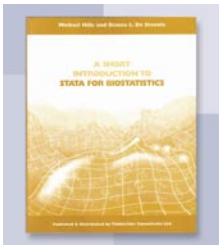
THE STATA NEWS

Volume 17 Number 3 July/August/September 2002 www.stata.com

From the Stata Bookstore

A Short Introduction to Stata for Biostatistics



Title: *A Short Introduction to Stata for Biostatistics*
Authors: M. Hills, B. De Stavola
Publisher: Timberlake Consultants Press
Copyright: 2002
ISBN: 0-9542603-0-9
Pages: 131, plus CD; paperback
Price: $32.75

A Short Introduction to Stata for Biostatistics bridges the information in the *Getting Started* manual and the *Reference Manuals* by providing a more detailed introduction to the most frequently used analytic methods in biomedical research. Although it is specifically written for biostatisticians, epidemiologists, and health professionals new to Stata, the book is useful to more experienced users wishing to have more in-depth knowledge of Stata commands and biostatistical issues. The book is hands-on, intended to be used while working with Stata, and it includes a CD-ROM containing the datasets and several author-written programs.

The first three chapters provide an overview of data entry and management commands, including those used to create, label, and drop variables, and sort observations. After a short introductory chapter on graphics, the bulk of the book details methods used in data description and analysis. Beginning with commands used to create frequency tables and summary statistics, the book proceeds to describe commands used for univariate and multivariate analyses, including linear regression, Poisson regression, logistic regression, and survival data analysis. The last few chapters include a chapter describing more advanced data management methods such as the merging and the appending of datasets, and a very helpful chapter detailing how to write your own Stata programs and commands.

A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/sisb.html**. You may also order using the enclosed bookstore order form.

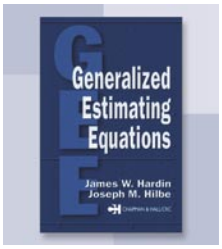
Logistic Regression: A Self-Learning Text (2nd edition)

Title: *Logistic Regression: A Self-Learning Text* (2nd edition)
Authors: D. Kleinbaum, M. Klein
Publisher: Springer-Verlag
Copyright: 2002
ISBN: 0-387-95397-3
Pages: 513; hardcover
Price: $79.75

This book is the second edition of the 1994 text, and succeeds in demonstrating that one need not be a mathematician in order fully to understand the underpinnings of logistic regression in all its forms. Ideally suited for a graduate course for students in the medical sciences, the text has the look and feel of a course textbook, with formulae, diagrams, and important points set off to the side margins for emphasis. Also, each chapter contains an end summary, detailed outline, practice exercises (with answers), and chapter test.

The mathematics are kept to the most basic level, but nevertheless, because of its completeness in coverage of logistic regression, this text would be a good reference for even the most theoretical statistician. In fact, those already well-versed in logistic regression methods will benefit from seeing advanced topics such as generalized estimating equations (GEE) explained from first principles.
A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/lr.html**. You may also order using the enclosed bookstore order form.

Generalized Estimating Equations



Title: *Generalized Estimating Equations*
Authors: J. Hardin, J. Hilbe
Publisher: Chapman & Hall/CRC
Copyright: 2003
ISBN: 1-58488-307-3
Pages: 222; hardcover
Price: $69.75

The method of Generalized Linear Models (GLM) is an integral part of the data analyst's toolkit, as it encompasses many models under one roof: logistic and probit regressions, ordinary least-squares, ordinal outcome regression, and regression models for the analysis of survival data, to name a few. Nominal GLM, however, is inadequate when the data are longitudinal or otherwise grouped such that it is expected that observations within the same group would be correlated. The method of generalized estimating equations (GEE) is a generalization of GLM that takes into account this "within-group" correlation.

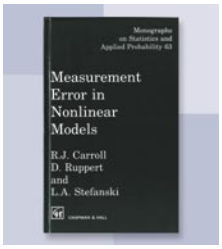
Inside this issue:

This text is the sequel to the 2001 text, *Generalized Linear Models and Extensions*, by the same authors, and provides the first complete treatment of GEE methodology. As with the previous text on GLM, this text is ripe with examples on how the methodology may be carried out using Stata. In fact, the principal author, James Hardin, developed much of the Stata software for fitting GEE models while he was a Senior Statistician at StataCorp.

