The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*, *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$ 98	1-year subscription	$138
2-year subscription	$165	2-year subscription	$245
3-year subscription	$225	3-year subscription	$345
1-year student subscription	$ 75	1-year student subscription	$ 99
1-year university library subscription	$125	1-year university library subscription	$165
2-year university library subscription	$215	2-year university library subscription	$295
3-year university library subscription	$315	3-year university library subscription	$435
1-year institutional subscription	$245	1-year institutional subscription	$285
2-year institutional subscription	$445	2-year institutional subscription	$525
3-year institutional subscription	$645	3-year institutional subscription	$765
Electronic only		**Electronic only**	
1-year subscription	$ 75	1-year subscription	$ 75
2-year subscription	$125	2-year subscription	$125
3-year subscription	$165	3-year subscription	$165
1-year student subscription	$ 45	1-year student subscription	$ 45

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The Stata Journal (2013)
13, Number 2, pp. 401–405

Stata tip 115: How to properly estimate the multinomial probit model with heteroskedastic errors

Michael Herrmann
Department of Politics and Public Administration
University of Konstanz
Konstanz, Germany
michael.herrmann@uni-konstanz.de

Models for multinomial outcomes are frequently used to analyze individual decision making in consumer research, labor market research, voting, and other areas. The multinomial probit model provides a flexible approach to analyzing decisions in these fields because it does not impose some of the restrictive assumptions inherent in the often used conditional logit approach. In particular, multinomial probit relaxes 1) the assumption of independent error terms, allowing for correlation in individual choices across alternatives, and 2) it does not impose the assumption of identically distributed errors, allowing unobserved factors to affect the choice of some alternatives more strongly than others (that is, heteroskedasticity).

By default, `asmprobit` relaxes both the assumptions of independence and homoskedasticity. To avoid overfitting, however, the researcher may sometimes wish to relax these assumptions one at a time.[1] A seemingly straightforward solution would be to rely on the options `stddev()` and `correlation()`, which allow the user to set the structure for the error variances and their covariances, respectively (see [R] **asmprobit**).

When doing so, however, the user should be aware that specifying `std(het)` and `corr(ind)` does not actually fit a pure heteroskedastic multinomial probit model. With J outcome categories, if errors are independent, $J-1$ error variances are identified (see below). Instead, Stata estimates $J-2$ error variances and, hence, imposes an additional constraint, which causes the model to be overidentified. As a result, the estimated model is not invariant to the choice of base and scale outcomes; that is, changing the base or scale outcome leads to different values of the likelihood function.

To properly estimate a pure heteroskedastic model, the user needs to define the structure of the error variances manually. This is easy to accomplish using the `pattern` or `fixed` option. The following example illustrates the problem and shows how to estimate the model correctly.

1. Another reason to relax them one at a time is that heteroskedasticity and error correlation cannot be distinguished from each other in the default specification. That is, one cannot simply look at the estimated covariance matrix of the errors and see whether the errors are heteroskedastic, correlated, or both. What Stata estimates is the normalized covariance matrix of error differences whose elements do not allow one to draw any conclusions on the covariance structure of the errors themselves.

