Tests for the Multinomial Logit Model[1]

Jeremy Freese and J. Scott Long
University of Wisconsin-Madison and Indiana University

June 6, 2000 Draft (sl)

Overview

There are several tests that are commonly used in association with the multinomial logit model (MNLM hereafter). First, we can test that all of the coefficients associated with an independent variable are simultaneously equal to zero (i.e., test that a variable has no effect). Second, we can test whether the independent variables differentiate between two outcomes; this test is commonly used to determine if two outcomes can be combined. Third, we can assess the assumption of the independence of irrelevant alternatives (IIA) using either a Hausman test or the LR test proposed by McFadden et. al. (1976) and improved by Small and Hsiao (1985). While each of these tests can be computed using either the `test`, `lrtest`, or `hausman` commands in Stata or the `smhsiao` command by Nick Winter (available at the SSC-IDEAS archive), in practice computing these tests can be awkward and/or tedious. The `mlogtest` command is designed to simplify the use of these tests. `mlogtest` is a post-estimation command that requires that `mlogit` is the last model estimated.

A note on specification searches: Given the difficulties of interpretation that are associated with the MNLM, it is tempting to search for a more parsimonious model constructed by excluding variables or combining outcome categories based on a series of statistical tests. While `mlogtest` facilitates computing tests that can be used in a specification search, great care is required. First, these tests all involve multiple coefficients. While the overall test might indicate that *as a group* the parameters are not significantly different from zero, an *individual* parameter can still be substantively and statistically significant. Accordingly, you need to carefully examine the individual coefficients involved in each test before deciding to revise your model. Second, as with all searches that use repeated, sequential tests, there is a danger of over-fitting the data. When models are constructed based on prior testing using the same data, significance levels should only be used as rough guidelines.

[1] For information on related programs and future updates to this program, please check www.indiana.edu/~jsl650/post.htm .

Syntax

```
mlogtest[, iia lr wald combine lrcomb
      set(varlist [\ varlist ...]) all full base]
```

Options

`all` requests that all available tests should be performed.

`lr` requests that LR tests for each independent variable should be performed.

`wald` requests that Wald tests for each independent variable should be performed.

`combine` requests Wald tests of whether dependent categories can be combined.

`lrcomb` requests LR tests of whether dependent categories can be combined. This option uses constrained estimation and overwrites constraint 999 if it is already defined.

`hausman` requests Hausman tests of the IIA assumption.

`smhsiao` requests Small-Hsiao tests of the IIA assumption.

`detail` reports the full **hausman** output for the IIA test. The default is to provide only a summary of the results.

`base` also conducts an IIA test omitting the base category of the original `mlogit` estimation. This is done by re-estimating the model using the largest remaining category as the base category, although the original estimates are restored to memory afterward.

`set(`*varlist* `[\` *varlist*`] ...)` specifies that a set of variables is to be considered together for the LR test or Wald test. The slash \ is used to specify multiple sets of variables. This option is particularly useful when a categorical independent variable is entered as a set of dummy variables.

Utility Procedures

`mlogtest` uses several utility ado files that are also used in other programs by the authors. The utility procedures are described briefly. For more details, you can type `help` *command-name* after the programs have been installed.

`_perhs.ado` returns the number of right-hand-side variables and their names for regression models.

`_pecats.ado` returns the names and values of the categories for models with ordinal, nominal, or binary outcomes. For `mlogit` it indicates the value of the reference category.

