suest — Seemingly unrelated estimation

Syntax

    suest namelist [, options]

where namelist is a list of one or more names under which estimation results were stored via estimates store; see [R] estimates store. Wildcards may be used. * and _all refer to all stored results. A period (.) may be used to refer to the last estimation results, even if they have not (yet) been stored.

options          Description

SE/Robust
        svy             survey data estimation
        vce(vcetype)    vcetype may be robust or cluster clustvar

Reporting
        _level(#)       set confidence level; default is level(95)
        dir             display a table describing the models
        eform(string)   report exponentiated coefficients and label as string
        display_options control column formats, row spacing, line width, display of omitted
                          variables and base and empty cells, and factor-variable labeling
        coeflegend      display legend instead of statistics

coeflegend does not appear in the dialog box.

Menu

Statistics > Postestimation > Tests > Seemingly unrelated estimation

Description

    suest is a postestimation command; see [U] 20 Estimation and postestimation commands.

    suest combines the estimation results—parameter estimates and associated (co)variance matrices—stored under namelist into one parameter vector and simultaneous (co)variance matrix of the sandwich/robust type. This (co)variance matrix is appropriate even if the estimates were obtained on the same or on overlapping data.

    Typical applications of suest are tests for intramodel and cross-model hypotheses using test or testnl, for example, a generalized Hausman specification test. lincom and nlcom may be used after suest to estimate linear combinations and nonlinear functions of coefficients. suest may also be used to adjust a standard VCE for clustering or survey design effects.
Different estimators are allowed, for example, a `regress` model and a `probit` model; the only requirement is that `predict` produce equation-level scores with the `score` option after an estimation command. The models may be estimated on different samples, due either to explicit `if` or `in` selection or to missing values. If weights are applied, the same weights (type and values) should be applied to all models in `namelist`. The estimators should be estimated without `vce(robust)` or `vce(cluster clustvar)` options. `suest` returns the robust VCE, allows the `vce(cluster clustvar)` option, and automatically works with results from the `svy` prefix command (only for `vce(linearized)`). See example 7 in [SVY] `svy postestimation` for an example using `suest` with `svy: ologit`.

Because `suest` posts its results like a proper estimation command, its results can be stored via `estimates store`. Moreover, like other estimation commands, `suest` typed without arguments replays the results.

**Options**

- `suest` specifies that estimation results should be modified to reflect the survey design effects according to the `svyset` specifications, see [SVY] `svyset`.

  The `svy` option is implied when `suest` encounters survey estimation results from the `svy` prefix; see [SVY] `svy`. Poststratification is allowed only with survey estimation results from the `svy` prefix.

- `vce(vcetype)` specifies the type of standard error reported, which includes types that are robust to some kinds of misspecification (`robust`) and that allow for intragroup correlation (`cluster clustvar`); see [R] `vce_option`.

  The `vce()` option may not be combined with the `svy` option or estimation results from the `svy` prefix.

**Reporting**

- `level(#)` specifies the confidence level, as a percentage, for confidence intervals of the coefficients; see [R] `level`.

- `dir` displays a table describing the models in `namelist` just like `estimates dir namelist`.

- `eform(string)` displays the coefficient table in exponentiated form: for each coefficient, $\exp(b)$ rather than $b$ is displayed, and standard errors and confidence intervals are transformed. `string` is the table header that will be displayed above the transformed coefficients and must be 11 characters or fewer, for example, `eform("Odds ratio")`.

**display_options:** `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] `estimation options`.

The following option is available with `suest` but is not shown in the dialog box:

`coeflegend`; see [R] `estimation options`.

**Remarks and examples**

Remarks are presented under the following headings:

- *Using suest*
- *Remarks on regress*
- *Testing the assumption of the independence of irrelevant alternatives*
- *Testing proportionality*
- *Testing cross-model hypotheses*
Using suest

If you plan to use suest, you must take precautions when fitting the original models. These restrictions are relaxed when using svy commands; see [SVY] svy postestimation.

1. suest works with estimation commands that allow predict to generate equation-level score variables when supplied with the score (or scores) option. For example, equation-level score variables are generated after running mlogit by typing

```
predict sc*, scores
```

2. Estimation should take place without the vce(robust) or vce(cluster clustvar) option. suest always computes the robust estimator of the (co)variance, and suest has a vce(cluster clustvar) option.

The within-model covariance matrices computed by suest are identical to those obtained by specifying a vce(robust) or vce(cluster clustvar) option during estimation. suest, however, also estimates the between-model covariances of parameter estimates.

