Emulating Experiments*

Matthew T. Panhans† John D. Singleton‡

Duke University

May 2015

[Preliminary draft. Please do not cite or circulate without permission.]

Abstract

While historians of economics have noted the transition toward empirical work in economics since the 1970s, less understood is the shift toward "quasi-experimental" methods in applied microeconomics. Angrist and Pischke (2010) trumpet the wide application of these methods as a "credibility revolution" in econometrics that has finally provided persuasive answers to a diverse set of questions. Particularly influential in the applied areas of labor, education, public, and health economics, the methods shape the knowledge produced by economists and the expertise they possess. First documenting their growth bibliometrically, this paper aims to illuminate the origins, content, and contexts of quasi-experimental methods, which seek natural experiments to justify causal inference. To highlight lines of continuity and discontinuity in the transition, the quasi-experimental program is situated in the historical context of the Cowles econometric framework and a case study from the economics of education is used to contrast the practical implementation of the approaches. Finally, significant historical contexts of the paradigm shift are explored, including the marketability of quasi-experimental methods and the 1980s crisis in econometrics.

JEL: B21, B23, B4, C1
Keywords: econometrics; quasi-experimental methods; natural experiments; applied economics

*We thank Jeff Biddle, Hsiang-Ke Chao, Pedro Duarte Garcia, Kevin Hoover, and Roy Weintraub for their comments. We also thank participants in the HOPE lunch group and the 2015 History of Economics Summer School in Latin America (Bogotá, Colombia).

†*matthew.panhans@duke.edu*
‡*john.singleton@duke.edu*

1 Introduction

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 had posited that econometricians project on themselves the false image of a "white coat" experimental scientist. More accurately,

> 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 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. (Leamer 1983, 31)

In fact, 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" (Leamer 1983, 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" (Angrist and Pischke 2010, 116). They single out the fields of industrial organization and macroeconomics as the sole exceptions to the shift.

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,b; Wolpin 2013). Nonetheless, Angrist and Pischke's article was less an argument than it was a victory lap:

[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 research finds pleasing are selected for reporting purposes" (Leamer 1983, 36).

For instance, of the nine economists awarded the John Bates Clark Award 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 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), and in the rhetoric and working discourse among applied practitioners.[2] 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 posses (Fourcade et al. 2014).

While historians of economics have noted the transition in the character of economic research since the 1970s toward applications (see for example the 2000 *History of Political Economy* conference volume), less understood is the shift toward quasi-experimental work. In particular, quasi-experimental methods can be usefully juxtaposed with structural applied work, with its roots in the Cowles econometric program.[3] 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 into 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). 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.

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. The structural methodology helps characterize quasi-experimental methods by what they do *not* do. In a number of fields, labor and education perhaps foremost among them, the growth of quasi-experimental work has displaced structural approaches, while in others, such as health, they have expanded the boundaries of economic science to areas detached from choice and optimizing frameworks. From this vantage, the shift toward quasi-experimental designs can be viewed as a transition from models to methods.

[2]It is not unusual for a research presentation in an applied economics field to focus heavily on the plausibility of exclusion restrictions and the validity of instruments.

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

This paper aims to illuminate the origins, content, and contexts of quasi-experimental methods and their applications in economics. First, bibliometric evidence for the shift is presented by tracing 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 within the top general interest journals, but also across a wide selection of field journals, where the particular techniques appears with different frequencies.

Second, to highlight the lines of continuity and discontinuity, quasi-experimental approaches are situated in the historical context of the Cowles econometric program. Although the individual methods that make up the quasi-experimental program were originally were developed well outside this framework, the experimental ideal behind causal economic relationships, elaborated by Haavelmo in *The Probability Approach* (1944), provides one link. A case study is used to contrast the practical implementation of structural and quasi-experimental methods for addressing 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 emphasized along with the subtle, but significant reorientation of the role of economic theory in applied work.

