

# A Structural Perspective on the Experimentalist School

Michael P. Keane

**I**n writing an appreciation of Leamer's (1983) classic "Taking the Con out of Econometrics," it would seem mandatory to start with a good joke. Unfortunately I'm a structural econometrician, so I don't know any. So I'll have to start with a bad one. Actually, I only know one econometrician joke. It goes something like this: An official at Treasury asks three experts, "What's 200 billion plus 200 billion?" The first expert, a mathematician, immediately responds, "Four hundred billion, of course." The second, an economist, kind of grimaces and says, "Well, that depends . . ." But the third expert, an econometrician, doesn't immediately answer. Instead, he gets up and quietly closes the office door. Once he's sure no one is listening, he leans over and whispers in the official's ear, "What do you want it to be?"

I never thought this joke was very deep, but thinking about Leamer's (1983) paper made me appreciate it more. Insightful jokes typically exaggerate to make a point, so let's assume what is really being asked is a hard question like "How will consumer spending be affected by \$200 vs. \$400 billion in fiscal stimulus?" The econometrician is well aware that by playing with assumptions—what control variables and instruments to use, what functional forms to pick—it's possible to obtain pretty much any desired coefficient on government spending in the consumption function. This is precisely the problem Leamer (p. 36) talked about: "The econometric art . . . involves fitting many, perhaps thousands, of

■ *Michael P. Keane is an Australian Research Council Federation Fellow and Distinguished Research Professor, Faculty of Business, University of Technology Sydney, Sydney, Australia. He is also Visiting Research Professor, Department of Economics, Arizona State University, Tempe, Arizona. His e-mail address is <Michael.Keane@uts.edu.au>.*

doi=10.1257/jep.24.2.47

statistical models. One or several that the researcher finds pleasing are selected for reporting purposes.”

What struck me for the first time upon rereading Leamer (1983) is that the economist is really the hero of this joke. He knows what the econometrician knows, but he’s willing to admit it. In Leamer’s words, “All knowledge is human belief; more accurately human opinion.” In contrast, it is the mathematician who is really misguided, by expressing a false degree of certainty. My view, like Leamer’s, or the economist in the joke, is that there is no way to escape the role of assumptions in statistical work, so our conclusions will always be contingent. Hence, we should be circumspect about our degree of knowledge. In the words of Maimonides: “Teach thy tongue to say ‘I do not know,’ and thou shalt progress.”

### **Does the Experimentalist School Provide the Answer?**

This brings me to the paper by Angrist and Pischke (this issue). What has always bothered me about the “experimentalist” school is the false sense of certainty it conveys. The basic idea is that if we have a “really good instrument” we can come up with “convincing” estimates of “causal effects” that are not “too sensitive to assumptions.” Elsewhere I have written an extensive critique of this experimentalist perspective, arguing it presents a false panacea, and that *all* statistical inference relies on some untestable assumptions (Keane, 2010b). I won’t repeat all those arguments here. Let me instead just give a couple of examples of why natural experiments do not resolve this problem.

Consider Angrist and Lavy (1999), who estimate the effect of class size on student performance by exploiting variation induced by legal limits. It works like this: Let’s say a law prevents class size from exceeding 30. Let’s further assume a particular school has student cohorts that average about 90, but that cohort size fluctuates between, say, 84 and 96. So, if cohort size is 91–96 we end up with four classrooms of size 22 to 24, while if cohort size is 85–90, we end up with three classrooms of size 28 to 30. By comparing test outcomes between students who are randomly assigned to the small versus large classes (based on their exogenous birth timing), we obtain a credible estimate of the effect of class size on academic performance. Their answer is that a ten-student reduction raises scores by about 0.2 to 0.3 standard deviations.

This example shares a common characteristic of natural experiment studies, which I think accounts for much of their popularity: At first blush, the results do seem incredibly persuasive. But if you think for awhile, you start to see they rest on a host of assumptions. For example, what if schools that perform well attract more students? In this case, incoming cohort sizes are not random, and the whole logic breaks down. What if parents who care most about education respond to large class sizes by sending their kids to a different school? What if teachers assigned to the extra classes offered in high enrollment years are not a random sample of all teachers?

One can name many other threats to internal validity, but another problem is more fundamental: In a child cognitive ability production function, class size is but one of many inputs. Others are teacher quality, school facilities, educational philosophy, parent involvement, peer effects, and so on. Besides school inputs we also have home inputs: “quality” time with parents; parenting style; books at home; educational toys/games; time spent doing homework versus watching TV; nutrition; and so on. Unfortunately, many of these inputs are difficult or impossible to measure. Thus, we must be aware that an estimate of the “causal effect” of class size obtained via a quasi-experimental design is not a *ceteris paribus* effect. It subsumes how class size influences all the factors we haven’t controlled, including parent/school reactions to mandated class size changes.