Example

The data for this example are from the 1993 and 1994 General Social Survey. The nominal variable (`kidvalue`) is the respondent's choice of which of the following is most important for a child to learn to prepare him or her for life: "to obey" (`kidvalue==1`), "to think for himself or herself" (`==2`), "to work hard" (`==3`), or "to help others when they need help" (`==4`). The fifth option, "to be popular", was excluded because it was very rarely chosen. The independent variables are respondent's sex (`female`), race (`black` and `othrrace`, with the reference category being `white`), education (`degree`), and whether the respondent has any children of her or his own (`anykids`). We begin by estimating the MNLM:

```
. mlogit kidvalue female black othrrace degree anykids, nolog

Multinomial regression                            Number of obs   =       2978
                                                  LR chi2(15)     =     300.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -3396.3518                       Pseudo R2       =     0.0423

------------------------------------------------------------------------------
kidvalue |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
obey     |
  female |  -.2605371   .1048637     -2.485   0.013      -.4660662   -.0550079
   black |   .3297048   .1452035      2.271   0.023       .0451112    .6142984
othrrace |   .5711209   .2872073      1.989   0.047       .0082049    1.134037
  degree |  -.7040498   .0577797    -12.185   0.000       -.817296   -.5908037
 anykids |  -.0401693   .1202552     -0.334   0.738      -.2758652    .1955265
   _cons |  -.0847716   .1376452     -0.616   0.538      -.3545513    .1850081
---------+--------------------------------------------------------------------
workhard |
  female |  -.4657661   .1104007     -4.219   0.000      -.6821476   -.2493846
   black |   .1975939   .1714529      1.152   0.249      -.1384475    .5336354
othrrace |   1.621659   .2233146      7.262   0.000       1.183971    2.059348
  degree |  -.1824923   .0479872     -3.803   0.000      -.2765455   -.0884391
 anykids |   .0052844   .1243124      0.043   0.966      -.2383635    .2489323
   _cons |  -.8719322   .1472885     -5.920   0.000      -1.160612   -.5832521
---------+--------------------------------------------------------------------
helpoth  |
  female |  -.3530656   .1165728     -3.029   0.002      -.5815441   -.1245871
   black |  -.1156104   .1892914     -0.611   0.541      -.4866148     .255394
othrrace |   .8759096   .2791998      3.137   0.002        .328688    1.423131
  degree |  -.3875589   .0549027     -7.059   0.000      -.4951661   -.2799517
 anykids |  -.1913028   .1286881     -1.487   0.137      -.4435269    .0609214
   _cons |  -.5615834   .1493388     -3.760   0.000      -.8542821   -.2688846
------------------------------------------------------------------------------
(Outcome kidvalue==thnkself is the comparison group)
```

In the following examples, we use a series of `mlogtest` commands to estimate several tests. Alternatively, we could have requested any combination of tests by combining options or requested all possible tests with the single command: `mlogtest, all` .

Tests of Independent Variables We first conduct a LR test for each independent variable:

```
. mlogtest, lr

**** Likelihood-ratio tests for independent variables

 Ho: All coefficients associated with given variable(s) are 0.

kidvalue |      chi2    df   P>chi2
---------+-------------------------
  female |     23.558    3    0.000
   black |      7.231    3    0.065
othrrace |     51.944    3    0.000
  degree |    211.133    3    0.000
 anykids |      2.323    3    0.508
-----------------------------------
```

For example, we can reject the hypothesis that gender does not affect the values considered important for children at the .01 level. Or, the effect of gender is significant (p<.01, df=3). Next, we conduct a Wald test for each independent variable. We also use the `set` option to test the hypothesis that the coefficients for the two dummy variables indicating race are simultaneously equal to zero:

```
. mlogtest, wald set(black othrrace)

**** Wald tests for independent variables

 Ho: All coefficients associated with given variable(s) are 0.

kidvalue |       chi2   df   P>chi2
---------+-------------------------
  female |     23.451    3    0.000
   black |      7.317    3    0.062
othrrace |     54.177    3    0.000
  degree |    174.002    3    0.000
 anykids |      2.359    3    0.501
---------+---------------------------
set_1:   |     60.988    6    0.000
   black |
othrrace |
-----------------------------------
```