Finally, a confluence of trends that fed into the stabilization of the "credibility revolution" are identified 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 able to be synthesized within the pre-existing 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 and yielding a number of rich historical questions.

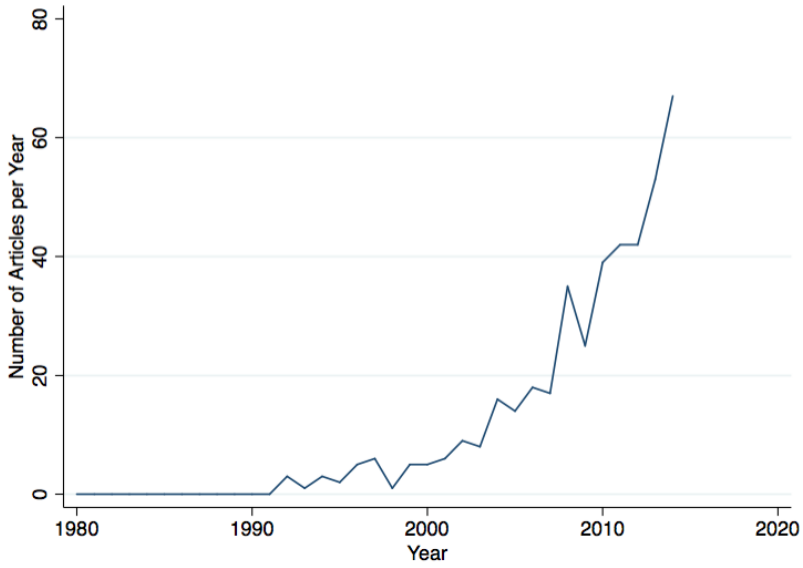
2 Bibliometric Evidence

This section presents bibliometric evidence of the shift towards 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 on an annual basis. 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 Economics Behavior & 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 counts the number 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 count of articles. In terms of absolute numbers, the count for 2010, the year of the *Journal of Economic Perspectives* symposium, represents around a 400% increase over year 2000. In Figure 2, the same exercise is performed for the term "instrumental variable." This similar upward trend not only corroborates the shift to applied work generally, but also to quasi-experimental research designs, as later sections elaborate.

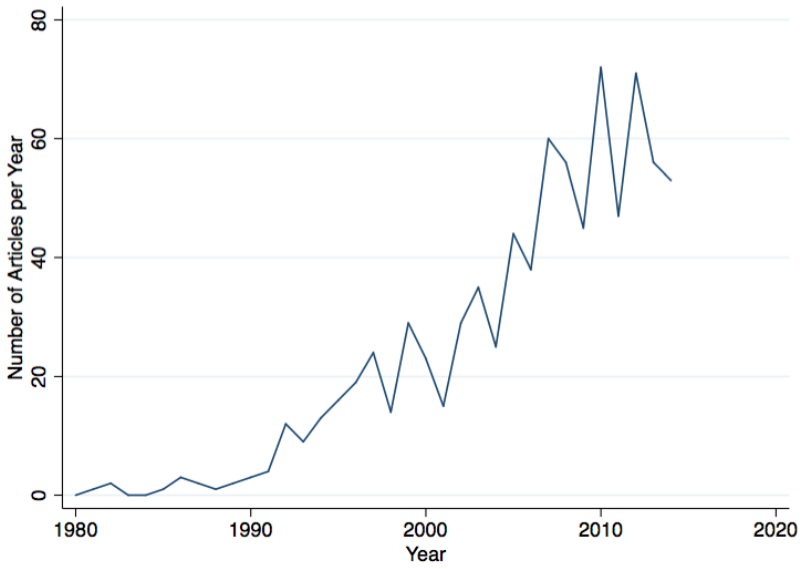
Figure 1: Uses of Quasi-experimental Terms in Top Economics Field Journals



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.

