The Empirical Economist's Toolkit: From Models to Methods

Matthew T. Panhans and John D. Singleton

In 2010, the *Journal of Economic Perspectives* hosted a symposium revisiting Edward E. Leamer's provocative 1983 article, "Let's Take the Con out of Econometrics." Taking aim at existing econometric practice, Leamer (1983, 31) had posited that econometricians project on themselves the false image of a "white coat" experimental scientist:

> The applied econometrician is like a farmer who notices that the yield is somewhat higher under trees where birds roost, and he uses this as evidence that bird droppings increase yields. However, when he presents this finding at the annual meeting of the American Ecological Association, another farmer in the audience objects that he used the same data but came up with the conclusion that moderate amounts of shade increase yields. A bright chap in the back of the room then observes that these two hypotheses are indistinguishable given the available data. He mentions the phrase "identification problem," which, though no one knows quite what he means, is said with such authority that it is totally convincing.

Correspondence may be addressed to Matthew T. Panhans, Duke University, Department of Economics, 213 Social Sciences, Campus Box 90097, Durham NC 27708; e-mail: mpanhans @gmail.com; and John D. Singleton, Department of Economics, 280 Hutchinson Road, Box 270156, University of Rochester, Rochester, NY 14627; e-mail: john. singleton@duke.edu.

History of Political Economy 49 (annual suppl.) DOI 10.1215/00182702-4166299

Rather, claimed Leamer, empirical researchers' inferences were troublingly subject to "whimsical assumptions" and subjective judgments.[1] Ending with a plea for systematic examination of the sensitivity of econometric results, Leamer concluded: "If it turns out that almost all inferences from economic data are fragile, I suppose we shall have to revert to our old methods lest we lose our customers in government, business, and on the boardwalk at Atlantic City" (43).

In their contribution to the *Journal of Economic Perspectives* symposium, Joshua D. Angrist and Jörn-Steffen Pischke argue that, nearly twenty years after Leamer's critique, "better research design is taking the con out of econometrics." They identify instrumental variables, regression discontinuity, and difference-in-differences analyses as "quasi-experimental" methods to justify casual inference that have "grown and become more self-conscious and sophisticated since the 1970s" (Angrist and Pischke 2010, 12). Pointing to examples from the economics of crime, education, and health, Angrist and Pischke trumpet the application of these methods as a "credibility revolution" in econometrics that has finally provided persuasive answers to a diverse set of questions: "It's no longer enough to adopt the language of an orthodox simultaneous equations framework, labeling some variables endogenous and other exogenous, without offering strong institutional or empirical support for these identifying assumptions" (116). They single out the fields of industrial organization and macroeconomics as the sole exceptions to the shift.[2]

Angrist and Pischke's article not only picked up on points in an ongoing methodological debate within applied economics (Heckman 1997; Angrist and Imbens 1999; Rosenzweig and Wolpin 2000); it precipitated a number of responses (Sims 2010; Nevo and Whinston 2010; Keane 2010a, 2010b; Rust 2010; Wolpin 2013). Nonetheless, Angrist and Pischke's article was less an argument than it was a victory lap. For instance, of the nine economists awarded the John Bates Clark Medal within the prior fifteen years, the research of at least five, most notably David Card and Steven Levitt, could be counted as quasi-experimental. The trend has only accelerated since, with four of the five most recent winners applying quasi-experimental

1. "The econometric art as it is practiced at the computer terminal involves fitting many, perhaps thousands, of statistical models. One or several that the researcher finds pleasing are selected for reporting purposes" (Leamer 1983, 36).

2. In both of these fields, however, an increasing amount of work employs quasi-experimental tools. See Fuchs-Schuendeln and Hassan 2015.

methods in their work. Moreover, the central importance of "credibly" identified econometric work is attested in the training of applied economists, where it is reinforced in chapters and handbooks (Angrist and Pischke 2009, 2014; Imbens and Lemieux 2008; Lee and Lemieux 2010; Lewbel 2016), and in the rhetoric and working discourse among applied practitioners.[3] This paradigm shift has also been institutionalized in schools of public policy, applied economics, and public health, in turn influencing the way economics engages with neighboring disciplines and with policymakers (Hirschman and Berman 2014). The methods thereby constitute a key feature of the transition to applied economics and shape the knowledge produced by economists and the expertise they possess (Fourcade, Ollion, and Algan 2014). As summarized recently by a graduate student: "We [economists] have the computing power, we have the data sets, we understand identification issues, and the combination of the three makes the analysis much more credible than in the past, and therefore more readily consumed by policy makers" (Colander 2005, 192–93).

While historians of economics have noted the transition in the character of economic research since the 1970s toward applications (see, e.g., the 2000 *History of Political Economy* conference volume), less understood is the shift toward quasi-experimental work. Duo Qin's (2013) history of recent econometrics, for instance, does not discuss the methods. However, quasi-experimental methods can be usefully juxtaposed with structural applied work, with its roots in the Cowles econometric program.[4] Quasi-experimental research designs employ the logic of experimentation. Working with observational data, researchers seek a "natural experiment" or a context in which subjects are assigned to treatment and control groups as if part of an experiment—"to exploit situations where the forces of nature or government policy have conspired to produce an environment somewhat akin to a randomized experiment" (Angrist and Krueger 2001, 73).[5] Then, by examining the differences in outcomes between groups, the researcher obtains estimates of causal effects. Common techniques exploit variation in the timing of policy changes across subjects (difference-in-differences) or known selection rules (regression discontinuity) to make such comparisons.

3. It is not unusual for a research presentation in an applied economics field to focus heavily on the plausibility of exclusion restrictions.

4. See Heckman 2010 and Angrist and Imbens 1999. Card (2012) labels the alternative approaches "design-based" and "model-based."

5. Note that we use the terms *natural experiments* and *quasi-experiments* interchangeably.

By contrast, structural research aims to link empirical work explicitly with economic theory by modeling agents' optimization problems and decision rules. The goal is to estimate the fundamental (structural) parameters determining outcomes and behavior (generally production and utility functions), which can then be used for welfare and counterfactual analyses. Typically taking the form of systems of equations, estimation often requires specifying the entire data-generating process. Economic theory, in the sense of explicitly stated behavioral assumptions, is important for both specifying the estimating equations and for justifying identification. In a number of fields, labor and education perhaps foremost among them, the growth of quasi-experimental work has displaced structural approaches, while in other areas, such as health, they have expanded the boundaries of economic science to subjects and questions detached from choice and optimizing frameworks.

It is from this vantage that the shift toward quasi-experimental designs can be viewed as a transition from models to methods. Taking seriously Jacob Viner's quip that "economics is what economists do" (Backhouse and Medema 2009), we explore the transition through a focus on the shifting *practices* in which empirical economists engage (Pickering 1992; Stapleford 2015). This article therefore aims to illuminate the origins, content, and contexts of quasi-experimental methods and their applications. To do this, we first present bibliometric evidence that traces the use of quasi-experimental terms in the titles and abstracts of articles published in top general interest and field journals in economics. Beginning in 1990, there is a sharp upward climb in the count of articles not only in the top general interest journals but also across a wide selection of field journals, where the particular techniques appear with different frequencies.

Second, to highlight the lines of continuity and discontinuity, we situate quasi-experimental approaches in the historical context of the Cowles econometric program. Although the individual methods that make up the quasi-experimental program were originally developed well outside this framework, the experimental ideal behind causal economic relationships, elaborated by Trygve Haavelmo in *The Probability Approach* (1944), provides one link. To contrast the change in empirical practices associated with the application of quasi-experimental methods, we then analyze a case study regarding a long-standing question in applied economics: the returns to education. Comparing Willis and Rosen 1979 and Angrist and Krueger 1991, the appeal of the quasi-experimental approach is empha-

sized along with the subtle but significant reorientation of the role of economic theory in applied work.

Finally, we identify a confluence of trends that fed into the stabilization of the "credibility revolution" to explore the significant historical contexts of the paradigm shift. Alongside the roles of vast increases in computing power and the availability of micro data, quasi-experimental methods were synthesized within the preexisting econometric framework and the established conceptual categories, such as exogeneity and endogeneity, providing stability. Additionally, quasi-experimental methods were suited to meet the demands of patrons of economic research, particularly policymakers. The emphasis on "transparency" of the techniques for obtaining "credible" causal effects for "evidence-based policy evaluation" underscores the marketability of quasi-experimental approaches. Furthermore, as Leamer's critique (1983) represents well, the perception of structural methods' inadequacies and failures into the 1980s, a discernible crisis in econometrics, is another key context. These contexts, both internal and external to economics, constitute only a partial account of the "credibility revolution" in applied economics and its implications, which continue to evolve, however. The narrative is thus intended to sketch the larger project of contextualizing the shift by circumscribing the issues, personalities, and institutions involved, yielding a number of rich historical questions.

