Handbook on impact evaluation: quantitative methods and practices, by S.R. Khandker, G.B. Koolwal and H.A. Samad

Richard Palmer-Jones


To link to this article: http://dx.doi.org/10.1080/19439342.2010.499188

Published online: 24 Sep 2010.

Submit your article to this journal

Article views: 130

View related articles
BOOK REVIEW

Handbook on impact evaluation: quantitative methods and practices, by
S.R. Khandker, G.B. Koolwal and H.A. Samad, Washington, DC, World Bank, 2009,

This is a short book that, according to the Foreword, is a ‘comprehensive overview of steps in designing and evaluating programs amid uncertain and potentially confounding conditions’ (p. xiii). It is based on ‘materials prepared for a series of impact evaluation workshops . . . sponsored by the World Bank Institute’ (p. xv). It consists of two parts: the first on Methods and Practices, and the second a set of exercises using the commercial data processing and statistical software Stata (for some exercises, version 8 or above is probably required). The book explains concepts and practices in simple terms, does not make use of any complex algebra and in Part 1 gives a wide range of examples of impact evaluation (IE) from different countries and programmes/projects in short boxes, with references that the reader can usefully follow up. Part 2 consists of a series of worked examples using data from a quite well-known example of an IE – that of microfinance programmes in Bangladesh, of which Khandker was a co-author (Pitt and Khandker 1998) or sole author (Khandker 2005).

The chapters in Part 1 (Chapters 1–10) have a common format, deriving, no doubt, from the course from which the material derives. Each chapter has a Summary, Learning Objectives, Introduction, some chapter-specific headings, Questions, Notes and References. The topics covered include a discussion of monitoring and evaluation, and the logic of ‘modern’ IE. The next six chapters outline different approaches to modern impact evaluation: Chapter 1 is a discussion of the randomised control trials (RCTs) approach (Chapter 2), which is strongly advocated in some quarters (Duflo and Kremer 2003, Miguel and Kremer 2004, Banerjee et al. 2009), although subject to criticism by others who assert the continuing value of observational data (Deaton 2009, Heckman and Urzua 2009, Imbens 2009, Pritchett 2009). Chapter 3 covers propensity score matching (PSM), Chapter 4 the double difference (DD) method, Chapter 5 instrumental variables (IV), Chapter 6 the regression discontinuity (RD) method and Chapter 7 the pipeline method. Following these presentations, Chapter 8 discusses measuring distributional effects, Chapter 9 model-based methods of policy evaluation, and then Chapter 10 concludes this part of the book. Each chapter includes box materials on studies illustrating points being made, and ends with a set of questions on the learning, most of which are multiple choice with answers clearly signalled in the preceding text; answers are also given.

Part 2 has five chapters on the use of various methods described in Part 1, using Stata and the dataset derived (mentioned above). These chapters are preceded by a brief introduction to Stata (Chapter 11). There are chapters on randomised evaluation, PSM, DD (which includes a DD approach using panel data), IV, and RD (Chapters 12–16). This part is followed by answers to questions and an appendix with the Stata code for Chapters 12–16. The Table of Contents lists all these chapters and sub-headings, and the boxes, with
their pages. There is a useful index but no comprehensive bibliography; however, there are references at the end of each of Chapters 1–9.

In general this is a very useful book for anyone wishing to tutor themselves in these IE methods, or prepare an IE course to teach them. The standard of knowledge required is quite basic for either task (a master’s, or indeed good undergraduate, degree, with a first econometrics course, or the ability to quickly master these materials). The text requires concentration and application as it is quite dense. Fortunately, perhaps, the Stata code can be downloaded, from a website that is active at the time of writing, as can the two data files. However, since all the codes are given in the appendix, with explicit instructions in the main text, it might be better not to allow students to cut and paste, but rather to think through the code for themselves. From my use of these materials they are correct, as is the text describing them. It is a pity that the book only uses Stata code since the cost of this software makes it prohibitive for many users. The free software ‘R’ has equivalent capabilities, so perhaps in a future edition the authors can also describe and present their case examples in R as well as Stata. As a minor point, in its example of PSM the text only uses the Stata pscore and nnmatch user-written commands – and not psmatch2, which is another popular PSM routine. The book is not particularly cheap. But it is a useful source for instructors in cash-strapped situations, and is uniquely useful in the context of IE of development interventions.