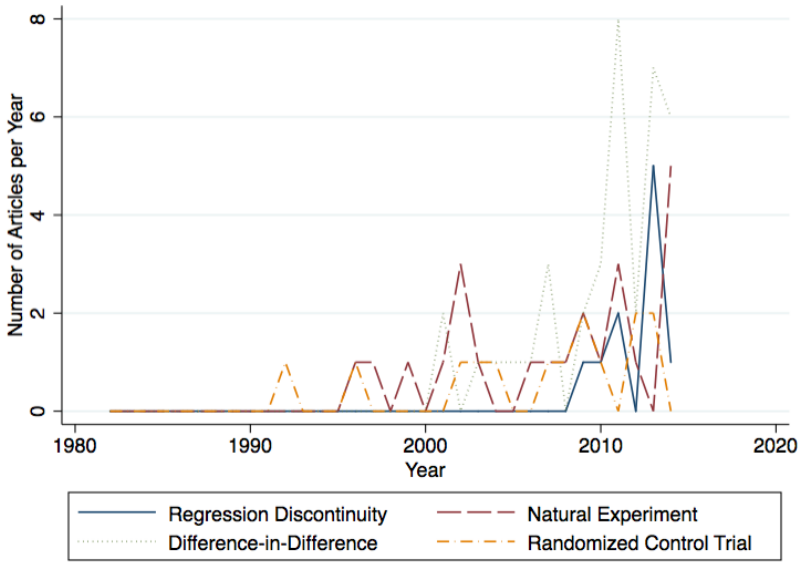
To gain a sense of how the application of quasi-experimental methods varies across fields, the same exercise is completed for specific selected field journals. Figure 3 illustrates that there is an increase in the use of the quasi-experimental techniques in the field of health economics around the year 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 United States

Figure 2: Uses of "Instrumental Variable" in Top Economics Field Journals



health care system and uses research designs that exploit variation in regulations across states. By contrast, Figure 4 verifies that randomized control trials are particularly widespread in development economics, and especially in just the last few years. The overall trend is also reflected in Figures 5 and 6, with a pronounced recent trend in work using regression discontinuity in the *Journal of Public Economics* as shown in Figure 5.

Figure 3: Uses of Quasi-experimental Terms in the *J Health Econ*



While the trends within field journals in economics are compelling, it is useful to also examine counts for the top four economics journals, which reflect the highest profile general interest research