This brings us to the key problem: Suppose we grant that experimental studies have found clear evidence that smaller class size leads to large improvements in student achievement. We still run up against the “so what?” test. Given that we don’t have estimates of the structural parameters of the cognitive ability production function, or the decision rules that parents and schools use to determine other inputs, we cannot determine if reduced class size would be a more cost effective way to improve student achievement than, say, higher teacher salaries or better nutrition and health care *in utero*.

Now, I’d be the first to admit that structural work that attempts to estimate the child cognitive ability production function also suffers from serious omitted input problems.<sup>1</sup> But if the experimental approach claims to be a “revolution,” it should be held to a high standard. As I said earlier, what bothers me is not the natural experiment approach *per se*, but rather the exaggerated claim that it enables us to attain relatively assumption-free statistical inference. In other words, I’m not dismissing the Angrist and Lavy (1999) result. I take it as one piece of evidence that may prompt us to update our priors about the relative importance of class size versus other inputs to child development. But I would not take it as definitive.

My view that we can’t escape assumptions is echoed in my favorite section of Leamer’s (1983, p. 36–37) paper, Section IV, entitled “Do we need prior information?” This has great quotations like: “[D]ata alone cannot reveal the relationship between yield and fertilizer intensity . . . we must resort to subjective prior information.” If that’s true of fertilizer, imagine the difficulty with something as complex as class size. I also like “The false idol of objectivity has done great damage to economic science” and “Because both the sampling distribution and the prior distribution are actually opinions and not facts, a statistical inference is and must forever remain an opinion.” Clearly, Leamer is rejecting the whole notion of “objective” or “assumption free” inference that the experimentalist school claims to provide.

<sup>1</sup> See Bernal and Keane (2009) for an extensive discussion of this issue in the related context of measuring effects of childcare on child outcomes using welfare rule changes as a “natural experiment.”

A second point in Angrist and Pischke that I'd like to address is the view that issues of functional form are just a "distraction" and that all we need are linear models. Actually, I have heard a pretty good joke about this (Cathcart and Klein, 2006):

Salesman: Ma'am, this vacuum cleaner will cut your work in half.

Customer: Terrific! Give me two!

A good example of why functional form matters is the Donohue and Wolfers (2005) study that Angrist and Pischke discuss. Look at Figure 1 of the Angrist and Pischke article, which is reproduced from that paper, and focus on the 1965 to 1971 period. Canada abolished the death penalty in 1965 while the U.S. retained it through 1971. During that period the homicide rate in the United States went from 5.0 to 8.5 per 100,000, while in Canada it went from 1.3 to 2.2 per 100,000. So the homicide rate goes up by 70 percent in each country, and the difference-in-difference estimate implies no effect of the death penalty. But how do we know that it is percentages, rather than levels, that matter? This is purely a functional form assumption. If the murder production function were in levels, rather than logs, we would conclude that abolishing the death penalty had lowered the murder rate by 2.6 per 100,000! (The increase in the United States with the death penalty (3.5) minus the increase in Canada without (0.9) equals 2.6.)<sup>2</sup>

## The Proof Is in the Pudding?

A main theme of the Angrist and Pischke paper is the "proof is in the pudding" argument. They claim that labor economics, utilizing experimental methods, has left other fields in the dust. Specifically, labor has developed wide consensus on a broad range of questions, while fields like macro and industrial organization remain in disarray. Hearing this claim makes me feel like Count Almaviva when he's told that Figaro is of noble blood: "Where am I? Who am I?"<sup>3</sup> As far as I can see, labor economists don't agree on much of anything.

For instance, I've been working on a survey of the labor supply literature—see Keane (2010a)—and it's clear that estimates of important quantities like Frisch, Hicks, and Marshallian labor supply elasticities are all over the map.<sup>4</sup> There is no

<sup>2</sup> Apropos of Leamer, as I'm criticizing death penalty research, I ought to be up front about my own views. My subjective prior is that the death penalty doesn't do much to crime rates. But vengeance is another justification. Personally, I'd like to see the death penalty maintained only for the most heinous crimes: For me, that includes child molestation but little else. Lately, however, I've been tempted to add the selling of credit default swaps and collateralized mortgage obligations to my list.

<sup>3</sup> "*Dove sono? Chi sono?*" *Marriage of Figaro*, Act 3, Scene 5.

