiques or statistical ones, support Kaye's agenda. The difficulty is that Kaye's paper does not tell us how the flaws in Leftwich's analysis are different in kind from the critiques that are leveled at articles published in refereed economics journals. 28 For example, consider some recent exchanges among academic economists:

(1) There has been a long-running dispute about proper specification in the estimation of the amount, if any, residential property values are reduced by local property taxes. A 1977 paper claims

19 Kaye, supra note 1, at 1996–97 n.292.
20 In some cases it may be difficult to draw a line between the elements of an economic model and the statistical method used to test the model, but the distinction between model and statistical method seems to be a useful one.
21 There are others, including, for example, seemingly unmotivated "modifications" to Conwood's market shares, Kaye, supra note 1, at 2003, and that a regression may not have been the best statistical technique to employ in testing the "resistance" theory, id. at 1995 n.289.
22 Id. at 1998–2000.
23 Id. at 2001.
24 Id. at 1993 n.283.
25 Id. at 1994 n.288.
26 Id. at 2012.
27 E.g., id. at 1994 n.290.
28 It would seem counterintuitive to suggest that a judge should exclude a study simply because such criticisms have some traction against it. The fact that published economics articles can later suffer sustained criticism need not suggest that such articles should have been inadmissible in court. Therefore, we need to be very cautious in assigning weight to these sorts of criticisms and using them to exclude a study.
that "[m]any studies of property values have used a clearly incor-
rect specification of the tax effect, one likely to introduce a bias of
uncertain direction and magnitude.... [In one case], it appears
that the original estimate of capitalization is about 40 percent too
great."\textsuperscript{29} A few years later, another paper appeared asserting that the "suggested estimation technique [in that 1977 paper] is flawed
by an error of econometrics."\textsuperscript{30}

(2) A 1981 study of the effectiveness of antitrust enforcement
concludes that Department of Justice enforcement against whole-
sale bakers had both remedial and deterrent effects.\textsuperscript{31} A subsequent paper challenged the results on three grounds: (1) the use of a price
index for bread that did not control for quality changes, (2) the
omission of a variable to control for retailer pricing decisions, and
(3) the sensitivity of the results to a single observation (out of
twenty) infected by the first two problems.\textsuperscript{32} The follow-up study
finds weaker results using "corrected" estimates.\textsuperscript{33}

(3) The value of environmental amenities has been estimated
through "[h]ypothetical referenda."\textsuperscript{34} A recent study questioned
the reliability of these estimates by suggesting that participants in
such referenda behaved differently than they would have if real
money had been changing hands.\textsuperscript{35} A subsequent paper criticized
that study, concluding that "[f]ailure to account for heteroskedas-
ticity when pooling real and hypothetical data results in insufficient
[sic] estimators and unreliable hypothesis tests.... lead[ing the
original study] to mistake heteroskedasticity for" differences in be-

der behavior in situations involving real and hypothetical choices.\textsuperscript{36}

\textsuperscript{29} A. Thomas King, Estimating Property Tax Capitalization: A Critical Comment, 85
\textsuperscript{30} Raymond M. Reinhard, Estimating Property Tax Capitalization: A Further
\textsuperscript{31} Michael Kent Block et al., The Deterrent Effect of Antitrust Enforcement, 89 J.
\textsuperscript{32} Craig M. Newmark, Is Antitrust Enforcement Effective?, 96 J. Pol. Econ. 1315,
1315 (1988).
\textsuperscript{33} Id. at 1322.
\textsuperscript{34} Ronald G. Cummings et al., Are Hypothetical Referenda Incentive Compatible?,
\textsuperscript{35} Id. at 615–16.
\textsuperscript{36} Timothy C. Haab et al., Are Hypothetical Referenda Incentive Compatible? A
Each of the original studies was published in the *Journal of Political Economy*, a well respected refereed economic journal—one of the oldest in the United States, which we would have thought constituted prima facie evidence of admissibility, indeed of scientific rigor. The criticisms in these comments seem to differ, at best, only in degree, not in kind, from many of the criticisms Kaye directs at Leftwich's work in *Conwood*. We are left uncertain as to how Kaye would guide (non-expert) judges in distinguishing between disputes about proper specification that are common in econometrics and those criticisms that put the challenged study beyond the pale. One might anticipate offsetting affidavits attacking or attesting to expert studies' acceptability "in an undergraduate econometrics class" or in "professional work."37