3. Finally, the estimation results to be combined should be stored by estimates store; see [R] estimates store.

After estimating and storing a series of estimation results, you are ready to combine the estimation results with suest,

```
suest name1 [ name2 ... ] [, vce(cluster clustvar)]
```

and you can subsequently use postestimation commands, such as test, to test hypotheses. Here an important issue is how suest assigns names to the equations. If you specify one model name, the original equation names are left unchanged; otherwise, suest constructs new equation names. The coefficients of a single-equation model (such as logit and poisson) that was estimate stored under name X are collected under equation X. With a multiequation model stored under name X, suest prefixes X_ to an original equation name eq, forming equation name, X_eq.

Technical note

Earlier we said that standard errors from suest are identical to those obtained by specifying the vce(robust) option with each command individually. Thus if you fit a logistic model using logit with the vce(robust) option, you will get the same standard errors when you type

```
suest .
```

directly after logit using the same data without the vce(robust) option.

This is not true for multiple estimation results when the estimation samples are not all the same. The standard errors from suest will be slightly smaller than those from individual model fits using the vce(robust) option because suest uses a larger number of observations to estimate the simultaneous (co)variance matrix.

Technical note

In rare circumstances, suest may have to truncate equation names to 32 characters. When equation names are not unique because of truncation, suest numbers the equations within models, using equations named X_#.
Remarks on `regress`

`regress` (see [R] `regress`) does not include its ancillary parameter, the residual variance, in its coefficient vector and (co)variance matrix. Moreover, while the `score` option is allowed with `predict` after `regress`, a score variable is generated for the mean but not for the variance parameter. `suest` contains special code that assigns the equation name `mean` to the coefficients for the mean, adds the equation `lnvar` for the log variance, and computes the appropriate two score variables itself.

Testing the assumption of the independence of irrelevant alternatives

The multinomial logit model and the closely related conditional logit model satisfy a probabilistic version of the assumption of the independence of irrelevant alternatives (IIA), implying that the ratio of the probabilities for two alternatives does not depend on what other alternatives are available. Hausman and McFadden (1984) proposed a test for this assumption that is implemented in the `hausman` command. The standard Hausman test has several limitations. First, the test statistic may be undefined because the estimated VCE does not satisfy the required asymptotic properties of the test. Second, the classic Hausman test applies only to the test of the equality of two estimators. Third, the test requires access to a fully efficient estimator; such an estimator may not be available, for example, if you are analyzing complex survey data. Using `suest` can overcome these three limitations.

Example 1

In our first example, we follow the analysis of the type of health insurance reported in [R] `mlogit` and demonstrate the `hausman` command with the `suest/test` combination. We fit the full multinomial logit model for all three alternatives and two restricted multinomial models in which one alternative is excluded. After fitting each of these models, we store the results by using the `store` subcommand of `estimates`. `title()` simply documents the models.

```
. use http://www.stata-press.com/data/r13/sysdsn4
(Health insurance data)
. mlogit insure age male
Iteration 0: log likelihood = -555.85446
Iteration 1: log likelihood = -551.32973
Iteration 2: log likelihood = -551.32802
Iteration 3: log likelihood = -551.32802
Multinomial logistic regression Number of obs = 615
LR chi2(4) = 9.05
Prob > chi2 = 0.0598
Log likelihood = -551.32802 Pseudo R2 = 0.0081

                      Coef.  Std. Err.      z    P>|z|     [95% Conf. Interval]
------------- -------- ------------- ------ -------- ------------------
Indemnity     (base outcome)
            age   -.0100251    .0060181   -1.67    0.096    -.0218204    .0017702
            male   .5095747    .1977893    2.58    0.010     .1219147    .8972346
            _cons   .2633838    .2787575    0.94    0.345   -.2829708    .8097383
Prepaid       age   -.0051925    .0113821   -0.46    0.648   -.0275011    .0171161
            male   .4748547    .3618462    1.31    0.189   -.2343508    1.184060
            _cons   -1.756843   .5309602   -3.31    0.001   -2.797506   -.7161803
Uninsure      age   -.0051925    .0113821   -0.46    0.648   -.0275011    .0171161
            male   .4748547    .3618462    1.31    0.189   -.2343508    1.184060
            _cons   -1.756843   .5309602   -3.31    0.001   -2.797506   -.7161803
```

. estimates store m1, title(all three insurance forms)
Having performed the three estimations, we inspect the results. estimates dir provides short descriptions of the models that were stored using estimates store. Typing estimates table lists the coefficients, displaying blanks for a coefficient not contained in a model.

```plaintext
name | command | depvar | npar | title
-----|----------|--------|------|------
m1 | mlogit | insure | 9 | all three insurance forms
m2 | mlogit | insure | 6 | insure != Uninsure :insure
m3 | mlogit | insure | 6 | insure != Prepaid :insure
```

