Journal of Official Statistics, Vol. 22, No. 2, 2006, pp. 351–355

Book and Software Review

Books for review are to be sent to the Book Review Editor Jaki S. McCarthy, USDA/NASS, Research and Development Division, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A.

Carl-Erik Särndal and Sixten Lundström. *Estimation in Surveys with Nonresponse*. Hoboken, NJ: Wiley, ISBN 0470011335, 194 pp, $99.95.

Overview of Focus for this Book

A John Wiley web page states that "Estimation techniques that account systematically for nonresponse and at the same time succeed in delivering acceptable accuracy are much needed. *Estimation in Surveys with Nonresponse* provides an overview of these techniques. . . " From my reading of this book, it does at least mention a variety of techniques, but focuses on the calibration approach to the reduction of nonresponse bias that may be found in a previous issue of this journal (Lundström and Särndal 1999).

The primary approach espoused in this book is as follows: (1) start by imputing for item nonresponse, perhaps by regression or a "donor-based" method, to fill out the response set, r, a subset of the sample, s; and then (2) adjust weights to cover both the unit nonresponse and that part of the population not sampled, calibrated in accordance with auxiliary data. The auxiliary data are very important for the methodology used by the authors, as is a stratification approach. Statisticians are familiar with the use of auxiliary data and stratification for improving estimation when there is full response to a sample, but the authors deal with how to make these techniques work to reduce nonresponse bias as well. They seem to generally assume a sample is taken, but the techniques should be applicable to adjust for nonresponse in a census as well. The authors consider that they are addressing a common situation confronted by national statistical agencies: large sample surveys with considerable nonresponse. They implicitly assume that reliable data may be obtained for both relevant auxiliary data and study variable data in all categories under a design-based sample. (This I have found problematic with regard to "small" establishments in the electric power industry, and I would think this would be true elsewhere as well. Low data quality can be a problem. Household surveys and establishment surveys can have different problems.)

For auxiliary data and stratification to help in reducing nonresponse bias, such data and groupings must be related to response rates. While I was reading this book, the ability to do this seemed like an assumption that might bear out well in some cases but not in many. After all, if we form what we think are "response homogeneity groups"

("RHGs"), we can see the average observed response rate in each group, but how do we know they are reasonably homogeneous? If this effect is likely to be commonly helpful, then perhaps this could be shown in more exhaustive examples and simulation studies than were to be found in the book, although there were some presentations given that did support the authors' contentions. However, for me the clearest example is shown in a table near the end of the book. In this very simple example, they show how the response rate rose with the increase in an auxiliary variable that was used for stratification for "Statistics Sweden's 2003 School Survey." (It appears that the more likely it is that a given student will attend a university, the more likely it is that he or she will respond to that particular survey, and the higher the score will be that is used for stratification.) The proof is in the improvement in results, however, and that was clear.

Throughout most of the book, it is assumed that not only auxiliary data can be found which "explains" (is well correlated with) the "main study variables," but also that one can find information that will "explain the inverse response probability, called the response influence." These are two of three principles given as guidelines for good auxiliary data. The third principle is that "The auxiliary vector should identify the most important domains." In attempting to implement these three principles, I would think it would often be most difficult to find auxiliary data that would be believed to sufficiently correspond to (inverse) response rates. However, apparently the authors have experienced otherwise. They develop performance indicators useful for comparing auxiliary data vectors, to help make the best choices. Further, they stress the use of "administrative registers" as a source of auxiliary data.

A theme that the authors press several times is that software exists which will implement the calibration estimation methodology that they promote, and that even when there is no simple closed form solution, one need only enter the auxiliary information. Early in the book, the authors list several names for software and provide references, but also state "Software development is ongoing. Existing software is updated. New products appear." Subsequent references to software are less thorough and do not reference the early part of the book which covers this. Thus, a researcher using this book as a reference, relying on the book's index, may not see the software list.

Because the authors concentrated on calibration for treating unit nonresponse, imputation was given less emphasis. My experience with official statistics has been quite different from theirs. I have found that with highly skewed electric power establishment surveys, regression imputation has been very useful. The authors state that "Mass imputation is not discussed in this book." That, however, is a topic of great interest to me (although it may require a different approach than that of imputation for nonignorable nonresponse, or it could possibly result from a nonignorable process itself). When the smallest establishments have enough difficulty responding to an annual census, they may not be able to respond well on a monthly sample, and may as well be nonrespondents. Data quality is an important issue. We have found evidence of disproportionately large nonsampling error for the smallest respondents (Knaub 2002, p. 17). Some small establishments, for example, may not reasonably be asked to respond monthly when they only read meters every three months. Thus mass imputation is sometimes a reasonable alternative to design-based sampling, in my opinion.

The authors also deemphasize ratio estimation. However, it appeared that simple regression estimation was made to perform better by a variable transformation. The authors also say that regression estimation "allows the freedom of a nonzero intercept." Contrast that with the opinion in Brewer (2002), page 110: "It is more often the case than not, in survey sampling, that the most appropriate supplementary variable is close to being proportional to its corresponding survey variable, and that their natural relationship or line of best fit is a straight line through the origin. If the range of the supplementary variable is limited . . . an intercept term permits the estimated relationship to stray well from the origin, with a consequent loss of efficiency." In addition, I have found ratio estimation, particularly the classical ratio estimator (Knaub 2005), to be extremely useful and robust for electric power establishment survey data.