In a possibly very large fraction of the litigation involving empirical social science, Kaye's guidelines might justify—it is harder to predict how judges would actually apply them—strict scrutiny of a much greater portion of the evidence than he seems to favor or expect. To a greater degree than in many physical sciences, useful social science models are necessarily flawed. To quote National Bureau of Economic Research President Martin Feldstein, "we must learn about the world by examining a variety of estimates, each with its own biases and measurement problems, and trying to draw inferences that take these problems into account."38 In fact, to put a slight gloss on Feldstein (and perhaps exaggerate a bit), use-

---

37 Kaye, supra note 1, at 2012 (quoting Brief of Amicus Curiae Dr. Daniel L. McFadden in Support of Defendants-Appellants, *Conwood* (No. 00-6267)). Judge Richard Posner, who had significant experience with economic testimony in litigation before becoming a judge, has noted that "[e]conomometrics is such a difficult subject that it is unrealistic to expect the average judge or juror to be able to understand all the criticisms of an econometric study, no matter how skillful the econometrician is in explaining the study to a lay audience." Richard A. Posner, The Law and Economics of the Economic Expert Witness, 13 J. Econ. Persp. 91, 96 (Spring 1999). According to Judge Posner, "[t]his Problem cannot be solved, but it could be elided, by more frequent appointment of court appointed experts." Id.

ful social science models are necessarily really flawed. Results that are reproduced in a series of studies employing different methods command some respect, even if there are recognized problems with each study’s method. In litigation, however, the luxury of considering a variety of perspectives developed over time will rarely be available. To quote Feldstein again:

[W]hen economists deal with large data sets and complex econometric operations, there will be mistakes. If anyone relies on one study, he runs the risk of being misled by an error or statistical fluke. Indeed, all models are untrue in the sense that they are crude approximations of the real world.39

If it is correct that the distinction between major and minor premises is largely illusive in social science research prepared for litigation, then it is probably unrealistic to seek to have judges screen social science models prepared for litigation on the basis of their scientific pedigree. While we may be able to borrow heavily from the scientific model to avoid perverse incentives and gross inefficiencies in litigation and the shadow of litigation, we should not pretend that we scrutinize the expert because we want litigation to be guided as much as possible by objective science. We suspect it is very unlikely that enough relevant social science exists or can be produced quickly enough between the time cases are filed and they are tried. Frequently something is going to have to be jerry-rigged for the occasion, and often it is not going to be real social science—but that does not mean it has to be junk.

We would not advocate ad hoc screening of social science statistics simply because a study may not be truly scientific by academic standards. It is necessary, however, to recognize the inherent limitations of litigation social science in order to formulate systematic standards for determining its admissibility.40 This is partly an em-

40 Cf. City of Tuscaloosa v. Harcros Chems., Inc., 158 F.3d 548, 566 n.25 (11th Cir. 1998) (“[A]lthough ‘an important aspect of assessing scientific validity (and therefore evidentiary reliability) is the ability of other scientists to test or retest a proponent’s theory,’ not every scientific or technical methodology applied by expert witnesses is susceptible to such an analysis.”) (citations omitted); Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 596–97 (1993) (“[T]here are important differences
Literary question, based on the nature of the market for expert testimony and the practices of the courts. But we will be probably largely unsuccessful in formulating useful criteria for admissibility if the criteria are based on academic standards unrealistic in the context of most litigation.

Of course, it is not obvious that the Supreme Court was able to provide much guidance either. This may explain its willingness to cede authority to district courts (for example, ordering appellate courts to review their determinations only under an "abuse of discretion" standard). Nonetheless, if it is possible to provide such guidance, it would certainly be welcomed by the courts. We cannot count on most judges to be as intrepid as Judge Easterbrook in Mister v. Illinois Central Gulf Railroad when he boldly performed his own econometric analysis in an attempt to shed further light on the parties' underlying dispute.

---

42 For example, certain conventions in published academic work may be driven as much by cost-related issues (for example, the cost of publication) as by any theoretical necessity. Id. at 1511 & n.72 (discussing the 5% level for statistical significance).
44 832 F.2d 1427 (7th Cir. 1987).
45 Id. at 1432–33.