Bibliometric Evidence

This section presents bibliometric evidence of the shift to quasi-experimental applied work in economics since the 1970s. This development is traced in both field journals and top general-interest journals using data from the Web of Science online database of academic journals. For each journal, the number of articles using a quasi-experimental term (*regression discontinuity*, *natural experiment*, *difference-in-differences*, and *randomized control trials*, as well as variants of these terms) in either the title or the abstract are counted annually. In restricting attention to only abstracts and titles, the absolute numbers will undercount the number of quasi-experimental papers, though the trends should remain representative.

To demonstrate that quasi-experimental methods are applied broadly in applied economics, eleven top-tier field journals were sampled: *Journal of Development Economics*, *Economic Policy*, *Journal of Health Economics*, *Journal of International Economics*, *Journal of Econometrics*, *Journal of*

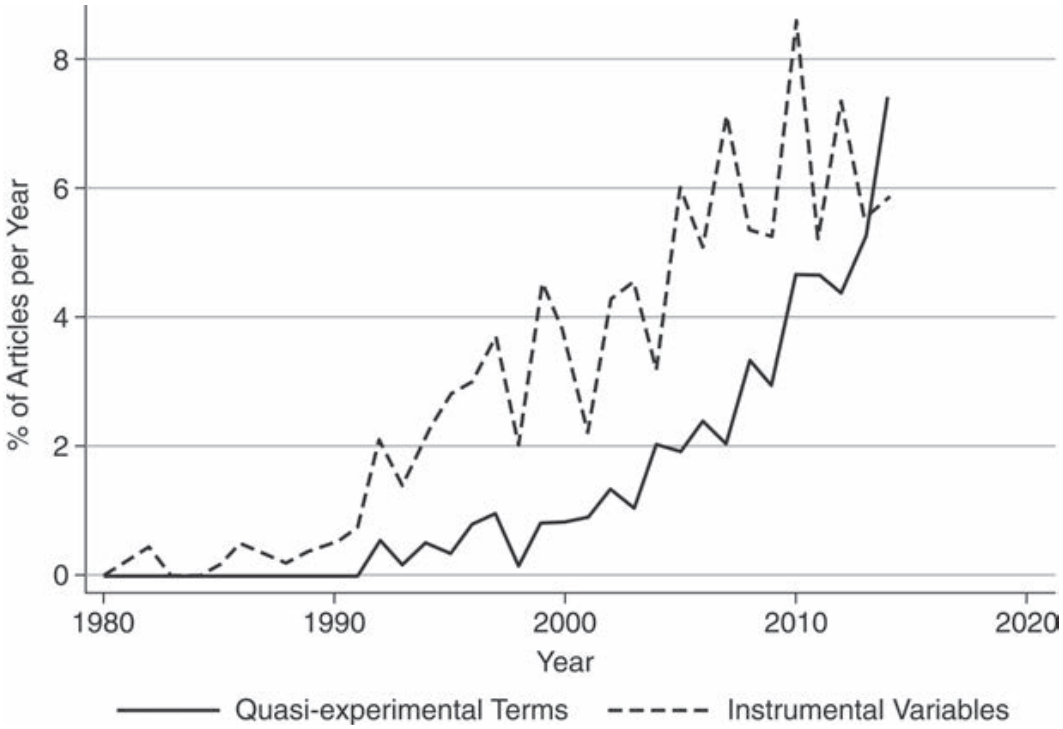

Figure 1 Uses of quasi-experimental terms and "instrumental variable" in top economics field journals

Note: "Quasi-experimental terms" include *difference-in-differences*, *regression discontinuity*, *natural experiment*, and *randomized control trial*. Data for this and following figures are drawn from Web of Science.

Economic Behavior and Organization, *Journal of Human Resources*, *Journal of Labor Economics*, *Journal of Public Economics*, *Journal of Urban Economics*, and *Review of Economics and Statistics*. Coincident with the growth of applied economics, many of these field journals were started in the 1970s. Figure 1 shows a count of the fraction of articles using a quasi-experimental term in the title or abstract from among these field journals. Beginning around 1990, there is a sharp upward climb in the percentage of articles using quasi-experimental terms. In terms of absolute numbers, the count for 2010, the year of the *Journal of Economic Perspectives* symposium, represents around a 400 percent increase over the year 2000. Figure 1 also shows the results of the same exercise performed for the term *instrumental variable*. Although instrumental variables have a long history independent of quasi-experimental methods, we find it useful to trace their history here because instrumental variables are

included in the applied economist's causal toolkit, as well as their role in stabilizing quasi-experimental methods in the discipline of economics, as described in later sections. This similar upward trend thus corroborates the shift not only to applied work generally but also to quasi-experimental research designs.

To gain a sense of how the application of quasi-experimental methods varies across fields, the same exercise is completed for specific selected field journals. Figure 2 illustrates that there is an increase in the use of the quasi-experimental techniques in health economics around 2000. Further, the application of difference-in-differences appears to be a commonly used research design—a finding that fits expectations, as the field deals largely with the US health care system and uses research designs that exploit variation in regulations across states. By contrast, figure 3 verifies that randomized control trials are particularly widespread in development economics, and especially in just the last few years.[6] The overall trend is also reflected in the *Journal of Public Economics*, as seen in figure 4, with a particularly pronounced recent trend in work using regression discontinuity.

While the trends in field journals in economics are compelling, it is useful to also examine percentages for the top four economics journals, which reflect the highest-profile general-interest research in the profession: the *American Economic Review, Econometrica, Quarterly Journal of Economics*, and the *Journal of Political Economy.* As figure 5 illustrates, the sharp upward turn in quasi-experimental terms in abstracts and titles remains evident. Likewise, the share for instrumental variable in figure 5 suggests 1990 as an inflection point.[7]

Together, the graphs suggest a significant and recent shift toward quasi-experimental methods. In addition, this shift appears not only in the top general interest journals but also across a wide selection of field journals, where the particular techniques appear with different frequencies. This documents both the timing and the extent of the "credibility revolution" in applied economics, as viewed through the publication record. In the following sections, the history of econometrics and a case study from the returns to education are examined to characterize the content of the

6. For debates about the roles of experimental and observational research in development, see Teele 2014.

7. This timing is consistent with Bowen et al. (2015), who use bibliometric methods to date the adoption of quasi-experimental techniques in finance to the mid-nineties.