Book Layout

This book contains a table of contents, a preface, 14 chapters, references and an index. A good index is essential if a book is to be used as a reference, and this index seems adequate. Within the book itself, I noticed that essential points were repeated, which may be good, for instance, if one is only reading certain parts while researching a narrow topic. I also noted that the authors would sometimes refer to other parts of the book, either ahead of or previous to that point, in a manner that might make one think one was reading an encyclopedia. (That is neither criticism nor endorsement, just an impression.) There are 70 references listed which further the use of this book as a reference itself. However, as indicated above, I felt that the focus of the book was rather narrow for the most part; perhaps more suitably considered a monograph.

Relationship to Other Texts

In the past, a book such as Cochran (1977) could be counted upon to cover design-based sampling error very well. However, the authors note that other sources of survey inaccuracy, such as nonresponse, need now to be given more serious consideration. (It seems to me to be particularly telling that popular estimators of the variance of totals have presupposed complete and fully accurate sample data collection, when in reality measurement errors have an effect which diminishes with increasing sample size, often through a "finite population correction factor.") This book may go a long way toward addressing an issue that previous books have not: nonresponse in the case of large-scale social surveys with available auxiliary data. Some users may find that the approach, outlined in the second paragraph of this review, is very useful. It is not, however, as comprehensive as the book's title might suggest.

This book could perhaps complement Särndal, Swensson and Wretman (1992), as it does assume a focus on model-assisted design-based sampling and inference. Särndal, Swensson, and Wretman (1992) is listed as a reference. They also mention Lohr (1999), which I think can be useful, but they do not include Brewer (2002), which provides an overall perspective. Also omitted is any reference to Valliant, Dorfman, and Royall (2000) on prediction theory. Lohr (1999) is quoted for definitions of ignorable and nonignorable nonresponse, and mention is made of some parts of the three-volume set, Madow, Olkin, and Rubin (eds) (1983), *Incomplete Data in Sample Surveys*, but traditional work is

generally given short shrift. The authors say that "Important as the terms MAR, MCAR, ignorable and nonignorable may be, the reasoning in this book is not significantly facilitated or clarified by a regular use of these terms." Instead, we are told to look for "new important concepts for classification." This is because they address response mechanism concerns through the use of auxiliary information and stratification. There is a very strong emphasis on bias reduction. The authors say that "The calibration weights . . . have embedded in them the notion of inverse response probability weighting." An appendix to a chapter seems to make their case a bit more convincing. Another method is described by way of developing reasoning, but calibration is the method chosen in this book.

Chapter Organization

Chapters 1–3 are introductory, "The importance of auxiliary information." Chapter 4 includes the Horvitz-Thompson estimator and then the generalized regression estimator (GREG), essential to the calibration estimation the authors propose to improve accuracy, even in the presence of nonresponse. Chapters 5–9 progressively deal with nonresponse, including simulation studies, and bring the reader to the important tenth chapter, "Selecting the Most Relevant Auxiliary Information." Bias is considered more important than variance in this book, but Chapter 11 does deal with variance. It is broken down into (1) variance due to sampling and (2) variance due to nonresponse. A good reference for that is included in the book: Lee, Rancourt, and Särndal (2001). It is also noted that even if variance estimates are good, confidence intervals are degraded by bias that makes them "off-centre." Chapter 12 is the only chapter really dedicated to imputation. As stated earlier, in this book the focus on imputation is only for use in completing data for units with some item nonresponse. There is only light mention of major methods of imputation. The book finishes with Chapter 13, "Variance Estimation in the Presence of Imputation," and Chapter 14, "Estimation Under Nonresponse and Frame Imperfections," which uses a Venn diagram to illustrate the effect of frame imperfections.

In Conclusion

This book has value, but the focus is narrower than I had expected. If you deal with large sample surveys with data of "good" quality, and have access to considerable "good" auxiliary data, and to software that "handles the computation of weights for survey data affected by nonresponse," then this book may be of interest to you. It may be of academic use as well. The reader may further judge what interest the book may have by visiting the *JOS* website, www.jos.nu (if you are not already there) and perusing Lundström and Särndal (1999).

References

Brewer, K.R.W. (2002). Combining Survey Sampling Inferences: The Weighing of Basu's Elephants, Arnold: London and Oxford University Press.

Cochran, W.G. (1977). Sampling Techniques, 3rd ed. John Wiley and Sons.

Knaub, J.R., Jr. (2002). Practical Methods for Electric Power Survey Data, InterStat, July, http://interstat.statjournals.net/. (Note another version in ASA Survey Research Methods Section Proceedings, 2002.).

Knaub, J.R., Jr. (2005). The Classical Ratio Estimator, InterStat, October. http://interstat.statjournals.net/.

Lee, H., Rancourt, E., and Särndal, C.-E. (2001). Variance Estimation from Survey Data Under Single Value Imputation, presented at the International Conference on Survey Nonresponse, Oct. 1999, published in Survey Nonresponse, edited by Groves, Dillman, Eltinge and Little, 2001, John Wiley and Sons, Inc., 315–328.

Lohr, S.L. (1999). Sampling: Design and Analysis. Duxbury Press.

Lundström, S. and Särndal, C.-E. (1999). Calibration as a Standard Method for Treatment of Nonresponse, Journal of Official Statistics, 15, 305–327.

Madow, W.G., Olkin, I., and Rubin, D.B. (eds) (1983). Incomplete Data in Sample Surveys, Three vVolumes, Academic Press.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York, NY: Springer-Verlag.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). Finite Population Sampling and Inference, A Predictive Approach. New York, NY: John Wiley and Sons.

James R. Knaub Jr
U.S. Department of Energy
Energy Information Administration
Forrestal Building, EI-53
1000 Independence Ave
Washington, DC 20585
U.S.A.
Phone: 202-287-1733
Email: james.knaub@eia.doe.gov