Book Review: “Mastering ’Metrics” by Joshua D. Angrist and Jörn-Steffen Pischke

Danilo Freire

2nd February 2015

The field of causal inference has seen a remarkable development in the last years. While social scientists have always devoted considerable effort to understand causal effects, their quest for the perfect identification strategy received a new impulse after the publication of Donald Rubin’s causal model in the late 1970s (Holland, 1986; Morgan & Winship, 2007). Although the “potential outcomes revolution” is still in its infancy, a myriad of statistical methods have been designed to help researchers untangle true relationships from spurious effects (Imbens & Rubin, 2009). However, whereas such methods already have an established place in cutting-edge publications, they remain relatively unknown to undergraduate students of economics and to academics of other areas.

Joshua D. Angrist and Jörn-Steffen Pischke try to fill this gap with their new book, *Mastering ’Metrics: The Path From Cause to Effect* (2014). Angrist and Pischke are household names of applied econometrics – *’metrics* in economics parlance – not only because of their many influential papers on labour markets and education, but mainly for their companion to the aspiring econometrician, *Mostly Harmless Econometrics* (2008). But in contrast with its famous sibling, *Mastering ’Metrics* was not written with a PhD candidate in mind. *’Metrics* is very light on mathematics and its many Kung Fu-themed jokes create a conversational tone which is lacking in more formal texts. This is perhaps the first econometrics book that can be read without much effort, making it suitable to those meeting the discipline for the first time or to scholars with little quantitative background.

’Metrics is focused exclusively on what the authors call “the furious five”, a set of statistical techniques to assess causal relationships from observational data. Namely, the methods are: randomised trials, regression with matching, instrumental variables/two-stage least squares, regression discontinuity design, and differences in differences. Angrist and Pischke dedicate one chapter to each technique, and all of them follow the same structure: they start with one or two empirical examples of how the method is used, continue with a theoretical discussion, and conclude with a brief appendix. The appendices include recommended readings, basic mathematical notation, and a short note on the researcher who developed the technique. Each chapter is self-contained and can be read independently.

The book begins with an exposition of potential outcomes and *experimental random assignment*, the foundations of Rubin’s causal model of inference (Rubin, 1974, 1978). Then the authors ask a puzzling question: does health insurance make people healthier? The answer is not as trivial as it seems. A simple comparison between those who are insured and those who are not shows that the second group tends to enjoy considerably better health standards than the former. The tricky part is that those groups cannot be directly compared because of *selection bias*: people who can afford private health insurance are usually richer than average, so they exercise more, have a well-balanced diet and other things. A reliable inference strategy should be able to mitigate such unobservable differences and estimate the true causal effect. Ideally, a researcher would measure the same individual in two hypothetical situations, when he or she has a health insure and when he or she does not, and then take the difference between both "states of the world". It is evident, however, that *only one of these outcomes is actually observable*. This is what Holland (1986) calls "the fundamental problem of causal inference": it is impossible to observe an individual with two characteristics at the same time. A reasonable solution to this problem is to have two randomly-chosen groups of people – one that receives the treatment and another one that does not –, average both of them and calculate the difference between the results. If the samples are large enough and the treatment condition has been randomly assigned, all biases will "average out" and the groups can thus be meaningfully compared. After presenting the assumptions required by the experimental design (such as *ceteris paribus* and constant effects), the authors mention that the state of Oregon involuntarily conducted a true randomised experiment when it decided to offer health insurance through lottery to about 75,000 people. Results show that the health insurance coverage did help people to get more medical treatment, but it had no effect when it comes to blood pressure or other medical conditions.

