

Design-Based Research in Empirical Microeconomics

By David Card
WORKING PAPER #654
PRINCETON UNIVERSITY
INDUSTRIAL RELATIONS SECTION
JANUARY 2022

James J. Heckman



Econ 312, Spring 2022

The Setting

- The late 1970s and early 1980s was a period of turmoil in applied economics.
- The previous decade of high inflation and high unemployment led to the scornful dismissal of the big macro models that were viewed as a key legacy of the Cowles Commission research program (Fair, 1992).
- The Negative Income Tax Experiments – which had begun with strong support from both liberals and conservatives – ended with little fanfare but much disagreement over the findings.
- And several high-profile papers – including Sims (1980), Hendry (1980), and Leamer (1983) – identified serious credibility gaps in many applied studies, particularly those based on macro data.

- Subsequent studies in the mid-1980s suggested that the credibility gap extended to the field of labor economics, where cross sectional and longitudinal micro data had already supplanted the use of macro data.
- LaLonde's (1986) analysis of the National Supported Work (NSW) demonstration project showed that when state of the art econometric methods were applied to the NSW treatment group – throwing away the randomized control group – they gave a wide range of estimates.
- And it was very hard to tell without peeking at the real control group which method or specification gave the 'best' answer – a problem that had been identified by Ashenfelter (1974, 1978) and Ashenfelter and Card (1985) in earlier attempts to evaluate subsidized training programs.

- Of course at the time very few economists used the term “research design”.
- My introduction to the phrase came when Alan Krueger, who had recently joined the Princeton faculty, started a subscription to the New England Journal of Medicine.
- Alan and I were intrigued by the abstracts of the papers, which typically had a sentence on “research design.”
- I noted that few of my papers had a research design that could be summarized in a sentence.

- Ashenfelter was not the only senior labor economist arguing for an emphasis on research design.
- Richard Freeman, one of Krueger's thesis advisors and head of the National Bureau of Economic Research's Labor Studies program, had long argued for the value of what he called 'natural experiments' (e.g., Freeman, 1989, p. x).
- He also pushed the idea of looking for 'big shocks', such as the imposition of the federal minimum wage on Puerto Rico (Castillo-Freeman and Freeman, 1992), to help cut through the statistical noise in labor market data.

Natural Experiments

- My first attempt to incorporate these ideas was my study of the Mariel Boatlift (Card, 1990).
- The idea of looking at the Boatlift arose in discussions with a talented undergraduate at Princeton, Constantine Alexandrakis, who was raised in Miami and conducted an analysis of the episode for his senior thesis.
- I had been working with fellow Princeton alumnus Joe Altonji on the labor market effects of immigration (Altonji and Card, 1991) and was unsatisfied with the evidence we had been able to assemble using a simple enclave-style instrumental variable.
- The Mariel Boatlift seemed like a useful case study to benchmark our findings.

- My approach to the analysis was very much influenced by my work on longitudinal earnings models with Ashenfelter and with John Abowd (Abowd and Card, 1989).
- Since the Boatlift occurred just prior to the 1983 recession it was important to build a comparison group of cities whose economic conditions had closely tracked those in Miami in previous years (the “parallel trends” condition identified by Ashenfelter, 1974).
- The design of the comparison group was a natural outgrowth of the models labor economists had been building to describe individual earnings over time, but applied to a city rather than a person.
- And the Boatlift certainly met Freeman’s advice to look for a big shock.

- While the conference proceedings were in press, Krueger and I realized that an impending increase in the minimum wage in New Jersey provided the opportunity for a prospective quasi-experimental study with both a treatment group (restaurants in New Jersey) and a control group (restaurants in Pennsylvania).
- Building on Alan's experience with the Katz and Krueger (1992) study – which lacked a pure control group -- we tried to design the study to mimic a classical randomized controlled trial (RCT).
- Perhaps the most compelling feature of the study was that our hands were tied once we conducted the first-wave survey: the sample and the survey questions were frozen.
- I have learned over the years that most people want other researchers' hands to be tied.

- The results of our New Jersey – Pennsylvania study confirmed the findings in our earlier studies, but created a much bigger stir, partly because our paper was published in the American Economic Review, and partly because the design was so easily explained to non-specialists.

- The findings were widely criticized by many economists – including Nobel prize winners like Gary Becker, Merton Miller, and James Buchanan.
- Buchanan (1996) wrote: “no self-respecting economist would claim that increases in the minimum wage increase employment.
- Such a claim ... becomes equivalent to a denial that there is even minimal scientific content in economics, and that, in consequence, economists can do nothing but write as advocates for ideological interests”.

- Of course Joan Robinson (1933) had shown that an employer with wage setting power will increase employment if faced with a small increase in the minimum wage – a prediction that was widely understood (e.g., Stigler, 1946).
- So Buchanan was really arguing that any consideration of wage setting power by employers was unscientific. That seems to me to be an unusual definition of “scientific”.

What's Good About Design-Based Studies

- Design-based studies have several strengths that have contributed to their spread in Labor Economics and related fields.
- The most important is that they put causality front and center.
- In a design-based study, the issue of “identification” – which was stated as a mathematical condition in older textbooks -- is equated with the research design.
- This was clarified in the path-breaking paper by Imbens and Angrist (1994).
- In many cases a precise research design also makes it easier to construct specification tests, like pre-treatment comparisons that are now widely used in event study designs.

