authors briefly review various methods and refer readers to works such as Little (1995) for details. The analyses presented are based on certain assumptions, such that the available GEE software can be applied. Chapter 4 gives a thorough discussion on model selection and testing and graphical methods for residual diagnostics.

Overall, *Generalized Estimating Equations* is a good introductory book for analyzing continuous and discrete correlated data using GEE methods. The authors discuss the differences among the four commercial software programs and provide suggestions and cautions for users. This book is easy to read, and it assumes that the reader has some background in GLM. Many examples are drawn from biomedical studies and survey studies, and so it provides good guidance for analyzing correlated data in these and other areas.

Edward C. CHAO
Insightful Corporation

### REFERENCES

Hardin, J. W., and Hilbe, J. M. (2001), *Generalized Linear Models and Extensions*, College Station, TX: Stata Press.

Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Little, R. J. A. (1995), "Modeling the Drop-Out Mechanism in Repeated-Measures Studies," *Journal of the American Statistical Association*, 90, 1112–1121.

Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988), "Models for Longitudinal Data: A Generalized Estimating Equation Approach," *Biometrics*, 44, 1049–1060.

**Regression Models for Time Series Analysis**, by Benjamin KEDEM and Konstantinos FOKIANOS, Hoboken, NJ: Wiley, 2002, ISBN 0-471-36355-3, xiv + 337 pp., $84.95.

The stated aim of this book is to introduce readers to new developments in the use of regression methods for time series analysis, with a target audience of mixed-level graduate students with a masters-level background in statistical inference and applied stochastic processes. The book can be divided into two distinct parts. The first four chapters deal with the use of generalized linear models (GLMs) in the context of modeling discrete-valued time series. Basic ideas underlying the use of GLMs for time series are introduced through a discussion of the partial likelihood concept, based on work of Cox (1975). Chapter 1 is devoted to a treatment of inference based on the partial likelihood asymptotic theory for the obtained estimators, as well as hypothesis testing and model diagnostics. Quasi-partial likelihood is also briefly discussed. The next three chapters go into much greater detail of these concepts in the framework of binary time series, categorical time series, and count time series. Thus the book's first half forms a coherent body of knowledge concerning estimation of discrete-valued time series using partial likelihood methods and would serve as a good reference for time series researchers interested in understanding the key theoretical concepts underlying this methodology. I found that these chapters serve as a nice introduction to an area of time series modeling relatively unfamiliar to me.

The book's second half covers a variety of alternative modeling approaches for time series and spatial data, not necessarily discrete-valued, and as such is much less focused than the first half. Chapter 5 includes a discussion of integer autoregressive (AR) and moving average (MA) models, discrete ARMA models, mixture transition distribution models, hidden Markov models, variable mixture models, autoregressive conditionally heteroscedastic (ARCH) models, sinusoidal regression models, and mixed models for longitudinal data. The inclusion of some of these topics—in particular, ARCH models—seems arbitrary, except for the fact that they represent fairly recent developments in time series modeling.

Chapter 6 discusses state-space models, with particular attention given to nonlinear and non-Gaussian state-space models, in which framework *dynamic* GLMs can be cast. The chapter includes an introduction to state-space models in the linear Gaussian framework, as well as a brief discussion of simulation-based methods for state-space models.

Chapter 7 deals with the time series prediction problem for "messy" data, for example, data that may be non-Gaussian, irregularly observed, have few observations, or contain data gaps, by casting the problem in the much broader context of Bayesian spatial prediction methods for stationary random fields. Two examples to time series data are given, along with some discussion of the Bayesian Transformed Gaussian (BTG) algorithm for predictive density approximation.

An appendix provides key concepts related to stationary processes, and the reference list is extensive. A number of data examples are given. However, little information is provided concerning how the example data may be obtained for, say, teaching purposes. Moreover, a student may find the lack of motivating examples before the leap into the discussion of partial likelihood in Chapter 1.1 to be a little intimidating. Each chapter includes exercises, almost entirely of a theoretical nature. Throughout the book, little mention is made of computational issues, except for a brief discussion in Chapter 1 noting that GLM estimation of time series can be routinely carried out using S-PLUS or SAS, and a brief tutorial on Markov chain Monte Carlo methods in Chapter 6.

Although I gladly recommend this book for researchers working in the area of nonlinear time series modeling who want to have information on a variety of modeling methods easily available in one convenient reference, I would have some reservations about using it as a textbook except for advanced graduate students aiming to do research in the field.

Bonnie K. RAY
IBM Watson Research Center

### REFERENCE

Cox, D. R. (1975), "Partial Likelihood," *Biometrika*, 62, 69–76.

**Statistical Analysis With Missing Data** (2nd ed.), by R. J. A. LITTLE and D. B. RUBIN, Hoboken, NJ: Wiley, 2002, ISBN 0-471-18386-5, xv + 381 pp., $94.95.

Before I started graduate school, I audited a masters-level course on missing data and the EM algorithm. I remember being amazed at the fact that there were methods for "filling in" the missing observations, and at the elegance of the algorithm itself. Even at that early stage in my career, and with my limited exposure to data analysis, I recognized the importance of these tools for the applied statistician, for it is an unfortunate fact of statistical practice that datasets often come to us with some values missing. Data can be incomplete for many reasons, but whatever the reason, the data analyst should not ignore the problem. For an analysis to be valid, any missing data must be dealt with appropriately. How to do this is the topic of this book. The book begins with an overview of the mechanisms that can lead to missing data, some common patterns of missingness, and traditional methods for analyzing datasets with missing data. These last are categorized in a rough taxonomy: procedures that use only the observations that are completely recorded, weighting procedures, imputation, and modeling. Most of the book focuses on model-based techniques, approached from the likelihood perspective.

This second edition expands considerably on the first, which was published more than 15 years ago. Essentially all of the material in the first edition is still present, albeit in slightly reorganized fashion. Little and Rubin have also added substantial amounts of material on the Bayesian approach to missing data, as well as on multiple imputation, data augmentation, and extensions of the basic EM algorithm (e.g., ECM, ECME, PX-EM), which are all amply demonstrated. Most of these new techniques were not available at the time of the first edition. Assessing uncertainty in the estimates of the missing data (an area that was also quite undeveloped when the first edition was published) similarly receives expanded attention in the second edition. The additions therefore reflect the rapid advances in both theory and computing technology since the late 1980s in the field of statistics as a whole, and in the analysis of missing data in particular. Several prototypical examples are included and analyzed in depth, with details provided on the implementation of the EM algorithm or its variations, illustrated using various datasets taken from the literature. The examples cover such topics as multivariate normal data, partially classified contingency tables, and mixed normal and nonnormal data, all ignoring the missing-data mechanism. There is also a chapter on models for nonignorable missing data. A brief discussion is devoted to settings such as factor analysis, where the data are fully observed but that can be thought of as missing-data problems by the introduction of unobserved latent variables.