<sup>4</sup> The Marshallian elasticity gives the effect of a wage change on labor supply holding nonlabor income fixed. The Hicks elasticity gives the effect of a wage change that is "compensated" by a change in nonlabor income in the opposite direction designed so the worker is no better or worse off than before. The Frisch elasticity shows how willing workers are to shift their labor towards periods when the wage is relatively high.

consensus on the central issue of how taxes affect labor supply. What's more, labor economists don't even agree on whether they agree. Many think there is a clear consensus that labor supply elasticities are small, despite the existence of many studies finding they are large.

Ironically, Angrist, Pischke, and I are among the minority who think the Frisch elasticity is large—me because I think most estimates are downward biased by ignoring human capital, they because of the behavior of bicycle messengers and stadium vendors.<sup>5</sup> If there really is consensus on key issues, then why do so few people agree with us? This notion of consensus is reminiscent of the old joke where one Upper East Side grande dame says to the other, “I voted for Willkie, you voted for Willkie, everyone we know voted for Willkie. How did Roosevelt win?”

Actually, the Angrist and Pischke case for broad consensus/progress in labor is essentially rhetorical. They list many experimental papers that have obtained “convincing” and “influential” results but rarely state what the results are—presumably because we'd see they are controversial. I only find a few *specific* results mentioned: 1) the Frisch elasticity is about one (few but me believe that); 2) neighborhood effects don't matter for earnings (do you think there is consensus on that?); 3) smaller class sizes increase achievement (direction not too controversial, magnitude certainly is); 4) the death penalty doesn't affect murder rates (what else is this controversial? Abortion? Gun control? Yankees vs. Red Sox?); 5) military service reduces civilian earnings (given the United States has an all-volunteer military, it seems that at least 1.4 million Americans may disagree).<sup>6</sup> If this is the body of “convincing” evidence the experimentalists have generated, I hardly think it constitutes a “revolution.”

## **Economics Can Learn Something from Marketing**

In contrast to labor economics, there is a field where broad consensus has actually been reached on many key issues over the past 20 years. I suspect most economists will be surprised to discover that this field is marketing. Marketing is characterized by three key features: 1) the structural paradigm is dominant, a trend that began with the dynamic demand model my coauthor and I presented in Erdem and Keane (1996); 2) the data are a lot better than in labor economics, due largely to the availability of consumer panels; and 3) there is great emphasis on external validation.

<sup>5</sup> I have an incentive to claim these experimental results are conclusive, as they support my own views. However, these studies consider monthly or daily wage fluctuations. Only under strong assumptions is this informative about intertemporal substitution at annual frequencies.

<sup>6</sup> Note that different people have different skills and will benefit from different types of training. It is likely that the types of training provided by military service increase civilian earnings for some types of people but not others. Of course, it is likely that some people choose military service primarily for other reasons (like patriotism or lack of civilian employment opportunities in the short run). Hence, for some people, military service may increase the present value of lifetime earnings or utility, even if it does not increase civilian earnings.

Interestingly, it is easy to do natural experiments in marketing. Historically, firms were quite willing to manipulate prices experimentally to facilitate study of demand elasticities. But it is now widely accepted by firms and academics that such exercises are of limited use. Just knowing how much demand goes up when you cut prices is not very interesting. The interesting questions are things like: Of the increase in sales achieved by a temporary price cut, what fraction is due to stealing from competitors vs. category expansion vs. cannibalization of your own future sales? How much do price cuts reduce your brand equity? How would profits under an every-day-low-price policy compare to a policy of frequent promotion? It is widely accepted that these kinds of questions can only be addressed using structural models—meaning researchers actually need to estimate the structural parameters of consumers’ utility functions. As a result, the “experimentalist” approach has never caught on.

Relying heavily on structural econometric models, good data collection, and serious attempts at model validation, the field of marketing has reached broad consensus on all the questions noted above, among others. Here I will just discuss one topic, which also happens to be relevant to the Angrist and Pischke paper.

Consider the demand for frequently purchased consumer goods. There is broad consensus that own-price elasticities (given temporary price cuts) are about  $-3$  to  $-4.5$ . But, as noted above, the dynamics are much more interesting. In Erdem, Imai, and Keane (2003) and Erdem, Keane, and Sun (2008), my coauthors and I estimate that roughly 20–30 percent of the increase in sales due to a temporary price cut is cannibalization of future sales. Of the remaining incremental sales, 70–80 percent is due to category expansion and only about 20–30 percent is due to brand switching. A remarkable consensus has emerged on these figures in recent years.<sup>7</sup> It is hard to exaggerate the importance of this three-way decomposition of the price elasticity of demand, as it determines the profitability of price promotion. The analogous situation in labor economics would be if there were broad consensus on all the key labor supply elasticities and the labor supply effects of tax cuts.