- A second strength is that a simple research design is easily understood and interpreted by other researchers, and by the policy community.
- Ashenfelter (1987) noted that when the results of an RCT are presented “... it is my experience that the discussion turns almost immediately to substantive, as opposed to methodological, matters”.
- A well-designed nonexperimental study can have the same effect. It can also be directly policy relevant.
- The evidence from a quasi-experimental analysis of one state-specific minimum wage increase, for example, can be understood by people in other states who are contemplating a similar choice.

- A third strength is portability of the design.
- This is important in economics because we are often unsure whether the results in a particular study are just a fluke, or hold up more generally.
- Within a decade the basic design of the Mariel Boatlift study was adopted to France (Hunt, 1992), Portugal (Carrington and de Lima, 1996); and Israel (Friedberg, 2001).
- Variants of the New Jersey-Pennsylvania “border design” have been implemented by Dube Lester and Reich (2010), Cenzig, Dube, Lindner and Zipperer (2018), Campos-Vazquez and Esquivel (2021), and others.
- By comparing estimates from the same design across settings it is possible to reach stronger conclusions and potentially convince the research community to take the results more seriously, even if they appear to overturn the conventional wisdom.

What's Not to Like About Design-Based Studies

- In most design-based studies, as in a classical RCT, the counterfactual state for “treatment” is the state experienced by the comparison group.
- In my opinion, this is a strength, since it helps provide a clear interpretation of the treatment effect.
- Other analysts, however, view this as a limitation (e.g., Heckman, 2005; Deaton and Cartwright, 2018).
- Some of their criticisms involve the philosophy of science. But three are worth discussing here.

- One concern is that the particular counterfactual identified by the design is too restrictive. For example, people often asked Alan Krueger and me whether we thought the results of our minimum wage study would hold if the minimum wage were raised to a much higher level.
- Obviously we could not say. But even if we had taken a model-based approach, and estimated consumer demand and restaurant production functions, I think it would have been unwise to extrapolate the results too far outside the range of the existing data.
- Any economic model is only a crude approximation, and the associated parameter estimates can only reflect a limited historical experience.
- This is more or less what went wrong in the 1970s, when the macro models fit to earlier decades broke down.

- A second criticism is that design-based studies are often conducted in the framework of a single-equation model for the outcome of interest, with no explicit consideration of the process that assigns treatment to different units, or of how the assignment process is related to the outcome.
- Of course there are well-developed templates for conducting such an extended analysis (see e.g., Wooldridge, 2010; Cornelissen, Dustmann, Raute, and Schoenberg, 2016; Kline and Walters, 2019), so this is really a criticism of how authors report their results, rather than of the design-based approach per se.
- I think this criticism is also over-stated.
- A paper with a straightforward analysis is still worthy of attention if the research question is interesting and the design and implementation are of high quality.

- A third criticism – closely related to the first two -- is that the results from a simple design-based study cannot be used to evaluate scenarios outside those observed for the treatment and comparison groups.
- This is not always true: if an analyst uses a logistic model to summarize the results of a design-based study of discrete choice, for example, it is possible to predict the demand for other choices, including choices that do not yet exist (see McFadden, 2001).
- Again, I think this is a criticism of the complexity of the analysis in some studies, rather than a statement about the utility of a good research design, which is always helpful in obtaining credible estimates for the parameters of interest.

Why Do We (Still) Need Design-Based Studies?

- Design-based studies are particularly useful for testing basic predictions of a theory, or testing between competing theories.
- Many economists see little need for such work.
- They have much greater confidence than me in simple working models that in many cases have never been put to a rigorous test.
- In such a conservative intellectual environment, transparent design-based studies can play an important role in “opening the door” to a new approach or a new class of models.

- That is arguably what happened as a result of the work on minimum wages done by Alan Krueger and me in the 1990s.
- At the time, even two decades after the initial formalization of search theory (McCall, 1970) and of demand systems for differentiated products (McFadden, 1974), most economists believed that market power in wage setting required a single buyer of labor.
- Although Alan and I thought of our work as trying to test between wage-taking and wage-setting behavior by employers, I think it's safe to say that virtually no one changed their mind on that question when our New Jersey Pennsylvania study was published.

- Another example is the study of displaced workers by Jacobson, Lalonde, and Sullivan (1993) (JLS).
- This paper was one of the first in Labor Economics to use an explicit “event study” design – in their case to evaluate the effect of losing a job in a mass layoff event.
- Comparing job losers in mass layoffs with coworkers who retained their jobs, JLS showed that prior to the event the two groups had very similar earnings for many years (verifying the ‘parallel trends’ assumption).

- JLS's findings were disruptive because they were seen as inconsistent with a model where each worker is paid her "market wage."
- Prior to the mass layoff the displaced and nondisplaced workers earned about the same pay. But after the event the job losers earned substantially less, just because they no longer worked at one particular firm.
- The basic findings of JLS have been verified in many other settings and countries, and their event study design has become the leading methodology used by labor economists to establish causality.
- And today it is more widely accepted that an individual's pay is determined in part by who they work for.
- But without JLS's innovative research design, that message would have taken far longer to get through.

- I hope that young researchers today can find some inspiration in these examples.
- Labor economist's interest in research designs evolved out of the realization that what we had been doing before to answer causal questions was just not working.
- In my opinion, we have made some progress.
- A compelling design-based study can sometimes provide convincing evidence and get economists thinking a little differently about a problem – albeit slowly and with no guarantee of success.