This text is heavy in mathematical and computational detail, but the mathematics is balanced by an array of "real-world" datasets and analyses. Thus, the text should appeal to a very wide audience, from the mathematical statistician wishing to glean the current state of the GEE literature to the professional researcher needing to fit a GEE model in order to solve a particular problem.

A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/gee.html**. You may also order using the enclosed bookstore order form.

Measurement Error in Nonlinear Models



Title: *Measurement Error in Nonlinear Models*
Authors: R. Carroll, D. Ruppert, L. Stefanski
Publisher: Chapman & Hall/CRC
Copyright: 1995
ISBN: 0-412-04721-7
Pages: 305; hardcover
Price: $69.75

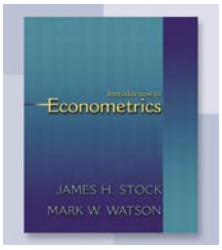
The bulk of statistical methodology concerns the development of a model for a random outcome variable conditional on a set of measured covariates. Sometimes, however, one or more of the covariates is measured with error, and the consequences of ignoring measurement error are well-established. In particular, ignoring the errors-in-variable leads to attenuation—which in linear models generally means an underestimation of the magnitudes of the effects of the covariates on the outcome.

The topic of measurement error has been the subject of a vast body of research in the area of linear models. When it comes to nonlinear models, however, until recently, the literature has remained sparse due to the mathematical and computational complexity involved. This text, published in 1995, is rare in that it was published right at the point in time when such a book could be written, that is, when the body of research in this field reached the point where it could be mined effectively to produce one coherent summary.

The methods covered are (by necessity) computationally intensive, but the authors have also led a project by which software, written in the Stata language, for the express purpose of fitting nonlinear models in the presence of measurement error, will become available.

A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/menm.html**. You may also order using the enclosed bookstore order form.

Introduction to Econometrics



Title: *Introduction to Econometrics*
Authors: J. Stock, M. Watson
Publisher: Addison Wesley
Copyright: 2003
ISBN: 0-201-71595-3
Pages: 696; hardcover
Price: $94.75

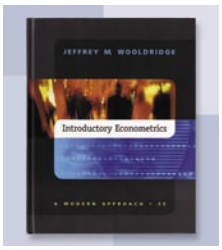
Introduction to Econometrics by James H. Stock and Mark W. Watson is a real page-turner. By ingeniously introducing statistical methods as means of answering four interesting empirical questions, the authors have written a rigorous text that one wants to keep reading to find out how the story ends. The authors use the excitement generated by the questions as a springboard for launching readers into an excellent introduction to estimation, inference, and interpretation in econometrics.

The extent to which it makes advanced statistical concepts easily understandable also makes this text noteworthy. For instance, the current econometric approach to analyzing linear models combines assumptions on the conditional moments of random variables and large sample theory to derive estimators and their properties. This textbook provides a very accessible introduction to this technique and its application to cross-sectional regression, panel-data regression, and time-series regression.

The coverage and level of this text make it an excellent choice for undergraduate study or as a supplement to advanced courses.

A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/ite.html**. You may also order using the enclosed bookstore order form.

Introductory Econometrics: A Modern Approach (2nd edition)



Title: *Introductory Econometrics: A Modern Approach* (2nd edition)
Author: J. Wooldridge
Publisher: South-Western College Publishing
Copyright: 2002
ISBN: 0-324-11364-1
Pages: 863; hardcover
Price: $104.75

The second edition of Jeffrey Wooldridge's text book, Introductory Econometrics: A Modern Approach, lives up to its subtitle. This book exemplifies a modern approach to econometrics in its choice of topics and its treatment of standard material.

Modern econometrics is much more than ordinary least squares (OLS) with a few extensions to handle the special cases commonly encountered in econometric data, and Wooldridge recognizes this point. In addition to chapters on OLS, he includes chapters on current techniques of estimation and inference for time-series data, panel data, limited dependent variables, and sample selection.

(Continued on page 5)

Second North American Stata Users' Group meeting

Call for papers

Monday, March 17, 2003

The second North American Stata Users' Group meeting will be held March 17, 2003 in Boston, Massachusetts at the Longwood Galleria Conference Center. It will be followed by a short course, Measurement error in nonlinear models, presented by Raymond Carroll and James Hardin.

This is the first announcement and call for offers of presentations for the North American Stata Users' Group meeting. Precise timings will be announced later.

The content of the meeting is being organized by a committee co-chaired by Kit Baum, Boston College (baum@bc.edu) and Marcello Pagano, Harvard School of Public Health (pagano@hsph.harvard.edu). The logistics are being organized by StataCorp.

