Jan de Leeuw
Erik Meijer
*Editors*

# Handbook of Multilevel Analysis

Springer

Jan de Leeuw and Erik Meijer (Eds.)

# Handbook of Multilevel Analysis

May 28, 2007

**Springer**
*Berlin Heidelberg New York*
*Hong Kong London*
*Milan Paris Tokyo*

# Preface

Social and medical researchers have long been concerned about the need properly to model complex data structures, especially those where there is a hierarchical structure such as pupils nested within schools or measurements nested within individuals. Statisticians, especially those involved in survey sampling, recognise that failure to take account of such structures in standard models can lead to incorrect inferences. What has been less well appreciated is that a failure properly to model complex data structures makes it impossible to capture that complexity that exists in the real world. It is only in the last 20 years or so, when appropriate and efficient model based methods have become available to deal with this issue, that we have come to appreciate the power that more complex models provide for describing the world and providing new insights. This book sets out to present some of the most recent developments in what has come to be known as multilevel modelling.

An introductory chapter by de Leeuw and Meijer gives a brief history and a standard exposition of the basic multilevel model involving random coefficients at level 2 and above, together with a discussion of some likelihood based estimation procedures. This is followed by a chapter by Draper that outlines a Bayesian approach to modelling multilevel structures using the MCMC algorithm, with a clear exposition of the rationale for such an approach and well worked through examples. This is as good an introduction as any to Bayesian analysis and MCMC estimation. The next chapter by Snijders and Berkhof deals with the important issue of diagnostics for multilevel models. It takes the reader carefully through the various model assumptions and how they can be examined, for example, making use of model elaborations and residual analysis. There is also a useful section on smoothing models. Moerbeek, van Breukelen and Berger look at ways of optimally sampling units in multilevel models. It includes clear examples for Normal and generalised linear models with useful discussions of repeated measures and schooling designs. Raudenbush contributes a chapter where he looks at the inferential problems that can arise when, in a 2-level model, the number of level 1 units per level 2 unit is small. He gives some examples, such as matched pairs and cluster randomised trials and explains how these can be interpreted and there is a brief discussion of issues in generalised linear models. The Chapter by Hedeker deals in detail with discrete responses, either ordered or nominal. It has a clear exposition with useful examples. Skrondal and Rabe-Hesketh discuss models for longitudinal repeated measures data, including those with serial dependency structures, for Normal and discrete responses. Well motivated examples are used for the exposition. Rasbash and Browne show how cross-classified and multiple membership structures can be modelled. They provide examples and a convincing exposition of why researchers should be looking beyond mere hierarchies when analysing real life data. Rodriguez looks at generalised linear models with particular reference to survival data and gives a detailed discussion of various estimation algorithms, together with a useful example. Longford provides a chapter on missing data, where he describes the use of the EM algorithm and random multiple imputation. Van der Leeden, Meijer and Busing, in a comprehensive account, take a careful look at bootstrap and jackknife procedures for studying bias and for obtaining valid standard errors and confidence intervals in multilevel models. Finally, the Du Toit's present an account of multilevel structural equation models with some useful examples and detailed derivations.

The book covers a great number of important topics and there is a useful amount of cross referencing with a good number of worked examples. The amount of methodological activity now underway is very impressive and as these become incorporated into software packages will hopefully persuade researchers to undertake data analysis that more closely reflects the structure of real world data than traditional methods assume. Most of the developments discussed leave room for further work. As hardware becomes more powerful so certain options will become more attractive. This is especially the case with resampling methods such as the bootstrap, multiple imputation and MCMC and these do seem to be where we may expect the most interesting future developments. In particular, given what is happening more generally, we should expect MCMC methods to become more and more prevalent. Not only do they allow proper Bayesian inference, especially for small samples, as emphasised by Draper, they also have great potential because of the modularity of the algorithm steps. This is clearly demonstrated in the Chapter by Rasbash and Browne, where, as they point out, certain kinds of data simply cannot be treated properly using maximum likelihood.