Tests of IIA Either the Hausman or Small-Hsiao tests can be used to test the IIA assumption. We begin with the Hausman test. The `base` option specifies that all tests should be computed using the most frequently observed remaining category as the base value (see Methods and Formulas for details). We do not use the `detail` option, which provides all of the output from the successive calls to Stata's `hausman` command.

```
. mlogtest, hausman base

**** Hausman tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |      chi2   df   P>chi2   evidence
---------+-----------------------------------
```

```
    obey |     7.764   12    0.803   for Ho
workhard |    -4.090   12     ---    for Ho
 helpoth |     9.154   12    0.690   for Ho
thnkself |   884.043   12    0.000   against Ho
----------------------------------------------
 Note: If chi2<0, the estimated model does not
 meet asymptotic assumptions of the test.
```

Note the considerably different results depending on the category considered. In our experience, negative test statistics are very common; Hausman and McFadden (1984:1226) note this possibility and conclude that a negative result is evidence that IIA has *not* been violated. When we run Small-Hsiao tests, we see that these results vary considerably from those of the Hausman tests:

```
. mlogtest, smhsiao base

**** Small-Hsiao tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |  lnL(full)  lnL(omit)    chi2   df   P>chi2   evidence
---------+---------------------------------------------------------
    obey |  -1041.535  -1039.193   4.683    6    0.585   for Ho
workhard |  -1107.167  -1103.476   7.381    6    0.287   for Ho
 helpoth |  -1178.179  -1175.128   6.101    6    0.412   for Ho
thnkself |   -744.697   -740.162   9.069    6    0.170   for Ho
-------------------------------------------------------------------
```

Since the Small-Hsiao test is based on the creation of random half-samples from one's data, the test may differ substantially with successive calls of the command. For example, when we run the tests again, we obtain:

```
. mlogtest, smhsiao base

**** Small-Hsiao tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |  lnL(full)  lnL(omit)    chi2   df   P>chi2   evidence
---------+---------------------------------------------------------
    obey |  -1098.851  -1089.556  18.589    6    0.005   against Ho
workhard |  -1164.440  -1153.210  22.459    6    0.001   against Ho
 helpoth |  -1169.482  -1165.634   7.695    6    0.261   for Ho
thnkself |   -786.601   -774.531  24.141    6    0.000   against Ho
-------------------------------------------------------------------
```

The `set seed` command can be used before `mlogtest` in a do file to have it produce the same results with each successive run. For example, `set seed 339487731` .

Tests for Combining Dependent Categories Finally, we test whether the independent variables differentiate pairs of outcome categories using a Wald test. Note that all pairs of outcomes have been evaluated:

```
. mlogtest, combine
```

```
**** Wald tests for combining outcome categories

 Ho: All coefficients except intercepts associated with given pair
     of outcomes are 0 (i.e., categories can be collapsed).

Categories tested |      chi2   df   P>chi2
------------------+------------------------
    obey-workhard |    81.629    5    0.000
    obey- helpoth |    31.332    5    0.000
    obey-thnkself |   167.265    5    0.000
workhard- helpoth |    19.637    5    0.001
workhard-thnkself |    79.317    5    0.000
 helpoth-thnkself |    65.716    5    0.000
-------------------------------------------
```

Alternatively, LR tests can be computed with the `lrcomb` option:

```
. mlogtest, lrcom

**** LR tests for combining outcome categories

 Ho: All coefficients except intercepts associated with given pair
     of outcomes are 0 (i.e., categories can be collapsed).

Categories tested |      chi2   df   P>chi2
------------------+------------------------
    obey-workhard |    89.431    5    0.000
    obey- helpoth |    32.089    5    0.000
    obey-thnkself |   212.672    5    0.000
workhard- helpoth |    20.523    5    0.001
workhard-thnkself |    80.259    5    0.000
 helpoth-thnkself |    70.485    5    0.000
-------------------------------------------
```

As with the Wald and LR tests for each independent variable, the two tests for combining categories generally provide very similar results, although many researchers prefer the LR test.

