STATA TECHNICAL BULLETIN

May 2001
STB-61

A publication to promote communication among Stata users

Subscriptions are available from Stata Corporation, email stata@stata.com, telephone 979-696-4600 or 800-STATAPC, fax 979-696-4601.

Previous Issues are available individually from StataCorp. See http://www.stata.com/bookstore/stbj.html for details.

Submissions to the STB, including submissions to the supporting files (programs, datasets, and help files), are on a nonexclusive, free-use basis. In particular, the author grants to StataCorp the nonexclusive right to copyright and distribute the material in accordance with the Copyright Statement below. The author also grants to StataCorp the right to freely use the ideas, including communication of the ideas to other parties, even if the material is never published in the STB. Submissions should be addressed to the Editor. Submission guidelines can be obtained from either the editor or StataCorp.

The *Stata Technical Bulletin* (ISSN 1097-8879) is published six times per year by Stata Corporation. Stata is a registered trademark of Stata Corporation.

Contents of this issue

an75	The Stata Journal begins publication fourth quarter 2001

H. Joseph Newton, Stata Technical Bulletin, editor@stata-journal.com

After ten years of continuous publication, the Stata Technical Bulletin is converting itself into The Stata Journal. The Stata Journal, Volume 1, Issue 1, will begin publication fourth quarter 2001. You should be getting your first copy around October. This, the 61st issue of the Stata Technical Bulletin—the first issue of its eleventh year—will be the last issue of the Bulletin.

The Journal will differ from the Bulletin:

- Whereas the Bulletin was published bimonthly, the Journal will be quarterly.
- Whereas the Bulletin was printed black-on-white on 8 1/2 by 11 inch paper, stapled, the Journal will be standard journal size (7 1/4 by 9 1/4), bound, and with a color cover.
- Whereas the Bulletin included only unreviewed inserts, the Journal will include reviewed articles together with columns.
- And whereas the Bulletin was always professional, the Journal will be even more professional, understated, and yes, academic.

I am very excited about this development and, once you see it, I know you will be, too.

As editor of the Journal, I will still be seeking articles from you, but I will now be able to be less restrictive on what kinds of articles are appropriate. Programs with documentation will always be welcome, but so will other kinds of articles, especially expository articles.

For those who want to follow along as the Journal develops, you can visit the website http://www.stata-journal.com. That website will become the official website of the new Journal and, among other things, will provide guidelines for authors together with subscription information. All STB subscribers will automatically have their subscriptions rolled over to Stata Journal subscriptions; see an76 for details.

I think the best way to understand the new Journal is to understand what was good and not so good about the Bulletin.

Background

For the past ten years, the STB has served as a means of distributing new commands and Stata updates, both user-written and "official". When the STB began under the editorship of Joseph Hilbe, there was no Internet or, at least, no Internet that was used to distribute software and updates, and for its first three years, the STB grew, fulfilling this need. The Stata software was, from the beginning, user programmable, and the STB became the way users distributed to other users what they had done. Users back then subscribed to the STB with diskettes, and these diskettes contained the updates.

For the next three years the STB continued under the editorship of Sean Becketti, growing and prospering even as the Internet grew and prospered.

For the past five years, the STB has continued under my editorship.

In the last five years, however, the growth of the Internet along with the growth in both the number and the diversity of Stata users have both led me and forced me gradually to introduce changes in the STB. In particular, the Internet, Stata's website, and the Statalist listserver (statalist@hsphsun2.harvard.edu) now allow instant communication among users and, moreover, recent improvements to Stata software actually allow it to search the Internet for desired statistical capabilities — whether written by Stata Corporation or by users — and instantly to install what it finds.

The result of this is that no longer is the STB the primary vehicle by which user-written programs are distributed and, with Stata 6, Stata Corporation itself stopped distributing official updates through the STB.

Meantime, with the growth of Stata, the number of application areas that Stata users are drawn from has also grown, with the result that the user base now has wider variation in statistical experience and techniques used.

The result of this has been that, over the past five years, STB "inserts" have become less announcements and short articles describing user-written programs and more longer articles describing complicated programs as well as more general articles about how Stata can be used to analyze interesting datasets.

This change is reflected in the fact that, between Volume 5 and Volume 9, the number of pages increased by 51% (238 to 360) while the number of articles remained almost the same (55 to 59).

In short, not only Stata but also the environment in which Stata users operate has changed enormously over the decade, 1991 to 2001, that the STB has been appearing. It is clear that we should do well to take stock and consider the best future directions for our journal.

The editors of the Stata Technical Bulletin (STB) met in Boston during the North American User Group Meeting in March to discuss the past and future of the STB. (Two editors were unable to attend: Patrick Royston and Joanne Garrett, but contributed to discussions by email.) Also attending was William Gould, president of Stata Corporation.

We came to the unanimous conclusion that the STB itself needed to change and that the changes needed to be substantial enough that a name change was also warranted.

When the STB began, timeliness was of primary importance. Nowadays, printed matter cannot compete with the Internet in that respect. Printed material, if it is to compete, outscores because it is more considered, more substantial, and more trustworthy. We want to focus on that and so came to the conclusion that the Journal should be printed less often (4 times per year rather than 6), thus allowing us to have more time to have articles reviewed, and so making the articles even more considered, substantial, and trustworthy.

We also knew that we needed to change the emphasis of the articles. As I mentioned, programs with documentation will always be welcome, but we believe that users want and need more expository articles, expository articles about statistics and using Stata, rather than about Stata.

And from that, one thing led to another and, by the end of the meeting, we had more or less designed The Stata Journal. There were and still are many details to be worked out. Nicholas Cox of the University of Durham has agreed to become Executive Editor of the Journal and he, along with the other Associate Editors of the STB, Kit Baum of Boston College, Joanne M. Garrett of the University of North Carolina, Marcello Pagano of the Harvard School of Public Health, J. Patrick Royston of the UK Medical Research Council Clinical Trials Unit, and Jeroen Weesie of Utrecht University in the Netherlands, will be working with me to design the Journal. I may soon be calling on others to assist.

If you have any thoughts, desires, or comments, please contact us during this design period. You can send email to editors@stata-journal.com and Nick and I will be sure to read it immediately and to pass along your comments to the other editors as appropriate.

Let me now summarize what we now plan for The Stata Journal.

The Stata Journal

The intention of the editorial board is that it will include the following types of content:

- A major focus will be on expository articles about statistics, especially related to (new) Stata commands, for example, introductions to cluster analysis using Stata (not necessarily about Stata), frailty modeling using Stata, generalized estimating equations, and so on.
- Articles about using Stata such as managing multi-level data in Stata or Stata programming.
- The Journal will have a number of interesting columns such as a column by Nicholas J. Cox on effective use of the Stata language, both interactive and programming, and the best of the FAQs (frequently asked questions) mounted under User Support on the Stata Corporation website, and so on.
- There will be reviews of books, usually those using Stata or about Stata, but occasionally other books in areas deemed of importance or of interest to Stata users as well.
- The Journal will still publish STB-type articles describing new commands by users along with unofficial commands written by Stata Corporation employees, but we will be asking that such articles contain more expository content.

Articles will be reviewed, which should lead to even better articles than those in the current STB. Naturally, this will be especially important to authors from academia, who are judged on their publication record in reviewed journals, but we intend that all readers will benefit. Reviewers will be attracted from the editorial board, previous authors of the Stata Journal and the STB, and experts from various fields of applied statistics.

The Journal will be published quarterly. We are currently designing the appearance of the Stata Journal, but it will have a glossy, color cover, have the same width and height as the Stata manuals, and be bound. The appearance of the articles will be similar to those in many scholarly journals. We are currently in the process of creating a style sheet for authors based on LaTeX, and that will be made available on the http://www.stata-journal.com website (although both ASCII and Word contributions will still be accepted).

Like the STB, the Stata Journal is intended to be for all Stata users, both novice and experienced, with various levels of expertise in statistics, research design, data management, graphics, reporting of results, and of Stata in particular. Each issue should have something of interest to all users. In addition, the Journal will attempt to attract new readers by demonstrating that their research interests are met well by Stata. We aim to disseminate excellent expository material on applied statistics to all researchers and students interested in statistics.

The numerous daily postings on Statalist illustrate very well the readership we have mind, as those who follow it will appreciate. As with many listservers, the style and content of Statalist discussions have evolved very much as an expression of members' interests and expertise. Statalist is centered on, but in no sense limited to, Stata users. Those members' questions and answers range back and forth through specifics on using Stata to general questions on data management, statistical data analysis and modeling, and what is and is not good practice, statistically, computationally and scientifically. Statalist is widely appreciated, not just as a relatively rapid and effective way of solving Stata problems, but also as a source of wisdom on statistical matters in the widest sense. It is this mix which we seek to emulate, although with more substantial and more durable contributions, in the Stata Journal.

Editorial board and more information

I will be the first Editor of the Stata Journal and Nicholas J. Cox of the University of Durham will be joining me as Executive Editor. We both can be reached by emailing editors@stata-journal.com

The current editorial board of the STB will act as the inaugural associate editors of the new journal. They are Kit Baum of Boston College, Joanne M. Garrett of the University of North Carolina, Marcello Pagano of Harvard School of Public Health, J. Patrick Royston of the UK Medical Research Council Clinical Trials Unit, and Jeroen Weesie of Utrecht University in the Netherlands.