I have, however, a number of reservations, centering on an apparently rather over-sanguine view of the robustness of these methods, and a tendency to assume that the data used in the evaluations discussed, and particularly those used in the examples of Chapters 12–16, are accurate and largely unproblematic. The fragility in practice of these methods is belied by the language used in the book. Thus, while acknowledging in its opening sentence that ‘[I]dentifying the precise effect of a policy is a complex and challenging task’ (p. xiii), the undoubted thrust of the book is that this (precise identification) is just what the methods described achieve. It would have been good to have some cautionary accounts of where these IE methods have been, more or less, unsuccessful in achieving this objective (see below).

In addition, a major deficiency of the book as an introduction to IE is its failure to discuss either research design for surveys for IE or the requirements for, and methods of, transforming raw data on survey forms into electronic data that can be confidently analysed with the procedures described. One might not expect this book to discuss the details of survey conduct, in particular, because this is dealt with reasonably in the World Bank’s three-volume work *Living Standards Measurement Surveys* (Grosh and Glewwe 2000), but issues of research design and of data management and processing are appropriate. Thus, discussion of some aspects of the sampling designs of surveys and of data production and transformation would seem warranted, since they are crucial to confident interpretation of findings. Discussion of the research design may be particularly important for RCTs since, notwithstanding the belief that they overcome the attribution problems, such social experiments have deep flaws that are not always recognised. The chief flaw of RCTs is that they cannot generally, ethically and or practically, be double-blinded; hence, compared with, say, most medical trials, both those who conduct the experiment and those who are participants generally know they are participating and this inevitably affects their behaviour compared with the real world, although it may not always be in the interests of proponents of RCTs to acknowledge this. Also, pseudo-random methods are often used during the process of assigning individuals to treatment and control groups, which many studies fail to describe accurately. This can have consequences for the reliability of the estimates obtained from RCTs. While problems related to this deficiency are mentioned in authoritative texts
(Duflo and Kremer 2003) as the Hawthorne and John Henry effects, these are not given the respect they warrant in most applications in developed countries (Scrivens 2008), let alone developing ones where the power of economists’ ideas arguably holds greater sway.

The relative neglect of data quality and the fragility of methods can be illustrated using the example from which the data used in Chapters 12–16 are drawn. Firstly, although the authors do not state that these data can be used to replicate the Pitt and Khandker (1998) study, the reader should be warned that the two datasets used are considerably simplified and the results of the exercises in the book cannot be taken as valid assessments of these microfinance interventions. This is a ‘cleaned’ subset of the data from the original study that are in large part, but not completely, available on the World Bank website.¹ It should have been possible for the purposes of helping the reader to understand the research design and data management issues alluded to above, to discuss the survey design in more detail than on pages 96–98, and to show how these ‘raw’ data could be manipulated to arrive at the data used in the book. However, these data are neither those used in Pitt and Khandker (1998), nor in Khandker (2005), nor even a proper and accurate subset of those data, as it seems the data do not give the same results when sympathetically applied to the original data. As far as I can see this fact is not mentioned anywhere. This might not be such a problem was it not for the fact that the original analyses that use these data (Pitt and Khandker 1998; Khandker 2005) have been comprehensively challenged in large part because of problems, if not errors, in the construction of the variables, as well as on methodological grounds (Morduch 1998, Roodman and Morduch 2009). The failure to mention doubts and qualifications with regard to the original studies (the authors cannot be blamed perhaps for not referring to recent publications given the usual delays in publication) lends support to the image that these methods support fairly clear conclusions, adding to the impression of authority that those possessing the skills that will be acquired through studying this book should be accorded. Some people may not see this as a problem for pedagogic purposes at the level of the student for whom it is appropriate, but I think it is. Those wishing for a more nuanced discussion of the observational methods of IE may like to consult Rosenbaum (2010) (which gives R examples for PSM), or the earlier standard reference Rosenbaum (2001). They may also consult Morgan and Winship (2007) and Shadish et al. (2002), while using this book to learn the basics of IE data analysis. Nichols (2007, 2008) offers a somewhat more critical discussion of IE methods with Stata.

Acknowledgements
With thanks to Maren Duvendack for helpful suggestions.

Note

References


Richard Palmer-Jones
School of International Development, University of East Anglia, UK
Email: r.palmer-jones@uea.ac.uk
© 2010, Richard Palmer-Jones