Figure 4: Uses of Quasi-experimental Terms in the *J Dev Econ*

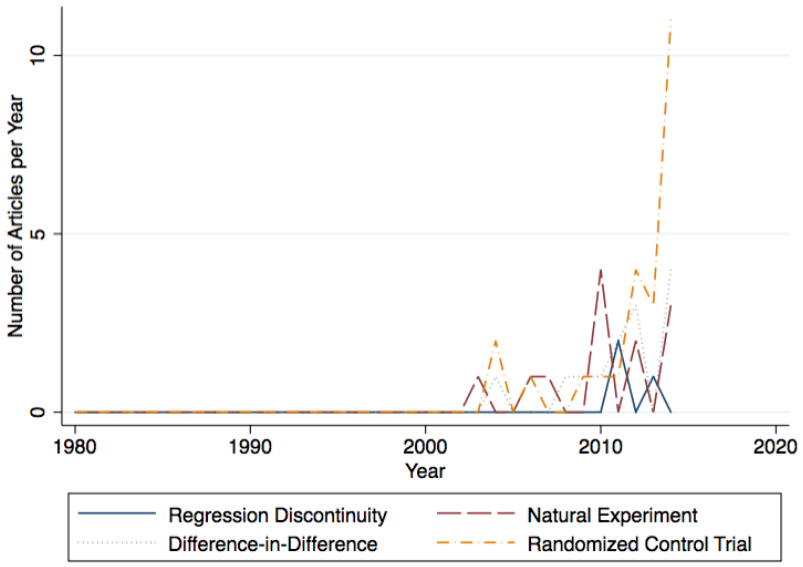
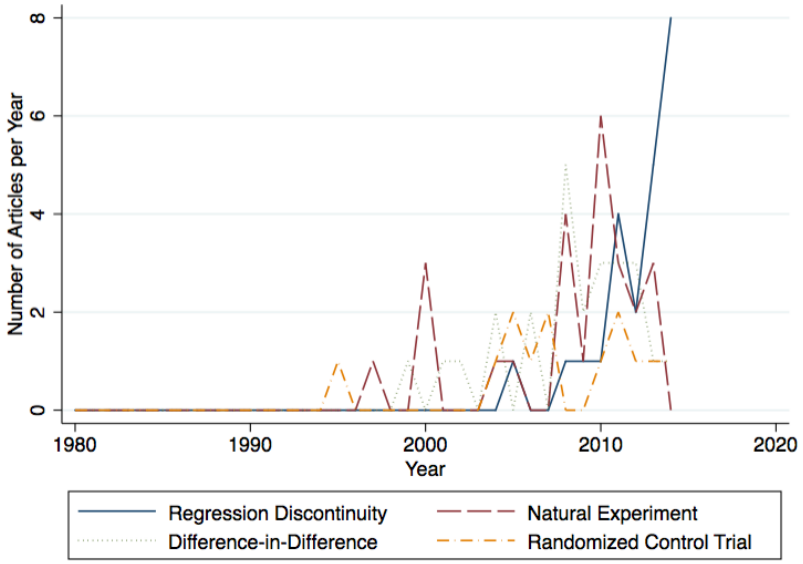


Figure 5: Uses of Quasi-experimental Terms in the *J Pub Econ*



in the profession: the *American Economic Review, Econometrica, Quarterly Journal of Economics*, and the *Journal of Political Economy*. As Figure 7 illustrates, the sharp upward turn in quasi-experimental terms in abstracts and titles remains evident. Likewise, the count for instrumental variable in Figure 8 suggests 1990 as an inflection point.

Together, the graphs suggest a significant and recent shift toward quasi-experimental methods. In addition, this shift appears not only within the top general interest journals, but also across a wide selection of field journals, where the particular techniques appear with different frequencies.