. estimates table m1 m2 m3, star stats(N ll) keep(Prepaid: Uninsure:)

```
<table>
<thead>
<tr>
<th>Variable</th>
<th>m1</th>
<th>m2</th>
<th>m3</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Prepaid</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>-.0100251</td>
<td>-.01015205</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>.50957468**</td>
<td>.51440033**</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>.26338378</td>
<td>.26780432</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Uninsure</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>-.00519249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>.47485472</td>
<td>.45910738</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>-1.7568431***</td>
<td>-1.8017743***</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistics</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>615</td>
<td>570</td>
<td>338</td>
</tr>
<tr>
<td>ll</td>
<td>-551.32802</td>
<td>-390.48643</td>
<td>-131.76807</td>
</tr>
</tbody>
</table>
```

Comparing the coefficients between models does not suggest substantial differences. We can formally test that coefficients are the same for the full model m1 and the restricted models m2 and m3 by using the hausman command. hausman expects the models to be specified in the order “always consistent” first and “efficient under $H_0$” second.

. hausman m2 m1, alleqs constant

```
<table>
<thead>
<tr>
<th></th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
<th>Diff.</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>-.0101521</td>
<td>-.0100251</td>
<td>-.0001269</td>
<td>.0001269</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>.51440033**</td>
<td>.50957468**</td>
<td>.0048256</td>
<td>.0123338</td>
<td>.0123338</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>.26780432</td>
<td>.26338378</td>
<td>.0044205</td>
<td>.0044205</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>
```

Test: Ho: difference in coefficients not systematic

```
chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 0.08
Prob>chi2 = 0.9944
(V_b-V_B is not positive definite)
```
. hausman m3 m1, alleqs constant

<table>
<thead>
<tr>
<th></th>
<th>Coefficients</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(b)</td>
<td>(B)</td>
<td>(b-B)</td>
<td>sqrt(diag(V_b-V_B))</td>
</tr>
<tr>
<td>m3</td>
<td>-0.0041055</td>
<td>-0.0051925</td>
<td>0.001087</td>
<td>0.0021355</td>
</tr>
<tr>
<td>m1</td>
<td>-0.4591074</td>
<td>0.4748547</td>
<td>-0.0157473</td>
<td>.</td>
</tr>
<tr>
<td>diff</td>
<td>-1.801774</td>
<td>-1.756843</td>
<td>0.0449311</td>
<td>0.1333421</td>
</tr>
</tbody>
</table>

b = consistent under Ho and Ha; obtained from mlogit

B = inconsistent under Ha, efficient under Ho; obtained from mlogit

Test: Ho: difference in coefficients not systematic

\[
\text{chi2}(3) = (b-B)'[(V_b-V_B)^{-1}](b-B)
\]

\[
= -0.18
\]

chi2<0 ==> model fitted on these data fails to meet the asymptotic assumptions of the Hausman test;

see suest for a generalized test

According to the test of m1 against m2, we cannot reject the hypothesis that the coefficients of m1 and m2 are the same. The second Hausman test is not well defined—something that happens fairly often. The problem is due to the estimator of the variance \( V(b-B) \) as \( V(b)-V(B) \), which is a feasible estimator only asymptotically. Here it simply is not a proper variance matrix, and the Hausman test becomes undefined.

suest m1 m2 estimates the simultaneous (co)variance of the coefficients of models m1 and m2. Although suest is technically a postestimation command, it acts like an estimation command in that it stores the simultaneous coefficients in \( e(b) \) and the full (co)variance matrix in \( e(V) \). We could have used the estat vce command to display the full (co)variance matrix to show that the cross-model covariances were indeed estimated. Typically, we would not have a direct interest in \( e(V) \).