Now let me turn to industrial organization. Angrist and Pischke argue that progress in industrial organization has been hindered by reliance on structural econometrics. The point they emphasize most is the failure of structural models of industry competition to predict accurately the effects of mergers on prices. But in my view, this failure is not surprising, because structural industrial organization models rely on static models of consumer demand. There is a broad consensus in marketing that static demand models greatly exaggerate cross-price elasticities,<sup>8</sup> as they attribute most of the incremental sales that accompany a price cut to brand switching and little to category expansion or cannibalization. And when

<sup>7</sup> Some key papers on cannibalization rates are van Heerde, Leeflang, and Wittink (2000, 2004) and Ailawadi, Gedenk, Lutzky, and Neslin (2007). Some important studies of brand switching are Pauwels, Hanssens, and Siddarth (2002), van Heerde, Gupta, and Wittink (2003), Sun, Neslin, and Srinivasan (2003), and Macé and Neslin (2004).

<sup>8</sup> See, for example, Keane (1997), where I first noted the problem theoretically, and Sun, Neslin, and Srinivasan (2003) and Erdem, Keane, and Sun (2008), papers which verified its importance empirically.

I say “exaggerate,” I’m not talking small potatoes—I’m talking factors of two to four. As cross-price elasticities of demand summarize the degree of competition between products, it’s obvious that such large biases will create serious problems in attempting to predict effects of mergers.

Thus, the problem with industrial organization—at least the part of the field dealing with models of industry competition—is not the use of structural models *per se*, but rather that the demand side of those models is typically static and hence, by consensus, badly misspecified. Given this, it isn’t surprising that industrial organization models do a poor job of forecasting effects of mergers.<sup>9</sup> I agree with Angrist and Pischke that it is puzzling there have been so few attempts to validate these models by looking at price effects of mergers. But I’m also puzzled by the failure of industrial organization to incorporate much of what marketers have learned about consumer demand. There is broad consensus that static demand models fit choice behavior terribly (for example, see Ching, Erdem, and Keane, 2009), so why does the field of industrial organization persist in using them?<sup>10</sup>

## Good Data Always Helps

There is another key point by Angrist and Pischke with which I agree: the experimentalist school has done a great service to empirical economics by forcing researchers to pay more attention to the sources of variation in data that identify their models. In my recent survey of the labor supply literature, I was struck by the cavalier approach to identification in many papers, even many of recent vintage. Just as Angrist and Pischke state (in discussing crime), “the use of instrumental variables . . . was typically mechanical, with little discussion of why the instruments affected the endogenous variables” or why we would expect them to be uncorrelated with the stochastic terms. As an example, to identify effects of wages on labor supply, one needs a source of wage variation that is uncorrelated with tastes for work. The typical labor supply paper deals with this problem via rather arbitrary exclusion restrictions. A common one is to assume that age and education affect wage offers for employment, but not tastes for work. This assumption is obviously debatable, yet in many papers it is made casually and without comment. Thanks to the experimentalist school, it is harder to get away with this sort of thing now.

However, the fact that we should pay close attention to sources of identifying variation in the data is not an argument for abandoning structural econometrics.

<sup>9</sup> In fact, in Erdem, Imai, and Keane (2003), having found that static models seriously exaggerate cross-price elasticities of demand, we predicted that existing models of competition would do a poor job of predicting effects of mergers. At the time, we were unaware of any papers that attempted to validate those models.

<sup>10</sup> Actually, I think there are two reasons that static demand models persist: First, the computational demands of solving an equilibrium model with dynamic consumer demand are substantial. Second, industrial organization models are often estimated using aggregate data, which makes individual demand dynamics essentially impossible to identify.



Plausibly exogenous variation in variables of interest is a desideratum in all empirical work—not an argument for one approach over another. Consider Erdem and Keane (1996). In that paper, my coauthor and I introduced the structural approach into marketing, where it rapidly became quite pervasive. But why was the paper so influential? One factor is that many found the structural model appealing; consumers learn about brand attributes via use experience and advertising signals, and brand equity (or “loyalty”) emerges naturally as risk-averse consumers are reluctant to buy unfamiliar products. Prior empirical work had treated brand equity as a black box and posited no structural mechanism for its development.<sup>11</sup>

But at least as important is that the paper produced a big result: it provided a reliable estimate of the long-run effect of advertising on brand equity and consumer demand. This had been a “holy grail” of marketing research, but prior work had failed to uncover reliable evidence that advertising affected demand at all—an embarrassing state of affairs for marketers!

Why did we find evidence of long-run advertising effects when others had not? Was it the use of a structural model? I think that helped, but the key reason is that we had great data. Specifically, we had scanner data where households were followed for years, and their televisions were monitored so we could see which commercials each household saw. If you are willing to believe that tastes for brands of detergent are uncorrelated with tastes for television shows (which seems fairly plausible), this is a great source of exogenous variation in ad exposures. I agree that all econometric work, *whether structural or not*, should ideally be based on such plausibly exogenous variation in the data.