Overall, these examples illustrate that `mlogtest` makes it very simple to compute many tests. At the risk of repetition, we note that it is not our intention to encourage researchers to combine categories or delete variables without careful consideration of the substantive issues related to the research.

Saved Results

`mlogtest` returns the following information which can be retrieved with `r()`.

`r(combine)` contains results of tests to combine categories. Rows represent all contrasts among categories; columns indicate the categories contrasted, the chi-square value, the degrees of freedom, and the *p*-value.

`r(iia)` contains results of Hausman tests of IIA assumption. Each row is one test. Columns indicate the omitted category of a given test, the chi-square value, the degrees of freedom, and the p-value.

`r(wald)` contains results of Wald test that all coefficients of an independent variable equal zero.

`r(lrtest)` contains results of LR test that all coefficients associated with an independent variable equal zero.

Methods and Formulas

This section provides brief descriptions of each of the tests. For further details, check the Stata manual for `mlogit`, `test`, and `hausman`. Full details along with citations to original sources are found in Long (1997). To make our discussion of the tests clear, we begin with a brief summary of the multinomial logit model (MNLM).

The multinomial logit model For simplicity, we consider a model with three outcomes and three independent variables. The MNLM can be thought of as simultaneously estimating binary logits among all pairs of the outcome categories. For example, with categories A, B, and C and independent variables x_1, x_2, and x_3, the MNLM is in effect simultaneously estimating three binary models:

$$\begin{aligned}
\ln\left[\frac{\Pr(A \mid \mathbf{x})}{\Pr(C \mid \mathbf{x})}\right] &= \beta_{0,A|C} + \beta_{1,A|C}x_1 + \beta_{2,A|C}x_2 + \beta_{3,A|C}x_3 \\
\ln\left[\frac{\Pr(B \mid \mathbf{x})}{\Pr(C \mid \mathbf{x})}\right] &= \beta_{0,B|C} + \beta_{1,B|C}x_1 + \beta_{2,B|C}x_2 + \beta_{3,B|C}x_3 \\
\ln\left[\frac{\Pr(A \mid \mathbf{x})}{\Pr(B \mid \mathbf{x})}\right] &= \beta_{0,A|B} + \beta_{1,A|B}x_1 + \beta_{2,A|B}x_2 + \beta_{3,A|B}x_3
\end{aligned}$$

Note that three more equations could be listed, comparing C to A, C to B, and B to A. Given that the sum of the probabilities for the outcomes must equal 1, there is an implicit constraint on the three logits. Specifically:

$$\ln\left[\frac{\Pr(A \mid \mathbf{x})}{\Pr(C \mid \mathbf{x})}\right] - \ln\left[\frac{\Pr(B \mid \mathbf{x})}{\Pr(C \mid \mathbf{x})}\right] = \ln\left[\frac{\Pr(A \mid \mathbf{x})}{\Pr(B \mid \mathbf{x})}\right]$$

In terms of the parameters:

$$\beta_{k,A|C} - \beta_{k,B|C} = \beta_{k,A|B}$$

`mlogit` estimates and prints only the non-redundant coefficients. Which set of coefficients is determined by the `basecategory()` option or by default is the category with the largest number of cases. The commands `mcross` (Rogers 1995, STB-23) and `listcoef` (Long and Freese, 2000) list coefficients for all comparisons of outcome categories.

Testing the effect of an independent variable With J dependent categories, there are $J-1$ non-redundant, coefficients associated with each independent variable x_k. The hypothesis that x_k does not affect the dependent variable can be written as:

$$H_0\colon \beta_{k,1|Base} = \cdots = \beta_{k,J|Base} = 0$$

where *Base* is the base category used in the comparison. Since $\beta_{k,Base|Base}$ is necessarily zero, the hypothesis imposes constraints on $J-1$ parameters. This hypothesis can be tested with either a Wald or a LR test.