Figure 6: Uses of Quasi-experimental Terms in the *ReStat*

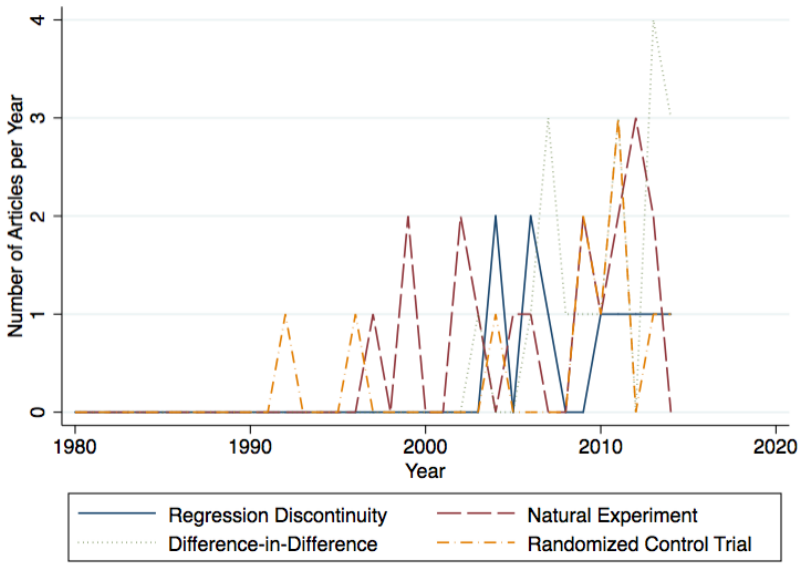
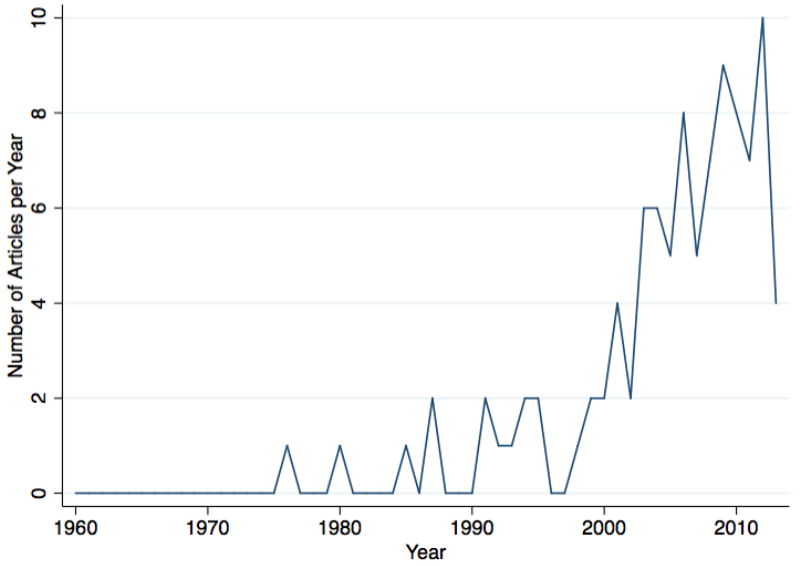
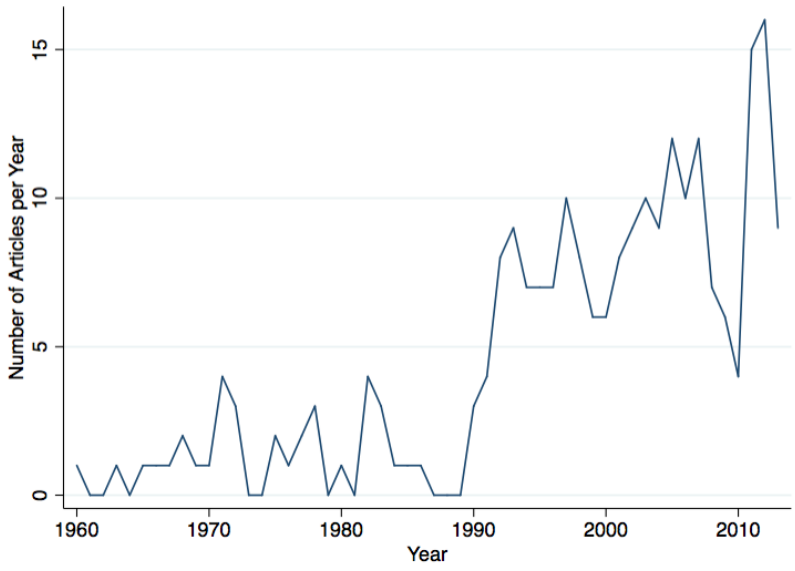


Figure 7: Uses of Quasi-experimental Terms in Top 4 Economics Journals



This documents both the timing and 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 shift before looking in greater detail at the stabilization and contexts of these changes.

Figure 8: Uses of "Instrumental Variable" in Top 4 Economics Journals



3 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. As James Heckman 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." (Heckman 2000, 46). This context is therefore important to understanding the elements of continuity and discontinuity in the growth of quasi-experimental methods.

3.1 The Probability Approach

Estimation of equations governing the economy was a goal of the early Cowles Commission. The setting for this research was macroeconomic policy as researchers wanted to to evaluate the effects of different policies. Ragner 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 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." (Frisch 1933, 185). 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 primarily 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. Hoover (2012) 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 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. (Haavelmo 1944, 27)

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" (Haavelmo 1944, 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 – i.e. 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.