## **The Ability to Do Controlled Experiments Does Not Obviate the Need for Theory**

Where I most strongly disagree with Angrist and Pischke is their notion that empirical work can exist independently from, or occur prior to, economic theory. In Keane (2010b), I argue that “we cannot even begin the systematic assembly of facts and empirical regularities without a preexisting theoretical framework that gives the facts meaning and tells us which facts we should establish.” I argue this is true not just in economics, but in all scientific disciplines. Thus, I found it interesting that Angrist and Pischke briefly extend their analysis outside of economics to the field of medicine and argue that an experimentalist approach has been fruitful there as well: “in medicine . . . clinical evidence of therapeutic effectiveness has often run ahead of doctors’ theoretical understanding of disease.”

I was left wondering how Dr. Owen Wangensteen would react to this observation. If you’ve not heard of Wangensteen, let me put it this way: if you can think

<sup>11</sup> The “Erdem–Swait” framework (Erdem and Swait, 1998) is now considered the canonical *economic* model of brand equity. (There are also a number of psychology-based models.) See Keller (2002) for an overview. The astute reader will notice that I obviously ought to go learn something about marketing.



of any famous surgeons, they were probably either his students or students of his students.<sup>12</sup> Wangensteen, who was surgeon-in-chief at the University of Minnesota Hospitals (1930–1967), played a key role in inventing modern medical education by reforming the curriculum, in concert with medical school dean Elias Lyon, to require that surgeons receive grounding in basic science (including biology, physiology, and other fields).

In 1928–32, Wangensteen embarked on a long series of controlled experiments designed to investigate the *mechanisms* that induced gaseous distension in obstructed intestines, a major cause of death following abdominal surgery. As far as I can tell, these experiments involved doing really odd things to dogs and seeing how long it took them to die (Edlich and Woods, 1997). Obviously, no one would do this sort of stuff unless they were either: 1) a sadist, or 2) had a theory in mind they were trying to test. In fact, Wangensteen did have a theory: that the mechanism causing gaseous distension was not primarily buildup of toxicity in the intestine but instead just swallowed air. His experiments showed his theory was correct. This led to his famous nasogastric suction procedure that is thought to have saved the lives of 100,000 U.S. troops with abdominal wounds in World War II. Visscher (1991) estimates that the Wangensteen procedure, lauded in a 1951 poem by Ogden Nash, had saved about a million lives by 1991.<sup>13</sup>

But along with his clinical contribution, Wangensteen also set an important example by his approach: the idea that understanding mechanisms is important for developing improved surgical procedures. Late in his career Wangensteen collaborated with his wife (an historian) on a history of surgery (Wangensteen and Wangensteen, 1978). This work is described by Visscher (1991) as “a comprehensive treatise on the emergence of surgery from primitive empiricism to the utilization of modern scientific and technological advances.” Economic empiricists regard the randomized clinical trials conducted in medicine as the “gold standard” towards which economics should strive. Yet, as should be obvious, a scientist doesn’t do experiments or run clinical trials as “accidental play, without pre-existence of more or less definite ideas about their meaning” (Einstein and Infeld, 1938). Theory forms the basis for empirical work in the science of medicine just like anywhere else.

## **Different Approaches to Model Validation**

Finally, let me return to Section IV of Leamer’s (1983) paper. I think the most important passage here (p. 38) is “the fundamental problem facing econometrics is how adequately to control the whimsical character of inference, how sensibly to base inferences on opinions . . .” I don’t think Leamer had any derogatory intent

<sup>12</sup> As Edlich (2007) notes, 110 full professors and 38 department heads were students of Wangensteen. Probably the most famous are the great cardiothoracic surgeons: C.W. Lillehei (often called “the father of open heart surgery”), Norman Shumway, and Christiaan Barnard.

<sup>13</sup> Ogden Nash wrote: “May I find my final rest in / Owen Wangensteen’s intestine / knowing that his masterly suction / will assure my resurrection.”

in choosing the word “whimsical.” He was simply stating the obvious: all inferences are based on some ultimately untestable assumptions that must, therefore, be made based on the “whim” (or opinion) of the investigator. He’s asking how, in such a subjective world, we can produce results that are credible (or at least useful) to others. Leamer’s answer was that we should report on sensitivity to specification. As an example, he reports estimates of a regression of murder rates on execution rates using sets of control variables that would be considered appropriate by, for example, 1) a bleeding-heart liberal,<sup>14</sup> 2) a social conservative, and 3) an economic determinist. The idea is that the researcher thus provides evidence of value to all three audiences, given their prior views, and also reveals to what extent estimates are determined by a set of prior beliefs. In effect, the Angrist–Pischke reaction is: “Don’t be so nihilistic. Let’s just come up with instruments that are so good our results will be convincing to everyone.” This is the false certainty to which I referred earlier. In response, I’d suggest reading Maimonides’ quote at the start of this article, or perhaps Plato’s *Apology*, 21–23.