A LR test First, estimate the full model M_F that contains all of the variables, with the resulting LR statistic LR_F^2. Second, estimate the restricted model M_R formed by excluding variable x_k, with the resulting LR statistic LR_R^2. This model has $J-1$ fewer parameters. Finally, compute the difference $LR^2_{\text{RvsF}} = LR_F^2 - LR_R^2$ which is distributed as chi-square with $J-1$ degrees of freedom if the hypothesis that x_k does not affect the outcome is true. `mlogtest, lr` computes this test for each of the K independent variables by making repeated calls to Stata's `lrtest`. Note that this requires estimating K additional multinomial logit models.

A Wald test While the LR test is generally considered to be superior, if the model is complex or the sample is very large, the computational costs of the LR test can be prohibitive. Alternatively, K Wald tests can be computed without estimating additional models. This test is defined as follows. Let $\widehat{\boldsymbol{\beta}}_k$ be the $J-1$ coefficients associated with x_k. Let $\widehat{Var}\left(\widehat{\boldsymbol{\beta}}_k\right)$ be the estimated covariance matrix. The Wald statistic for the hypothesis that all of the coefficients associated with x_k are simultaneously zero is computed as: $W_k = \widehat{\boldsymbol{\beta}}'_k \widehat{Var}\left(\widehat{\boldsymbol{\beta}}_k\right)^{-1} \widehat{\boldsymbol{\beta}}_k$. If the null hypothesis is true, then W_k is distributed as chi-square with $J-1$ degrees of freedom.

Testing multiple independent variables This logic of the Wald or LR tests can be extended to simultaneously test that the effects of two or more independent variables are zero. For example, the hypothesis to test that x_k and x_ℓ have no effects is:

$$H_0\colon \beta_{k,1|Base} = \cdots = \beta_{k,J|Base} = \beta_{\ell,1|Base} = \cdots = \beta_{\ell,J|Base} = 0$$

The `set(`*`varlist`*` [\ `*`varlist`*` ...])` option in `mlogtest` specifies which variables are to be simultaneously tested. This is particularly useful when a series of dummy variables are used to code a nominal or ordinal independent variable.

Testing that two outcomes can be combined If none of the x_k's significantly affect the odds of outcome m versus outcome n, we say that m and n are *indistinguishable* with respect to the variables in the model (Anderson 1984). If $\beta_{1,m|n}, \ldots, \beta_{K,m|n}$ are the coefficients for x_1 through x_K from the logit of m versus n, then the hypothesis that outcomes m and n are indistinguishable corresponds to:

$$H_0\colon \beta_{1,m|n} = \cdots \beta_{K,m|n} = 0$$

Note that if the base category used by Stata is not n, these coefficients are not directly available. However, this hypothesis can be rewritten equivalently using the coefficients with respect to the base category:

$$H_0\colon \left(\beta_{1,m|Base} - \beta_{1,n|Base}\right) = \cdots = \left(\beta_{K,m|Base} - \beta_{K,n|Base}\right) = 0$$

A Wald test for this hypothesis can be computed with Stata's `test` command. `mlogtest, combine` executes and summarizes the results of $J \times (J-1)$ calls to `test` for all pairs of outcome categories.

An LR test of this hypothesis can be computed by first estimating the full model that contains all of the variables, with the resulting LR statistic LR^2_F. Then estimate a restricted model M_R in which category m is used as the base category and all the coefficients (except the constant) in the equation for category n are constrained to 0, with the resulting chi-square statistic LR^2_R. The test statistic is the difference $LR^2_{\text{RvsF}} = LR^2_F - LR^2_R$ which is distributed as chi-square with K degrees of freedom. `mlogtest, lrcomb` summarizes the results of the $J \times (J-1)$ LR tests for all pairs of outcome categories.

Independence of Irrelevant Alternatives The MNLM assumes that the odds for any pair of outcomes are determined without reference to the other outcomes that might be available. This is known as the *independence of irrelevant alternatives* property or simply *IIA*. Hausman and McFadden (1984) proposed a Hausman-type test of this hypothesis. Basically, this involves the following steps.