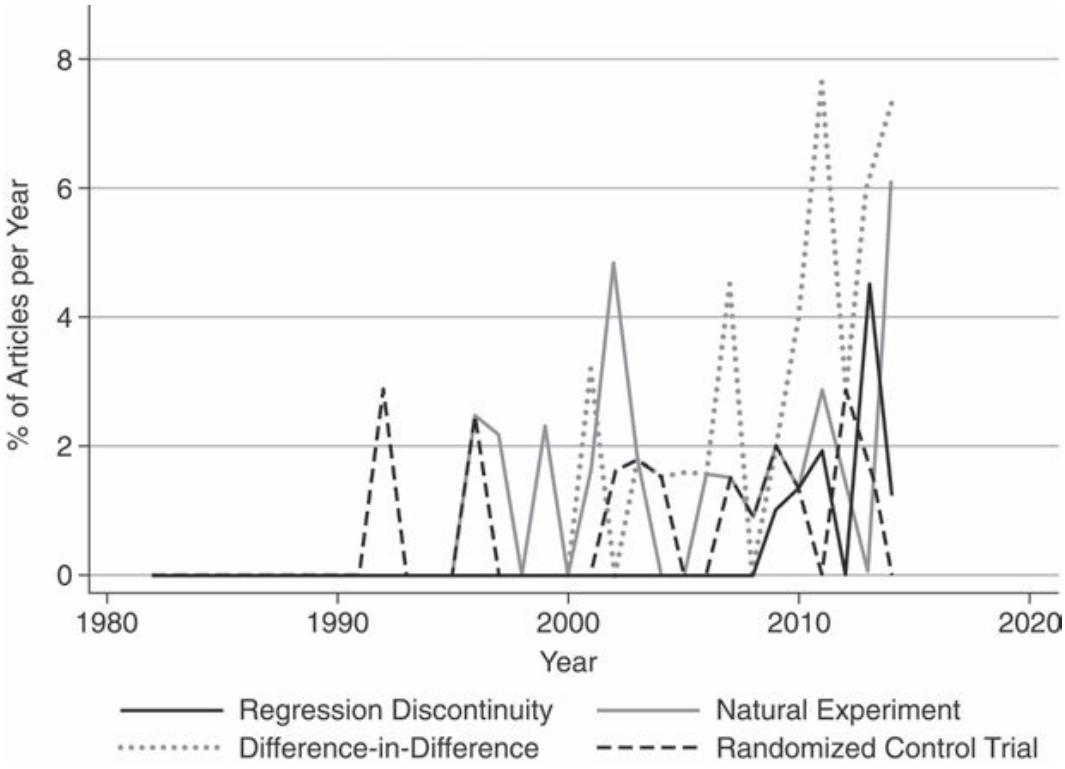

Figure 2. Uses of quasi-experimental terms in the *Journal of Health Economics*

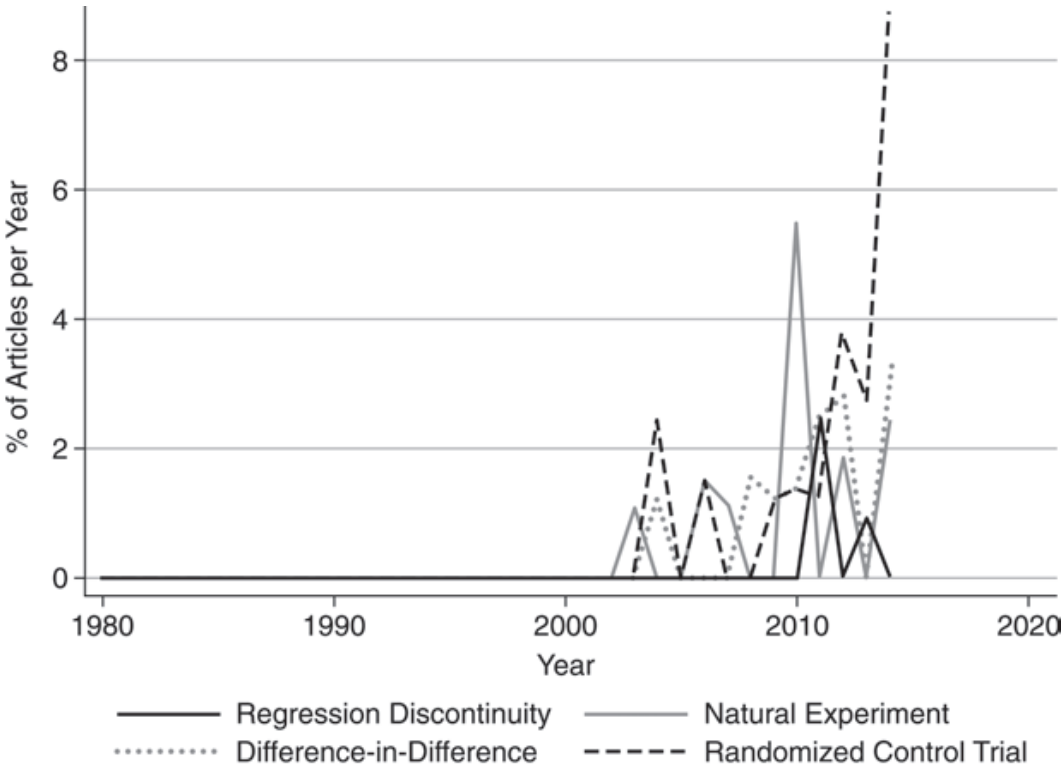

Figure 3. Uses of quasi-experimental terms in the *Journal of Development Economics*

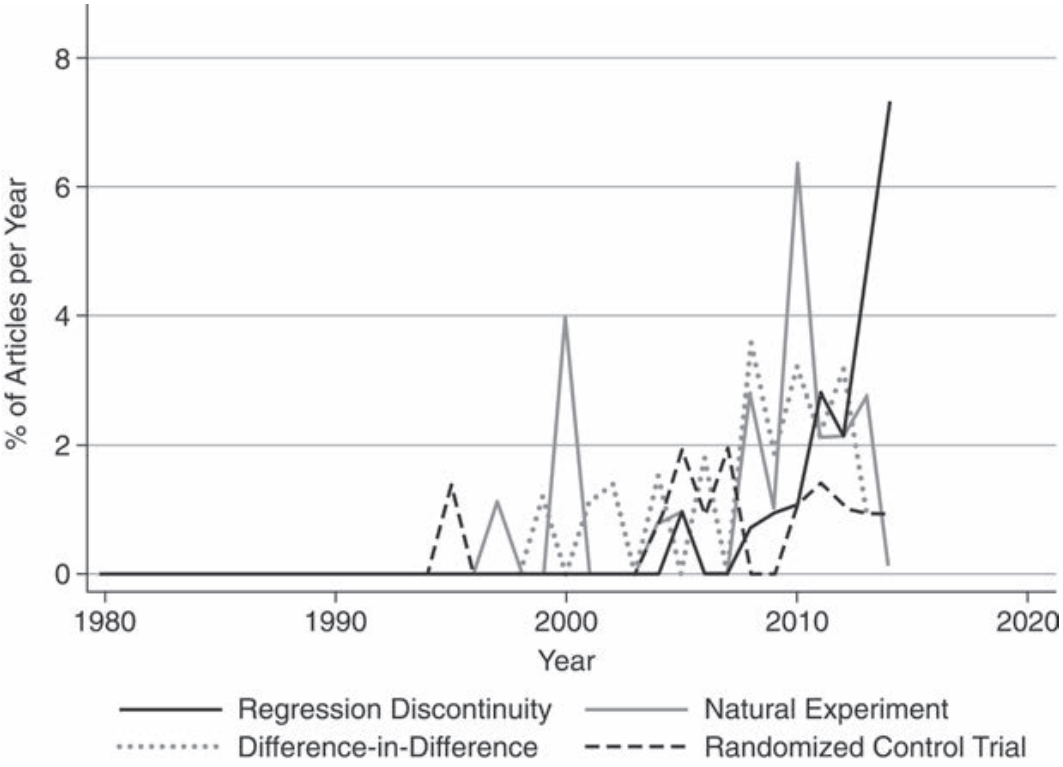

Figure 4. Uses of quasi-experimental terms in the *Journal of Public Economics*

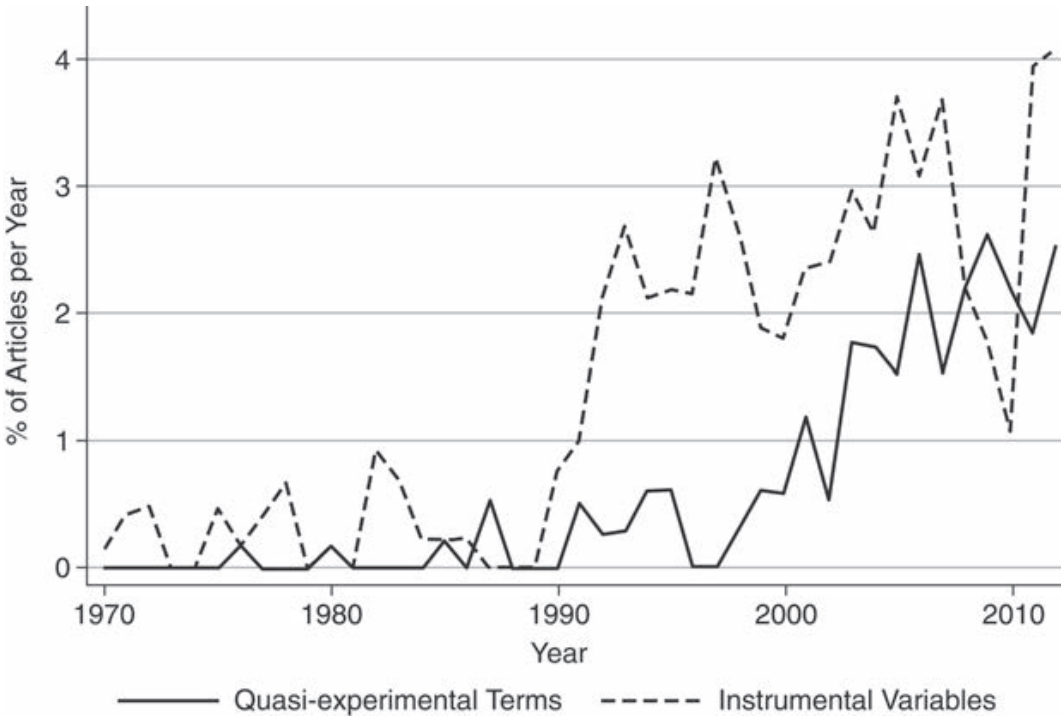

Figure 5. Uses of quasi-experimental terms and "instrumental variable" in top four economics journals

shift before looking in greater detail at the stabilization and contexts of these changes.

The Structure of Econometrics

While quasi-experimental work has become influential in a number of applied fields, econometrics was shaped historically by the aspirations of researchers connected to the Cowles Commission, which was founded in 1932 (Qin 2015). As James Heckman (2000, 46) summarizes: "The lasting legacy of this research program includes the concepts of exogenous (external) and endogenous (internal) variables, and the notions of 'policy invariant parameters' and 'structural parameters' which have entered everyday parlance inside and outside of economics." This context is therefore important for understanding the elements of continuity and discontinuity in the growth of quasi-experimental methods.