Previous Stata Users' Group meetings have been held in Boston (2001), Britain, Spain, and the Netherlands. All meetings are open to anyone interested, without regard for level of expertise or geographic location. StataCorp will be represented. There will be the usual "wishes and grumbles" session at which you may air your thoughts to Stata developers. The organizing committee invites offers of presentations. Presentations should be Stata-related in some manner. We encourage submission on any topic to do with the use of Stata in research, teaching, or any other field. The following headings are illustrative, but not restrictive.

- Using Stata for modeling and analysis
- User-written Stata programs
- Case studies of research or teaching using Stata
- Using Stata to import and manage data
- Using Stata for graphics
- Teaching Stata
- Teaching statistics with Stata
- Surveys or critiques of Stata facilities in specific fields
- Software comparisons

Meeting format

Presentation proposals will be evaluated by the organizing committee. Those proposals accepted by the committee will require either 25 minutes of presentation followed by five minutes of discussion or ten minutes of presentation followed by five minutes of discussion. Please state the length of presentation preferred on the abstract submission.

Submission information

Please submit an abstract (200 words maximum, ASCII text, no math symbols) to Kit Baum by December 15, 2002. Conference fees will be waived for presenters. Abstracts and contact information (name, affiliation, postal address, email address, fax, and phone numbers) should be entered on the online submission form at **http://fmwww.bc.edu/nasug2003/submission.html**.

Presenters will be asked to provide materials related to their talk (slides and programs/datasets, where applicable) in electronic form, preferably prior to the meetings, to the organizing committee, so that they may be posted on StataCorp's web site and in the Stata Users' Group RePEc archive.

Registration and cost

You can register online at **http://www.stata.com/support/meeting/2nasug/register.html** or by completing the enclosed registration form. Registration is limited, so we suggest that you enroll early. The costs are

	Professional	Student
North American Stata Users' Group meeting and Measurement error in nonlinear models course	**$85**	**$55**
North American Stata Users' Group meeting only	**$45**	**$20**
Measurement error in nonlinear models course only	**$45**	**$45**

The costs above are correct; Stata Corporation is subsidizing the course. Costs include lunch.

There is no cost for presenters either day, nor for the dinner on March 17th. Presenters should email Lisa Gilmore (**lgilmore@stata.com**) to tell her whether you will be able to join us for dinner, and whether you will be attending the Measurement error in nonlinear models course.

The North American Users' Group meeting will be held on March 17, 2003, and the Measurement error in nonlinear models course will be held the next day. Both events will be held at

Longwood Galleria Conference Center
342 Longwood Avenue
Boston, Massachusetts

A block of sleeping rooms has been reserved at the Best Western Boston – the Inn at Longwood Medical for March 15-19, 2003. The hotel is offering a limited number of sleeping rooms at a discount rate of $149 plus $12.45% tax per room per night. Please call the hotel direct at 617-731-4700 or toll free at 800-468-2378 and ask for in-house reservations. Note that the last day to make reservations at the discounted rate is February 22, 2003. For more information, visit the hotel's web site at **http://www.bestwestern.com/prop_22028** or Best Western's web site at **http://www.bestwestern.com**.

Please feel free to contact the organizers with any relevant questions.

The North American Stata Users' Group Meeting Organizing Committee

Elizabeth Allred
lizard@hsph.harvard.edu
Harvard School of Public Health

Kit Baum
baum@bc.edu
Boston College

Nicholas J. Cox
n.j.cox@durham.ac.uk
University of Durham

Marcello Pagano
pagano@hsph.harvard.edu
Harvard School of Public Health

See reverse side for details of the Measurement error in nonlinear models course

Measurement error in nonlinear models course

Tuesday, March 18, 2003

Raymond Carroll and James Hardin will be presenting a one-day short course on Measurement error in nonlinear models. It will follow the second North American Stata Users' Group meeting.

The Measurement error in nonlinear models short course is based on the book

> *Measurement Error in Nonlinear Models*
> R. J. Carroll, D. Ruppert, and L.A. Stefanski
> Chapman and Hall/CRC Press 1995
> ISBN 0-412-04721-7

and is an outgrowth of recent work with Stata Corporation by the presenters along with Joseph Newton, Henrik Schmiediche, and Tamara Stoner to develop new statistical techniques along with software for estimation in the presence of measurement error.