1. Estimate the full model with all J outcomes included; these estimates are contained in $\widehat{\boldsymbol{\beta}}_F$.

2. Estimate a restricted model by eliminating one or more outcome categories; these estimates are contained in $\widehat{\boldsymbol{\beta}}_R$.

3. Let $\widehat{\boldsymbol{\beta}}^*_F$ be a subset of $\widehat{\boldsymbol{\beta}}_F$ after eliminating coefficients not estimated in the restricted model . The Hausman test of IIA is defined as:

$$H_{\text{IIA}} = \left(\widehat{\boldsymbol{\beta}}_R - \widehat{\boldsymbol{\beta}}^*_F\right)' \left[\widehat{Var}\left(\widehat{\boldsymbol{\beta}}_R\right) - \widehat{Var}\left(\widehat{\boldsymbol{\beta}}^*_F\right)\right]^{-1} \left(\widehat{\boldsymbol{\beta}}_R - \widehat{\boldsymbol{\beta}}^*_F\right)$$

   H_{IIA} is asymptotically distributed as chi-square with degrees of freedom equal to the rows in $\widehat{\boldsymbol{\beta}}_R$ if IIA is true. Significant values of H_{IIA} indicate that the IIA assumption has been violated.

Hausman and McFadden (1984:1226) note that H_{IIA} can be negative when $\widehat{Var}\left(\widehat{\boldsymbol{\beta}}_R\right) - \widehat{Var}\left(\widehat{\boldsymbol{\beta}}^*_F\right)$ is not positive semidefinite and suggest that a negative H_{IIA} is evidence that IIA *holds*.

To compute Small and Hsiao's test, the sample is divided into two random subsamples of approximately equal size. The unrestricted MNLM is estimated on both subsamples. The weighted average of the coefficients from the two samples is defined as follows:

$$\widehat{\boldsymbol{\beta}}^{S_1S_2}_u = \left(\frac{1}{\sqrt{2}}\right)\widehat{\boldsymbol{\beta}}^{S_1}_u + \left[1 - \left(\frac{1}{\sqrt{2}}\right)\right]\widehat{\boldsymbol{\beta}}^{S_2}_u$$

where $\widehat{\boldsymbol{\beta}}_u^{S_1}$ is a vector of estimates from the unrestricted model on the first subsample and $\widehat{\boldsymbol{\beta}}_u^{S_2}$ is its counterpart for the second subsample. Next, a restricted sample is created from the second subsample by eliminating all cases with a chosen value of the dependent variable. The MNLM is estimated using the restricted sample yielding the estimates $\widehat{\beta}_r^{S_2}$ and the likelihood $L(\widehat{\beta}_r^{S_2})$. The Small-Hsiao statistic is the difference:

$$SH = -2 \left[L(\widehat{\boldsymbol{\beta}}_u^{S_1 S_2}) - L(\widehat{\boldsymbol{\beta}}_r^{S_2}) \right]$$

SH is asymptotically distributed as a chi-square with the degrees of freedom equal to $K+1$, where K is the number of independent variables.

For both the Hausman test and the Small-Hsiao test, multiple tests of IIA are possible. Assuming that the MNLM is estimated with base category *Base*, $J-1$ tests can be computed by excluding each of the remaining categories to form the restricted model. By changing the base category, a test can also be computed that excludes *Base*. Note that results differ depending on which base category was used to estimate the model.

References

Long, J. Scott. (1996). *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage.

Hausman, J. A. and McFadden, D. (1984). "Specification tests for the multinomial logit model." *Econometrica*, *52*, 1219-1240.

McFadden, D., Tye, W., & Train, K. (1976). "An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model." *Transportation Research Board Record*, *637*, 39-45.

Small, K. A. & Hsiao, C. (1985). "Multinomial logit specification tests." *International Economic Review*, *26*, 619-627.