. suest m1 m2, noomitted

Simultaneous results for m1, m2

|        | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|-------|----------------------|
| m1_Indemnity |       |           |       |       |                      |
| m1_Prepaid      |       |           |       |       |                      |
| age             | -0.0100251 | .0059403 | -1.69 | 0.091 | -.0216679             | .0016176 |
| male            | .5095747   | .1988159 | 2.56  | 0.010 | .1199027              | .8992467 |
| cons            | .2633838   | .277307  | 0.95  | 0.342 | -.280128              | .8068956 |
| m1_Uninsure     |       |           |       |       |                      |
| age             | -0.0051925 | .0109005 | -0.48 | 0.634 | -.0265571             | .0161721 |
| male            | .4748547   | .3677326 | 1.29  | 0.197 | -.2458879             | 1.195597 |
| cons            | -1.756843  | .4971383 | -3.53 | 0.000 | -2.731216             | -.78247  |
| m2_Indemnity    |       |           |       |       |                      |
| m2_Prepaid      |       |           |       |       |                      |
| age             | -0.0101521 | .0058988 | -1.72 | 0.085 | -.0217135             | .0014094 |
| male            | .5144003   | .1996133 | 2.58  | 0.010 | .1231654              | .9056352 |
| cons            | .2678043   | .2744019 | 0.98  | 0.329 | -.2700134             | .8056221 |
suest created equation names by combining the name under which we stored the results using estimates store with the original equation names. Thus, in the simultaneous estimation result, equation Prepaid originating in model m1 is named m1_Prepaid. According to the McFadden–Hausman specification of a test for IIA, the coefficients of the equations m1_Prepaid and m2_Prepaid should be equal. This equality can be tested easily with the test command. The cons option specifies that the intercept _cons be included in the test.

```
.suest m1 m3, noomitted
```

Simultaneous results for m1, m3

```
Number of obs = 615
```

|                | Robust Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|----------------|--------------|-----------|-----|-----|----------------------|
| m1_Indemnity   |              |           |     |     |                      |
| m1_Prepaid     |              |           |     |     |                      |
| age            | -.0100251    | .0059403  | -1.69| 0.091| -.0216679  | .0016176 |
| male           | .5095747     | .1988159  | 2.56| 0.010| .1199027  | .8992467 |
| _cons          | .2633838     | .277307   | 0.95| 0.342| -.280128  | .8068956 |
| m1_Uninsure    |              |           |     |     |                      |
| age            | -.0051925    | .0109005  | -0.48| 0.634| -.0265571  | .0161721 |
| male           | .4748547     | .3677326  | 1.29| 0.197| -.2458879 | 1.195597 |
| _cons          | -1.756843    | .4971383  | -3.53| 0.000| -2.731216 | -.78247 |
| m3_Indemnity   |              |           |     |     |                      |
| m3_Uninsure    |              |           |     |     |                      |
| age            | -.0041055    | .0111185  | -0.37| 0.712| -.0258974  | .0176865 |
| male           | .4591074     | .3601307  | 1.27| 0.202| -.2467357 | 1.164951 |
| _cons          | -1.801774    | .5226351  | -3.45| 0.001| -2.82612  | -.7774283 |

```
.test [m1_Uninsure = m3_Uninsure], cons
( 1) [m1_Uninsure]age - [m3_Uninsure]age = 0
( 2) [m1_Uninsure]male - [m3_Uninsure]male = 0
( 3) [m1_Uninsure]_cons - [m3_Uninsure]_cons = 0
    chi2( 3) = 1.49
    Prob > chi2 = 0.6845
```

Although the classic Hausman test computed by hausman is not defined here, the suest-based test is just fine. We cannot reject the equality of the common coefficients across m1 and m3.

A second advantage of the suest approach is that we can estimate the (co)variance matrix of the multivariate normal distribution of the estimators of the three models m1, m2, and m3 and test that the common coefficients are equal.
. suest m*, noomitted

Simultaneous results for m1, m2, m3

Number of obs = 615

| Coef. Std. Err. z   P>|z|   [95% Conf. Interval] |
|-------------------|-------------|-------------|-------------|
| **m1_Indemnity** |             |             |             |
| age               | -.0100251   | .0059403    | -1.69       | 0.091       | -.0216679 - .0016176 |
| male              | .5095747    | .1988159    | 2.56        | 0.010       | .1199027 .8992467   |
| _cons             | .2633838    | .277307     | 0.95        | 0.342       | -.280128 .8068956   |

| **m1_Uninsure**  |             |             |             |
| age               | -.0051925   | .0109005    | -0.48       | 0.634       | -.0265571 .0161721  |
| male              | .4748547    | .3677326    | 1.29        | 0.197       | -.2458879 1.195597  |
| _cons             | -1.756843   | .4971383    | -3.53       | 0.000       | -2.731216 -.78247   |

| **m2_Indemnity** |             |             |             |
| age               | -.0101521   | .0058988    | -1.72       | 0.085       | -.0217135 .0014094  |
| male              | .5144003    | .1996133    | 2.58        | 0.010       | .1231654 .9056352   |
| _cons             | .2678043    | .2744019    | 0.98        | 0.329       | -.2700134 .8056221  |

| **m3_Indemnity** |             |             |             |
| age               | -.0041055   | .0111185    | -0.37       | 0.712       | -.0258974 .0176865  |
| male              | .4591074    | .3601307    | 1.27        | 0.202       | -.2467357 1.164951  |
| _cons             | -1.801774   | .5226351    | -3.45       | 0.001       | -2.82612 -.7774283 |