This course concerns analysis strategies for nonlinear regression problems in which predictors are measured with error. Such problems are commonly known as measurement-error or errors-in-variables modeling, a topic for which there exists a large amount of literature in the case of linear models.

Nonlinear errors-in-variables modeling began in earnest in the early 1980s with the publication of a series of papers on diverse topics: Prentice (1982) on survival analysis; Carroll, Spiegelman, Lan, Bailey, and Abbott (1984) and Stefanski and Carroll (1985) on binary regression; Armstrong (1985) on generalized linear models; Amemiya (1985) on instrumental variables; and Stefanski (1985) on estimating equations. In 1987, David Byar and Mitchell Gail organized a workshop on the topic at the National Institutes of Health. In 1989, the results of the workshop were published as a special issue in Statistics in Medicine. Since these early papers, the field has grown dramatically.

The course will cover the following:

- Applications in which measurement error is of concern
- Definition of basic terminology of the error structure
- Illustration of consequences of measurement error in linear regression
- Effects of measurement error on hypothesis testing
- Regression calibration and simulation-extrapolation
- Instrumental variables
- Likelihood-based approaches to estimation and inference

In addition to showing how to obtain estimates, methods for obtaining bootstrap, jackknife, and robust ("sandwich") standard errors will be discussed.

A Stata module for estimating these models will be introduced and made available.

Registration and cost

You can register online at **http://www.stata.com/support/meeting/2nasug/register.html** or by completing the enclosed registration form. Registration is limited, so we suggest that you enroll early. The costs are

	Professional	Student
North American Stata Users' Group meeting and Measurement error in nonlinear models course	**$85**	**$55**
Measurement error in nonlinear models course only	**$45**	**$45**
North American Stata Users' Group meeting only	**$45**	**$20**

The costs above are correct; Stata Corporation is subsidizing the course. Costs include lunch.

There is no cost for presenters either day, nor for the dinner on March 17th. Presenters should email Lisa Gilmore (**lgilmore@stata.com**) to tell her whether you will be able to join us for dinner, and whether you will be attending the Measurement error in nonlinear models course.

The North American Users' Group meeting will be held on March 17, 2003, and the Measurement error in nonlinear models course will be held the next day. Both events will be held at

> Longwood Galleria Conference Center
> 342 Longwood Avenue
> Boston, Massachusetts

A block of sleeping rooms has been reserved at the Best Western Boston – the Inn at Longwood Medical for March 15-19, 2003. The hotel is offering a limited number of sleeping rooms at a discount rate of $149 plus $12.45% tax per room per night. Please call the hotel direct at 617-731-4700 or toll free at 800-468-2378 and ask for in-house reservations. Note that the last day to make reservations at the discounted rate is February 22, 2003. For more information, visit the hotel's web site at **http://www.bestwestern.com/prop_22028** or Best Western's web site at **http://www.bestwestern.com**.

About the presenters

Ray Carroll is a Distinguished Professor, a Professor of Statistics, and a Professor of Nutrition and Toxicology at Texas A&M University. He is also Director of Biostatistics Research at the Center for Environmental and Rural Health (NIEHS) and Director of the Training Program in Biology, Bioinformatics, and Nutrition for the National Cancer Institute, both at Texas A&M University.

Dr. Carroll is the author of 3 books and over 200 professional papers, including papers on measurement error modeling, regression variance functions and transformations, nutrition, toxicology, and bioinformatics. Dr. Carroll received his Ph.D. in Statistics from Purdue University in 1974.

James Hardin is Lecturer and Assistant Research Scientist at Texas A&M University and previously was a Senior Statistician at StataCorp, where he developed Stata's cross-sectional time-series capabilities. He is also the author of Stata's current GLM command. He is the author of two books and ten refereed papers, and he has recently been working with Henrik Schmiediche, developing the software, in Stata, for the estimation of measurement-error models. Dr. Hardin received his Ph.D. in Statistics from Texas A&M University in 1992.

See reverse side for details of the North American Stata Users' Group meeting

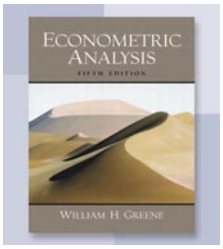
(Continued from page 2)

In his treatment of standard OLS and two-stage least squares, Wooldridge breaks new ground by concentrating the reader's energy on understanding advanced concepts instead of matrix algebra. A traditional approach to introductory econometrics will use its advanced materials sections to give courses on matrix algebra and its applications in econometrics. In contrast, Wooldridge uses the advanced sections to introduce recently developed statistical concepts and techniques. This leads to a text with greater breadth than is usual in most books of this type. This book is equally useful for advanced undergraduate study, as the basis of a survey course at the graduate level, or as a conceptual supplement to advanced courses.