Chapter two deals with regression methods. The motivating example is the impact of schooling on wages. But how to estimate the effect correctly when people have gone to colleges as different as the University of Massachusetts and Harvard? Angrist and Pischke take that question as a starting point to offer a detailed explanation of a common problem in regression analysis: *ommited variable bias* (OVB). This type of bias occurs when the researcher does not include an important independent variable in the model he wants to estimate, thus biasing the coefficients of the other factors. As the authors note, "regression is a way to make other things equal, but equality is generated only for variables that are included as controls on the right-side of the model" (Angrist & Pischke, 2014, 67). In the school example above, the authors suggest that family size or income may have determined if a student would go to college A or B, and therefore influenced his future income. Without such controls, there is no *ceteris paribus*, the assumption that "all other things should be equal". Curiously, the authors do not show present any modern matching technique to pre-process the data before running a regression model (e.g. Ho et al., 2007; Rosenbaum & Rubin, 1985), although this is a widespread practice amongst quantitative scholars (Stuart, 2010). A short mention would have solved the issue.

The next chapter introduces the reader to instrumental variables (IV) and the two-stage least squares regression framework (2SLS). Instrumental variables serve a useful purpose in applied research, enabling scholars to estimate unbiased coefficients and assess causality when randomisation is unfeasible. A reliable instrument should possess three distinct characteristics:

1) it has a causal effect on the independent variable one is trying to evaluate (the "treatment"); 2) it is unrelated with omitted variables the researcher is willing to control for; 3) it affects the dependent variable though a single channel, the treatment. If those conditions are met, the estimation can be considered "as good as if it was obtained with an experiment" (Angrist & Pischke, 2014, 106). Two-stage least squares, in turn, gives flexibility to IV estimations by allowing the analyst to add other independent variables to control for possibly omitted factors in his or her model. In this regard, 2SLS can substantially increase the accuracy of the estimations and allow for continuous, dichotomous and ordinal instruments to be included in a regression model. Two examples are used to illustrate the method: the first one is on the impact of charter school on wages (taking a lottery as an instrument), and the second one explains the influence of family size and wealth using the random incidence of twins as an IV.

The authors proceed to a discussion on regression discontinuity design (RDD). This method is particularly adequate to analyse natural experiments where an intervention caused a (sharp or fuzzy) cutoff point in the independent variable of interest. In their book example, Angrist & Pischke (2014, 150) employ a RDD to evaluate the causal effect of legal access to alcohol on death rates. There is a clear threshold that indicates legal drinking age in the United States (21 years old), and the authors use that jump to address their question. The theory behind RDD suggests that if one takes the cases that lie closely to the discontinuity, all of them have very similar characteristics apart from the causal effect one is trying to estimate. Thus, the difference between the death rate of youngsters who are about to turn 21 and those who have just passed that date are due to legal alcohol consumption. The results, for those who are curious, is that there is indeed a huge spike in death rates for youngsters who are 21 years old. The authors also discuss fuzzy RDD, that is, when the probability of an event happening increases as an individual approaches the cutoff. In this sense, RDD is similar to an IV where the discontinuity becomes an instrumental variable for treatment status. This concept is not very intuitive and it is difficult to explain without resorting to mathematics, but the authors do a good job by presenting the intuition behind sharp and fuzzy RDD with graphs. Although one sometimes cannot be sure if there is a jump by merely looking at covariate plots, it is probably the best way to approach the problem.

The last methodological chapter in the book analyses differences in differences (DID). This method is based upon the *common trends* assumption, which states that "sometimes, treatment and control outcomes move in parallel in the absence of treatment" (Angrist & Pischke, 2014, 178). When this is the case – and it is possible to test it visually and mathematically – one can attribute the shifting in the expected trend to a causal effect, estimated by subtracting the average gain over time in the control group is from the gain over time in the exposed group (Imbens & Wooldridge, 2009, 67). Such procedure removes eventual selection biases from the analysis. A study of the impact of the impact of different monetary policies on two of the twelve US Federal Reserve districts in 1929 is the authors' case study of choice. The DID results indicates that looser monetary policies led to less trouble in the financial sector, with fewer banks going bankrupt in posterior years. The effect of monetary policy can be estimated with a dummy variable that captures the state which had the treatment, a time indicator for post-treatment periods, and an interaction term composed by them both. The coefficient resulting from that multiplication is the DID causal effect (Angrist & Pischke, 2014, 186). The framework is flexible enough to include multiple years and comparison groups, as the authors note. Angrist and Pischke analyse a case of multiple legal drinking age being implemented in different states and estimate their impact on death rates, the same question presented in the earlier chapter on RDD. After controlling for state and time effects, the authors can obtain unbiased results with DID.