3.2 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. 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 were particularly influential in the human capital program's development. In addition to Becker and Schultz, applied economists like H. Gregg Lewis studied 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.[4]

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, 8). 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, X_i, and unobservables, τ_i. Writing this formally and generally,

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

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

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, i.e. the choice of educational attainment:

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

where i belongs to j if $V_{ij} = \max(V_{i1}, ..., V_{in})$. The person thus chooses the education level to maximize their 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, 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 that applied for the Army Air Corps and were eligible for the GI Bill.

Before presenting their results, Willis and Rosen 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" (Willis and Rosen 1979, 18). 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" (Willis and Rosen 1979, 19)). However, they concede that "easy distinctions" between the vectors are "increasingly difficult, if not impossible, to make" (Willis and Rosen 1979, 19).[5]

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.[6] The Willis and Rosen paper thus provides an illustrative case study of how the Cowles approach was practi-

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

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

cally implemented in applied research to contrast, in the next section, with a quasi-experimental approach.

4 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 were held constant. Similarly, Haavelmo 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. (Haavelmo 1944, 6; original emphasis)

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.

4.1 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 had pre-histories situated in different places and communities, from agricultural economics to psychology and statistics. For instance, Morgan (1990) attributes the invention of the instrumental variable regression to Philip G. Wright in 1928.[7] 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 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, developed tests and heuristics for

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

what constituted an instrument that was "strong enough" or what evidence might suggest that an instrument is valid, such as the Sargan test.

Regression discontinuity design, which takes advantage of known selection rules such as cutoffs or thresholds, has probably the most unusual pre-history. As detailed by 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 would continue 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.

Perhaps the oldest of quasi-experimental methods is difference-in-differences, which simply compares a treatment and control group over time. If assignment into treatment is as good as randomly assigned, or at least unrelated to the outcome of interest, 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 further 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 to 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 would come 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 due to the influence of Orley Ashenfelter.

Ashenfelter received his Ph.D. in economics from Princeton University in 1970, and soon after was appointed as the Director of the Office of Evaluation at the U.S. 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 would eventually return to a faculty position at Princeton. In addition to his research promoting the use of quasi-experimental methods in economics, he would also advise some of the most influential practitioners that would follow in this tradition, including 1995 John Bates Clark Medal winner David Card and one of the most well-known proponents of quasi-experimental methods, Joshua 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 individuals would play instrumental roles in the growth of quasi-experimental work and its stabilization as a self-conscious program and toolkit.

4.2 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 posits three reasons for this resurgence in interest (Card 2001). 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 vain and published in the *Quarterly Journal of Economics*, the second line of Angrist and Krueger (1991) 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" (Angrist and Krueger 1991, 979). 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 16 or 17. 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, i.e. the month in which an individual was born is as if it were randomly assigned. In other words, birth month is uncorrelated with an individual's other characteristics, and thus 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 of dealing 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 very little differences between ordinary least squares and two-stage least squares estimates, suggesting little, if any, bias.[8]

Despite a common longstanding 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 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 re-emphasized 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.

[8] 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.

4.3 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 Figure 9) 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 due to Maimonides' 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 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" (Card 1995, 201). The Table's story fits the

Figure 9: Table 1 from Angrist and Krueger (2001)

Table 1

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)

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 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 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." (Ashenfelter 2014, 577)

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, as examples, McClellan et al. (1994) – the effect of heart attack surgery on health – or Evans and Ringel (1999) – the effect of maternal smoking on birth weight. In these and similar instances, it is the training and, more to the point, the tools of the researchers that identifies the work as within economics.

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). 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 Nevo and Whinston (2010) echo Frisch to make the important point that current structural models are derived from 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)).

5 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, 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.[9] And the aforementioned National Longitudinal Surveys administered by the U.S. Census Bureau began in 1966. These types of survey data opened up research opportunities that simply were not available with only aggregate national accounting statistics.

