Book Review/Compte rendu

Robert Andersen, *Modern Methods for Robust Regression. Quantitative Applications in the Social Sciences*. London and Los Angeles: Sage, 2007, 128 pp. $US 16.95 paper (978-1-4129-4072-6)

Researchers employing linear models typically rely on estimates obtained via least squares, which is the optimum estimating procedure when the data are well behaved and all of the regression assumptions are met. However, a common problem confronting researchers is that real data are not typically "well behaved." For example, social researchers frequently encounter variables with heavy tailed distributions and/or observations that stand apart from the general pattern of the bulk of the data. Sometimes these are miscoded observations; other times they are simply unique cases that may even represent the most intriguing part of the data. In either case they have the potential to severely distort regression estimates and their corresponding standard errors, not to mention tests of statistical significance and measures of fit. Fortunately, modern methods are available which limit the influence of unusual observations on regression estimates. These methods have been available in statistics and econometrics for decades, and now, thanks to Robert Andersen, they are finally accessible to social scientists outside of economics.

Modern Methods for Robust Regression discusses the detection and proper handling of influential observations. It begins with an introduction to "modern methods" and background on the methodology, followed by a discussion of how to use graphical and numerical methods for detecting influential data points. When diagnostic tests reveal that one or more data points are problematic for traditional methods, those observations require special consideration. Unfortunately, however, many researchers are unaware of the implications of influential observations, or ignore the issue because they lack the appropriate training on how to deal with them. Outliers may also frequently be overlooked in survey research because of the common misconception that they are only a concern in small samples.

Robust methods are viable options for dealing with influential observations and will be especially attractive to researchers who are uncomfortable with simply discarding influential data points without sound justification. In general, these techniques work by employing complex

methods for reweighting influential observations. Hence, they are "robust" or "resistant" to outliers. The book covers the evolution of robust methods, devoting special attention to the latest techniques available, along with their strengths and limitations. The discussion extends to standard errors for robust regression models (not to be confused with robust standard errors); a particularly useful technique for drawing conclusions about estimates that are generated from smaller samples. Also covered are robust methods for regression models that accommodate response variables that are not normally distributed, or that take on discrete values. Practical examples are provided for robust logistic regression and robust Poisson regression. These examples reveal that robust methods are less practical for models involving dichotomous response variables (e.g., logistic regression), but can be useful for neutralizing influential observations for other types of generalized linear models.

The statistical analyses in the book are employed using R. This is a logical choice as R has tremendous facilities available for the procedures utilized in the book, and is available free of charge to anybody with a computer and internet connection. The web-appendix also includes the R-code for all of the examples in the book, and the text lists some procedures for robust regression that are available in SAS.

The book contains a considerable number of mathematical expressions, some of which employ matrix notation. There is also a fair amount of technical information regarding estimating procedures that often involve advanced (for social scientists) mathematical concepts. Whether this level of detail is necessary to convey most of the key ideas presented in the book is a matter of debate. Nevertheless, the frequent use of mathematical expressions could make it difficult to attract a key audience for the book — academic researchers who would benefit tremendously from the material, but are discouraged from reading sources perceived to be heavy in mathematical content. For this reason, I expect that this book will be most attractive to a readership with advanced training in quantitative methods.

Still, the topics are covered with much less "technical detail" than the other sources that I have read, making it considerably more accessible to social scientists than most, if not all, of the other books available on this topic. Moreover, the material is presented in a very clear and succinct manner, and includes an excellent use of practical examples that are easy to follow. Thus, equations aside, most of this book should be accessible to the average PhD student with at least one graduate-level course in regression.

This book is especially attractive as a teaching resource, as it provides a very thorough treatment of advanced topics for a short monograph. As

well, many sections include insightful suggested readings for ambitious readers who would like to explore these topics further. It would fit nicely as a required or recommended reading for a second year graduate level course involving advanced topics in quantitative methods, or on a reading list for a PhD comprehensive examination on research methods.

In sum, *Modern Methods for Robust Regression* does an excellent job at expanding the options available to social scientists for dealing with influential observations, and I am confident that those who read this book will benefit immensely from the material. At the very least readers will become better informed and more proficient social researchers. Robert Andersen couldn't have found a more appropriate venue to present this material than a "little green book."

David Walters   University of Guelph

David Walters is Associate Professor in the Department of Sociology and Anthropology at the University of Guelph. His areas of interest and research include quantitative methodology, sociology of education, health and wellbeing, and immigration. His most recent publications have appeared in the *Canadian Review of Sociology*, *Canadian Public Policy*, and *Journal of Health and Social Behaviour*. dwalters@uoguelph.ca