For more information on the Stata Journal, see http://www.stata-journal.com.

an76 Stata Journal subscriptions for STB subscribers

Patricia Branton, Stata Corporation, stata@stata.com

STB subscriptions will automatically be converted to Stata Journal subscriptions using the following rules:

For subscriptions expiring before April 2002 (and after June 2001), an additional three months will be added to your expiration date; e.g., if your subscription expires in February 2002 (after the January 2002 STB issue), your subscription to the Stata Journal will expire in May 2002. Instead of receiving the July, September, November, and January issues of the STB, you will receive 2001 quarter 4, 2002 quarter 1, and 2002 quarter 2 issues of the Stata Journal.

All other STB subscriptions will be converted to Stata Journal subscriptions with the same expiration date.

If any STB subscriber does not want their subscription to the STB automatically converted to a subscription to the Stata Journal, email us at stata@stata.com to arrange for a prorated refund of your STB subscription.

Subscriptions to the Stata Journal will be priced as follows:

For subscriptions mailed to US and Canadian addresses:

3-year subscription (includes printed and electronic copy)	$153
2-year subscription (includes printed and electronic copy)	$110
1-year subscription (includes printed and electronic copy)	$ 59

For subscriptions mailed to other countries:

3-year subscription (includes printed and electronic copy)	$225
2-year subscription (includes printed and electronic copy)	$158
1-year subscription (includes printed and electronic copy)	$ 83
3-year subscription (electronic only)	$153

New subscriptions may be ordered on-line at http://www.stata.com/bookstore/sj.

an77 Past issues of the Stata Technical Bulletin

Patricia Branton, Stata Corporation, stata@stata.com

Previous issues of the Stata Technical Bulletin will continue to be available both in printed and electronic form. A convenient way to acquire past issues is via the Stata Technical Bulletin Reprints, which are bound books each containing one year of the Stata Technical Bulletin.

Stata Technical Bulletin Reprints Volume 10, the latest volume, contains STB-55 through STB-61. Reprints Volume 10 is available from StataCorp for $30, plus shipping. Authors of inserts in STB-55 through STB-61 will automatically receive the book at no charge and need not order.

Since STB-61 is the last issue of the Stata Technical Bulletin, we have included this issue in Volume 10, along with the previous six issues.

This book of reprints includes everything that appeared in issues 55–61 of the STB. As a consequence, you do not need to purchase the reprints if you saved your STBs. However, many subscribers find the reprints useful since they are bound in a convenient volume.

Reprint Volume 10 will be available in late July and may be ordered online at http://www.stata.com/bookstore/stbr.html or with the enclosed order form.

All ten Reprint Volumes may be ordered as a set for $195 on-line at http://www.stata.com/bookstore/stbr.html.

dm73.3 Contrasts for categorical variables: update

John Hendrickx, University of Nijmegen, Netherlands, J.Hendrickx@mailbox.kun.nl

Abstract: Bug-fixes to `desmat`, `desrep` and `destest` are described.

Keywords: contrasts, interactions, categorical variables.

`desmat` has been enhanced to automatically encode string variables and will now work with the programs `stcox` and `streg` in which a dependent variable is not specified. Problems with pzat characteristics with the maximum length (8 in version 6, 32 in version 7) have been fixed. A compatibility problem with version 7 has been fixed by renaming the subprogram class.

`desrep` now works properly if equation names contain spaces, e.g., in the value labels of the dependent variable in `mlogit`. A bug in which the standard errors were not printed when the `all` option was specified has been fixed. A compatibility problem with version 7 has been fixed by specifying the main program first, and then the subprograms.

`destest` now works in version 7 as well. A compatibility problem arose here as well due to the fact that a subprogram had been specified before the main program.

The latest version of the `desmat` package is available at http://baserv.uci.kun.nl/~johnh/desmat/stata/ or using

```
. net from http://baserv.uci.kun.nl/~johnh/
```

References

Hendrickx, J. 1999. dm73: Using categorical variables in Stata. *Stata Technical Bulletin* 52: 2–8. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 51–59.

——. 2000. dm73.1: Contrasts for categorical variables: update. *Stata Technical Bulletin* 54: 7. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 60–61.

——. 2001. dm73.2: Contrasts for categorical variables: update. *Stata Technical Bulletin* 59: 2–5.

dm91 Patterns of missing values

Jeroen Weesie, Utrecht University, Netherlands, J.Weesie@fss.uu.nl

Abstract: The command `mvpatterns` lists patterns of missing values.

Keywords: missing values, patterns.

Introduction

This insert describes a command to list patterns of missing values. This is a command that is easily understood by an example using a slightly altered version of Stata's auto data.

(*Continued on next page*)

```
. mvpatterns
variables with no mv's: make
```

Variable	type	obs	mv	variable label
price	int	64	10	Price
mpg	int	71	3	Mileage (mpg)
rep78	int	59	15	Repair Record 1978
headroom	float	65	9	Headroom (in.)
trunk	int	71	3	Trunk space (cu. ft.)
weight	int	61	13	Weight (lbs.)
length	int	68	6	Length (in.)
turn	int	70	4	Turn Circle (ft.)
displacement	int	65	9	Displacement (cu. in.)
gear_ratio	float	69	5	Gear Ratio
foreign	byte	68	6	Car type

```
Patterns of missing values
```

_pattern	_mv	_freq
++++++++++	0	25
++.++++++++	1	6
.++++.+++++	2	4
+++++.+++++	1	3
+++.++++++.	2	3
++.+++++.++	2	3
++++++++.++	1	2
+++++++.+++	1	2
++++++.++++	1	2
++++.++++++	1	2
+++.+++++++	1	2
++++++++++.	1	1
+++++++++.+	1	1
+.+++++++++	1	1
.++++++++++	1	1
++++++++..+	2	1
++++++..+++	2	1
+++++.++++.	2	1
+++++.+.+++	2	1
++.++++++.+	2	1
++.++.+++++	2	1
.++++++++.+	2	1
.+++++++.++	2	1
.++.+++++++	2	1
+++.+.+++.+	3	1
++.+++.+.++	3	1
++.++.++.++	3	1
+.++..+++++	3	1
+.+.++.++++	3	1
.+.+++.++++	3	1
.+..++++++.	4	1

In the output, a + marks a nonmissing value and a period (.) a missing value. With string variables, empty strings are treated, as usual in Stata, as missing values.

The output can be made more readable with a number of options. First, the `skip` option specifies that lines and spaces are displayed after every fifth variable. The `sort` option specifies that the variables are sorted into order of decreasing missingness. Thus, the first variable is missing most frequently, and the last variable is missing least often. Finally, the `minfreq()` option sets the patterns to be restricted to those that occur at least `minfreq` times in the data.

(*Continued on next page*)

```
. mvpatterns, skip sort minfreq(2)
variables with no mv's: make
```

Variable	type	obs	mv	variable label
rep78	int	59	15	Repair Record 1978
weight	int	61	13	Weight (lbs.)
price	int	64	10	Price
displacement	int	65	9	Displacement (cu. in.)
headroom	float	65	9	Headroom (in.)
foreign	byte	68	6	Car type
length	int	68	6	Length (in.)
gear_ratio	float	69	5	Gear Ratio
turn	int	70	4	Turn Circle (ft.)
mpg	int	71	3	Mileage (mpg)
trunk	int	71	3	Trunk space (cu. ft.)

```
Patterns of missing values (freq >= 2)
     _pattern   _mv    _freq
+++++ +++++ +     0       25
.++++ +++++ +     1        6
+..++ +++++ +     2        4
+.+++ +++++ +     1        3
++++. .++++ +     2        3
.++.+ +++++ +     2        3
+++++ +++++ .     1        2
+++++ +++.+ +     1        2
+++++ +.+++ +     1        2
++++. +++++ +     1        2
+++.+ +++++ +     1        2
In addition:  20 observations with unique missing values patterns
```

The number of variables to be analyzed by **mvpatterns** is limited by two factors. First, it cannot list results for more than 80 variables with missing values. Second, the number of variables is limited by the linesize.

The inspiration for this command was found in the SOLAS program and the Stata command **pattern** written by Goldstein (1996a; 1996b). We welcome suggestions for additions to this command.

Syntax

mvpatterns *varlist* [**if** *exp*] [**in** *range*] [, **minfreq(#) skip sort notable nodrop**]

Description

mvpatterns lists the missing value patterns of the variables and their frequency. In a pattern, + denotes a nonmissing value, and a period (.) denotes a missing value. For string variables, empty strings are treated as missing values. Patterns are sorted by frequency among the selected observations, along with the frequency and number of missing values in the pattern.

mvpatterns is sensitive to the linesize, affecting both the display of variable labels and the number of variables to be analyzed simultaneously.

Options

minfreq(*#*) specifies the minimal frequency of a missing value pattern for being listed. The number of unlisted patterns and the number of associated observations are described after the list. **minfreq** defaults to 1, that is, all patterns are listed.

skip specifies that spaces are inserted in the missing value pattern after every 5 variables, and similarly, lines in the table of variables are displayed, both to enhance readability.

sort specifies that the variables are sorted into the order of decreasing "missingness".

notable specifies that the header table describing the variables is suppressed.

nodrop specifies that variables with no missing values are included in the table and in the listing of missing value patterns.

References

Goldstein, R. 1996a. sed10: Patterns of missing data. *Stata Technical Bulletin* 32: 12–13. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, p. 115.