In the second edition, the chapter on limited dependent variables models has been rewritten, producing an exceptional introduction to the topic. Small additions and rewrites are also sprinkled throughout the text. The result is that an excellent introductory book has been made even better.

A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/ie.html**. You may also order using the enclosed bookstore order form.

Econometric Analysis (5th edition)



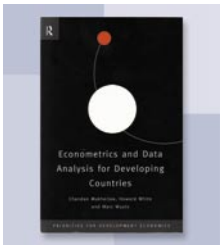
Title: *Econometric Analysis* (5th edition)
Author: W. Greene
Publisher: Prentice-Hall
Copyright: 2003
ISBN: 0-13-066189-9
Pages: 1,056; hardcover
Price: $104.75

Econometric Analysis by William H. Greene is the standard first year graduate econometrics textbook because its accessible style and outstanding coverage have made it so. Since the first edition was published, many new methods have become part of the standard econometric toolkit. The new fifth edition continues the tradition of its predecessors of incorporating new methods. In addition, the fifth edition is based on a much improved organization.

The book achieves its goal of providing a firm foundation in applied econometrics. The introduction to econometric theory in the book provides sufficient theoretical background to support the foundation in applied econometrics or to prepare readers for further work. The book contains many examples that help illustrate how to perform estimation and inference and how to interpret the results.

A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/ea.html**. You may also order using the enclosed bookstore order form.

Econometrics & Data Analysis for Developing Countries



Title: *Econometrics & Data Analysis for Developing Countries*
Authors: C. Mukherjee, H. White, M. Wuyts
Publisher: Routledge
Copyright: 1998
ISBN: 0-415-09400-3
Pages: 496; paperback
Price: $44.75

Chandan Mukherjee, Howard White, and Marc Wuyts provide an interesting introduction to econometrics in *Econometrics and Data Analysis for Developing Countries*.

The authors span an impressive set of topics in this book. In doing so, they emphasize the importance of model specification. The authors also provide a critical evaluation of different methods for how to seek an acceptable specification in practice.

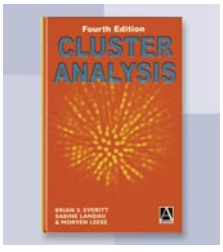
The authors present a number of graphical methods that are frequently overlooked in modern treatments of econometrics. They also point out that graphical methods can be quite informative, even if they are not accompanied by a *p*-value.

The book does not discuss methods that are robust to distributional assumptions, and, in this vein, does not make clear that the normality assumption of linear regression is required only for small-sample statistics to be valid (asymptotically, the standard errors are consistent and the test statistics converge to standard distributions without assuming that the disturbances are normally distributed). Readers interested in a bit more formal treatment or more emphasis on robust methods might want to supplement this book with *Introductory Econometrics: A Modern Approach*, Wooldridge, 2002.

What this book does, and does well, is provide practical approaches to data analysis and model development, with an emphasis on development economics. It will be a valuable tool for those tasked with providing policy analysis based on quantitative information.

A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/edadc.html**. You may also order using the enclosed bookstore order form.

Cluster Analysis (4th edition)



Title: *Cluster Analysis* (4th edition)
Authors: B. Everitt, S. Landau, M. Leese
Publisher: Oxford University Press
Copyright: 2001
ISBN: 0-340-76119-9
Pages: 237; hardcover
Price: $74.75

The fourth edition of *Cluster Analysis*, by Brian S. Everitt, Sabine Landau, and Morven Leese, is a popular and well-written introduction and reference. It introduces the topic and discusses a wide variety of cluster analysis methods.

Especially important is the practical information contained in the book. Examples include how to best visualize clusters, how (and if) to select and transform variables, choosing between the various clustering methods, and comparing the results of different cluster analyses. The book has many examples illustrating the discussion.

A complete table of contents, as well as online ordering information, can be found at **http://www.stata.com/bookstore/cla.html**. You may also order using the enclosed bookstore order form.