Second, despite the quasi-experimental tools having separate histories situated in different communities, quasi-experimental methods were able to be synthesized within the pre-existing econometric framework and the established conceptual categories, including exogeneity and endogeneity. For instance, regression discontinuity can be understood as a kind of instrumental variable. 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 researchers have also largely embraced.[10] This synthesis provided immediate stability and contributes a sense of internality and continuity to the "revolution." 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 structural methods' inadequacy. 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 investigate 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 policy-markers demanded evaluations of government programs, the quasi-experimental toolkit came to be an appealing way to provide simple yet satisfying answers to pressing questions. The emphasis on "transparency" of the methods for obtaining "credible" causal effects for "evidence-based policy evaluation" highlights the marketability of quasi-experimental approaches. A representative reflection of this 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 together as we cannot have credible policy recommendations without strong causal inference" (de Janvry and Sadoulet (2012) in an online interview).

While marketability and demand for policy evaluation are two likely factors in the success of quasi-experimental approaches, the disillusion with and perception of structural methods' failures

[9] This study has now followed families for over 40 years, consisting of 65,000 family members over four generations.

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

into the 1980s is another key context. Not the least testament to this is Leamer (1983) critique of econometric practice or David Hendry's "Econonmetrics: Alchemy or Science?" (1980), both raising concerns with robustness and researchers' subjectivity. A perhaps more serious indictment, though, was Robert LaLonde's comparison of structural estimates of training program effects with those obtained from an experiment (LaLonde 1986). Noting that "econometricians intend their empirical studies to reproduce the results of experiments" (LaLonde 1986, 604), LaLonde 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 divergence. Combined with the sensitivity of structural estimates, the findings, labeled a "landmark" by Angrist and Pischke (2010), further contributed to a discernible crisis in econometrics.[11]

6 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' definition of economics as "the science which studies human behavior as a relationship between ends and scarce means which have alternative uses" (Robbins 1935, 16), Stigler asked rhetorically "does [the definition] not make economics the study of all purposive behavior?" (Stigler 1984, 302). He proceeded to identify four areas in particular "in which the economist-missionaries have ventured, often against apprehensive and hostile natives" (Stigler 1984, 304): 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 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" (Stigler 1984, 308). 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, 309), the divorce rate declines with the age of marriage through the late 20s. In Stigler's view, this "invasion" of economics in each area had fundamentally altered the fields ("Political science will not be the same again" (Stigler 1984, 311) and held great future promise ("I predict that economic

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

theory will be used extensively in sociology" (Stigler 1984, 309). 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, 311).

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 academia, industry, and government and recruit researchers from multiple disciplines. However, there appears to be 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 also be usefully viewed as a transition from models to methods: quasi-experimental methods and methodology form core components of current applied economics practice that both altered the role of economic theorizing in applied work and expanded the boundaries of economic science. At the same time, this paradigm shift, some of whose continuities, discontinuities, exceptions, personalities, and institutions this paper has sought to highlight, influences economics' engagement with its disciplinary neighbors and with policymakers, yielding a number of rich historical questions.

References

Angrist, J. D. and G. W. Imbens (1999, October). Comment on James J. Heckman, "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations". *The Journal of Human Resources 34*(4), 823–827.

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

Angrist, J. D. and A. B. Krueger (2001, October). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives 15*(4), 69–85.

Angrist, J. D. and V. Lavy (1999, May). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

Angrist, J. D. and J.-S. Pischke (2009, January). *Mostly Harmless Econometrics: An Empiricist's Companion* (1 edition ed.). Princeton: Princeton University Press.

Angrist, J. D. and J.-S. Pischke (2010, April). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *The Journal of Economic Perspectives 24*(2), 3–30.

Angrist, J. D. and J.-S. Pischke (2014, December). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton ; Oxford: Princeton University Press.

Ashenfelter, O. C. (2014, May). The Early History of Program Evaluation and the Department of Labor. *Industrial & Labor Relations Review 67*(3 suppl), 574–577.