——. 1996b. sed10.1: Update to pattern. *Stata Technical Bulletin* 33: 2. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 115–116.

dm92	Simulating disease status and censored age

Jisheng Cui, University of Melbourne, Australia, j.cui@unimelb.edu.au

Abstract: The command `phenotyp` for simulating disease status and censored age of each person in two- or three-generation families is introduced and illustrated.

Keywords: disease status, family data, simulation.

Introduction

In Cui (2000), I proposed two programs, `simuped2` and `simuped3`, for simulating two- or three-generation families. However, to make them of general use, I did not include the function of simulating the disease status in these programs, because special statistical models are required for the risk of developing a disease. In addition, the natural death prior to the onset of a disease also needs to be taken into consideration. Here I present a program, `phenotyp`, which is used to simulate the disease status and censored age of each person in the two- or three-generation families on the basis of age, sex and genotype that are generated by `simuped2` or `simuped3` previously.

Syntax

`phenotyp` *hr* [`, type( d | r ) alpha(#) lambda(#) maxage(#) gamma(#) sex( b | f | m )`]

Description

`phenotyp` is a command used to generate the disease status and censored age of individuals in simulated families on the basis of age (hereafter referred to as calendar age), sex, and genotype that are generated by `simuped2` or `simuped3` previously (Cui 2000). A censored age is the age at onset of a disease if a person is affected, or the age of death if a person has died before developing the disease. Otherwise, it is the calendar age simulated by `simuped2` or `simuped3` previously.

`phenotyp` expects the genetic hazard ratio, *hr*, to be specified as a nonnegative value.

Options

`type`(*inheritance*) specifies the type of inheritance. `d` (the default) specifies dominant inheritance, and `r` specifies recessive inheritance.

`alpha`(#) specifies the shape parameter of the Weibull distribution. The default value is 4.21.

`lambda`(#) specifies the scale parameter of the Weibull distribution. The default value is 9.95×10^{-10}.

`maxage`(#) specifies the upper bound of age of death. The default value is 100.

`gamma`(#) specifies the parameter in the log-power density function of death. The default value is 15.

`sex`(*text*) specifies the sex affected by a disease. `b` (the default) specifies that both sexes are affected, `f` specifies that only females are affected, and `m` specifies that only males are affected by the disease.

Remarks

Two modes of inheritance, dominant and recessive, are allowed in this program. They are represented by `d` and `r` in the `type` option, respectively. Use of other letters will produce an error message. If a disease affects one gender only, the disease status is generated as 0 for a person with an opposite gender of the effect. Further details about the statistical models can be found in the *Methods and formulas* section below.

Example

Based on the calendar age, sex, and genotype of individuals in the 1,000 two-generation families simulated by `simuped2` (Cui 2000), we generate the disease status and censored age for each person in these families.

```
. simuped2 70 10 40 10, reps(1000) sav(output) alle(0.05) sib(5)
. use output
. phenotyp 10, type(d) alpha(4.21) lambda(9.95e-10) maxage(100) gamma(15) sex(f)
```

```
. describe
Contains data from output.dta
  obs:         6,818
 vars:            10                          21 Dec 2000 11:41
 size:       265,902 (74.5% of memory free)
-------------------------------------------------------------------------------
   1. famid     float  %9.0g
   2. id        float  %9.0g
   3. degree    float  %9.0g
   4. female    float  %9.0g
   5. age       float  %9.0g
   6. genotype  str2   %9s
   7. age_dth   float  %9.0g
   8. age_dis   float  %9.0g
   9. age_cen   float  %9.0g
  10. disease   byte   %8.0g
-------------------------------------------------------------------------------
Sorted by:  famid  id
     Note:  dataset has changed since last saved
. list famid id female age age_dth age_cen disease
          famid         id     female        age    age_dth    age_cen   disease
   1.         1          1          0         73         61         61         0
   2.         1          2          1         75         87         63         1
   3.         1          3          0         28         98         28         0
   4.         1          4          1         43         97         43         0
   5.         1          5          1         25         97         25         0
   6.         2          1          0         64         74         64         0
   7.         2          2          1         58         89         43         1
   8.         2          3          0         38         20         20         0
   9.         2          4          1         46         94         46         0
  10.         2          5          0         51         66         51         0
  (output omitted)
6809.       999          2          1         51         52         51         0
6810.       999          3          0         50         88         50         0
6811.       999          4          1         38         85         38         0
6812.       999          5          0         37         87         37         0
6813.       999          6          1         41         59         41         0
6814.      1000          1          0         74         38         38         0
6815.      1000          2          1         70         68         68         0
6816.      1000          3          1         38         57         38         0
6817.      1000          4          1         41         89         34         1
6818.      1000          5          1         38         99         38         0
```

In this example, the hazard ratio is assumed to be 10 for a dominantly inherited disease. The variable `disease` indicates whether an individual develops a disease (1 means yes, 0 mean no) at the censored age `age_cen`. The variable `age` is the calendar age generated by `simuped2`, and `age_dth` is the expected age of natural death. In this example, we specify that only females are affected by the disease, so all males are disease free, indicated by `disease` $= 0$.

Methods and formulas

In accordance with Cui and Hopper (2000), age at death (t_d) of a person is simulated by the log-power density function

$$f(t_d) = \gamma\{\ln(t_d)\}^{\gamma-1}/[t_d\{\ln(M)\}^{\gamma}]$$

where $1 \leq t_d \leq M$ and M is the maximum age of death. When $\gamma = 15$, this function gives a median age of death of 81 years, consistent with Australian female population data (Australian Bureau of Statistics 1997).

The baseline hazard function, which is the hazard of developing a disease at age t for individuals with genotypes AA, is assumed to be $h_0(t) = \alpha\lambda t^{\alpha-1}$, a Weibull distribution with shape parameter α and scale parameter λ. When $\alpha = 4.21$ and $\lambda = 9.95 \times 10^{-10}$, the cumulative probability of a disease at age 70 years is 6%, which is consistent with breast cancer incidence rates in Australia (McCredie et al. 1995).

Under dominant inheritance, the hazard of disease for individuals with genotype Aa or aa is given by $h_1(t) = h_0(t) \times HR$, where HR is the hazard ratio of mutation carriers compared with that of non-carriers. Under recessive inheritance, the hazard rate for individuals with genotype Aa is $h_0(t)$ and the hazard rate for individuals with genotype aa is $h_1(t)$ (Cui and Hopper 2000).

In accordance with Li and Thompson (1997), the age at onset of a disease (t_s) is simulated from the population hazard $h_0(t)$ and the genotype-specific HR, according to

$$\Pr(\text{disease at or before age } t_s) = 1 - \exp(\lambda_g t_s^{\alpha})$$

where, under dominant inheritance, $\lambda_g = \lambda$ for genotype AA and λHR for genotype Aa or aa; whereas, under recessive inheritance, $\lambda_g = \lambda$ for genotype AA or Aa and λHR for genotype aa.

The censored age $t_c = \min(t, t_d, t_s)$ where t is the calendar age generated by `simuped2` or `simuped3`. When $t_s \leq \min(t, t_d)$, a person develops the disease at the censored age t_c; otherwise, a person died or is still alive at age t_c depending on whether $t_d \leq t$. The disease status is 1 if a person develops the disease; otherwise, it is 0 (Cui and Hopper 2000).

References

Australian Bureau of Statistics. 1997. *Deaths, Australia, 1997.* Canberra, Australian Capital Territory: Australian Bureau of Statistics.

Cui, J. 2000. dm82: Simulating two- and three-generation families. *Stata Technical Bulletin* 58: 2–5.

Cui, J. and J. L. Hopper. 2000. Why are the majority of hereditary cases of early-onset breast cancer sporadic? A simulation study. *Cancer Epidemiology, Biomarkers & Prevention* 9: 805–812.

Li, H. and E. A. Thompson. 1997. Semiparametric estimation of major gene and family-specific random effects for age of onset. *Biometrics* 53: 282–293.

McCredie, M. R. E., J. L. Hopper, and J. N. Cawson. 1995. Risk factors and preventive strategies for breast cancer. *Medical Journal of Australia* 163: 435–440.

gr33.1	Violin plots for Stata 6 and 7

Thomas J. Steichen, RJRT, steicht@rjrt.com

Abstract: This insert updates `violin` to run correctly under Stata version 7; nonetheless, it is a version 6 program. A full description of the method and the operation of the original, version 5, command was given in Steichen (1998).

Keywords: violin plots, box plots, kernel density, exploratory data analysis.

Description

`violin`, a program that produces "violin" plots, as described by Hintze and Nelson (1998), was published in the STB as a version 5 program. It was upgraded to version 6 but that version was never released via the STB. Now, due to the sort-stability changes in Stata version 7, it is being officially published as a version 6 program that functions correctly under version 7. `violin` continues to operate as documented in Steichen (1998) but with slightly enhanced documentation and a minor change to display and save summary results for each violin when a multi-variable or a grouped plot is generated.

References

Hintze, J. L. and R. D. Nelson. 1998. Violin plots: a box plot—density trace synergism. *The American Statistician* 52(2): 181–184.

Steichen, T. J. 1998. gr33: Violin plots. *Stata Technical Bulletin* 46: 13–18. Reprinted in *The Stata Technical Bulletin Reprints*, vol. 8, pp. 57–65.

gr42.1	Quantile plots, generalized: update to Stata 7.0

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

Abstract: `quantil2`, a variant on `quantile` originally published in STB 51, is updated to Stata 7.0. The earlier version depended on `egen` leaving the data in a certain sort order, which no longer is the case, necessitating a small fix.

Keywords: quantile plots, graphics.