So, apart from the increasing adoption of MCMC methods what might be useful future directions for research? Several of these areas are described in this volume. I would single out cross classified and multiple membership models that move us on from the consideration of simple hierarchies. It is very rare in the real world to find structures that are purely hierarchical. In education students will typically belong to a school hierarchy at the same time as a neighbourhood hierarchy, both of which may influence the outcome variable of interest. In addition students may move among neighbourhoods or schools so that assignment to a single higher level unit may be misleading and lead to important biases. As Rasbash and Browne describe, many other areas of the biological and social sciences have these structures and this provides an exciting and fruitful challenge for multilevel techniques.

I would also emphasise missing data procedures with non random missingness, since problems such as non-response in surveys are becoming acute in many places. As Longford suggests, the existence of additional or 'auxiliary' information in surveys can be especially useful in allowing the application of existing missing data procedures to handle informative non-response.

We also need good diagnostic procedures to test the assumptions of our models and more work here would be very useful, for example in testing the validity of the standard assumption of multivariate Normality. It is particularly important that these procedures are brought within existing modelling packages so that their use is encouraged.

Likewise, another large area of interest is in latent variable models of all kinds, including complex ones such as latent growth trajectory models. The application of multilevel latent structure models with binary and ordered responses is an important area for psychometrics where much current activity under the heading of item response modelling often ignores the inherent hierarchical structures.

Despite the wide coverage of the topics that are dealt with, there are also areas that are not so well covered in this book, which is inevitable in a rapidly changing field.

Thus, measurement and misclassification errors, while mentioned briefly, are not treated in depth, yet we know that ignoring them can have profound effects on inferences. In educational and medical research, for example, they abound and are often correlated, and we need research on both how to estimate measurement error variances and covariances and misclassification probabilities and then how to incorporate these estimates into our models.

Multivariate models are not as well covered as I would wish, since they are becoming more extensively used. An interesting problem is where there are multiple responses at more than one level together. Such models have important applications to prediction problems, multi-process modelling and multiple imputation. An example of the first case is where we have both repeated measures data on individuals and subsequent individual level measures we wish to predict, as in growth studies. Likewise, in multi-process models we may wish to jointly model, say, pupil responses together with teacher or school level variates, for prediction or adjustment purposes, as well as moving us towards better causal understandings. For imputation procedures we often need jointly to model responses at several levels if these variables have missing values. Additionally, in all these cases our responses may be mixtures of continuous and discrete variables, and this presents an additional challenge.

Whilst all these methodological developments are exciting and important, the methodological community still has the task of communicating them to potential users. As with all new techniques this requires a combination of clear exposition together with suitable software tools. Many of the authors of chapters in this book have themselves provided such combinations, but more is needed as the methodology advances. Nevertheless we do need to be careful that we are not promoting multilevel modelling as a kind of magic wand that can transmute bad data into good or turn a poor design into a highly efficient one. Sensitivity to assumptions and accessible ways of investigating those assumptions are things we need continually to emphasise.

Finally, the editors are to be congratulated on bringing together a distinguished group of authors all of whom have interesting things to say. This volume gives us an insight into much current research and will hopefully attract others into this important area of activity.

November 2006

*Harvey Goldstein*
*Professor of Social Statistics*
*University of Bristol*

# Contents

## List of Contributors

**Martijn P. F. Berger**
Maastricht University
Department of Methodology
and Statistics
PO Box 616
6200 MD Maastricht
THE NETHERLANDS
martijn.berger@stat.unimaas.nl

**Johannes Berkhof**
VU University Medical Center
Department of Clinical Epidemiology
& Biostatistics
P.O. Box 7057
1007 MB Amsterdam
THE NETHERLANDS
h.berkhof@vumc.nl

**William J. Browne**
University of Bristol
Department of Clinical Veterinary
Science
Langford House
Langford
North Somerset BS40 5DU
UNITED KINGDOM
William.Browne@bristol.ac.uk

**Frank M. T. A. Busing**
Leiden University
Department of Psychology
PO Box 9555
2300 RB Leiden
THE NETHERLANDS
busing@fsw.leidenuniv.nl

**Jan de Leeuw**
Department of Statistics
University of California
at Los Angeles
Los Angeles, CA 90095-1554
USA
deleeuw@stat.ucla.edu

**David Draper**
Department of Applied Mathematics
and Statistics
Baskin School of Engineering
University of California
1156 High Street
Santa Cruz, CA 95064
USA
draper@ams.ucsc.edu

**Mathilda du Toit**
Scientific Software International
7383 N. Lincoln Ave., Suite 100
Lincolnwood, IL 60712-1747
USA
mdutoit@aol.com