The Probability Approach

Estimation of equations governing the economy was a goal of the early Cowles Commission (De Marchi and Gilbert 1989). The setting for this research was macroeconomic policy, as researchers wanted to evaluate the effects of different policies. Ragnar Frisch pioneered this approach with his work on business cycles in the early 1930s. The models included economic variables ("consumption," "capital starting," and "carry-on-activity"), which were composed of component cycles, and "the period and form of these component cycles were dependent in turn on the structural parameters of the system he had set up" (Morgan 1990, 94). In this early work, Frisch (1933, 185) himself did not estimate the parameters but calibrated values: "At present I am only guessing very roughly at these parameters, but I believe that it will be possible by appropriate statistical methods to obtain more exact information about them." In the late 1930s it was Jan Tinbergen, following Frisch's suggestion of using data to estimate structural parameters, who developed the first macroeconometric models and used them to conduct a wide range of policy simulations (Morgan 1990, 100).

By the 1950s researchers at the Cowles Commission had made breakthroughs in two areas of economics: general equilibrium theory and theoretical econometrics. Those associated with the econometrics revolution

included Jacob Marschak, Tjalling Koopmans, and, perhaps the most influential, Trygve Haavelmo (Christ 1994). Haavelmo's 1944 monograph *The Probability Approach in Econometrics* set out to encourage economists to recast their questions in ways amenable to the tools of modern statistics. This required thinking of economic models and relationships in terms of probability models, to which statistical inference could be applied. Kevin D. Hoover (2014) demonstrates how central Haavelmo was to the Cowles econometrics program; he was credited by colleagues at Cowles as launching their research program, and later researchers viewed their own influential work as largely extensions of Haavelmo's ideas.

For Haavelmo, developing empirically meaningful economic models demanded careful consideration of the economic relationships. To explain why, he applied terminology from Frisch on "confluent" and "autonomous" relationships. The degree of autonomy of a set of relationships indicated how invariant they were to changes in the system. Confluent relationships, on the other hand, were those derived from other relationships and thus could only decrease in autonomy. Haavelmo (1944, 27) used an example from engineering to illustrate this in *The Probability Approach*:

> If we should make a series of speed tests with an automobile, driving on a flat, dry road, we might be able to establish a very accurate functional relationship between the pressure on the gas throttle (or the distance of the gas pedal from the bottom of the car) and the corresponding maximum speed of the car. And the knowledge of this relationship might be sufficient to operate the car at a prescribed speed. But if a man did not know anything about automobiles, and he wanted to understand how they work, we should not advise him to spend time and effort in measuring a relationship like that. Why? Because (1) such a relation leaves the whole inner mechanism of a car in complete mystery, and (2) such a relation might break down at any time, as soon as there is some disorder or change in any working part of the car.

Accordingly, the goal of econometrics was to elucidate relationships in the economy with the highest degree of autonomy, akin to the highly invariant laws of physics: "On the other hand, the general laws of thermodynamics, the dynamics of friction, etc., etc., are highly autonomous relations with respect to the automobile mechanism, because these relations describe the functioning of some parts of the mechanism irrespective of what happens in some other parts" (27). Though a relationship might be

causal and estimated precisely, it is less useful if it depends on many other particular relationships holding. In the example of the gas pedal and car speed, for example, as soon as the car is on a different type of road or the car breaks down, the previously estimated relationship becomes less meaningful. Likewise, a researcher must think carefully in order to uncover "autonomous" or invariant—that is, structural—economic relationships.

With respect to economic models, this often meant starting from a system of simultaneous equations, such as supply and demand equations that jointly determine prices and quantities. This structure introduced complications to estimation, and the Cowles econometric program revealed the biases inherent in linear regression estimation that did not take into account the simultaneity in such situations. They focused on estimation methods based on specifying a probabilistic model derived from the joint distribution of the two (or more) variables: "And then, to avoid inconsistencies . . . all formulae for estimating the parameters involved should be derived on the basis of this joint probability law of all the observable variables involved in the system" (Haavelmo 1943, 7). Estimation of the parameters typically entailed maximum likelihood techniques.[8]

Estimating the Returns to Education

In the late 1950s, concurrent with the postwar expansion of higher education in the United States, economists became interested in education with its role in economic growth and the distribution of earnings in mind. The human capital program, which traces its heritage to early contributions by Jacob Mincer (1958), Gary Becker (1964), and Theodore Schultz (1961), developed in many ways around estimating the private return to education and the ensuing econometric difficulties encountered.[9] Further, widely available longitudinal data on labor market outcomes, such as the National Longitudinal Surveys first conducted by the Census Bureau in the 1960s, expanded the research opportunities.

A collection of researchers at the University of Chicago was particularly influential in the human capital program's development. In addition to Becker and Schultz, applied economists like H. Gregg Lewis studied

8. Epstein (1989) details the origins of LIML in the Cowles program and "fall" of ordinary least squares.

9. Teixeira 2000 and 2011 present this history.

labor market outcomes related to labor force participation, job training, and unionization, while econometricians, most notably Heckman, worked on the problems of estimation and identification. One difficulty of wide applicability concerned selection, in which those choosing an alternative are unrepresentative of the population, such that naive estimates of rates of return, the underlying parameters of interest, are biased.[10]

In the education setting, selection is particularly salient, and Sherwin Rosen, a Lewis student who returned to Chicago as a professor in 1977, aimed to apply frontier techniques in a paper with Robert Willis "to estimate life earnings conditioned on actual school choices that are purged of selection bias" (Willis and Rosen 1979, S8). In their joint paper, published in the *Journal of Political Economy* in 1979, Rosen and Willis specify a model in which students weigh the costs and benefits, in terms of their expected lifetime income, to attending college. The paper, and the ways it connects with the Cowles program, is useful as a case study to build contrasts with later quasi-experimental investigations of the returns to schooling.

In Willis and Rosen's model, expected lifetime earnings, Y_{ij}, of person i choosing schooling level j are determined by the individual's observed talent or ability indicators, $\boldsymbol{X}_i$, and unobservables, τ_i.Writing this formally and generally,

$$Y_{ij} = y_j (X_i, t_i), \quad j = 1,...n \tag{1}$$

This equation embeds the rates of return to different levels of education. Similarly, family background and taste indicators of person i are split into observable components in vector Z_i and unobserved components ω_i. Then, letting V_{ij} denote the value of choosing schooling level j for person i produces a second equation for selection, that is, the choice of educational attainment:

$$V_{ij} = g(Y_{ij}, Z_i, \omega_i) \tag{2}$$

where i belongs to j if $V_{ij} = \max(V_{iI}, \ldots, V_{in})$. The person thus chooses the education level to maximize his or her lifetime earnings, and the empirical specification of the model is derived from that dynamic optimization problem. The model is closed by also specifying a parametric distribution of the unobserved components in each equation. The model's structure,

10. As an example, parents who send their children to private school may also be more highly educated and of higher income.

consisting primarily of two equations whose parameters correspond to costs and rate of return, thereby fits comfortably within the ambitions of the Cowles program. Willis and Rosen estimate the model in three steps using the NBER-Thorndike sample of male World War II veterans who applied for the Army Air Corps and were eligible for the GI Bill.

Before presenting their results, Willis and Rosen (1979, 18) discuss identification of the model parameters. In particular, they note, as can be seen by substituting (1) into (2), that "estimation of the selection rule or structural probit equation is possible only if the vector *X* and *Z* have elements that are not in common." If the same variables enter both equations, the direct contribution to selection cannot be separated from the indirect, through lifetime earnings, contribution. In their discussion, they "tentatively associated" *X* with "measured abilities" and *Z* with "measured financial constraints" and tastes ("corresponding to the Beckerian distinction between factors that shift the marginal rate of return to the investment schedule and those that shift the supply of funds schedule" (19). However, Willis and Rosen concede that "easy distinctions" between the vectors are "increasingly difficult, if not impossible, to make" (19).[11]

To resolve this impasse, Willis and Rosen make a consciously extreme assumption: there are no commonalities between the vectors. Thus, in their empirical work, *X* includes only ability indicators, given by test scores, while *Z* contains only family-background variables, including parents' education and religious background. Though they do not phrase it in this way, the variables in *Z* are assumed to be instruments wherein lifetime earnings are only influenced by the family-background variables through their effect on the selection of schooling levels—an exclusion restriction. Willis and Rosen's justification for this assumption is instructive. They state that "it provides a test of the theory in its strongest form. Certainly, if the theory is rejected in this form there is little hope for it" (Willis and Rosen 1979, 19). The data, in principle, can reject the model.[12] The Willis and Rosen paper thus provides an illustrative case study of how the Cowles approach was practically implemented in applied research to contrast, in the next section, with a quasi-experimental approach.

11. They add: "If *X* and *Z* are indistinguishable, the economic theory of school choice has no empirical content" (Willis and Rosen 1979, 19).

12. Notably, just three years prior, Mark Blaug (1976) had taken the human capital program to task in the *Journal of Economic Literature* for failing to develop testable hypotheses.