Mastering 'Metrics concludes with an thorough investigation of a hot topic in labour and education economics, the effect of schooling on earnings. The authors present a myriad of studies that tackle that question using all methods shown in the book apart from experiments. There has never been a truly randomised controlled trial on returns to schooling, so no example of that technique is provided. The chapter begins with Jacob Mincer's (1974) and Zvi Griliches' (1977) regressions on returns to schooling, presenting the pros and cons of those pioneering studies. Angrist and Pisckhe then move to an evaluation of the work of Ashenfelter & Krueger (1994) and Ashenfelter & Rouse (1998) using twins as an IV for the difference in education. The results seem to confirm Mincer's regressions. Next, the authors show how differences in differences and instrumental variables can be combined. They present two papers published by Acemoglu & Angrist (2001) and Angrist & Krueger (1991) on the effects of compulsory attendance laws on returns to schooling. By using both methods one can ameliorate OVB issues. Finally, the authors discuss a paper by Clark & Martorell (2014) that employs a clever fuzzy RDD strategy using last-chance exams in Texas. The book ends with a short comment on measurement errors and notes regarding the data sources.

How can a political scientist benefit from reading *Mastering 'Metrics*? As the discipline moves towards a more rigorous approach to causality, the lessons offered by Angrist and Pischke are well timed. Moreover, the approachable style of the book can be stimulating to students, so it is a good recommendation for those teaching quantitative methods at undergraduate level.

Nevertheless, the book has its shortcomings, too. *'Metrics* has a much more limited scope than traditional textbooks such as Stock & Watson (2003) or Wooldridge (2012). Angrist and Pischke's discussion of causality is solely based upon the linear model, and all techniques presented in the book are applied only to ordinary least squares. In this regard, the authors do not provide any hint on how to employ their causality toolbox to dichotomous, ordinal, count or nominal variables. Even time series are absent from the book. Those looking for an introduction to those topics should therefore refer to one of the volumes mentioned above. Furthermore, *'Metrics* does not include any exercises, and although the authors claim that additional explanatory material will soon be available online[1], so far students cannot practice the concepts they have learnt in the book.

Despite its flaws, *Mastering 'Metrics* provides an enriching and even entertaining introduction to five important econometrics methods. Its discussion of causality is very clear, the

[1]See: http://masteringmetrics.com/resources/. Access: 2nd February 2015.

examples are relevant and the appendices present the mathematical notation readers may need to follow more advanced texts. Although the book alone will not make one a Shaolin master of econometrics, it will lead the hopeful apprentice onto the righteous path. And this is not a small feat.

References

Acemoglu, D. & Angrist, J. (2001). How large are human-capital externalities? evidence from compulsory-schooling laws. In *NBER Macroeconomics Annual 2000, Volume 15* (pp. 9–74). MIT Press.

Angrist, J. D. & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106(4), 979–1014.

Angrist, J. D. & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

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

Ashenfelter, O. & Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review*, (pp. 1157–1173).

Ashenfelter, O. & Rouse, C. (1998). Income, schooling, and ability: Evidence from a new sample of identical twins. *The Quarterly Journal of Economics*.

Clark, D. & Martorell, P. (2014). The signaling value of a high school diploma. *Journal of Political Economy*, 122(2), 282–318.

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, (pp. 1–22).

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3), 199–236.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396), 945–960.

Imbens, G. & Rubin, D. B. (2009). *Causal inference in statistics, and in the social and biomedical sciences*. Cambridge University Press.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings.* National Bureau of Economic Research.

Morgan, S. L. & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research.* Cambridge University Press.

Rosenbaum, P. R. & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33–38.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5), 688.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, (pp. 34–58).

Stock, J. H. & Watson, M. W. (2003). *Introduction to econometrics.* Addison Wesley Boston.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1), 1–29.

Wooldridge, J. (2012). *Introductory econometrics: A modern approach.* Cengage Learning.