Observational Studies.

Paul R. ROSENBAUM. New York: Springer-Verlag, 1995. xv + 230 pp. $44.95.

This book is timely, illuminating, applicable, and well written. The problem under consideration, simply stated, is as follows. We want to assess the effect of a "treatment" on some population of "subjects" (often people); to this end we compare "responses" of the subjects in the treated and nontreated (control) groups. But for whatever reason—ethical considerations, practical constraints, and so on—we are not able to assign subjects randomly to treatment and control. Thus we cannot be sure that the observed effect (if any) in the treatment group is actually due to the treatment or rather to some *hidden bias;* that is, to some other, unobserved difference between the groups. How can we detect hidden biases? How can their importance be assessed? How much hidden bias would it take to reverse the conclusions of the study? The author proposes a formal statistical framework to address these formidable problems.

We statisticians are quite confident in our purely mathematical derivations, but our science becomes increasingly uncertain as we leave this comfortable territory to confront the seemingly imponderable relationship of statistical analysis to underlying reality. Robust statistics (and robustness studies) represent an a priori attempt to deal with some problems of the latter sort. But in observational studies, even the minimal assumptions of robust procedures may fail. One of the main topics of this book is post hoc analysis of the sensitivity of results (expressed, for example, as a p value) to departures from assumptions, chief among these being the assumption that treatment and control were randomly assigned within groups of comparable subjects. Behind such formal techniques lie fundamental epistemological questions, and indeed philosophy of science is discussed throughout; among the works cited are Popper's *The Logic of Scientific Discovery* (1959) and Wittgenstein's *On Certainty* (1969).

I approached the book not as an expert in the research area, but rather as a statistical consultant charged with testing certain hypotheses in a large and messy observational data set (the U.S. Environmental Protection Agency's *Toxics Release Inventory*). In my problem there is essentially one treatment group, one control group, and several categorical covariates. I had settled on rank-based analysis of variance (ANOVA) to compare treatment and control nonparametrically while adjusting for covariates, but although assignment to treatment and control is undoubtedly nonrandom in my problem, I had envisioned only a rudimentary analysis of sensitivity to hidden biases. Having read this book, I have a much better understanding of the specific problems presented by observational studies, my use of the relevant statistical procedures is better informed, and I can carry out a formal sensitivity analysis with interpretable results.

The book is organized as follows. Chapter 1, "Observational Studies," gives an overview and several interesting examples of successful studies, such as the establishment of the link between smoking and lung cancer in the 1964 U.S. Surgeon General's report, and of unsuccessful studies, such as Cameron and Pauling's 1976 finding that vitamin C is an effective treatment for advanced cancer, which was later overturned by a randomized controlled experiment. Chapter 2, "Randomized Experiments," reviews this topic and serves as a reference point for the rest of the book, the idea being that "a theory of observational studies must have a clear view of the role of randomization, so it can have an equally clear view of the consequences of its absence" (p. 13). In addition to conceptual issues, hypothesis tests and confidence intervals using several classes of statistics (sum, arrangement-increasing, order-preserving, and sign-score) are discussed. I found the author's approach here somewhat unexpected; under the null hypothesis of no treatment effect, the responses (observed values) of the experimental units are taken to be fixed, and randomness enters only through the random assignment of units to treatment and control groups. (This is analogous to "design-based" finite population sampling theory, where the population values are taken to be fixed and randomness enters only through selection of the sample.) This approach is easy to understand when considering the behavior, under the null hypothesis, of a statistic with a fixed finite sample space for any population (such as the Mann–Whitney–Wilcoxon two-sample statistic); the arguments regarding alternative hypotheses and confidence intervals take a bit more effort to follow. Nonetheless, the discussion is coherent and useful; I particularly liked Section 2.10, which discusses the set of treatment assignments as a finite distributive lattice. Still, I would recommend that the reader also have a good grounding in the "model-based" nonparametric theory of experiments, under which the observations are typically taken to be independent random variables from a location family (see, e.g., Dean and Wolfe 1996).