Emulating Experiments

Cowles researchers operationalized Alfred Marshall's notion of ceteris paribus in a regression framework as the effect on a variable *Y* when other variables *X* are held constant. Similarly, Haavelmo (1944, 6) argued in *The Probability Approach* that econometric researchers should have experiments in mind when developing models:

> What makes a piece of mathematical economics not only mathematics but also economics is, I believe, this: When we set up a system of theoretical relationships and use economic names for the otherwise purely theoretical variables involved, we have in mind some actual *experiment*, or some *design of an experiment*, which we could at least imagine arranging, in order to measure those quantities in real economic life that we think might obey the laws imposed on their theoretical namesakes.

Although, as described in this section, the individual methods that make up the quasi-experimental program were largely originally developed outside the Cowles econometric framework, this experimental ideal or standard for causal relationships provides an enduring continuity with the Cowles program. The implementation of quasi-experimental methods, though, reveals the elements of discontinuity.

The Search for Identification

While quasi-experimental methods—instrumental variables, difference-in-differences, and regression discontinuity—are jointly implicated in the "credibility revolution," the methods have prehistories situated in different places and communities, from agricultural economics to psychology and statistics. For instance, Mary Morgan (1990) attributes the invention of the instrumental variable regression to Philip G. Wright in 1928.[13] Wright's 1928 work investigated the market for flaxseed, and in appendix B he used an instrumental variables regression to estimate the elasticities of supply and demand for flaxseed. Perhaps because it was tucked away in an appendix of a detailed 285-page examination of the flaxseed industry, the method did not receive citations in the literature at

13. There is some debate, however, over whether the instrumental variables regression was invented by Wright's son Sewall Wright, a distinguished biostatistician. James Stock and Francesco Trebbi (2003) use stylometrics to investigate this question of authorship and conclude it was indeed Philip who wrote appendix B.

the time, and it was not until the 1950s that the method would be revived and extended. Henri Theil first developed the method of estimation using instrumental variables known as two-stage least squares, which facilitated estimation with multiple instruments, in 1953. Later work, which synthesized this heritage and the Cowles program in econometrics, formalized identification (Qin 1989) and developed tests and heuristics for what constituted a "strong enough" instrument or evidence an instrument is valid, such as the Sargan test.

Regression discontinuity designs, which take advantage of known selection rules such as cutoffs or thresholds, probably have the most unusual prehistory. As detailed by Thomas Cook (2008), this technique was developed by a group of psychologists at Northwestern University in the 1960s. The first publication using the approach was Thistlethwaite and Campbell 1960, which studied the effect of receiving recognition from a national merit scholarship program on students' ambitions and attitudes. Donald Campbell, who continued to work in developing the regression discontinuity approach at Northwestern for some time after this publication, was initially concerned with selection, which can bias causal estimates. In this 1960 paper, the authors used test scores to compare students just above and below the cutoff for receiving national merit scholarship recognition. Although the psychologists at Northwestern collaborated with mathematicians and statisticians to work out many of the problems with regression discontinuity, according to Cook (2008) the method soon fell out of favor in the field of psychology and never really caught the attention of statisticians. It would not be until the 1990s that the method would be revived, as it gained popularity in the field of economics.

At the same time, it is natural to attempt to situate the history of quasi-experimental work with respect to the long history of social or field experiments in economics. Levitt and John List (2008) trace the history of experiments to Jerzy Neyman and Ronald Fisher, whose work in the 1920s and 1930s laid the conceptual foundations for how randomization could lead to insights from experimentation with agricultural plots. The late 1960s and early 1970s saw field experiments gain a new role in social policy, with a series of income maintenance (or negative income tax) experiments conducted by the US federal government. The New Jersey experiment from 1968 to 1972 "was the first large-scale attempt to test a policy initiative by randomly assigning individuals to alternative programs" (Munnell 1986). Another touchstone is the RAND Health Insurance Experiment, started in 1971, which lasted for fifteen years. Nonetheless, the history of social and field experiments in economics is nonlinear

with respect to the epistemology that characterizes the quasi-experimental program. The "second period" of interest in experiments (in Levitt and List's 2008 classification) represented interventions designed to produce estimates of structural parameters—subordinating the method to the priorities embedded in an economic model (Breslau 1998, 102). This context of the history is returned to in the last section.

Perhaps the oldest of quasi-experimental methods is difference-in-differences, which simply compares a treatment and control group over time. If each individual in the sample is assigned to the treatment or control group in a way that at least resembles random assignment, the effect of the treatment can be estimated as the difference in the outcome before and after treatment between the treatment and control groups. The assumptions, limitations, and conditions necessary for a valid design were clarified in the 1990s, but the use of the framework goes back much farther even in economics. A 1915 Bureau of Labor Statistics report, for example, investigated the implementation of the minimum wage in Oregon, where the minimum wage was set to different levels in Portland compared with the rest of the state, to understand the effects on wages and employment of women working in retail (Obenauer and Nienburg 1915, cited in Angrist and Pischke 2009). However, it would not be until much later that this method came to be known as a difference-in-differences (or "diff-in-diff") and an important part of every empirical economist's training, in no small part because of the influence of Orley Ashenfelter.

Ashenfelter received his PhD in economics from Princeton University in 1970 and soon after was appointed as the director of the Office of Evaluation at the US Department of Labor. His office was tasked with evaluating job training programs that had been funded by the government, including the Manpower Development and Training Act (MDTA). A key difficulty encountered was that, even with employment and earnings histories of the program participants, it was difficult to evaluate the effect of job training without a suitable comparison group (Ashenfelter 2014). Government survey data allowed for a random sample of the population to serve as a comparison group, but there was the further challenge that labor market conditions could change over time. It was the longitudinal nature of the data that allowed for a solution: "In short, the difference from the pre to the post period in earnings for the treatment group could be compared against the difference from the pre to the post period for the comparison group. This is the origin of the 'difference-in-differences' method that has come to dominate discussions in labor economics and, in fact, is found in much of the empirical study of economics more generally" (Ashenfelter 2014, 576).

After his time at the Department of Labor, Ashenfelter eventually returned to a faculty position at Princeton. In addition to his promoting the use of quasi-experimental methods in economics through both his research and his role as editor of the *American Economic Review* beginning in 1985, he also advised some of the most influential practitioners who followed in this tradition, including 1995 John Bates Clark Medal winner David Card and, another of the foremost promoters of quasi-experimental methods, Angrist. While faculty at Princeton, Card and then Princeton colleague Alan Krueger conducted what has become the most famous use of a difference-in-differences design in economics, the Card and Krueger (1994) minimum wage study. The authors compared the employment of fast-food workers in New Jersey before and after the 1992 minimum wage increase, and used fast-food workers in the state of Pennsylvania as a control. Their paper suggested the still-controversial finding that increasing the minimum wage did not decrease employment in the fast-food industry. These scholars played instrumental roles in the growth of quasi-experimental work and its stabilization as a self-conscious program and toolkit.

The Causal Effect of an Additional Year of Schooling

In the early 1990s researchers returned to examining the returns to education, but began employing quasi-experimental designs. In his Fisher-Schultz Lecture to the Econometric Society, Card (2001) posits three reasons for this resurgence in interest. Card, like the early motivations for the human capital program, first points to a recognized disparity between more- and less-educated workers and mentions interest in the determinants of economic growth. However, he adds that the interest was "heightened by the belief that some progress has been made—and more may follow—in the very difficult task of uncovering the causal effect of education in labor market outcomes" (Card 2001, 1127). That is to say, the existing answers were unsatisfying and, with the application of new methods, credible answers might be forthcoming.

A seminal paper in this vein published in the *Quarterly Journal of Economics* by Angrist and Krueger (1991, 979) has a familiar ring over two decades on: "This paper exploits an unusual natural experiment to estimate the effect of compulsory schooling laws in the United States." As the authors detail, the experiment rests on the interaction of two institutional features: school-entry requirements and compulsory schooling laws. School-entry

requirements stipulate that students born before a given date in the calendar year must enter schooling. As a consequence, students on the other side of the date are not so compelled. Compulsory schooling laws, on the other hand, require school attendance up to a certain age, usually sixteen or seventeen. The joint effect is that students of the same age on either side of the entry date are able to drop out of school at the same age, but one will have completed one more year of schooling. These two groups thereby represent "control" and "treatment" groups, and the difference in their labor market outcomes is the causal effect of an additional school year.

