

Book Reviews

Causal Inference: The Mixtape

Scott Cunningham. 2021. New Haven, CT: Yale University Press. ISBN 978-0300251685 \$35.00 (paperback)

Reviewed by Sarah Hamersma, Syracuse University

Causal Inference: The Mixtape is a book on causal inference methods that speaks with a unique voice. I see this book as, ideally, something to help students who are formally learning these methods from more technical sources to feel like they have someone talking to them more like a human being and giving them the intuitive lay of the land. I believe this style of writing conveys the author's confidence that people from a wide variety of disciplines and backgrounds can become well-informed and competent on methods of causal inference by walking through the logic of the models. While the technical material is included as needed, the key to the book is that anyone can learn when and how to use different modeling techniques, and can practice this with the materials in the book (which address both Stata and R users). While the book is available in an online version, it's a great volume to hold in your hand, and this book (\$35) is worth the price just for the introductory chapter and the opportunity to mark it up to your heart's content (as I have).

My PhD advisor taught me to write my research papers in the first person, instead of the weirdly disconnected way in which I had previously thought "objective" science was to be written. I learned to write, "In this paper I estimate..." instead of the passive "This model was estimated..." First person makes clear that the researcher is in fact the decider and the do-er, and I have become convinced that it helps us take responsibility for those decisions and to acknowledge that we are just people, not robots (maybe not even *homo economicus*).

In this book, Scott Cunningham of Baylor University takes the first person vantage point to a whole new level. Throughout the book he is talking *to you*: about his life, his ideas, how he understands the purpose and methods of causal inference, and, most importantly how excited he is about them. He writes in conversational language that interacts with the reader, peppered with questions like "Now isn't that just a little bit surprising?" and "But what might be wrong with this reasoning?" and "So what's the big deal? Why is [this method] so special?" Cunningham

doesn't pretend to be objective about this material – he loves it! He actually uses that word, *love*. He talks about how it changed his life – he really uses those words. He refers to empirical investigations as exciting, he describes being enamored, he talks about his eyes being opened. His experience discovering not just the utility but the *beauty and brilliance* of this work is palpable and infectious. He has the energy of an evangelist, claiming no credit for the methods themselves but being convinced that this is really good news that he wants to share with you.

Cunningham does a wonderful job laying out the fundamental problem of causal inference – that is, the absence of an observable counterfactual to compare actual outcomes to. But beyond that, he puts his finger squarely on the depth and even irony of the challenge that an economist in particular faces: we model the world with the idea that choices are made with great purpose, and so being able to treat observational data as a quasi-experiment goes directly against the grain of our view of how the world works. In his introduction, Cunningham writes the following (p. 10):

To put it as bluntly as I can, economic theory says choices are endogenous, and therefore since they are, the correlations between those choices and outcomes in the aggregate will rarely, if ever, represent a causal effect. Now we are veering into the realm of epistemology...

This bold admission goes beyond “correlation isn't necessarily causation” – it is closer to “it would be very unusual if correlation reflected only causation.” He uses a wonderful example to illustrate this. Instead of showing us a spurious correlation to make the point (my go-to strategy), he shows us the opposite: a zero correlation that actually masks a lot of causality. The illustration is literal – a drawing by Seth Hahne that he commissioned for Figure 1 in the book – showing a person steering a boat through stormy waters. The boat is going straight, while the rudder is moving back and forth. There is no correlation between the movement of the rudder and the direction of the boat – no matter what the rudder is doing, the boat keeps going straight. The lack of correlation masks a tremendous amount of causality, with human decisions made optimally to create what appears to be no relationship between the steering and the direction. It is a profound lesson in thinking both about decision making and about the key role of the counterfactual: if you are

not accounting for the direction of the wind or waves to know where the boat would have gone, you would never be able to uncover the causality between the rudder and the boat's direction. It is all about getting the counterfactual right. One of my students was reading the book online and said that by the time he got to this illustration (on p. 7) he knew he wanted to invest in the hard copy.

Alongside framing the fundamental problem of causal inference, Cunningham unapologetically tells us what he sees as the right reasons for doing this work in the first place. As he puts it:

Identifying causal effects involves assumptions, but it also requires a particular kind of belief about the work of scientists. Credible and valuable research requires that we believe that it is more important to do our work **correctly** than to try and achieve a certain outcome (e.g., confirmation bias, statistical significance, asterisks)... True scientists do not collect evidence in order to prove what they want to be true or what others want to believe. That is a form of deception and manipulation called **propaganda**, and propaganda is not science... Scientific methodologies allow us to accept unexpected, and sometimes undesirable, answers.

While he does not reference explicitly his Christian identity as a foundation for the beliefs that transparency, honesty, and truth-seeking are of primary importance, surely these commitments should particularly resonate with us as Christian economists. Credible investigation of a question cannot be constrained by culture-war political incentives, and every question must be approached with a humility that recognizes that we are constantly learning ever more about the social world God has created.