It’s interesting to ask how Leamer’s (1983) ideas about specification testing apply to structural econometrics. Speaking for myself, it usually takes about two years to program up and estimate a structural model. Writing a program to solve the optimization problem faced by economic agents, and to estimate the parameters of their utility functions, is a lot more work than just running regressions! Thus, if I find my model fits poorly, or produces parameter values that seem odd *a priori*, I know that making some changes to the model and reestimating will take a few months, as it typically involves substantial rewriting of the original code. In this way structural work is fundamentally different from regression analysis. It’s not possible to fit “thousands” of models and report the one you like. Depending on how many years one wants to devote, it might be feasible to estimate five or ten specifications, none of which can differ too dramatically from the one with which you started. Thus, I would contend that specification searches are not a big problem in structural work—it just takes too long to do them. Faster computers won’t change this situation, as much of the time involved is programming time.

For this reason, the best structural work has not involved extensive specification testing, but rather careful external validation exercises designed to persuade the audience to take the researcher’s model seriously. Examples of such validation exercises can be found in Keane and Moffitt (1998) and Keane and Wolpin (2007). Both papers fit structural models of welfare program participation and labor supply. The former uses the model to attempt to forecast (really backcast) behavior prior to a significant regime change, while the latter fits the model to a subset of U.S. states and attempts to predict behavior in a holdout state with a very different policy regime. In Keane (2010b), I discuss a range of other validation techniques. A key point is that structural econometricians do not perform these exercises to

<sup>14</sup> For younger readers, “bleeding-heart” was a term used in the 1970s to distinguish bad liberals, like Eugene McCarthy, from good liberals, like Hubert Humphrey. Now we know that all liberals are bad, so the phrase has fallen into disuse.

persuade the audience that the model is “true.” We know perfectly well that our models aren’t true. Validation exercises are used purely as a way to persuade the audience (and ourselves) that a model may be a useful tool for prediction and policy evaluation.

To conclude, I once heard Noam Chomsky say at a public lecture that if a field spends a lot of time debating methodology it’s a sure sign it’s not getting anywhere. Maybe we should all get back to work.

■ *I’d like to thank the editors David Autor and Timothy Taylor, along with three anonymous referees, for helpful comments, and assistant editor Ann Norman for exceptional assistance in preparing the manuscript (in particular, for detecting an error in one of my examples).*

## References

- Ailawadi, Kusum L., Karen Gedenk, Christian Lutzky, and Scott A. Neslin. 2007. “Decomposition of the Sales Impact of Promotion-induced Stockpiling.” *Journal of Marketing Research*, 44(3): 450–67.
- Angrist, Joshua D., and Victor Lavy. 1999. “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics*, 114(2): 533–75.
- Bernal, Raquel, and Michael Keane. 2009. “Child Care Choices and Children’s Cognitive Achievement: The Case of Single Mothers.” Available at: [http://economia.uniandes.edu.co/profesores/planta/Bernal\\_Raquel/documentos\\_de\\_trabajo](http://economia.uniandes.edu.co/profesores/planta/Bernal_Raquel/documentos_de_trabajo).
- Cathcart, Thomas, and Daniel Klein. 2006. *Plato and a Platypus Walk Into a Bar: Understanding Philosophy through Jokes*. New York: Abrams Image.
- Ching, Andrew, Tülin Erdem, and Michael P. Keane. 2009. “The Price Consideration Model of Brand Choice.” *Journal of Applied Econometrics*, 24(3): 393–420.
- Donohue, John J., and Justin Wolfers. 2005. “Uses and Abuses of Empirical Evidence in the Death Penalty Debate.” *Stanford Law Review*, vol. 58, pp. 791–845.
- Edlich, Richard F. 2007. “In Memoriam: A Tribute to Dr. Owen H. Wangensteen, the Greatest Teacher of Surgery During the 20th Century (1898–1981).” *Journal of Surgical Research*, 138(2): 241–53.
- Edlich, Richard F., and Julia Woods. 1997. “Wangensteen’s Transformation of the Treatment of Intestinal Obstruction from Empiric Craft to Scientific Discipline.” *Journal of Emergency Medicine*, 15(2): 235–41.
- Einstein, Albert, and Leopold Infeld. 1938. *The Evolution of Physics*. New York: Simon and Schuster.
- Erdem, Tülin, and Joffre Swait. 1998. “Brand Equity as a Signaling Phenomenon.” *Journal of Consumer Psychology*, 7(2): 131–57.
- Erdem, Tülin, Susumu Imai, and Michael P. Keane. 2003. “Brand and Quantity Choice Dynamics under Price Uncertainty.” *Quantitative Marketing and Economics*, 1(1): 5–64.
- Erdem, Tülin, and Michael P. Keane. 1996. “Decision Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets.” *Marketing Science*, 15(1): 1–20.
- Erdem, Tülin, Michael P. Keane, and Baohong Sun. 2008. “A Dynamic Model of Brand Choice when Price and Advertising Signal Product Quality.” *Marketing Science*, 27(6): 1111–25.
- Keane, Michael P. 1997. “Current Issues in Discrete Choice Modeling.” *Marketing Letters*, 8(3): 307–22.
- Keane, Michael P. 2010a. “Labor Supply and Taxes: A Survey.” Available at: <http://www.bus.uts.edu.au/fin&econ/staff/MichaelK/research.html>.
- Keane, Michael P. 2010b. “Structural vs. Atheoretic Approaches to Econometrics.” *Journal of Econometrics*, 156(1):3–20.