Description

`quantil2` produces a plot of the ordered values of a *varlist* against the so-called plotting positions, which are essentially quantiles of a uniform distribution on $[0, 1]$ for the same number of values. `quantil2` generalizes the `quantile` command in five ways:

1. One or more variables may be plotted.
2. The sort order may be reversed, so that values decrease from top left.
3. A single variable may be classified by another single variable, specified by the `by()` option.
4. There is support for graphical choices other than the set `s(o) c(.) xsca(0,1) xlab(0,.25,.5,.75,1)` wired into `quantile`.
5. The plotting position is in general $(i - a)/(n - 2a + 1)$, in contrast to the more specific $(i - 0.5)/n$ wired into `quantile`. This is a minor point graphically, but may be useful to some users.

This insert updates `quantil2` to Stata 7.0 to fix one specific problem. `quantil2` uses the `egen` function `pp()` to calculate plotting positions previously published by Cox (1999a). (A copy of the program is included on the media associated with this insert.) As explained in [P] **intro** (p. 3), `egen` under Stata 7.0 leaves the data in the sort order in which it found them. The previous version of `quantil2` relied on the previous behavior whereby the data could emerge after a call to `egen` with a different sort order. This is now fixed.

For more details on `quantil2`, see Cox (1999b).

References

Cox, N. J. 1999a. dm70: Extensions to generate, extended. *Stata Technical Bulletin* 50: 9–17. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 34–45.

——. 1999b. gr42: Quantile plots, generalized. *Stata Technical Bulletin* 51: 16–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 113–116.

sbe19.4 | Update to metabias to work under version 7

Thomas J. Steichen, RJRT, steicht@rjrt.com

Abstract: This insert updates `metabias` to run correctly under Stata version 7; it remains a version 6 program.

Keywords: meta-analysis, publication bias, Egger, Begg.

Description

This insert updates `metabias` to run correctly under Stata version 7; it remains a version 6 program. A full description of the method and of the operation of the original command and options was given in Steichen (1998). A few revisions were documented later in Steichen et al. (1998) and the Stata 6 version and enhancements were provided in Steichen (2000a) and Steichen (2000b).

`metabias` required a few changes to function correctly under version 7 due to the new sort-stability of Stata. These changes have been implemented. `metabias` remains a version 6 program and continues to operate as documented in prior inserts.

References

Steichen, T. J. 1998. sbe19: Tests for publication bias in meta-analysis. *Stata Technical Bulletin* 41: 9–15. Reprinted in *Stata Technical Bulletin Reprints*,vol. 7, pp. 125–133.

——. 2000a. sbe19.2: Updates of tests for publication bias in meta-analysis. *Stata Technical Bulletin* 57: 4.

——. 2000b. sbe19.3: Tests for publication bias in meta-analysis: erratum. *Stata Technical Bulletin* 58: 8.

Steichen, T. J., M. Egger, and J. Sterne. 1998. sbe19.1: Tests for publication bias in meta-analysis. *Stata Technical Bulletin* 44: 3–4. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 84–85.

sbe39.2 | Update of metatrim to work under version 7

Thomas J. Steichen, RJRT, steicht@rjrt.com

Abstract: This insert updates `metatrim` (Steichen 2000a, 2000b) to run correctly under Stata version 7; it remains a version 6 program. `metatrim` implements the Duval and Tweedie (2000) nonparametric "trim and fill" method of accounting for publication bias in meta-analysis.

Keywords: meta-analysis, publication bias, nonparametric, data augmentation.

Description

`metatrim`, a Stata version 6 program, required a few changes to function correctly under version 7 due to the new sort-stability of Stata. These changes have been implemented. `metatrim` remains a version 6 program and continues to operate as documented in prior inserts.

References

Duval, S. and R. Tweedie. 2000. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* 95(449): 89–98.

Steichen, T. J. 2000a. sbe39: Nonparametric "trim and fill" analysis of publication bias in meta-analysis. *Stata Technical Bulletin* 57: 8–14.

——. 2000b. sbe39.1: Nonparametric "trim and fill" analysis of publication bias in meta-analysis: erratum. *Stata Technical Bulletin* 58: 8–9.

sg158.1	Update to random-effects ordered probit

Guillaume R. Frechette, Ohio State University, gurst1@econ.ohio-state.edu

Abstract: This insert fixes some bugs that sometimes lead the command `reoprob` to abort maximization unexpectedly.

Keywords: random-effects ordered probit.

Description

`reoprob` estimates a random-effects ordinal probit model, a full description of the program is provided in Frechette (2001).

I corrected a bug which sometimes forced the optimization process to stop when the parameter estimate attempted by `ml` was such that the probabilities of some events became negative.

Acknowledgments

I am grateful to David Cooper for alerting me of the problem, and to Vince Wiggins for giving me the solution to it.

Reference

Frechette, G. R. 2001. sg158: Random-effects ordered probit. *Stata Technical Bulletin* 59: 23–27.

sg163	Stereotype ordinal regression

Mark Lunt, ARC Epidemiology Unit, University of Manchester, UK, mdeasml2@fs1.ser.man.ac.uk

Abstract: This insert describes an alternative form of ordinal regression model, the Stereotype Ordinal Regression (SOR) Model, which can be thought of as imposing ordering constraints on a multinomial model. The command `soreg` is introduced and illustrated.

Keywords: ordinal regression, stereotype ordinal regression, proportional odds model.

Introduction

There are a number of reasonable approaches to analyzing an ordinal outcome variable. One common approach, known as the Proportional Odds (PO) Model, is implemented in Stata as `ologit`. If the assumptions of the PO model are not satisfied, an alternative is to treat the outcome as categorical, rather than ordinal, and use multinomial logistic regression (`mlogit`) in Stata. This insert describes an alternative form of ordinal regression model, the Stereotype Ordinal Regression (SOR) Model, which can be thought of as imposing ordering constraints on a multinomial model. The multinomial model provides the best possible fit to the data, at the cost of a large number of parameters which can be difficult to interpret. Stereotype regression aims to reduce the number of parameters by imposing constraints, without reducing the adequacy of the fit.

Distinguishability and dimensionality

In introducing his stereotype ordinal regression model, Anderson (1984) justified it in terms of the concepts of distinguishability and dimensionality.

Dimensionality

A fundamental assumption of the Grouped Continuous Model is that given a set of predictor variables x_{ij}, where i indexes the subject $1, \ldots, n$ and j indexes the variable, $1, \ldots, p$, the same combination of variables, $\sum_{j=0}^{p} x_{ij}\beta_j$, can be used to distinguish between all levels of the outcome variable. If, however, one combination can distinguish between levels 1 and 2, but a different one is required to distinguish between levels 2 and 3, the relationship is two dimensional.

Distinguishability

Two outcome categories are *indistinguishable* with respect to a variable x_j if x_j is not predictive between the two categories.

The PO model, implemented in `ologit` assumes that the association between the predictor variables x_j and the outcome variable y is one dimensional. In addition, if there are k possible outcome categories, it assumes that the odds ratio for being in group s or higher, relative to the odds of being in group $s-1$ or lower, associated with each variable x_j, is the same for all $s, 2 \leq s \leq k$. If either of these assumptions are untrue, the PO model is inappropriate.

The stereotype ordinal regression model

A full multinomial model model is of the form

$$\Pr(y_i = s|x_{i1} \dots x_{ip}) = \frac{\exp\left(\beta_{0s} + \sum_{j=1}^{p} x_{ij}\beta_{sj}\right)}{\sum_{t=1}^{k} \exp\left(\beta_{0t} + \sum_{j=1}^{p} x_{ij}\beta_{tj}\right)}$$

As it stands, this model is not identified since adding a fixed constant to every β will give exactly the same predicted probabilities. To identify the model, constraints need to be placed on the parameters. Commonly β_k and β_{0k} are constrained to be 0. If the x_j variables are all categorical, this model is then saturated.

The multinomial model assumes that different linear combinations of the predictor variables are required to discriminate between different pairs of levels of the outcome variable. All of these combinations need to be estimated separately. However, if the outcome is ordinal, rather than categorical, it may be that a single linear combination can discriminate between all levels. If this is the case, the above model can be simplified to

$$\Pr(y_i = s|x_{i1} \dots x_{ip}) = \frac{\exp\left(\beta_{0s} + \phi_s \sum_{j=1}^{p} x_{ij}\beta_j\right)}{\sum_{t=1}^{k} \exp\left(\beta_{0t} + \phi_t \sum_{j=1}^{p} x_{ij}\beta_j\right)}$$

This is a one-dimensional stereotype model. In this model, the β parameters no longer differ between the different levels of the outcome. The combination that best discriminates between the outcome variables is given by $\sum_{j=1}^{p} x_{ij}\beta_j$ and the distance between the outcome levels in terms of this linear predictor is given by the ϕ_j parameters. If the relationship between the predictors and outcome is ordinal, then $\phi_1 \geq \phi_2 \geq \dots \geq \phi_k$.

Again, constraints are needed to make the model identifiable. Anderson recommended setting $\phi_k = 0$ and $\phi_1 = 1$, but other constraints are possible.