Chapter 3, "Overt Bias in Observational Studies," deals with stratification based on an observed covariate vector $\mathbf{x}$, the idea being to assess the effect of the treatment within apparently homogeneous groups of experimental units. For the jth unit, the author defines the probability of being assigned to treatment, π_j, and the *propensity score,* $\lambda(\mathbf{x}_j)$, which is the conditional probability of receiving the treatment given $\mathbf{x}_j$. A study is free of hidden bias when $\pi_j = \lambda(\mathbf{x}_j)$ for all units j. In this case, if we could stratify *exactly* on $\mathbf{x}$ (e.g., if $\mathbf{x}$ had only a few values), then the observational study could be regarded as a randomized experiment (given the numbers of treated units in each stratum). Indeed, the same argument holds for stratification on the propensity score alone. The practical aspects of stratifying exactly or approximately on $\mathbf{x}$ or on λ are discussed here and in Chapter 9. But in fact we do believe that there is hidden bias, so that $\pi_j \neq \lambda(\mathbf{x}_j)$, and this is the subject of the remainder of the book.

Chapter 4, "Sensitivity to Hidden Bias," is the book's longest and to me most important chapter. Suppose that although the probability of assignment to treatment is not exactly a function of the observed covariate $\mathbf{x}$, the odds ratio for two units with the same value of $\mathbf{x}$ is bounded, so that for all units j and k with $\mathbf{x}_j = \mathbf{x}_k$, $1/\Gamma \leq \text{ODDS}_j/\text{ODDS}_k \leq \Gamma$, where $\text{ODDS}_j = \pi_j/(1-\pi_j)$ and $\Gamma \geq 1$ is fixed. This is equivalent to a model in which $\log(\text{ODDS}_j) = \kappa(\mathbf{x}_j) + \gamma u_j$, where $\kappa(\cdot)$ is an unknown function, γ is an unknown parameter, and u_j is a variable taking values in [0, 1] that essentially represents the effect on π_j of any unobserved (hence uncontrolled) covariates. We then watch what happens to the statistical analysis, typically summarized as a (maximum) p value, as Γ increases. For example, if $p \geq .05$ when $\Gamma \geq 3$, then we would say that it would take a hidden bias that tripled the odds of getting the treatment to invalidate the conclusion of the study. In the rest of the chapter these computations are worked out in detail for various classes of statistics and in several concrete examples. It is remarkable that by this measure, studies differ widely in their degree of sensitivity to hidden bias, so reporting the sensitivity analysis along with the original results would add substantial new information. It is also interesting to note that most of the examples discussed (theoretical or concrete) already use robust statistics—nonnormality, outliers, and so forth are not the main issues here.

Chapters 5–7 deal with methods of detecting the presence of hidden bias. The main idea in Chapter 5, "Known Effects," is that if we record an extra variable known to be unaffected by the treatment and a test shows a significant difference between the treatment and control groups for this variable, then there is evidence of hidden bias. Alternatively, if the extra variable must either be affected positively or not at all but the treatment group shows a negative effect for this variable, then hidden bias is again indicated. Chapter 6, "Multiple Reference Groups in Case-Referent Studies," focuses on the special case of case-referent studies, which compare "the frequency or intensity of exposure to the treatment among cases and among referents or noncases" (p. 154). (These appear in several places throughout the book.) Here the interplay of hidden bias and bias in selecting the cases and referents is discussed. In Chapter 7, "Multiple Control Groups," the question is whether evidence of hidden bias can be obtained by comparing outcomes among several different control groups.

Chapter 8, "Coherence and Focused Hypotheses," considers the idea that evidence that specifically fits an elaborate theory of the treatment effect may be less sensitive to hidden bias. In particular, the author discusses reducing multivariate to univariate data by means of a partial order that "points in the direction" of the anticipated treatment effect. The resulting "poset statistics" are studied, along with the associated sensitivity analyses. Chapter 9, "Constructing Matched Sets and Strata," begins by discussing stratification on the propensity score, which may be estimated from the sample. Next, the author lays out a theory of "optimal stratification" based on minimizing the distance between each treated unit and each control in terms of their observed covariate vectors, that is, in $\mathbf{x}$ space. In the following sections the problem of optimal stratification or matching is mapped to the problem of finding a minimum cost flow in a network, and existing techniques and software for this are discussed. Finally, the brief Chapter 10, "Some Strategic Issues," discusses the relationship between a study and its audience.