 st0302

Consider an individual's choice of travel mode with the alternatives being air, train, bus, and car and predictor variables, including general cost of travel, terminal time, household income, and traveling group size. One might suspect the choice of some alternatives to be driven more by unobserved factors than the choice of others. For example, there might be more unobserved reasons related to an individual's decision to travel by plane than by train, bus, or car. Allowing the error variances associated with the alternatives to differ, we fit the following model:

```
. use http://www.stata-press.com/data/r12/travel

. asmprobit choice travelcost termtime, casevars(income partysize)
> case(id) alternatives(mode) std(het) corr(ind) nolog

Alternative-specific multinomial probit       Number of obs      =       840
Case variable: id                             Number of cases    =       210

Alternative variable: mode                    Alts per case: min =         4
                                                             avg =       4.0
                                                             max =         4

Integration sequence:      Hammersley
Integration points:               200            Wald chi2(8)    =     71.57
Log simulated-likelihood = -181.81521            Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
mode         |
  travelcost |   -.012028   .0030838    -3.90   0.000    -.0180723   -.0059838
    termtime |   -.050713   .0071117    -7.13   0.000    -.0646517   -.0367743
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |    -.03859   .0093287    -4.14   0.000    -.0568739   -.0203062
   partysize |   .7590228    .190438     3.99   0.000     .3857711    1.132274
       _cons |  -.9960951   .4750053    -2.10   0.036    -1.927088   -.0651019
-------------+----------------------------------------------------------------
bus          |
      income |  -.0119789   .0081057    -1.48   0.139    -.0278658     .003908
   partysize |   .5876645   .1751734     3.35   0.001     .2443309     .930998
       _cons |  -1.629348   .4803384    -3.39   0.001    -2.570794   -.6879016
-------------+----------------------------------------------------------------
car          |
      income |   -.004147   .0078971    -0.53   0.599     -.019625     .011331
   partysize |   .5737318    .163719     3.50   0.000     .2528485    .8946151
       _cons |  -3.903084    .750675    -5.20   0.000     -5.37438   -2.431788
-------------+----------------------------------------------------------------
 /lnsigmaP1  |  -1.097572   .7967201    -1.38   0.168    -2.659115    .4639704
 /lnsigmaP2  |  -.3906271   .3468426    -1.13   0.260    -1.070426    .2891719
-------------+----------------------------------------------------------------
      sigma1 |          1  (base alternative)
      sigma2 |          1  (scale alternative)
      sigma3 |   .3336802   .2658497                      .0700102    1.590376
      sigma4 |   .6766324   .2346849                      .3428624    1.335321
------------------------------------------------------------------------------
(mode=air is the alternative normalizing location)
(mode=train is the alternative normalizing scale)
```

As can be seen, two of the four error variances are set to one. These are the base and scale alternatives. While choosing a base and scale alternative is necessary to identify the model, the problem here is that because errors are uncorrelated, fixing the variance of the base alternative is not necessary to identify the model. As a result, an additional constraint is imposed, which leads to a different model structure depending on the choice of base and scale alternatives. For example, changing the base alternative to car produces a different log likelihood:

```
. quietly asmprobit choice travelcost termtime, casevars(income partysize)
> case(id) alternatives(mode) std(het) corr(ind) nolog base(4)
. display e(ll)
-181.58795
```