The stereotype model can be considered as a constrained multinomial model, in which the ratios β_{sj}/β_{tj} are constant for all variables x_j. In fact, if there is only one predictor, the stereotype model is simply a reparameterization of the multinomial model; the goodness-of-fit, predicted values, and so on, are all identical. If there is more than one predictor variable, there are two comparisons to be made when deciding if a stereotype model is adequate: it should be compared to the null (intercept only model) to see if its ability to predict outcome is greater than would be expected by chance. This is equivalent to the chi-squared test supplied by most regression programs. Secondly, it should be compared to the full (multinomial) model, to see if there has been a significant loss of predictive ability when the simplifying constraints of the stereotype model were imposed. This can be thought of as a "goodness-of-fit" test, analogous to the Pearson or Hosmer–Lemeshow test after logistic regression (`lfit`) or the Poisson goodness-of-fit test (`poisgof`) after Poisson regression.

Syntax

`soreg` *varlist* [`if` *exp*] [`in` *range*] [`, maxdim(#) constraints(`*numlist*`)`]

Description

`soreg` implements the stereotype ordinal regression model proposed by Anderson. Constraints may be defined to perform constrained estimation.

Options

`maxdim(#)` defines the dimension of the stereotype model to be fitted.

`constraints(`*numlist*`)` specifies the linear constraints to be applied during estimation. The default is to perform unconstrained estimation. Constraints are defined with the `constraint` command. `constraints(1)` specifies that the model is to be constrained according to constraint 1; `constraints(1-4)` specifies constraints 1 through 4; `constraints(1-4,8)` specifies 1 through 4 and 8. It is not considered an error to specify nonexistent constraints as long as some of the constraints exist. Thus, `constraint(1-999)` would specify that all defined constraints be applied.

Two simple examples

Consider the nausea data presented by Farewell (1982) and reanalyzed by Anderson. The file `nausea.dta` contains this dataset, with a single entry for each of the 219 subjects. Typing

```
. soreg nausea treat
```

will produce the following output:

```
. soreg nausea treat
iteration 0: Log Likelihood = -371.4567
iteration 1: Log Likelihood = -371.4567
iteration 2: Log Likelihood = -371.4567
Stereotype Logistic Regression                      Number of obs =       219
Comparison to null model                            LR Chi2(5)    =     18.01
                                                    Prob > chi2   =    0.0029
Comparison to full model                            LR Chi2(0)    =      0.00
                                                    Prob > chi2   =         .
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   phi11 |          1          .          .       .            .           .
   phi21 |   .9101804   .5096764      1.786   0.074       -.088767    1.909128
   phi31 |   .6789914   .6490182      1.046   0.295       -.593061    1.951044
   phi41 |  -.1123247   .4708013     -0.239   0.811      -1.035078     .810429
   phi51 |  -.6699209   .6525803     -1.027   0.305      -1.948955    .6091131
   phi61 |  (dropped)
  beta11 |  -1.087051   .5215595     -2.084   0.037      -2.109289   -.0648137
------------------------------------------------------------------------------
beta1 = treat
```

The ϕ_i variables are labeled `phi11` to `phi61` (in a two-dimensional model, there would be two sets of ϕ_i variables, `phi11` to `phi61` and `phi12` to `phi62`). The predictor variables for the first dimension are labeled `beta11` to `betak1`, for the second dimension `beta12` to `betak2`, and so on.

There are no standard errors for `phi11` or `phi61`, since they are constrained to be 1 and 0 respectively. Compared to the null (intercept only model), the stereotype model fits significantly better (`chi2 = 18.0` on 5 df, `p = 0.003`). However, there is no difference in the fit when compared to the full (multinomial) model, since the SOR model is simply a reparameterization of the multinomial model.

A more complex example, also analyzed by Anderson, concerns the prognosis for suffers of backpain originally presented by Doran and Newell (1975). In this case, the outcome was scored from 1 ("worse") to 6 ("complete relief"), and there were three predictor variables, `x1`, `x2`, and `x3`. Fitting the stereotype model to these data gives the following output:

```
. soreg pain x1 x2 x3
iteration 0: Log Likelihood = -151.1720
iteration 1: Log Likelihood = -151.4538
iteration 2: Log Likelihood = -151.5506
iteration 3: Log Likelihood = -151.5501
iteration 4: Log Likelihood = -151.5501
iteration 5: Log Likelihood = -151.5501
Stereotype Logistic Regression                      Number of obs =       101
Comparison to null model                            LR Chi2(7)    =     39.96
                                                    Prob > chi2   =    0.0000
Comparison to full model                            LR Chi2(8)    =      4.09
                                                    Prob > chi2   =    0.8494
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   phi11 |          1          .          .       .            .           .
   phi21 |   .3092677   .1294594      2.389   0.017       .0555321    .5630034
   phi31 |   .3465159   .1308123      2.649   0.008       .0901285    .6029034
   phi41 |   .5097036   .1608969      3.168   0.002       .1943514    .8250558
   phi51 |   .1414456   .0933218      1.516   0.130      -.0414618     .324353
   phi61 |  (dropped)
  beta11 |   5.374471   2.000401      2.687   0.007       1.453757    9.295185
  beta21 |    3.08102   1.055979      2.918   0.004       1.011339      5.1507
  beta31 |   2.712501   1.180846      2.297   0.022       .3980843    5.026917
------------------------------------------------------------------------------
beta1 = x1
beta2 = x2
beta3 = x3
```

In this case, the stereotype model requires 8 fewer parameters than the full multinomial model, although the multinomial model does not fit significantly better ($\chi^2 = 4.1$ on 8 df, $p = 0.85$). However, the stereotype model does fit significantly better than the null model ($\chi^2 = 40.0$ on 7 df, $p < 0.0001$). Thus, the simplification can be considered successful. However, both of the above models can be simplified further, using constraints.

Constraints

Constraints play two important roles in Anderson's ordinal regression strategy. First, it is necessary to impose constraints on the parameters, in order to make the model identifiable. In addition, questions of distinguishability can be addressed through constraints.

Identifiability

We have already seen that the basic stereotype model is not identified (multiplying all of the β parameters by a constant c and dividing all the ϕ parameters by the same constant would produce exactly the same fitted values). To identify the model, constraints must be applied. Anderson recommended setting ϕ_k to 0 and ϕ_1 to 1, and these constraints are implemented in `soreg` by default. However, the defaults can be easily overridden by user-defined constraints. For example, consider the analysis by Greenland (1994) of the pneumoconiosis data from Ashford (1959). Greenland fitted the constraints $\phi_1 = 0$ and $\phi_2 = 1$, and estimated ϕ_3. The fitting of this model is shown below.

```
. constraint define 1 phi11 = 0
. constraint define 2 phi21 = 1
. soreg pneum lyears, c(1 2)
 ( 1)  phi11 = 0.0
 ( 2)  phi21 = 1.0
iteration 0: Log Likelihood = -204.4344
iteration 1: Log Likelihood = -204.4344
iteration 2: Log Likelihood = -204.4344
Stereotype Logistic Regression                  Number of obs =       371
Comparison to null model                        LR Chi2(2)    =     96.29
                                                Prob > chi2   =    0.0000

Comparison to full model                        LR Chi2(0)    =      0.00
                                                Prob > chi2   =         .
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   phi11 |  (dropped)
   phi21 |          1          .         .       .            .           .
   phi31 |     1.4166   .3649241      3.882   0.000       .7013616    2.131838
  beta11 |   2.165373   .4574869      4.733   0.000       1.268715    3.062031
------------------------------------------------------------------------------
beta1 = lyears
```

It can be seen that after setting the constraints, `phi11` is fixed at 0 (that is, dropped from the model) and `phi21` is fixed at 1. With these constraints, the estimates for `phi31` and `beta11` from `soreg` match those given by Greenland.

A two-dimensional model requires additional constraints; again defaults are available, although they can be overridden. However, for models of greater than two dimensions, defaults are not provided, since the appropriate constraints will often depend on the context. If no constraints are defined in such a model, an error message will be printed and no model estimated.

Distinguishability

The ϕ parameters give a measure of the distinguishability of the various categories with respect to the predictors: if the ϕ parameters of two categories are similar, it is likely that the categories are indistinguishable. In this case, the model can be simplified by constraining the corresponding ϕ parameters to be equal. For example, consider the nausea data analyzed above. Looking at the phi parameters and their standard errors suggested that there were only two distinguishable groups, and that `phi11` = `phi21` = `phi31`, and `phi41` = `phi51` = `phi61`. We can therefore define the following constraints.

```
. constraint define 1 phi11 = phi21
. constraint define 2 phi11 = phi31
. constraint define 3 phi41 = phi51
. constraint define 4 phi41 = phi61
. constraint define 5 phi11 = 1
. constraint define 6 phi61 = 0
```

Fitting the stereotype model with these constraints applied gives the following output:

```
. soreg nausea treat, c(1/6)
 ( 1)  phi11 - phi21 = 0.0
 ( 2)  phi11 - phi31 = 0.0
 ( 3)  phi41 - phi51 = 0.0
 ( 4)  phi41 - phi61 = 0.0
 ( 5)  phi11 = 1.0
 ( 6)  phi61 = 0.0
iteration 0: Log Likelihood = -371.4567
iteration 1: Log Likelihood = -372.7759
iteration 2: Log Likelihood = -372.7759
Stereotype Logistic Regression                  Number of obs =       219
Comparison to null model                        LR Chi2(1)    =     15.37
                                                Prob > chi2   =    0.0001
Comparison to full model                        LR Chi2(4)    =      2.64
                                                Prob > chi2   =    0.6200
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   phi11 |          1          .          .       .              .           .
   phi21 |          1          .          .       .              .           .
   phi31 |          1          .          .       .              .           .
   phi41 |  (dropped)
   phi51 |  (dropped)
   phi61 |  (dropped)
  beta11 |  -1.244581   .3299873     -3.772   0.000      -1.891344   -.5978177
------------------------------------------------------------------------------
beta1 = treat
```

This model is significantly better than the null model, and not significantly worse than the full model. Thus, it provides a good fit to the data, despite having only one parameter. It says that the odds of being in group 1, 2 or 3 rather than 4, 5 or 6 decrease by a factor of 3.5 (exp(1.24)) for those treated with cisplatinum relative to those untreated. The interpretation of this model is therefore extremely simple while explaining the meaning of each of the parameters in the unconstrained model would have been far more difficult.

