

Book Review: Causal Inference: The Mixtape by Scott Cunningham

In Causal Inference: The Mixtape, Scott Cunningham offers a new guide to methods for determining cause and effect in the social sciences. In summarising, systematising and prioritising methodological tools for researchers, this book will be of use to all social scientists looking to validate their quantitative findings, recommends Simeon Mitropolitski.

Causal Inference: The Mixtape. Scott Cunningham. Yale University Press. 2021.

Find this book (affiliate link): [amazon](#)

Many of us have been told that 'correlation does not automatically mean causation'. In other words, that covariation between variables representing cause and effect is not a sufficient condition for drawing valid causal conclusions based on observable data. Sometimes a correlation is a spurious one. Two variables, such as shoe size and IQ, may correlate very well, yet there is not always direct causation from the former to the latter. The proverbial rooster announcing the sunrise may disappear, yet a new day will come as always.

But what about another issue, no less important for the advancement of knowledge, this time dealing with the presence of no-good correlation which in fact hides the presence of real causation? What about a central bank frequently increasing and decreasing main interest rates, thus counterbalancing economic cycles, without this having any apparent impact on real economy growth rates? Or on consumer price indexes, for that matter? Does this suggest that interest rates do not affect GDP rates?

These are questions that look for answers in Scott Cunningham's book, [Causal Inference: The Mixtape](#). The author defines causal inference as the impact of events and choices on a given outcome of interest (4). The first few chapters present the problem itself: namely the apparent lack of correlation while causation is present; the remaining chapters offer possible methodological solutions.

Among the latter, one stands out for its applicability to most social sciences: regression discontinuity. Cunningham dedicates an entire chapter to it, claiming that the popularity of regression discontinuity has increased in recent years. He uses Google Scholar to show that it has made the quantitative jump between 2000 and 2020 from close to zero studies to almost 6000 (242). In brief, the problem of 'having no correlation but still having causation' comes from amalgamating two or more qualitatively different groups within one statistical population, thus setting the table for unconvincing statistical regressions regarding the strength and character of the relations between main variables.

Just for illustration purposes, it would be like putting university graduates and elementary school children as elements in the same dataset, looking for causation between different factors in education regarding their individual performance. Factors showing good results within one of these subgroups may not be so promising within the other. Showing what the problem is automatically leads to suggestions of how to solve it. Instead of having one comprehensive statistical regression that may hide correlation between variables, statisticians may subdivide populations before applying regression analysis. Instead of one rather weak correlation, they will have two or more strong correlations, which will lead them to control for possible causations.



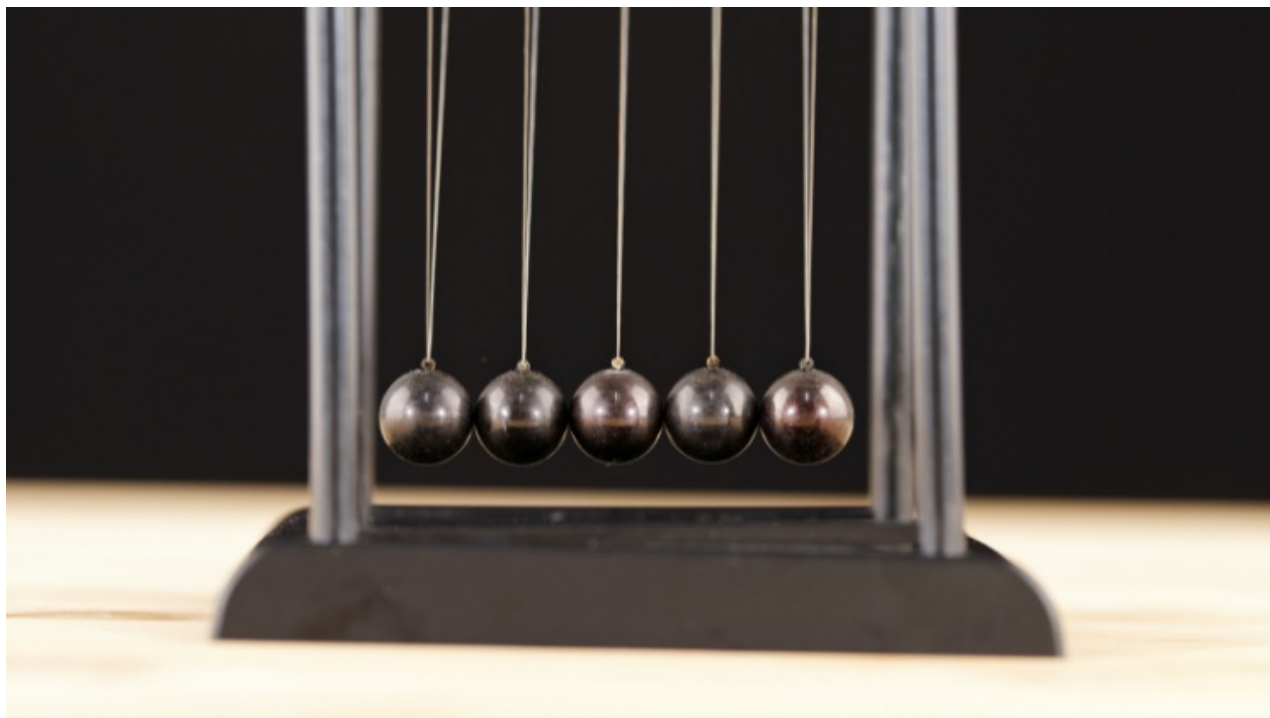


Photo by [engin akyurt](#) on [Unsplash](#)

Regression discontinuity is not the only tool presented by Cunningham in order to reveal causation when there is no-good or no correlation. Other instruments at the disposal of statisticians deal with matching and classifying elements in the dataset, with the use of instrumental variables, with panels and in estimations of difference-in-differences. These tools applied at different stages of research, from planning new studies to analysing data, are expected to decrease the probability of finding no statistical correlation, and thus to increase the probability of suggesting possible causation between variables. Even if the author mostly refers to economics as the potential field of application for these methods, all social sciences using statistical methods can benefit from them.

Cunningham's book is interesting not only regarding its main topic, but also as far as its methodology is concerned. This is not just a theoretical narrative, far from it. About half of the entire study represents tables, graphs, formulas and also protocols for producing statistical analysis using either Stata (paid) or R (free) programming languages. The technical language of the book requires prior knowledge beyond basic introductory courses in statistics. Some quite sophisticated terms, like 'directed acyclic graphs', are not sufficiently explained to the reader before solving the main theoretical problem. Others, far less complicated, like 'panel data' and 'randomization', are properly defined. The narrative style frequently changes from highly formalised technical notes (formulas, Stata and R protocols) to unnecessary biographical notes regarding some of the researchers who have made important contributions to the subject matter.

Causal Inference does not contain new theories and concepts. It summarises, systematises and prioritises already existing methodological tools that aim to validate statistical results. The main intended audience is economists, but all social scientists using statistical methods can benefit from Cunningham's book. I wholeheartedly recommend *Causal Inference* to lecturers in methodology classes as well as graduate students and researchers looking to validate their quantitative findings.

Note: This review gives the views of the author, and not the position of the LSE Review of Books blog, or of the London School of Economics and Political Science. The LSE RB blog may receive a small commission if you choose to make a purchase through the above Amazon affiliate link. This is entirely independent of the coverage of the book on LSE Review of Books.