- Keane, Michael P., and Robert Moffitt.** 1998. "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review*, 39(3): 553–89.
- Keane Michael P., and Kenneth I. Wolpin.** 2007. "Exploring the Usefulness of a Non-Random Holdout Sample for Model Validation: Welfare Effects on Female Behavior." *International Economic Review*, 48(4): 1351–78.
- Keller, Kevin.** 2002. "Branding and Brand Equity." In *Handbook of Marketing*, ed. B. Weitz and R. Wensley, 151–78. London: Sage Publications.
- Leamer, Edward.** 1983. "Let's Take the Con Out of Econometrics." *American Economic Review*, 73(1): 31–43.
- Macé, Sandrine, and Scott A. Neslin.** 2004. "The Determinants of Pre- and Postpromotion Dips in Sales of Frequently Purchased Goods." *Journal of Marketing Research*, 41(3): 339–50.
- Pauwels, Koen, Dominique M. Hanssens, and S. Siddarth.** 2002. "The Long-Term Effects of Price Promotions on Category Incidence, Brand Choice, and Purchase Quantity." *Journal of Marketing Research*, 39(4): 421–39.
- Sun, Baohong, Scott A. Neslin, and Kannan Srinivasan.** 2003. "Measuring the Impact of Promotions on Brand Switching under Rational Consumer Behavior." *Journal of Marketing Research*, 40(4): 389–405.
- van Heerde, Harald J., Sachin Gupta, and Dick R. Wittink.** 2003. "Is 75% of the Sales Promotion Bump Due to Brand Switching? No, Only 33% Is." *Journal of Marketing Research*, 40(4): 481–91.
- van Heerde, Harald J., Peter S. H. Leeflang, and Dick R. Wittink.** 2000. "The Estimation of Pre- and Postpromotion Dips with Store-Level Scanner Data." *Journal of Marketing Research*, 37(3): 383–95.
- van Heerde, Harald J., Peter S. H. Leeflang, and Dick R. Wittink.** 2004. "Decomposing the Sales Promotion Bump with Store Data." *Marketing Science*, 23(3): 317–34.
- Visscher, Maurice B.** 1991. "Owen Harding Wangenstein." Chap. 18 in *Bibliographic Memoirs*, National Academy of Sciences. Washington, DC: National Academy Press. Available at [http://books.nap.edu/catalog.php?record\\_id=6061#toc](http://books.nap.edu/catalog.php?record_id=6061#toc).
- Wangensteen, Owen, and Sarah Wangenstein.** 1978. *The Rise of Surgery—from Empiric Craft to Scientific Discipline*. University of Minnesota Press.