Backhouse, R. and J. Biddle (Eds.) (2000). *Toward a history of applied economics*, Volume 32(Suppl 1) of *History of Political Economy*. Durham: Duke University Press,.

Becker, G. S. (1964). *Human capital : a theoretical and empirical analysis, with special reference to education* (2d ed. ed.). New York: National Bureau of Economic Research; distributed by Columbia University Press,.

Blaug, M. (1976, September). The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey. *Journal of Economic Literature 14*(3), 827–855.

Blaug, M. (1992, July). *The Methodology of Economics: Or, How Economists Explain*. Cambridge University Press.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In L. Christofides, E. K. Grant, and R. Swidinsky (Eds.), *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*, pp. 201–222. Toronto: University of Toronto Press.

Card, D. (2001, September). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica 69*(5), 1127–1160.

Card, D. (2012, April). Model-Based or Design-Based? Methodological Approaches in Empirical Micro. The University of Michigan's 2012 Woytinsky Lecture.

Card, D. and A. B. Krueger (1994, September). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review 84*(4), 772–793.

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics 1*(1), 451–488.

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

Cook, T. D. (2008, February). "Waiting for Life to Arrive": A history of the regression-discontinuity design in Psychology, Statistics and Economics. *Journal of Econometrics 142*(2), 636–654.

de Janvry, A. and E. Sadoulet (2012, February). A Q&A with the editors of the WBER, Alain de Janvry and Elisabeth Sadoulet.

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

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

Fourcade, M., E. Ollion, and Y. Algan (2014, November). The Superiority of Economists. *maxpo discussion paper no. 14/3*.

Frisch, R. (1933). *Propagation problems and impulse problems in dynamic economics*.

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

Haavelmo, T. (1944, July). The Probability Approach in Econometrics. *Econometrica 12*, iii–115.

Heckman, J. (1997, July). Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *The Journal of Human Resources 32*(3), 441–462.

Heckman, J. J. (2000, February). Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective. *The Quarterly Journal of Economics 115*(1), 45–97.

Heckman, J. J. (2008, April). Econometric Causality. *International Statistical Review 76*(1), 1–27.

Heckman, J. J. (2010, June). Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature 48*(2), 356–398.

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

Hirschman, D. and E. P. Berman (2014, April). Do economists make policies? On the political effects of economics. *Socio-Economic Review*, mwu017.

Hoover, K. D. (2012, February). On the Reception of Haavelmo's Econometric Thought. SSRN Scholarly Paper ID 2001485, Social Science Research Network, Rochester, NY.

Imbens, G. W. and T. Lemieux (2008, February). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

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

Keane, M. P. (2010b, May). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics 156*(1), 3–20.

LaLonde, R. J. (1986, September). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review 76*(4), 604–620.

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

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

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995, June). *Microeconomic Theory* (1st edition ed.). New York: Oxford University Press.

Mincer, J. (1958, August). 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 University Press.

Nevo, A. and M. D. Whinston (2010, April). Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference. *The 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*. U.S. Government Printing Office.

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

Rosenzweig, M. R. and K. I. Wolpin (2000, December). Natural "Natural Experiments" in Economics. *Journal of Economic Literature 38*(4), 827–874.

Rubin, D. B. (2005, March). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association 100*(469), 322–331.

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

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

Stigler, G. J. (1984, September). Economics: The Imperial Science? *The Scandinavian Journal of Economics 86*(3), 301–313.

Stock, J. H. and F. Trebbi (2003, July). Retrospectives Who Invented Instrumental Variable Regression? *The Journal of Economic Perspectives 17*(3), 177–194.

Thistlethwaite, D. L. and D. T. Campbell (1960, December). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology 51*(6), 309–317.

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

Wolpin, K. I. (2013, April). *The Limits of Inference without Theory.* Cambridge, Mass: The 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. [Preface by Harold G. Moulton.].* New York: the Macmillan Company.