One of my favorite features of this book is the way that it digs into the history and context of the use of estimators. Cunningham has a curiosity about this whole enterprise and its history and development that is essentially never covered in methodology textbooks. You come away with the idea that, indeed, necessity is the mother of invention. For instance, physician John Snow suspected the water supply in the Thames was to blame for England's 19th-century cholera epidemic. His creativity in assessing this hypothesis led to what we now call a difference-in-differences analysis. Throughout the book, Cunningham uses these kinds of context-setting stories and studies of important issues to inspire the need for the methods he teaches.

Speaking of the methods: while my review is primarily about the unique tone and added value of this book, I do also want to review a bit about the topics covered. Cunningham begins with an introduction that is worth the price of admission and ought to inform how every one of us begins a class in econometrics. He follows with a review of probability and regression, which is a hefty 80 pages and admittedly the one part of the book that I felt might be only helpful for a more select audience, although I wasn't sure who – at times, I felt it was nicely pitched at students with very little statistical or math background but then, at other times, there were proofs with integrals or “argmin” notation. If I had to make a suggestion for the second edition, I would trim down this chapter quite substantially and point students to other texts to prepare them for the potential outcomes notation on which this book relies.

The first chapter after this review explains directed acyclic graphs (DAGs), which Cunningham finds crucial to laying out our ideas about potential causal relationships before we even try to determine a research design. While this approach was new to me, I do appreciate the value of diagrams that force one to explicitly express hypotheses being made about relationships (and assumptions about non-relationships) in the population before working with data. Perhaps because this was so new to me, however, as soon as the diagrams went beyond the most intuitive structures, I found myself starting to get confused. The examples helped me gain some clarity but I know I would need to invest more time to understand deeply the full value of DAGs.

Cunningham's book really hits its full stride with the fourth chapter: the potential outcomes causal model. This sets the frame for the rest of the book, which has chapters on matching and subclassification, regression discontinuity, instrumental variables, panel data, difference-in-differences, and synthetic control. For my own students in Applied Econometrics for Policy Analysis, it made sense to recommend the Introduction and then to kick off the course with the potential outcomes chapter. In fact, I had recently thought about rearranging my course to start with matching instead of difference-in-differences (after listening to a talk by Pedro Sant'Anna describing some alternatives to regression) and this aligned perfectly with the organization of Cunningham's book. Essentially, if we begin with the potential outcomes framework, we see the fundamental problem of missing counterfactuals for individual outcomes. A natural way to think of imputing these counterfactuals is first at

the individual level, which is how matching methods work. The methods that follow use regression to add structure (and assumptions) to get at average treatment effects, and the reader sees how every kind of estimated counterfactual outcome is, at its roots, some choice of weighted average of the comparison group members. Regression discontinuity, instrumental variables, panel data methods, difference-in-differences, and synthetic control are suddenly not just a list of estimators, but really a *family* of estimators. This text wonderfully conveys the coherence of this wide variety of causal inference methods by explicitly grounding them in a common framework.

Overall, this book does an important service through both its tone and its content – it does not make these methods easy but it does make them easier to understand. Perhaps more importantly, it makes you believe that with some effort, *you* will be able to understand and implement them. The book is not oversimplified or boiled down to a lowest common denominator. When it gets technical, it is rigorous. But a huge added value of the book is the way that the context is laid around each method – where it came from, what kinds of problems it can address, how the literature has developed. From a pedagogical perspective, this book is much more likely than a usual textbook to help a student (or you) be able to see which methods might make sense for a specific project. In addition, one of Cunningham's key innovations is the explicit use of both Stata and R code throughout the book, combined with a library of datasets available for free online, so that a reader can not only choose a method but actually see how to implement it with real data. He walks through the intuition of what the code is doing and explains the tables and figures produced. Cunningham is a big believer in a hands-on approach to teaching these models, and his courses involve significant lab time. Using this book, one could modify an existing, more traditional course to include more in-person practicum time, reflecting a flipped-classroom approach to learning.

For my course (taught to PhD students in Public Administration), this book still needs a companion text but I recommend such texts for reference rather than as our main guides through the material. For example, a classic panel data text may be needed if the students want to really nail down some methodological details. More generally, if students haven't seen the methods before, other articles or texts may need to come first. My course combines a daily research article

discussion with more explicitly methodological instruction, and I point students to Cunningham's chapters frequently. Beyond the classroom, Cunningham's book does well in helping my students think through the nature of the causal inference problem they are attacking, determining a research design that could uncover the effect, and – importantly – implementing it! Suddenly, the proofs in other textbooks take their proportional place as only *one* part of what it means to be trained in econometrics as an empirical researcher. Cunningham's book helps not just by showing you the models and their assumptions but by convincing you that excellent analytical work is truly good work and worthy of our pursuit. This book is as good as it gets: it makes me want to be a better researcher. ■