Note that the default constraints were redefined. This is necessary since the default constraints are only applied if *no* user-defined constraints are given. If insufficient constraints are given, the model may be estimated and the likelihood-ratio tests give exactly the same results, but the parameters will have arbitrary values, which makes interpretation difficult.

Constraints can cause some difficulties. First, the first predictor variable has a special role in the model-fitting process. it is therefore not possible to constrain `beta11`, `beta12`, and so on, to be equal to another variable. These variables can, however, be constrained to a constant value. Secondly, constraining parameters to have a fixed value is better, if possible, than constraining them to be equal. For example, the pair of constraints

```
. constraint define 1 phi11 = 1
. constraint define 2 phi21 = phi11
```

should be equivalent to the constraints

```
. constraint define 1 phi11 = 1
. constraint define 2 phi21 = 1
```

However, there can be circumstances in which the model will not converge using the former constraints but will with the latter. This is a weakness of the constraint handling code within `soreg`.

(*Continued on next page*)

Saved results

soreg saves in e():

Scalars

e(N)	number of observations
e(ll_0)	log likelihood, constant only model
e(ll)	log likelihood, stereotype model
e(ll_full)	log likelihood, full model
e(df_m)	model degrees of freedom
e(df_full)	model degrees of freedom for multinomial model
e(k_cat)	number of categories

Macros

e(cmd)	soreg
e(chi2type)	type of model χ^2 test (LR)
e(depvar)	name of dependent variable

Matrices

e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators
e(cat)	category values

Methods and formulas

Since the stereotype model is nonlinear (it contains the product of parameters), it cannot be fitted by standard generalized linear model methods. The model is fitted using a technique due to Box and Tidwell (1962), described by McCullagh and Nelder (1989). Briefly, a linear model containing a nonlinear function, say

$$\eta = \alpha + \beta g(x|\theta)$$

can be estimated by iteratively fitting instead

$$\eta = \alpha + \beta u_i + \gamma v_i$$

where

$$u_i = g(x|\theta_{i-1})$$
$$v_i = \left[\frac{\partial g}{\partial \theta}\right]_{\theta=\theta_{i-1}}$$

and θ_{i-1} is the value of θ calculated from the $(i-1)$th iteration. Then $\theta_i = \theta_{i-1} + \widehat{\gamma}/\widehat{\beta}$, and the process iterates to convergence. A final iteration with $v = \widehat{\beta}v$ gives the variance of θ and its covariances with α and β. Unfortunately, this iteration process is not guaranteed to converge. Since the first variable in the list has a special role in this process, it can happen that a model that will not converge can be forced to by changing the order of the predictor variables.

Further remarks

There already exist two Stata commands, mclgen and mclest which can be used to estimate one-dimensional stereotype ordinal regression models. However, soreg has a number of advantages:

- soreg can estimate models with more than one dimension, which mclest cannot. This was the motivation for writing soreg: I needed to fit a two-dimensional model.
- soreg allows the use of constraints, which mclest does not.
- mclest does not produce standard errors for the ϕ_{ij} parameters, which are important when considering distinguishability.
- The standard errors for the β_{ij} parameters produced by mclest are conditional on the ϕ_{ij} parameters and hence underestimated. The standard errors produced by soreg allow for the uncertainty in estimating the ϕ_{ij} parameters.

It should be pointed out that while there are models that can be estimated with soreg but not mclest, there are also models that can be estimated with mclest but not soreg.

Care needs to be taken in interpreting the results of significance tests on stereotype ordinal regression models. Unlike generalized linear models, it cannot be shown that the log likelihood follows a χ^2 distribution asymptotically. Hence the likelihood ratio chi-squared tests and corresponding p-values should be treated with care.

In addition, it should be remembered that when considering the distinguishability of k groups, there are implicitly $k-1$ tests being performed. This multiple testing should be borne in mind in interpreting the resultant model. However, this is true in any case in which data-based variable selection takes place.

Finally, the standard errors reported by `soreg` differ, in some cases, from those reported by Anderson in his paper. However, the differences are small. In addition, in some cases the standard errors reported by Anderson are smaller than those reported by `mclest`, despite the fact that those reported by `mclest` are conditional on the estimated values of the phi parameters and therefore should be underestimated. The standard errors reported by `soreg` are always greater than those reported by `mclest` and are similar to those reported by Anderson. Where it is possible to constrain a multinomial model to produce the same parameterization as a stereotype model, the standard errors are the same by both methods.

However, the significance of an individual parameter reported by the Wald test (i.e., using the standard error) may be quite different from that reported by the likelihood-ratio test. This is due to the correlations between estimates of the ϕ parameters and the β parameters. The likelihood-ratio tests are likely to be a more reliable guide to inference. For example, the overall test of the model in the first nausea example is highly significant, but only one of the individual parameters has a p-value of less than 0.05 (`beta11` $= 0.04$).

Acknowledgment

`soreg` owes a great deal to `mclest`, by John Hendrickx (J.Hendrickx@mailbox.kun.nl).

References

Anderson, J. A. 1984. Regression and ordered categorical variables. *Journal of the Royal Statistical Society, Series B* 46: 1–30.

Ashford, J. R. 1959. An approach to the analysis of data for semi-quantal responses in biological assay. *Biometrics* 26: 535–546.

Box, G. E. P. and P. W. Tidwell. 1962. Transformation of the independent variables. *Technometrics* 4: 531–550.

Doran, D. M. L. and D. J. Newell. 1975. Manipulation in treatment of low back pain: A multicentre trial. *British Medical Journal* 1: 161–164.

Farewell, V. T. 1982. A note on regression analysis of ordinal data with variability of classification. *Biometrika* 69: 533–538.

Greenland, S. 1994. Alternative models for ordinal logistic regression. *Statistics in Medicine* 13: 1665–1677.

McCullagh, P. and J. Nelder. 1989. *Generalized Linear Models*. London: Chapman & Hall.

sg164	Specification tests for linear panel data models

Walter Sosa-Escudero, National University of La Plata, Argentina, wsosa@feedback.net.ar
Anil K. Bera, Department of Economics, University of Illinois

Abstract: `xttest1` computes seven specification tests for balanced error component models. It is an extension of `xttest0` and it is used exactly in the same way except that panels must be balanced.

Keywords: specification test, panel data

Syntax

`xttest1` is used after estimating a random-effects model with `xtreg, re`, and presents specification tests for balanced error component models, all of them based solely on OLS residuals. It includes the Breusch and Pagan (1980) Lagrange multiplier test for random effects, the Baltagi and Li (1995) test for first order serial correlation, the Baltagi and Li (1991) joint test for serial correlation and random effects, and the family of adjusted tests in Bera, Sosa-Escudero, and Yoon (2001).

Description

Consider a simple one-way error component model which allows for possible random individual effects and first-order autocorrelation in the disturbance term

$$
\begin{aligned}
y_{it} &= x'_{it}\beta + u_{it}, \qquad i = 1, 2, \ldots, N, t = 1, 2, \ldots, T \\
u_{it} &= \mu_i + \nu_{it} \\
\nu_{it} &= \rho\nu_{i,t-1} + \epsilon_{it}, \qquad |\rho| < 1
\end{aligned}
$$

where β is a $(k \times 1)$ vector of parameters including the intercept, $\mu_i \sim \text{IIDN}(0, \sigma_\mu^2)$ is a random individual component, and $\epsilon_{it} \sim \text{IIDN}(0, \sigma_\epsilon^2)$. The μ_i and ν_{it} are assumed to be independent of each other with $\nu_{i,0} \sim N(0, \sigma_\epsilon^2/(1-\rho^2))$. N and T denote the number of individual units and the number of time periods, respectively.

Researchers are typically interested in testing the nulls of no random effects ($H_0 : \sigma_\epsilon^2 = 0$) and/or no first-order serial correlation ($H_0 : \rho = 0$). The standard Breusch and Pagan (1980) statistic is used to test the null of no random effects, assuming that there is no serial correlation. Similarly, the statistic derived by Baltagi and Li (1995) tests the null of no serial correlation, assuming no random effects.

Recently, Bera, Sosa-Escudero, and Yoon (2001, BSY hereafter) showed that the presence of first-order serial correlation makes the Breusch and Pagan (1980) test reject the null of no random effects even when it is correct. They propose an adjusted version which is not affected by the presence of serial correlation. A similar adjusted version is derived by BSY for the Baltagi and Li (1995) test for serial correlation, which is invalid under the presence of random effects.

Baltagi and Li (1991) propose a simple test for the joint null of no serial correlation and random effects. Recognizing the one-sided nature of the problem of testing for random effects, Honda (1985) proposes a one-sided version of the Breusch-Pagan test which is also invalid in the presence of first order serial correlation. BSY propose a corrected version of this one-sided test.