Observational Studies will be extremely useful to researchers and graduate students in the biomedical and social sciences. It will also be very interesting for statisticians, because quite a bit of the material appears here in book form for the first time and because new research areas are proposed, such as developing optimal tests for use in sensitivity analysis. (Its most direct predecessor is probably Cochran's *Planning and Analysis of Observational Studies* [1983], but Cochran did not complete his book, and the extant version, though conceptually illuminating, is relatively elementary from a technical viewpoint.) The book is structured as a textbook, with some problems for solution at the end of each chapter. For applied students, I can envision basing a course around the book and using actual studies from the scientific literature in addition to the problems; for statistics students, I would probably want to teach this material after a course in general nonparametrics, with some supplementary mathematical problems added. The book is well written in an area where clarity is difficult, and it successfully stresses the role of *reasoning* in science. There are not too many typographical errors in this first printing. In short, I highly recommend this book both for its interest and its utility; I believe it will set a new standard for the analysis of observational studies.

John BUNGE
Cornell University

REFERENCES

Cochran, W. (1983), *Planning and Analysis of Observational Studies*, New York: John Wiley.

Dean, A., and Wolfe, D. (1996), "Nonparametric Analysis of Experiments," in *Handbook of Statistics*, Vol. 13 eds. G. P. Patil and C. R. Rao, New York: North-Holland.

Popper, K. (1959), *The Logic of Scientific Discovery*, New York: Harper and Row.

Wittgenstein, L. (1969), *On Certainty*, New York: Harper and Row.

Estimation, Inference and Specification Analysis.

Halbert WHITE, Cambridge, U.K.: Cambridge University Press, 1994. x + 380 pp. $44.95.

In this monograph the author addresses the following questions:

- Under what conditions can standard parametric statistical techniques such as maximum likelihood be meaningfully applied to estimate the parameters of a misspecified model?
- What happens if these conditions are not met?
- How can one tell if these conditions are not met?

The starting point for answering these queries, in Chapter 2, is the Kullback–Leibler information criterion (KLIC) as a motivation for quasi–maximum likelihood estimation (QMLE). If the likelihood function is misspecified, then the QMLE of the parameter vector will converge, in probability or almost surely, to the parameter vector that minimizes the KLIC, provided that certain conditions are met. Some of these conditions concern the existence and measurability of QMLE, which are stated in Chapter 2. Conditions for convergence of the QMLE to some limit are set forth in Chapter 3. Chapter 4 is concerned with conditions for correctness of the specification of the likelihood and conditional likelihood function. In particular, the misspecification of the conditional density function and dynamic misspecification are treated separately, because the latter might be less harmful than the former; for example, in the case of a linear regression with serially correlated errors. Chapter 5 focuses on conditions for correctness of conditional expectation specifications and specifications of other conditional attributes. Chapter 6 provides generic conditions for asymptotic normality of the QMLE; in particular, the effect of misspecification on the asymptotic variance matrix is analyzed. Chapter 7 sets forth conditions for asymptotic efficiency of ML estimators. Chapter 8 is concerned with hypothesis testing with QMLE and the estimation of the asymptotic variance matrix of the QMLE. Chapter 9 discusses the m-testing framework of Newey (1985) and Tauchen (1995). This is a general framework for testing model *mis*specification, based on the simple idea that if the model is correctly specified, then certain moment conditions should hold. On the basis of the sample counterparts of these moment conditions, one can construct a Wald-type χ^2 test. Chapter 10 considers special cases of the m test, such as the LM test and Cox's test of nonnested hypotheses, to mention a few. Chapter 11 is devoted to White's information matrix test, which of course is also a special case of the m test. Finally, Chapter 12 contains some concluding remarks.

Each chapter except 1 and 12 has its own appendix containing the proofs, and three separate appendixes review the elementary concepts of measure theory and the Radon–Nikodym theorem, uniform laws of large numbers, and central limit theorems.