Crucial to the research design is that the groups are otherwise the same, that is, it is as if the month in which an individual was born were randomly assigned. Birth month can serve as an instrument that shifts schooling but can be excluded from the estimating equation. If this assumption holds, then this approach provides an alternative way to deal with omitted ability and selection biases. To support validity, Angrist and Krueger argue that month of birth is uncorrelated with other personal attributes. Their estimation then becomes quite straightforward, accomplished via two-stage least squares. In the results, they find little difference between ordinary least squares and two-stage least squares estimates, suggesting little, if any, bias.[14]

Despite a common long-standing question in the return to schooling, the Angrist and Krueger research design displays illustrative contrasts with Willis and Rosen 1979. In the first place, the empirical specification is not derived from an explicit optimization problem or economic model. In particular, unlike Willis and Rosen 1979, the paper does not contain the system of equations structure, wherein selection is modeled and estimated simultaneously with the outcome equation. Instead, economic theory is reserved for motivating the question or interpreting the result, though even little of this is in the paper; it is the effect of compulsory schooling laws themselves that is estimated without reference to costs or rates of return.

The second key point of contrast concerns the justification for identification. Whereas Willis and Rosen separate the determinants of outcome and of selection by reference to human capital theory and to falsificationism, the institutional features that interact to create the natural experiment in Angrist and Krueger 1991 provide the required instrument. In a loose sense, the instrument, rather than an internal element explicitly modeled, is outside even the background model, not unlike an experimenter's intervention. As a result, the behavioral assumptions needed for validity are seemingly less

14. In other words, some of the very considerations that motivated Willis and Rosen's approach appear not to actually matter at all, at least in this context.

demanding. It is important not to understate the perceived advantages of this; the validity of the Angrist-Krueger design does not rest on believing arbitrary exclusion restrictions implied by an economic theory, a point Angrist and Krueger reemphasized ten years later: "We would argue that this approach contrasts favorably with studies that provide detailed but abstract theoretical models, followed by identification based on implausible and unexamined choices about which variables to exclude from the model and assumptions about what statistical distribution certain variables follow" (Angrist and Krueger 2001, 76). In other words, the "con" is removed. In sum, as the case study reveals, the application of the quasi-experimental design represents a subtle but significant reorientation of the role of economic theory in applied work on two levels. Rather than objects belonging in an economic model, the question itself is framed around a specific historical intervention, and the intervention replaces the role of a model in the empirical practice.

The Empirical Economist's Toolkit

Angrist and Krueger 1991 also symbolically marks the beginning of the surge in research using natural experiments and, thereby, the quasi-experimental program's conscious existence. Table 1 from Angrist and Krueger's 2001 review article (reproduced here as fig. 6) displays the multitude of subsequent studies, the outcomes examined, and the corresponding natural experiment. In addition to quarter of birth, other notable papers include Angrist and Lavy 1999 exploiting a regression discontinuity because of Maimonides's rule, Card 1995 instrumenting with growing up near a college or university, and Duflo 2001 making use of school construction in Indonesia. With reference to the returns to education, Card's (1995, 201) declaration could not be more representative of the revolution: "A convincing analysis of the causal link between schooling and earnings requires an exogenous source of variation in education outcomes."

The table's story fits the picture painted from the bibliometric evidence of, beginning in the 1990s, researchers across the profession finding contexts (law, policies, institutions, natural disasters, etc.) to which these new methods could fruitfully be applied. As Ashenfelter (2014, 577) reflected on his own career:

> When I first became interested in the credible, transparent evaluation of social programs, very few others shared these interests or carefully

Examples of Studies that Use Instrumental Variables to Analyze Data from Natural and Randomized Experiments

Outcome Variable	*Endogenous Variable*	*Source of Instrumental Variable(s)*	*Reference*
		1. Natural Experiments	
Labor supply	Disability insurance replacement rates	Region and time variation in benefit rules	Gruber (2000)
Labor supply	Fertility	Sibling-Sex composition	Angrist and Evans (1998)
Education, Labor supply	Out-of-wedlock fertility	Occurrence of twin births	Bronars and Grogger (1994)
Wages	Unemployment insurance tax rate	State laws	Anderson and Meyer (2000)
Earnings	Years of schooling	Region and time variation in school construction	Duflo (2001)
Earnings	Years of schooling	Proximity to college	Card (1995)
Earnings	Years of schooling	Quarter of birth	Angrist and Krueger (1991)
Earnings	Veteran status	Cohort dummies	Imbens and van der Klaauw (1995)
Earnings	Veteran status	Draft lottery number	Angrist (1990)
Achievement test scores	Class size	Discontinuities in class size due to maximum class-size rule	Angrist and Lavy (1999)
College enrollment	Financial aid	Discontinuities in financial aid formula	van der Klaauw (1996)
Health	Heart attack surgery	Proximity to cardiac care centers	McClellan, McNeil and Newhouse (1994)
Crime	Police	Electoral cycles	Levitt (1997)
Employment and Earnings	Length of prison sentence	Randomly assigned federal judges	Kling (1999)
Birth weight	Maternal smoking	State cigarette taxes	Evans and Ringel (1999)

Figure 6. Table 1 from Angrist and Krueger 2001

thought through the key elements in an evaluation design. Today, it has become commonplace to see literally hundreds of studies that follow the steps many of us stumbled onto—data collection, an empirical appraisal of whether a program exists, and an attempt to define an exogenous treatment—which is now called "evidence-based policy evaluation."

Angrist and Krueger's list of studies, though, also suggests another element to the "credibility revolution." While in Angrist and Krueger 1991 an economic model can be read into the background motivating the question and interpreting the results, this is harder to see with, for examples, McClellan et al. 1994 on the effect of heart attack surgery on patient health or Evans and Ringel 1999 on the effect of maternal smoking on a child's birth weight. While these questions certainly have economic

implications, they are not questions elicited from an economic model. In this way, the quasi-experimental toolkit facilitates answers to questions removed from choice and optimizing behavioral frameworks and into neighboring disciplinary subjects. A recent illustration of this is provided by the literature in economics examining the "fetal origins hypothesis" (Almond and Currie 2011). Highlighted in the 2016 annual report of the Council of Economic Advisers, these papers use in utero natural experiments, such as the timing of natural disasters and other shocks, to identify effects on childhood health and later life outcomes.[15] In these and similar cases, it is the training and, more to the point, the *tools* of the researchers that identify the work as within economics rather than the subject matter; this is what applied microeconomists "do."

The new toolkit did not universally or entirely supplant the earlier Cowles-like approach of developing structural, probabilistic models, however. In fact, across fields, journals, and even departments, the penetration of quasi-experimental methods can vary greatly, and their growing application was accompanied by no small amount of methodological debate (Heckman 1997; Angrist and Imbens 1999; Rosenzweig and Wolpin 2000; Rust 2014). One example of the variation, also singled out by Angrist and Pischke (2010), is the field of industrial organization (and its cousin quantitative marketing), where structural methods continue to dominate. In their response to Angrist and Pischke (2010), industrial organization economists Aviv Nevo and Michael Whinston (2010) echo Frisch to make the important point that current structural models are derived from a theory of *individual* consumer or firm behavior: "Structural modeling attempts to use data to identify the parameters of an underlying economic model, based on models of individual choice or aggregate relations derived from them." More recently, much frontier work in applied economics aims to synthesize quasi-experimental identification and structural ambitions (Chetty 2009; Heckman 2010). How, where, and why these cleavages appeared, shifted, or converged are rich historical questions, which we begin to explore in the next section. At the same time, the credibility "revolution" has had an indisputable impact on the practice and rhetoric of economic science. In the words of one economist, "the modern thing is: 'it's not identified.' God, when your causality is not identified, that's the end of you" (quoted in Fourcade 2009, 91).

15. Douglas Almond and Janet Currie (2011) conclude their review that while the "fetal origins hypothesis" emerges from epidemiology, it "has not only survived contact with economics, but has flourished."

The Structure of Econometric Revolution

To begin to explore some of the significant historical contexts, this section identifies a confluence of trends that fed into the growth and stabilization of quasi-experimental approaches. These contexts, both internal and external to economics, constitute only a partial account of the "credibility revolution" in applied economics and its implications, which continue to evolve. However, the narrative is intended to sketch the larger historical project of contextualizing this paradigm shift by circumscribing key issues, personalities, and institutions.

In the first place, the last forty years experienced remarkable improvements in the ease of computation and the accessibility of data. With statistical software and personal computers today, a researcher is able to run millions of regressions with considerable ease. In addition, micro-level data collected by government agencies and research organizations allowed for wide investigation of the effects of various treatments on numerous outcomes, including wages, education, and health. Large surveys tracked individuals and entities over time, and this longitudinal structure was important for research designs applying difference-in-differences and panel methods. For example, the University of Michigan Institute for Social Research initiated in 1968 the Panel Study of Income Dynamics, which collected individuals' employment, income, expenditures, and wealth.[16] The aforementioned National Longitudinal Surveys administered by the US Census Bureau began in 1966. Ariane Dupont-Kieffer and Alain Pirotte (2011) analyze the early years of panel data econometrics. While not a sufficient condition for the surge in quasi-experimental research, the availability of longitudinal survey and administrative data from a wide variety of contexts able to be easily linked to policy changes satisfied at least a necessary condition for its growth.