Introducción a Stata (in Spanish)

Title: *Introducción a Stata*
Author: Alfonso Ibarra Uriarte
Publisher: MultiON Consulting
Copyright: 2002
ISBN: none
Pages: 75; paperback
Price: $39.00

Este materical es un extracto de los cursos NetCourse 101 y NetCourse 151 impartidos por Stata Corporation en Inglés por Internet. El materical fue seleccionado y traducido al español con el permiso de Stata Corporation.

Una tabla de contenidos completa, así como la información necesaria para ordenar en línea se puede encontrar en **http://www.stata.com/bookstore/mult.html**. También puede hacer su pedido utilizanda el formulario para órdenes de nuestra librería que se encuentra adjunto.

Stat/Transfer for Macintosh OS X available soon

Stat/Transfer, already available for Windows and Unix, will soon be available for Macintosh! Circle Systems will release Stat/Transfer 6 for Macintosh OS X on October 21st. Stat/Transfer facilitates the transfer of data between different spreadsheets and statistical programs by providing a fast and reliable way to convert files from one format to another.

Stat/Transfer for Macintosh OS X supports the following files:

1-2-3	OSIRIS (read-only)
ASCII – Delimited	Paradox
dBASE and compatible formats	Quattro Pro
Epi Info	SAS for Windows and OS/2
Excel	SAS for Unix and the Mac
FoxPro	SAS for Macintosh
Gauss	SAS Transport
HTML tables	S-PLUS
JMP	SPSS Data
LIMDEP	SPSS Portable
Matlab	Stata
Mineset	SYSTAT
Minitab	

You may order Stat/Transfer for Macintosh, as well as Windows and Unix, online at **http://www.stata.com/info/products/ordertr.html**.

THE STATA NEWS is published 4 times a year. It is free to all registered users of Stata.

HOW TO REACH US

Stata Corporation
4905 Lakeway Drive
College Station TX 77845
USA

PHONE 979-696-4600
FAX 979-696-4601
EMAIL stata@stata.com
WEB www.stata.com

Please include your Stata serial number with all correspondence.

INTERNATIONAL DISTRIBUTORS

Chips Electronics
Serving Brunei, Indonesia, Malaysia, Singapore
tel: 62 - 21 - 452 17 61
email: puyuh23@indo.net.id

Dittrich & Partner Consulting
Serving Austria, Czech Republic, Germany, Hungary, Poland
tel: +49 2 12 / 26 066 - 0
email: sales@dpc.de

IEM
Serving Botswana, Lesotho, Namibia, Mozambique, SouthAfrica, Swaziland, Zimbabwe
tel: +27-11-8286169
email: instruments@mweb.co.za

Ixon Technology Company Ltd
Serving Taiwan
tel: +886-(0)2-27045535
email: hank@ixon.com.tw

JasonTech Inc
Serving Korea
tel: +82-(0)2-420-6700
email: info@jat.co.kr

MercoStat Consultores
Serving Argentina, Brazil, Paraguay, Uruguay
tel: 598-2-613-7905
email: mercost@adinet.com.uy

Metrika Consulting
Serving the Baltic States, Denmark, Finland, Iceland, Norway, Sweden
tel: +46-708-163128
email: sales@metrika.se

MultiON Consulting S.A. de C.V.
Serving Mexico, Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama
tel: 52 (55) 55 59 40 50
email: multion@multion.com.mx

Ritme Informatique
Serving Belgium, France, Luxembourg
tel: +33 (0)1 42 46 00 42
email: info@ritme.com

Scientific Solutions S.A.
Serving Switzerland
tel: 41 (0)21 711 15 20
email: info@scientific-solutions.ch

Smit Consult
Serving The Netherlands
tel: +31 416-378 125
email: info@smitconsult.nl

Survey Design & Analysis Services
Serving Australia, New Zealand
tel: +61 (0)3 9878 7373
email: sales@survey-design.com.au

Timberlake Consultants
Serving Eire, U.K.
tel: +44 (0)208 697 3377
email: info@timberlake.co.uk

Timberlake Consulting S.L.
Serving Spain
tel: +34 (9) 5 560 14 30
email: timberlake@timberlakeconsulting.com

Timberlake Consultores, Lda.
Serving Portugal
tel: +351 (2)14307340
email: timberlake.co@mail.telepac.pt

TStat S.r.l.
Serving Italy
tel: +39 0864 210101
email: tstat@tstat.it

Vishvas Marketing-Mix Services
Serving India
tel: 91-22-5405146
email: vishvas@vsnl.com