To properly estimate an unconstrained heteroskedastic model, one needs to define a vector of variance terms in which one element (the scale alternative) is fixed and pass this vector on to the estimation command. For example, to set the error variance of the second alternative to unity, define a vector of missing values, `stdpat`, whose second element is 1, and then call this vector from inside `asmprobit` using the option `std(fixed)` (see [R] **asmprobit** for details):

```
. matrix define stdpat = (.,1,.,.)

. asmprobit choice travelcost termtime, casevars(income partysize)
> case(id) alternatives(mode) std(fixed stdpat) corr(ind) nolog base(1)

Alternative-specific multinomial probit        Number of obs      =       840
Case variable: id                              Number of cases    =       210

Alternative variable: mode                     Alts per case: min =         4
                                                              avg =       4.0
                                                              max =         4

Integration sequence:      Hammersley
Integration points:               200             Wald chi2(8)    =     26.84
Log simulated-likelihood = -180.01839             Prob > chi2     =    0.0008

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
mode         |
  travelcost |  -.0196389   .0067143    -2.92   0.003    -.0327988   -.006479
    termtime |  -.0664153   .0140353    -4.73   0.000     -.093924   -.0389065
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
train        |
      income |  -.0498732   .0154884    -3.22   0.001      -.08023   -.0195165
   partysize |   1.126922   .3651321     3.09   0.002     .4112761    1.842568
       _cons |  -1.072849    .680711    -1.58   0.115    -2.407018    .2613198
-------------+----------------------------------------------------------------
bus          |
      income |  -.0210642   .0139892    -1.51   0.132    -.0484826    .0063542
   partysize |   .8678651   .3179559     2.73   0.006      .244683    1.491047
       _cons |  -1.831363   .7345686    -2.49   0.013    -3.271091   -.3916349
-------------+----------------------------------------------------------------
car          |
      income |   -.010205   .0131711    -0.77   0.438    -.0360199     .01561
   partysize |   .8708577   .3202671     2.72   0.007     .2431458    1.49857
       _cons |  -4.971594   1.261002    -3.94   0.000    -7.443112   -2.500075
-------------+----------------------------------------------------------------
  /lnsigmaP1 |    .558377   .3076004     1.82   0.069    -.0445087    1.161263
  /lnsigmaP2 |    -1.0078   1.116358    -0.90   0.367    -3.195822    1.180223
  /lnsigmaP3 |  -.0158072   .3593511    -0.04   0.965    -.7201225    .6885081
-------------+----------------------------------------------------------------
      sigma1 |   1.747833   .5376342                      .9564673    3.193964
      sigma2 |          1  (scale alternative)
      sigma3 |   .3650213   .4074946                      .0409329    3.255099
      sigma4 |   .9843171   .3537155                      .4866926    1.990743
------------------------------------------------------------------------------
(mode=air is the alternative normalizing location)
(mode=train is the alternative normalizing scale)
```

Now the model is properly normalized, and the user may verify that changing either the scale alternative (that is, changing the location of the `1` in `stdpat`) or the base alternative leaves results unchanged. Note that while, in theory, the only restriction necessary to identify the heteroskedastic probit model is to fix one of the variance terms, in the Stata implementation of the model, the base and scale outcomes must be different. That is, Stata does not allow the same alternative to be the base outcome and the scale outcome. However, this is more of an inconvenience than a restriction: such a model would be equivalent to one in which the base and scale outcomes differed.

Finally, to show that independence of errors indeed implies $J-1$ estimable error variances, we must verify that the error variances can be calculated directly from the variance and covariance parameters of the normalized error differences. Only the latter are identified and, hence, estimable (Train 2009). Suppose, without loss of generality, $J=3$, and let $j=1$ be the base outcome.

Following the normalization approach advocated by Train (2009, 100f.), the normalized covariance matrix of error differences is given by

$$\widetilde{\Omega}_1^* = \begin{pmatrix} 1 & \theta_{23}^* \\ & \theta_{33}^* \end{pmatrix}$$

with elements θ^* relating to the actual error variances σ_{jj} and covariances σ_{ij} as follows:

$$\begin{aligned} \theta_{23}^* &= \frac{\sigma_{23}+\sigma_{11}-\sigma_{12}-\sigma_{13}}{\sigma_{22}+\sigma_{11}-2\sigma_{12}} \\ \theta_{33}^* &= \frac{\sigma_{33}+\sigma_{11}-2\sigma_{13}}{\sigma_{22}+\sigma_{11}-2\sigma_{12}} \end{aligned}$$

Under independence, $\sigma_{ij}=0$. Fixing $\sigma_{22}=1$ (that is, choosing $j=2$ as the scale outcome) yields $\theta_{23}^*=\sigma_{11}/(1+\sigma_{11})$ and $\theta_{33}^*=(\sigma_{33}+\sigma_{11})/(1+\sigma_{11})$. Obviously, σ_{11} can be calculated from θ_{23}^*, and subsequent substitution produces σ_{33} from θ_{33}^*. The same is true if we choose to fix either σ_{11} or σ_{33} because in each case, we would obtain two equations in two unknowns. Similar conclusions follow when there are four or more outcome categories. Thus, with independent errors, $J-1$ variance parameters are estimable.

Reference

Train, K. E. 2009. *Discrete Choice Methods with Simulation.* 2nd ed. Cambridge: Cambridge University Press.