**Stephen H. C. du Toit**
Scientific Software International
7383 N. Lincoln Ave., Suite 100
Lincolnwood, IL 60712-1747
USA
sdutoit@ssicentral.com

**Donald Hedeker**
University of Illinois at Chicago
School of Public Health
Division of Epidemiology
and Biostatistics
1603 West Taylor Street, Room 955
Chicago, IL 60612-4336
USA
hedeker@uic.edu

**Nicholas T. Longford**
SNTL
99 Swallows Croft
Reading RG1 6EH
United Kingdom
ntl@sntl.co.uk

**Erik Meijer**
RAND Corporation
1776 Main St
PO Box 2138
Santa Monica, CA 90407-2138
USA
meijer@rand.org

**Mirjam Moerbeek**
Utrecht University
Department of Methodology
and Statistics
PO Box 80140
3508 TC Utrecht
The Netherlands
m.moerbeek@fss.uu.nl

**Sophia Rabe-Hesketh**
Graduate School of Education
& Graduate Group in Biostatistics
University of California
3659 Tolman Hall
Berkeley, CA 94720
USA
sophiarh@berkeley.edu

**Jon Rasbash**
University of Bristol
Graduate School of Education
Helen Wodehouse Building
35 Berkeley Square
Clifton
Bristol BS8 1JA
United Kingdom
j.rasbash@bristol.ac.uk

**Stephen W. Raudenbush**
University of Chicago
Department of Sociology
1126 East 59th Street
Room SSR 416
Chicago, IL 60637
USA
sraudenb@uchicago.edu

**Germán Rodríguez**
Office of Population Research
Wallace Hall
Princeton University
Princeton, NJ 08544-2091
USA
grodri@princeton.edu

**Anders Skrondal**
Department of Statistics
and Methodology Institute
Columbia House
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom
a.skrondal@lse.ac.uk

**Tom A. B. Snijders**
Department of Statistics, University of Oxford
Nuffield College
New Road
Oxford OX1 1NF
United Kingdom
tom.snijders@nuffield.ox.ac.uk

**Gerard J. P. Van Breukelen**
Maastricht University
Department of Methodology and Statistics
PO Box 616
6200 MD Maastricht
The Netherlands
gerard.vbreukelen@stat.unimaas.nl

**Rien van der Leeden**
Leiden University
Department of Psychology
PO Box 9555
2300 RB Leiden
The Netherlands
vanderleeden@fsw.leidenuniv.nl

# Handbook of Multilevel Analysis

Jan de Leeuw and Erik Meijer *Editors*

Multilevel analysis is the statistical analysis of hierarchically and non-hierarchically nested data. The simplest example is clustered data, such as a sample of students clustered within schools. Multilevel data are especially prevalent in the social and behavioral sciences and in the biomedical sciences. The models used for this type of data are linear and nonlinear regression models that account for observed and unobserved heterogeneity at the various levels in the data.

This book presents the state of the art in multilevel analysis, with an emphasis on more advanced topics. These topics are discussed conceptually, analyzed mathematically, and illustrated by empirical examples. The authors of the chapters are the leading experts in the field.

Given the omnipresence of multilevel data in the social, behavioral, and biomedical sciences, this book is useful for empirical researchers in these fields. Prior knowledge of multilevel analysis is not required, but a basic knowledge of regression analysis, (asymptotic) statistics, and matrix algebra is assumed.

**Jan de Leeuw** is Distinguished Professor of Statistics and Chair of the Department of Statistics, University of California at Los Angeles. He is former president of the Psychometric Society, former editor of the *Journal of Educational and Behavioral Statistics*, founding editor of the *Journal of Statistical Software*, and editor of the *Journal of Multivariate Analysis*. He is coauthor (with Ita Kreft) of *Introducing Multilevel Modeling* and a member of the Albert Gifi team who wrote *Nonlinear Multivariate Analysis*.

**Erik Meijer** is Economist at the RAND Corporation and Assistant Professor of Econometrics at the University of Groningen. He is coauthor (with Tom Wansbeek) of the highly acclaimed book *Measurement Error and Latent Variables in Econometrics*.

STATISTICS

ISBN 978-0-387-73183-4

EAN

9 780387 731834 >