Expressions of the test statistics

Let I_N be an identity matrix of dimension N, e_T a vector of ones of dimension T, let

$$u' = (u_{11}, \ldots, u_{1T}, \ldots, u_{N1}, \ldots, u_{NT})$$

and u_{-1} an $(NT \times 1)$ vector containing $u_{i,t-1}$. Define A and B as in Baltagi and Li (1991):

$$A = 1 - \frac{\widetilde{u}'(I_N \otimes e_T e_T')\widetilde{u}}{\widetilde{u}'\widetilde{u}}$$

and

$$B = \frac{\widetilde{u}'\widetilde{u}_{-1}}{\widetilde{u}'\widetilde{u}}.$$

where $\widetilde{u}$ are the OLS residuals from the standard linear model $y_{it} = x_{it}'\beta + u_{it}$ without the random effects and serial correlation.

The LM test, also known as the Rao (1948) score test for random effects is given in Breusch and Pagan (1980):

$$\texttt{LM(Var(u)=0)} = \frac{NTA^2}{2(T-1)}$$

and the adjusted version in BSY (2001) is

$$\texttt{ALM(Var(u)=0)} = \frac{NT(A+2B)^2}{2(T-1)(1-\frac{2}{T})}$$

The one-sided versions of the previous tests are given by

$$\texttt{LMO(Var(u)=0)} = -\sqrt{\frac{NT}{2(T-1)}}A$$

and

$$\texttt{ALMO(Var(u)=0)} = -\sqrt{\frac{NT}{2(T-1)(1-\frac{2}{T})}}(A-2B)$$

The LM statistic to test the null of no serial correlation assuming no random effects is given in Baltagi and Li (1991):

$$\texttt{LM(rho=0)} = \frac{NT^2B^2}{T-1}$$

and the adjusted version by Bera, et al. (2000), valid under random effects, is

$$\texttt{ALM(rho=0)} = \frac{NT^2(B+\frac{A}{T})^2}{(T-1)(1-\frac{2}{T})}.$$

Baltagi and Li (1991, 1995) derived a joint LM test for serial correlation and random individual effects which is given by

$$\texttt{LM(Var(u)=0,rho=0)} = \frac{NT^2}{2(T-1)(T-2)}[A^2+4AB+2TB^2]$$

It is interesting to note that this joint test statistic is related to the one-directional adjusted and unadjusted tests as follows:

```
LM(Var(u)=0,rho=0) = ALM(Var(u)=0) + LM(rho=0) = LM(Var(u)=0) + ALM(rho=0)
```

which implies that the adjusted tests could be computed as

```
ALM(Var(u)=0) = LM(Var(u)=0, rho=0) - LM(rho=0)
ALM(rho=0) = LM(Var(u)=0, rho=0) - LM(Var(u)=0)
```

Example

This example illustrates the use of **xttest1** and the interpretation of the statistics computed, and it is taken from BSY(2001). It is based on the well-known Grunfeld (1958) investment dataset for five US manufacturing firms measured over 20 years which is frequently used to illustrate panel issues. It has been used in the illustration of misspecification tests in the error-component model in Baltagi, et al. (1992), and in recent books such as those by Baltagi (1995, 20) and Greene (2000, 592). The equation to be estimated is a panel model of firm investment using the real value of the firm and the real value of capital stock as explanatory variables:

$$I_{it} = \beta_0 + \beta_1 F_{it} + \beta_2 C_{it} + u_{it}$$

where I_{it} denotes real gross investment for firm i in period t, F_{it} is the real value of the firm and C_{it} is the real value of the capital stock, $i = 1, 2, \ldots, 5$, and $t = 1, 2, \ldots, 20$.

First, we estimate the parameters of a one-way error component model with random effects using **xtreg**:

```
. xtreg i f c, i(firm)

Random-effects GLS regression                   Number of obs      =       100
Group variable (i) : firm                       Number of groups   =         5

R-sq:  within  = 0.8003                         Obs per group: min =        20
       between = 0.7696                                        avg =      20.0
       overall = 0.7781                                        max =        20

Random effects u_i ~ Gaussian                   Wald chi2(2)       =    384.93
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       i |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       f |   .1048856   .0147972      7.088   0.000       .0758835    .1338876
       c |   .3460156   .0242535     14.267   0.000       .2984796    .3935517
   _cons |  -60.29048   54.48389     -1.107   0.268      -167.0769    46.49599
---------+--------------------------------------------------------------------
 sigma_u |   104.6527
 sigma_e |  69.117979
     rho |  .69628405   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Then, the **xttest1** command computes the seven tests described before: the Breusch and Pagan test for random effects (**LM(Var(u)=0)**), the BSY adjusted version (**ALM(Var(u)=0)**), the corresponding one sided versions (**LMO(Var(u)=0)** and **ALMO(Var(u)=0)**), the Baltagi and Li serial correlation test (**LM(rho=0)**), the corresponding adjusted version (**ALM(rho=0)**), the Baltagi and Li joint test for serial correlation and random effects (**LM(Var(u)=0,rho=0)**), the Honda one-sided test for random effects (**LMO(Var(u)=0)**) and the adjusted version (**ALMO(Var(u)=0)**). The output of **xttest1** is as follows:

```
.xttest1
Tests for the error component model:
        i[firm,t] = Xb + u[firm] + v[firm,t]
           v[firm,t] = rho v[firm,(t-1)] + e[firm,t]
        Estimated results:
                                 Var     sd = sqrt(Var)
                 ---------+-----------------------------
                        i |    71751.9       267.8654
                        e |   4777.295      69.117979
                        u |   10952.19      104.6527
        Tests:
           Random Effects, Two Sided:
           LM(Var(u)=0)        =  453.82   Pr>chi2(1) =  0.0000
           ALM(Var(u)=0)       =  384.18   Pr>chi2(1) =  0.0000
           Random Effects, One Sided:
           LMO(Var(u)=0)       =   21.30   Pr>N(0,1)  =  0.0000
           ALMO(Var(u)=0)      =   19.60   Pr>N(0,1)  =  0.0000
           Serial Correlation:
           LM(rho=0)           =   73.35   Pr>chi2(1) =  0.0000
           ALM(rho=0)          =    3.71   Pr>chi2(1) =  0.0540
           Joint Test:
           LM(Var(u)=0,rho=0)  =  457.53   Pr>chi2(2) =  0.0000
```

The unadjusted tests for serial correlation (`LM(rho=0)`) and for random effects (`LM(Var(u)=0)` and `LMO(Var(u)=0)`) reject the respective null hypothesis of no serial correlation and no random effects, and the omnibus test (`LM(Var(u)=0,rho=0)`) rejects the joint null. But the adjusted tests suggest that in this example the problem seems to be the presence of random effects rather serial correlation. The adjusted versions of the random effect tests (`ALM(Var(u)=0)` and `ALMO(Var(u)=0)`) also reject the null but the adjusted serial correlation test (`ALM(rho=0)`) barely fails to reject the null at the 5% significance level. It is interesting to note the substantial reduction of the autocorrelation test statistic, from 73.35 to 3.71. So in this example the misspecification can be thought to come from the presence of random effects rather than serial correlation.

In spite of the small sample size of the datasets, this example seems to illustrate clearly the usefulness of BSY tests: the adjusted versions are more informative than a test for serial correlation or random effect that ignores the presence of the other effect. In this case, the presence of a random effect seems to spuriously induce rejection of the no-serial correlation hypothesis. The joint test (`LM(Var(u)=0,rho=0)`) rejects the joint null but is not informative about the direction of the misspecification.

Saved results

`xttest1` saves in global macros `S_1`, `S_2`, . . ., `S_7` the test statistics in the following order: `LM(Var(u)=0)`, `ALM(Var(u)=0)`, `LMO(Var(u)=0)`, `ALMO(Var(u)=0)`, `LM(rho=0)`, `ALM(rho=0)`, and `LM(Var(u)=0,rho=0)`.

References

Baltagi, B. 1995. *Econometric Analysis of Panel Data.* New York: John Wiley & Sons.

Baltagi, B., Y. Chang, and Q. Li. 1992. Monte Carlo results on several new and existing tests for the error component model. *Journal of Econometrics* 54: 95–120.

Baltagi, B. and Q. Li. 1991. A joint test for serial correlation and random individual effects. *Statistics & Probability Letters* 11: 277–280.

——. 1995. Testing AR(1) against MA(1) disturbances in an error component model. *Journal of Econometrics* 68: 133–151.

Bera, A., W. Sosa-Escudero, and M. Yoon. 2001. Tests for the error component model in the presence of local misspecification. *Journal of Econometrics* 101: 1–23.

Breusch, T. S. and A. R. Pagan. 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47: 239–253.

Greene, W. 2000. *Econometric Analysis*. 4th ed. New Jersey: Prentice–Hall.

Grunfeld, Y. 1958. The Determinants of Corporate Investment. Unpublished Ph.D thesis. Department of Economics, University of Chicago.

Honda, Y. 1985. Testing the error component model with non-normal disturbances. *Review of Economic Studies* 52: 681–690.

Rao, C. R. 1948. Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* 44: 50–57.

snp15.3	Update to somersd

Roger Newson, Guy's, King's and St Thomas' School of Medicine, London, UK, roger.newson@kcl.ac.uk

Abstract: `somersd` calculates confidence intervals for rank order statistics. It has been updated for compatibility with Stata 7.

Keywords: Somers' D; Kendall's tau; rank correlation; confidence intervals; nonparametric methods.

`somersd` was introduced in Newson (2000a) and updated in Newson (2000b) and Newson (2000c). The program calculates confidence intervals for the rank order statistics Somers' D, Kendall's τ_a and Greiner's ρ, saving estimates and covariances as estimation results. The latest version is still in Stata 6, but has been updated for compatibility with Stata 7. (The previous version sometimes failed under Stata 7 owing to an obscure problem with nested `if` statements.) Also, the output has been clarified, and the program has been extensively certified.

References

Newson, R. 2000a. snp15: somersd—Confidence limits for nonparametric statistics and their differences. *Stata Technical Bulletin* 55: 47–55.

——. 2000b. snp15.1: Update to somersd. *Stata Technical Bulletin* 57: 35.

——. 2000c. snp15.2: Update to somersd. *Stata Technical Bulletin* 58: 30.

snp16.1	Update to cendif

Roger Newson, Guy's, King's and St Thomas' School of Medicine, London, UK, roger.newson@kcl.ac.uk

Abstract: `cendif` calculates robust confidence intervals for median (and other percentile) differences (or ratios) between two groups. It has been updated for compatibility with Stata 7.

Keywords: robust, confidence interval, median, percentile, difference, ratio, rank-sum, Wilcoxon, Hodges–Lehmann.

`cendif` was introduced in Newson (2000). The program calculates robust confidence intervals for median (and other percentile) differences (or ratios) between two groups, including the Hodges–Lehmann median difference (Hodges and Lehmann 1963). `cendif` requires `somersd` (Newson 2001) in order to work. The latest version is still in Stata 6, but has been updated for compatibility with Stata 7, as the previous version sometimes failed under Stata 7 owing to a similar problem to the old version of `somersd` (Newson 2001). `cendif` has been extensively certified.

References

Hodges, J. R. and E. L. Lehmann. 1963. Estimates of location based on rank tests. *Annals of Mathematical Statistics* 34: 598–611.

Newson, R. 2000. snp16: Robust confidence intervals for median and other percentile differences between two groups. *Stata Technical Bulletin* 58: 30–35.

——. 2001. snp15.3: Update to `somersd`. *Stata Technical Bulletin* 61: 22.

zz11	Cumulative index for STB-55–STB-61

[an] Announcements

[dm] Data Management

[gr] Graphics

[ip] Instruction on Programming

[sbe] Biostatistics & Epidemiology

[sg] General Statistics

[snp] Nonparametric Methods

[ssa] Survival Analysis

[sts] Time-series, Econometrics

[sxd] Experimental Design

[zz] Not elsewhere classified

STB categories and insert codes

Inserts in the STB are presently categorized as follows:

General Categories:

an	announcements
cc	communications & letters
dm	data management
dt	datasets
gr	graphics
in	instruction
ip	instruction on programming
os	operating system, hardware, & interprogram communication
qs	questions and suggestions
tt	teaching
zz	not elsewhere classified

Statistical Categories:

sbe	biostatistics & epidemiology
sed	exploratory data analysis
sg	general statistics
smv	multivariate analysis
snp	nonparametric methods
sqc	quality control
sqv	analysis of qualitative variables
srd	robust methods & statistical diagnostics
ssa	survival analysis
ssi	simulation & random numbers
sss	social science & psychometrics
sts	time-series, econometrics
svy	survey sampling
sxd	experimental design
szz	not elsewhere classified

In addition, we have granted one other prefix, *stata*, to the manufacturers of Stata for their exclusive use.

Guidelines for authors

We welcome submissions to the new Stata Journal. Guidelines for authors may be found at http://www.stata-journal.com.

International Stata Distributors

International Stata users may also order subscriptions to the *Stata Technical Bulletin* from our International Stata Distributors.

Company:	Applied Statistics & Systems Consultants Ltd.	Countries served:	Israel
		Phone:	+972 (0)4 6100101
Address:	P.O. Box 5	Fax:	+972 (0)4 6100101
	38900 CAESAREA, Israel	Email:	stata@watchwise.net
Company:	Axon Technology Company Ltd	Countries served:	Taiwan
Address:	9F, No. 259, Sec. 2	Phone:	+886-(0)2-27045535
	Ho-Ping East Road	Fax:	+886-(0)2-27541785
	TAIPEI 106, Taiwan	Email:	hank@axon.axon.com.tw
Company:	Chips Electronics	Countries served:	Indonesia, Brunei, Malaysia, Singapore
Address:	Kelapa Puyuh IV KB 23	Phone / Fax:	62 - 21 - 452 17 61
	Kelapa Gading Permai	Mobile Phone:	62 - 81 - 884 95 84
	JAKARTA 14240, Indonesia	Email:	puyuh23@indo.net.id
Company:	Dittrich & Partner Consulting	Countries served:	Germany, Austria, Czech Republic, Hungary, Poland
Address:	Kieler Straße 17		
	5. floor	Phone:	+49 2 12 / 26 066 - 0
	D-42697 Solingen, Germany	Fax:	+49 2 12 / 26 066 - 66
URL:	http://www.dpc.de	Email:	sales@dpc.de
Company:	IEM	Countries served:	South Africa, Botswana, Lesotho, Namibia, Mozambique, Swaziland, Zimbabwe
Address:	P.O. Box 2222		
	PRIMROSE 1416, South Africa		
		Phone:	+27-11-8286169
		Fax:	+27-11-8221377
		Email:	iem@hot.co.za
Company:	JasonTech Inc.	Countries served:	Korea
Address:	181-3 Hansang B/D, Bangyidong	Phone:	+82-(0)2-420-6700
	Songpaku	Fax:	+82-(0)2-420-8600
	Seoul 138-050, Korea	Email:	info@jat.co.kr
Company:	MercoStat Consultores	Countries served:	Uruguay, Argentina, Brazil, Paraguay
Address:	9 de junio 1389	Phone:	598-2-613-7905
	CP 11400 MONTEVIDEO, Uruguay	Fax:	598-2-613-7905
		Email:	mercost@adinet.com.uy
Company:	Metrika Consulting	Countries served:	Sweden, Baltic States, Denmark, Finland, Iceland, Norway
Address:	Mosstorpsvagen 48		
	183 30 Taby STOCKHOLM, Sweden	Phone:	+46-708-163128
URL:	http://www.metrika.se	Fax:	+46-8-7924747
		Email:	sales@metrika.se
Company:	MultiON Consulting, SA de CV	Countries served:	Mexico
Address:	Insurgentes Sur 1236-301	Phone:	52 (5) 559-4050 Ext 190
	Mexico, DF, 03200, Mexico	Fax:	52 (5) 559-4048
		Email:	multion@multion.com.mx

(*List continued on next page*)

International Stata Distributors

(*Continued from previous page*)

Company:	Ritme Informatique	Countries served:	France, Belgium, Luxembourg
Address:	34, boulevard Haussmann	Phone:	+33 (0)1 42 46 00 42
	75009 Paris, France	Fax:	+33 (0)1 42 46 00 33
URL:	http://www.ritme.com	Email:	info@ritme.com
Company:	Scientific Solutions S.A.	Countries served:	Switzerland
Address:	Avenue du Général Guisan, 5	Phone:	41 (0)21 711 15 20
	CH-1009 Pully/Lausanne, Switzerland	Fax:	41 (0)21 711 15 21
		Email:	info@scientific-solutions.ch
Company:	Smit Consult	Countries served:	Netherlands
Address:	Doormanstraat 19	Phone:	+31 416-378 125
	5151 GM Drunen, Netherlands	Fax:	+31 416-378 385
URL:	http://www.smitconsult.nl	Email:	info@smitconsult.nl
Company:	Survey Design & Analysis Services Pty Ltd	Countries served:	Australia, New Zealand
Address:	PO Box 1206	Phone:	+61 (0)3 9878 7373
	Blackburn North VIC 3130, Australia	Fax:	+61 (0)3 9878 2345
URL:	http://survey-design.com.au	Email:	sales@survey-design.com.au
Company:	Timberlake Consultants	Countries served:	United Kingdom, Eire
Address:	Unit B3 Broomsleigh Bus. Park	Phone:	+44 (0)208 697 3377
	Worsley Bridge Road	Fax:	+44 (0)208 697 3388
	LONDON SE26 5BN, United Kingdom	Email:	info@timberlake.co.uk
URL:	http://www.timberlake.co.uk		
Company:	Timberlake Consultants Srl	Countries served:	Italy
Address:	Via Baden Powell, 8	Phone:	+39 (0)864 210101
	67039 SULMONA (AQ), Italy	Fax:	+39 (0)864 206014
URL:	http://www.timberlake.it	Email:	timberlake@arc.it
Company:	Timberlake Consulting S.L.	Countries served:	Spain
Address:	Calle Mendez Nunez, 1, 3	Phone:	+34 95 421 9369
	41011 Sevilla, Spain	Fax:	+34 95 422 0648
		Email:	timberlake@zoom.es
Company:	Timberlake Consultores, Lda.	Countries served:	Portugal
Address:	Praceta Raúl Brandao, n°1, 1°E	Phone:	351 (0)1 471 73 47
	2720 ALFRAGIDE, Portugal	Fax:	+351 (0)1 471 73 47
		Email:	timberlake.co@mail.telepac.pt
Company:	Vishvas Marketing-Mix Services	Countries served:	India
Address:	C\O S. D. Wamorkar	Phone:	+91-251-440087
	"Prashant" Vishnu Nagar, Naupada	Fax:	+91-22-5378552
	THANE - 400602, India	Email:	vishvas@vsnl.com