Second, despite the quasi-experimental tools having separate histories situated in different communities, quasi-experimental methods were able to be synthesized within the preexisting econometric framework and its established conceptual categories, such as exogeneity and endogeneity. For instance, regression discontinuity can be understood as a kind of instrumental variable, an object whose properties were already well-understood in econometrics. Likewise, as the preceding history emphasized, causality was well-defined within econometrics as the effect of X on Y holding all else constant, a formulation that quasi-experimental

16. This study has now followed families for over forty years, consisting of sixty-five thousand family members over four generations.

researchers have also largely embraced.[17] In an important way, quasi-experimental research designs *are* instruments. This synthesis provided immediate stability and contributes a sense of internality and continuity to the "revolution": doing quasi-experimental work is still doing econometrics. This could also partly explain why this consolidation occurred in economics rather than in another discipline.

Finally, quasi-experimental methods were suited to meet the demands of patrons of economic research, alongside an at least perceived sense of the inadequacy of structural methods. In the welfare state context, there was an increasing interest in evaluating government programs, of which the experience of Ashenfelter at the Department of Labor during the 1970s is illustrative. He recalled about his time on the project using difference-in-differences to evaluate job training that "a key reason why this procedure was so attractive to a bureaucrat in Washington, D.C., was that it was a transparent method that did not require elaborate explanation and was therefore an extremely credible way to report the results of what, in fact, was a complicated and difficult study" (Ashenfelter 2014, 576). He continued: "It was meant, in short, not to be a method, but instead a way to display the results of a complex data analysis in a transparent and credible fashion." Thus, as policymakers demanded evaluations of government programs, the quasi-experimental toolkit became an appealing (and low cost) way to provide simple yet satisfying answers to pressing questions. The rhetorical emphasis on "transparency" of the methods for obtaining 'credible' causal effects for "evidence-based policy evaluation" highlights the marketability of quasi-experimental approaches. As Daniel Breslau (1998, 104–5) summarizes, in contrast to the structural approach's subordination of measurement to model building and academic priorities, "the nonacademic and government-linked researchers adopt an epistemological stance that subordinates the entire research to the task of providing clients in the state with unassailable measures of the effect of a particular historical and local instance of intervention." A representative reflection of this stance can be seen in recent comments of the editors of *World Bank Economic Review*: "Our main two criteria in selecting papers for publication are rigorous identification and policy relevance. The two go

17. That said, there are distinctions. See Rubin 2005 and Heckman 2008. For the statistician Donald Rubin (2005), the bridge between experimental work and parametric modeling becomes almost automatic when viewed through what Rubin calls the potential outcomes framework.

together as we cannot have credible policy recommendations without strong causal inference" (de Janvry and Sadoulet 2012).

While marketability and demand for policy evaluation are factors in the success of the quasi-experimental program, the disillusion with and perception of structural methods' weaknesses in the 1980s is another key context. Not the least testament to this is the Leamer (1983) critique of econometric practice or David Hendry's "Econometrics: Alchemy or Science?" (1980), both raising concerns with robustness and researchers' subjectivity and stimulating methodological debate among econometricians.[18] However, this context is additionally linked with the history of social experimentation itself in economics and for social policy, which pitted structural approaches against experimental evaluations (Breslau 1997a, 1997b). A serious indictment, therefore, was Robert LaLonde's (1986) comparison of structural estimates of training program effects with those obtained from an experiment. Noting that "econometricians intend their empirical studies to reproduce the results of experiments" (604), LaLonde, an Ashenfelter student, compared the experimental effects of a job training program (the National Supported Work Demonstration) with results obtained by estimating a simultaneous equation model of wages and participation, finding significant differences. Combined with the sensitivity of estimates produced from existing econometric practices, the findings, labeled a "landmark" by Angrist and Pischke (2010), further contributed to the discernible crisis in econometrics.[19] Claiming the social and epistemological high ground won in these debates, the quasi-experimental program emerged in the wake of this crisis.

The New Economic Imperialism?

In 1984, two years after receiving the Nobel Prize in economics, George Stigler pondered the "imperialism" of economics in a lecture published in the *Scandinavian Journal of Economics*. Reflecting on Lionel Robbins's (1935, 16) definition of economics as "the science which studies human behavior as a relationship between ends and scarce means which have alternative uses," Stigler (1984) asked rhetorically, "Does [the definition] not make economics the study of all purposive behavior?" He proceeded

18. Boumans and Dupont-Kieffer (2011) discuss this *Methodenstreit* and the diverse responses, such as general-to-specific, Bayesian, and VAR approaches. See Qin 2013 for more.

19. Card refers to it as such in his Woytinsky lecture (2012), but the phrase also appears in Blaug's (1992) preface to his revised *The Methodology of Economics*.

to identify four areas in particular "in which the economist-missionaries have ventured, often against apprehensive and hostile natives": the economics of the law; "new" economic history; the "economic analysis of social structure and behavior" including crime, discrimination, and marriage; and the economic analysis of politics. The key evangelists named by Stigler in each mission field, from Ronald Coase and Richard Posner (law) to Robert Fogel (history), Becker (sociology), and James Buchanan (politics), bore University of Chicago connections.

Despite the diverse subject matters, what unified the work for Stigler was the application of a common behavioral model. In other words, what made the analyses "economic" was the postulate of rational pursuit of goals. For instance, turning to Becker's work on the family ("Becker's theory of marriage"), Stigler (1984) summarized: "Marriage is viewed as a rational act by which the partners seek to maximize their incomes relative to remaining single or choosing different mates." He highlighted two results from this approach. First, Becker demonstrates that polygamy is a "practice favorable to increasing the income of women derived from marriage" because of intensified demand-side bidding. Second, consistent with the "received economic theory that the longer you search in a market, the better the price you will find at which to buy or sell" Stigler (1984), the divorce rate declines with the age of marriage through the late twenties. In Stigler's (1984) view, this "invasion" of economics in each area had fundamentally altered the fields ("Political science will not be the same again") and held great future promise ("I predict that economic theory will be used extensively in sociology"). He concluded the lecture that, indeed, "economics is an imperial science: it has been aggressive in addressing central problems in a considerable number of neighboring social disciplines without any invitations" (Stigler 1984).

From a superficial point of view, Stigler's predictions appear prophetic. Not only has work by economists in the applied areas of crime, education, and politics expanded greatly, in large part marking "the applied turn," the fields of health and development economics are today at the forefront of the profession in prestige and the public view. Further, the disciplinary interactions and applied orientation have been institutionalized in schools of public policy, applied economics, public health, and others that train graduate students for academe, industry, and government, and recruit researchers from multiple disciplines. However, there is a key distinction with how this imperialism proceeds: rather than the application of a behavioral model of purposive goal-seeking, "economic" analysis is

increasingly the empirical investigation of causal effects for which the quasi-experimental toolkit is essential. To return to Stigler's analogy, the missionary's Bible is less Mas-Colell et al. 1995 and more *Mostly Harmless Econometrics* (Angrist and Pischke 2009).

It is this sense in which the transition to applied economics can be viewed as a transition from models to methods: quasi-experimental methods and methodology form core components of current applied economics practice, becoming the "off-the-shelf equipment of the laboratory" (Breslau and Yonay 1999, 325). The application of these methods has both altered the role that economic theorizing plays in applied work, as illustrated by the case study, and expanded the boundaries of economic science. If indeed, as Viner quipped, economics is what economists do, then applied microeconomics is increasingly about the answers that quasi-experimental research designs can provide. As a result, this paradigm shift, whose continuities, discontinuities, personalities, and institutions this article has sought to outline, in turn influences economics' engagement with its disciplinary neighbors and with policymakers.

References

Almond, D., and J. Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25 (3): 153–72.

Angrist, J. D., and G. W. Imbens. 1999. Comment on James J. Heckman, "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 34 (4): 823–27.

Angrist, J. D., and A. B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106 (4): 979–1014.

———. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15 (4): 69–85.

Angrist, J. D., and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (2): 533–75.

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, N.J.: Princeton University Press.

———. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3–30.

———. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton, N.J.: Princeton University Press.

Ashenfelter, O. C. 2014. "The Early History of Program Evaluation and the Department of Labor." *Industrial & Labor Relations Review* 67 (3 suppl.): 574–77.

Backhouse, R., and J. Biddle, eds. 2000. *Toward a History of Applied Economics*. Supplemental issue to volume 32 of *History of Political Economy*. Durham, N.C.: Duke University Press.

Backhouse, R. E., and S. G. Medema. 2009. "Retrospectives: On the Definition of Economics." *Journal of Economic Perspectives* 23 (1): 221–33.

Becker, G. S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. 2nd ed. New York: National Bureau of Economic Research.

Blaug, M. 1976. "The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey." *Journal of Economic Literature* 14 (3): 827–55.

Blaug, M. 1992. *The Methodology of Economics, or How Economists Explain*. Cambridge: Cambridge University Press.