I wonder what kind of audience that the author had in mind when he wrote this book. Because this book is published as an Econometric Society monograph, I guess that he had also a subset of econometricians in mind. Indeed, in the introduction he writes, "I have attempted to make this book accessible to graduate students well-trained in econometrics or statistics; it is by no means self-contained." However, the treatment of the various topics uses an excessive amount of abstract probability and measure theory, generally far beyond the level of the average Ph.D. student in economics, even those who have taken a field in econometrics (except perhaps at UCSD). On the other hand, in Chapter 12 the author states that "the results of the foregoing chapters are intended to provide empirical researchers with an appreciation of the dangers of taking one's explanatory models too literally, and with the tools for coping with the necessity of using models, which by their very nature as human artifacts, may be misspecified to greater or lesser extent." If this was the intention, then the book should have been written in a different way, without the excessive use of measure theory, because I am sure that the number of empirical economics researchers (or empirical econometricians, for that matter) able to read this book is very limited. This is a pity, because the book does provide useful insight into the nature, consequences, and detection of model misspecification, but only after spending quite a lot of effort in translating the measure-theoretical jargon into plain English. Most of the topics in this book could have been treated in a much less formal way, without substantial loss of generality.

A large part of the contents of this book consists of previously published material by the author and others, which as usual has been subjected to severe page constraints imposed by the editors of econometric and statistical journals. So what is the purpose of reviewing this material in a less accessible and more compact way than the original articles? Admittedly, reviewing the literature is probably only part of the reason that the author wrote this book. Much of the material in Chapters 2–8 is a unification of the theory of QMLE, by stating generic conditions that cover a wide range of special cases. But this generality could have been achieved without the present mathematical purism.

Finally, the book would have been strengthened if a chapter on consistent model misspecification tests had been included. The usual m tests based on a finite number of moment conditions are not consistent, because it is always possible to construct an alternative data-generating process for which a given finite set of moment conditions hold. By using infinitely many moment conditions, or by comparing parametric functional specifications with nonparametric estimates, one can construct consistent m tests for misspecification of functional form. Note that both the author and I have worked in that area (see Bierens 1982, 1984, 1990; Bierens and Ploberger 1995; De Jong and Bierens 1994; Hong and White 1991; Stinchcombe and White 1991; White 1989; and White and Hong 1993, among others).

In summary, the actual contents of *Estimation, Inference, and Specification Analysis* deserve a larger readership than the book's presentation will likely permit.

Herman J. BIERENS
Southern Methodist University and Tilburg University

REFERENCES

Bierens, H. J. (1982), "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

——— (1984), "Model Specification Testing of Time Series Regressions," *Journal of Econometrics*, 26, 323–353.

——— (1990), "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

Bierens, H. J., and Ploberger, W. (1994), "Asymptotic Theory of Integrated Conditional Moment Tests," working paper, Southern Methodist University, Dept. of Economics.

De Jong, R. M., and Bierens, H. J. (1994), "On the Limiting Behavior of a Chi-Square Type Test if the Number of Conditional Moments Tested Approaches Infinity," *Econometric Theory*, 9, 70–90.

Hong, Y., and White, H. (1991), "Consistent Specification Testing via Nonparametric Series Regression," mimeo, University of California, San Diego, Dept. of Economics.

Newey, W. K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070.

Stinchcombe, M. B., and White, H. (1991), "Consistent Specification Tests Using Duality," mimeo, University of California, San Diego, Dept. of Economics.

Tauchen, G. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443.

White, H. (1989), "An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks," in *Proceedings of the International Joint Conference on Neural Networks*, Washington, DC, San Diego: SOS Printing, pp. 451–455.

White, H., and Hong, Y. (1993), "M-Testing Using Finite and Infinite Dimensional Parameter Estimators," mimeo, University of California, San Diego, Dept. of Economics.

Social Network Analysis: Methods and Applications.

Stanley WASSERMAN and Katherine FAUST. New York: Cambridge University Press, 1994. xxxi + 825 pp. $64.95 (hard cover); $29.95 (paperback).