**This article has been cited by:**

1. John K. Dagsvik. 2017. Invariance axioms and functional form restrictions in structural models. *Mathematical Social Sciences* . [[Crossref](#)]
2. Hakan Seckinelgin. Evidence-Based Policy: Randomised Controlled Trials' Knowledge Claims to AIDS Policy 105-124. [[Crossref](#)]
3. Carlianne Patrick, Amanda Ross, Heather Stephens. Designing Policies to Spur Economic Growth: How Regional Scientists Can Contribute to Future Policy Development and Evaluation 119-133. [[Crossref](#)]
4. 2016. The Psychology of Human Risk Preferences and Vulnerability to Scare-Mongers: Experimental Economic Tools for Hypothesis Formulation and Testing. *Journal of Cognition and Culture* **16**:5, 383-414. [[Crossref](#)]
5. Michael P. Keane, Nada Wasi. 2016. How to model consumer heterogeneity? Lessons from three case studies on SP and RP data. *Research in Economics* **70**:2, 197-231. [[Crossref](#)]
6. Don Ross. 2016. Introduction to discussion forum on Glenn W. Harrison's 'field experiments and methodological intolerance'. *Journal of Economic Methodology* **23**:2, 127-129. [[Crossref](#)]
7. Richard A. Ashley, Christopher F. Parmeter. 2015. Sensitivity analysis for inference in 2SLS/GMM estimation with possibly flawed instruments. *Empirical Economics* **49**:4, 1153-1171. [[Crossref](#)]
8. Richard A. Ashley, Christopher F. Parmeter. 2015. When is it justifiable to ignore explanatory variable endogeneity in a regression model?. *Economics Letters* **137**, 70-74. [[Crossref](#)]
9. Thor O. Thoresen, Trine E. Vattø. 2015. Validation of the discrete choice labor supply model by methods of the new tax responsiveness literature. *Labour Economics* **37**, 38-53. [[Crossref](#)]
10. Judea Pearl. 2015. TRYGVE HAAVELMO AND THE EMERGENCE OF CAUSAL CALCULUS. *Econometric Theory* **31**:01, 152-179. [[Crossref](#)]
11. Franz Wirl. 2015. Output adjusting cartels facing dynamic, convex demand under uncertainty: The case of OPEC. *Economic Modelling* **44**, 307-316. [[Crossref](#)]
12. Rolf Aaberge, Ugo Colombino. Labour Supply Models 167-221. [[Crossref](#)]
13. Bas van der Klaauw. 2014. From micro data to causality: Forty years of empirical labor economics. *Labour Economics* **30**, 88-97. [[Crossref](#)]
14. FRANZ WIRL, SEBASTIAN CABAN. 2014. A RATIONALIZATION OF UPS AND DOWNS OF OIL PRICES BY SLUGGISH DEMAND, UNCERTAINTY, AND NONCONCAVITY. *Natural Resource Modeling* **27**:2, 178-196. [[Crossref](#)]
15. A. Agrawal. 2014. Matching and mechanisms in protected area and poverty alleviation research. *Proceedings of the National Academy of Sciences* . [[Crossref](#)]
16. G. W. Harrison. 2014. Cautionary notes on the use of field experiments to address policy issues. *Oxford Review of Economic Policy* **30**:4, 753. [[Crossref](#)]
17. Andrew T. Ching, Tülin Erdem, Michael P. Keane. 2013. Invited Paper —Learning Models: An Assessment of Progress, Challenges, and New Developments. *Marketing Science* **32**:6, 913-938. [[Crossref](#)]
18. Yves Breitmoser. 2013. Estimation of social preferences in generalized dictator games. *Economics Letters* . [[Crossref](#)]
19. Glenn W. Harrison. 2013. Field experiments and methodological intolerance. *Journal of Economic Methodology* **20**:2, 103-117. [[Crossref](#)]
20. Franz Wirl. 2012. OPEC's Strategies. *Zeitschrift für Energiewirtschaft* **36**:3, 227-237. [[Crossref](#)]

21. Peter C. Reiss. 2011. Structural Workshop Paper —Descriptive, Structural, and Experimental Empirical Methods in Marketing Research. *Marketing Science* **30**:6, 950-964. [[Crossref](#)]
22. Kevin Milligan. 2011. The design of tax policy in Canada: thoughts prompted by Richard Blundell's 'Empirical evidence and tax policy design'. *Canadian Journal of Economics/Revue canadienne d'économie* **44**:4, 1184-1194. [[Crossref](#)]
23. François Claveau. 2011. Evidential variety as a source of credibility for causal inference: beyond sharp designs and structural models. *Journal of Economic Methodology* **18**:3, 233-253. [[Crossref](#)]
24. G. W. Harrison. 2011. Randomisation and Its Discontents. *Journal of African Economies* **20**:4, 626-652. [[Crossref](#)]
25. Judea Pearl. 2011. Statistics and Causality: Separated to Reunite-Commentary on Bryan Dowd's "Separated at Birth". *Health Services Research* **46**:2, 421-429. [[Crossref](#)]
26. John A. List, Imran Rasul. Field Experiments in Labor Economics 103-228. [[Crossref](#)]
27. Robert J. Sampson. 2010. Gold Standard Myths: Observations on the Experimental Turn in Quantitative Criminology. *Journal of Quantitative Criminology* **26**:4, 489-500. [[Crossref](#)]
28. James J. Heckman. 2010. Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature* **48**:2, 356-398. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]