Boumans, M., and A. Dupont-Kieffer. 2011. "A History of the Histories of Econometrics." In *Histories on Econometrics*, edited by M. Boumans, A. Dupont-Kieffer, and D. Qin. *History of Political Economy* 43:5–31.

Bowen, D., L. Fresard, and J. P. Taillard. 2015. "What's Your Identification Strategy? Innovation in Corporate Finance Research." Working paper.

Breslau, D. 1997a. "Contract Shop Epistemology: Credibility and Problem Construction in Applied Social Science." *Social Studies of Science* 27 (3): 363–94.

———. 1997b. "The Political Power of Research Methods: Knowledge Regimes in U. S. Labor-Market Policy." *Theory and Society* 26 (6): 869–902.

———. 1998. *In Search of the Unequivocal: The Political Economy of Measurement in U.S. Labor Market Policy*. Westport, Conn.: Praeger.

Breslau, D., and Y. Yonay. 1999. "Beyond Metaphor: Mathematical Models in Economics as Empirical Research." *Science in Context* 12 (2): 317–32.

Card, D. 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*, edited by L. Christofides, E. K. Grant, and R. Swidinsky, 201–22. Toronto: University of Toronto Press.

———. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5): 1127–60.

———. 2012. *Model-Based or Design-Based? Methodological Approaches in Empirical Micro*. Woytinsky Lecture, University of Michigan, April.

Card, D., and A. B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (4): 772–93.

Chetty, R. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods." *Annual Review of Economics* 1 (1): 451–88.

Christ, C. F. 1994. "The Cowles Commission's Contributions to Econometrics at Chicago, 1939–1955." *Journal of Economic Literature* 32 (1): 30–59.

Colander, D. 2005. "The Making of an Economist Redux." *Journal of Economic Perspectives* 19 (1): 175–98.

Cook, T. D. 2008. "'Waiting for Life to Arrive': A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics." *Journal of Econometrics* 142 (2): 636–54.

De Janvry, A., and E. Sadoulet. 2012. *A Q&A with the Editors of the WBER, Alain de Janvry and Elisabeth Sadoulet.* "Development Impact" (blog), http://blogs.worldbank.org/impactevaluations/a-qa-with-the-editors-of-the-wber-alain-de-janvry-and-elisabeth-sadoulet.

De Marchi, N., and C. Gilbert. 1989. "Introduction." *Oxford Economic Papers* 41 (1): 1–11.

Duflo, E. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91 (4): 795–813.

Dupont-Kieffer, A., and A. Pirotte. 2011. "The Early Years of Panel Data Econometrics." In *Histories on Econometrics*, edited by M. Boumans, A. Dupont-Kieffer, and D. Qin. *History of Political Economy* 43: 258–82.

Epstein, R. J. 1989. "The Fall of OLS in Structural Estimation." *Oxford Economic Papers* 41 (1): 94–107.

Evans, W. N., and J. S. Ringel. 1999. "Can Higher Cigarette Taxes Improve Birth Outcomes?" *Journal of Public Economics* 72 (1): 135–54.

Fourcade, M. 2009. *Economists and Societies: Discipline and Profession in the United States, Britain, and France, 1890s to 1990s.* Princeton, N.J.: Princeton University Press.

Fourcade, M., E. Ollion, and Y. Algan. 2014. "The Superiority of Economists." Maxpo Discussion Paper No. 14/3.

Frisch, R. 1933. *Propagation Problems and Impulse Problems in Dynamic Economics.* Oslo: Universitetets Økonomiske Institutt.

Fuchs-Schuendeln, N., and T. A. Hassan. 2015. "Natural Experiments in Macroeconomics." NBER Working Paper No. 21228, National Bureau of Economic Research.

Haavelmo, T. 1943. "The Statistical Implications of a System of Simultaneous Equations." *Econometrica* 11 (1): 1–12.

———. 1944. "The Probability Approach in Econometrics." *Econometrica* 12:iii–115.

Heckman, J. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32 (3): 441–62.

———. 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." *Quarterly Journal of Economics* 115 (1): 45–97.

———. 2008. "Econometric Causality." *International Statistical Review* 76 (1): 1–27.

———. 2010. "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature* 48 (2): 356–98.

Hendry, D. F. 1980. "Econometrics: Alchemy or Science?" *Economica* 47 (188): 387–406.

Hirschman, D., and E. P. Berman. 2014. "Do Economists Make Policies? On the Political Effects of Economics." *Socio-Economic Review* 12 (4): 779–811.

Hoover, K. D. 2014. "On the Reception of Haavelmo's Econometric Thought." *Journal of the History of Economic Thought* 36 (1): 45–65.

Imbens, G. W., and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

Keane, M. P. 2010a. "A Structural Perspective on the Experimentalist School." *Journal of Economic Perspectives* 24 (2): 47–58.

———. 2010b. "Structural vs. Atheoretic Approaches to Econometrics." *Journal of Econometrics* 156 (1): 3–20.

LaLonde, R. J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76 (4): 604–20.

Leamer, E. E. 1983. "Let's Take the Con out of Econometrics." *American Economic Review* 73 (1): 31–43.

Lee, D. S., and T. Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

Levitt, S. D., and J. A. List. 2008. "Field Experiments in Economics: The Past, the Present, and the Future." NBER Working Paper No. 14356, National Bureau of Economic Research.

Lewbel, A. 2016. "The Identification Zoo—Meanings of Identification in Econometrics." Working paper.

Mas-Colell, A., M. D. Whinston, and J. R. Green. 1995. *Microeconomic Theory*. New York: Oxford University Press.

McClellan, M., B. J. McNeil, and J. P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *JAMA* 272 (11): 859–66.

Mincer, J. 1958. "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy* 66 (4): 281–302.

Morgan, M. S. 1990. *The History of Econometric Ideas*. Cambridge: Cambridge University Press.

Munnell, A. H., ed. 1986. *Lessons from the Income Maintenance Experiments*. Conference Series No. 30. Federal Reserve Bank of Boston.

Nevo, A., and M. D. Whinston. 2010. "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives* 24 (2): 69–81.

Obenauer, M. L., and B. M. v. d. Nienburg. 1915. *Effect of Minimum-Wage Determinations in Oregon: July, 1915*. Washington, D.C.: US Government Printing Office.

Pickering, A., ed. 1992. *Science as Practice and Culture*. Chicago: University of Chicago Press.

Qin, D. 1989. "Formalization of Identification Theory." *Oxford Economic Papers* 41 (1): 73–93.

———. 2013. *A History of Econometrics: The Reformation from the 1970s*. Oxford: Oxford University Press.

———. 2015. "Consolidation of the Haavelmo-Cowles Commission Research Program." *Econometric Theory* 31 (2): 275–93.

Robbins, L. 1935. *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.

Rosenzweig, M. R., and K. I. Wolpin. 2000. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature* 38 (4): 827–74.

Rubin, D. B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100 (469): 322–31.

Rust, J. 2010. Comments on: "Structural vs. Atheoretic Approaches to Econometrics" by Michael Keane. *Journal of Econometrics* 156 (1): 21–24.

———. 2014. "The Limits of Inference with Theory: A Review of Wolpin (2013)." *Journal of Economic Literature* 52 (3): 820–50.

Schultz, T. W. 1961. "Investment in Human Capital." *American Economic Review* 51 (1): 1–17.

Sims, C. A. 2010. "But Economics Is Not an Experimental Science." *Journal of Economic Perspectives* 24 (2): 59–68.

Stapleford, T. A. 2015. "Historical Epistemology and the History of Economics: Views through the Lens of Practice." SSRN Scholarly Paper ID 2692611, Social Science Research Network, Rochester, N.Y.

Stigler, G. J. 1984. "Economics: The Imperial Science?" *Scandinavian Journal of Economics* 86 (3): 301–13.

Stock, J. H., and F. Trebbi. 2003. "Retrospectives: Who Invented Instrumental Variable Regression?: *Journal of Economic Perspectives* 17 (3): 177–94.

Teele, D. L., ed. 2014. *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*. New Haven, Conn.: Yale University Press.

Teixeira, P. N. 2000. "A Portrait of the Economics of Education, 1960–1997." In *Toward a History of Applied Economics*, edited by R. Backhouse and J. Biddle. *History of Political Economy* 32 (supplement): 257–88.

———. 2011. "A Reluctant Founding Father: Placing Jacob Mincer in the History of (Labor) Economics." *European Journal of the History of Economic Thought* 18 (5): 673–95.

Thistlethwaite, D. L., and D. T. Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology* 51 (6): 309–17.

Willis, R. J., and S. Rosen. 1979. "Education and Self-Selection." *Journal of Political Economy* 87 (5): S7–S36.

Wolpin, K. I. 2013. *The Limits of Inference without Theory.* Cambridge, Mass.: MIT Press.

Wright, P. G. 1928. *The Tariff on Animal and Vegetable Oils, by Philip G. Wright, with the Aid of the Council and Staff of the Institute of Economics*. New York: Macmillan.