. test [m1_Prepaid = m2_Prepaid], cons notest
( 1) [m1_Prepaid]age - [m2_Prepaid]age = 0
( 2) [m1_Prepaid]male - [m2_Prepaid]male = 0
( 3) [m1_Prepaid]_cons - [m2_Prepaid]_cons = 0

. test [m1_Uninsure = m3_Uninsure], cons acc
( 1) [m1_Prepaid]age - [m2_Prepaid]age = 0
( 2) [m1_Prepaid]male - [m2_Prepaid]male = 0
( 3) [m1_Prepaid]_cons - [m2_Prepaid]_cons = 0
( 4) [m1_Uninsure]age - [m3_Uninsure]age = 0
( 5) [m1_Uninsure]male - [m3_Uninsure]male = 0
( 6) [m1_Uninsure]_cons - [m3_Uninsure]_cons = 0

chi2(  6) =  1.95
Prob > chi2 =  0.9240

Again we do not find evidence against the correct specification of the multinomial logit for type of insurance. The classic Hausman test assumes that one of the estimators (named B in hausman) is efficient, that is, it has minimal (asymptotic) variance. This assumption ensures that $V(b) - V(B)$ is an admissible, viable estimator for $V(b - B)$. The assumption that we have an efficient estimator is a restrictive one. It is violated, for instance, if our data are clustered. We want to adjust for clustering via a vce(cluster clustvar) option by requesting the cluster-adjusted sandwich estimator of variance. Consequently, in such a case, hausman cannot be used. This problem does not exist with the suest version of the Hausman test. To illustrate this feature, we suppose that the data are clustered by city—we constructed an imaginary variable cityid for this illustration. If we plan to apply suest, we would not specify the vce(cluster clustvar) option at the time of estimation.
. suest m1 m2, vce(cluster cityid) noomitted

Simultaneous results for m1, m2

Number of obs   =        615

(Std. Err. adjusted for 260 clusters in cityid)

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
</tr>
<tr>
<td>m1_Indemnity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m1_Prepaid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>-0.0100251</td>
<td>0.005729</td>
<td>-1.75</td>
<td>0.080</td>
</tr>
<tr>
<td>male</td>
<td>0.5095747</td>
<td>0.1910496</td>
<td>2.67</td>
<td>0.008</td>
</tr>
<tr>
<td>_cons</td>
<td>0.2633838</td>
<td>0.2698797</td>
<td>0.98</td>
<td>0.329</td>
</tr>
<tr>
<td>m1_Uninsure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>-0.0051925</td>
<td>0.0104374</td>
<td>-0.50</td>
<td>0.619</td>
</tr>
<tr>
<td>male</td>
<td>0.4748547</td>
<td>0.3774021</td>
<td>1.26</td>
<td>0.208</td>
</tr>
<tr>
<td>_cons</td>
<td>-1.756843</td>
<td>0.4916613</td>
<td>-3.57</td>
<td>0.000</td>
</tr>
<tr>
<td>m2_Indemnity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m2_Prepaid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>-0.0101521</td>
<td>0.0057164</td>
<td>-1.78</td>
<td>0.076</td>
</tr>
<tr>
<td>male</td>
<td>0.5144003</td>
<td>0.3774021</td>
<td>1.36</td>
<td>0.176</td>
</tr>
<tr>
<td>_cons</td>
<td>0.2678043</td>
<td>0.2682193</td>
<td>1.00</td>
<td>0.318</td>
</tr>
</tbody>
</table>

. test [m1_Prepaid = m2_Prepaid], cons

( 1)  [m1_Prepaid]age - [m2_Prepaid]age = 0
( 2)  [m1_Prepaid]male - [m2_Prepaid]male = 0
( 3)  [m1_Prepaid]_cons - [m2_Prepaid]_cons = 0

chi2( 3) = 0.79
Prob > chi2 = 0.8529

suest provides some descriptive information about the clustering on cityid. Like any other estimation command, suest informs us that the standard errors are adjusted for clustering. The Hausman-type test obtained from the test command uses a simultaneous (co)variance of m1 and m2 appropriately adjusted for clustering. In this example, we still do not have reason to conclude that the multinomial logit model in this application is misspecified, that is, that IIA is violated.

The multinomial logistic regression model is a special case of the conditional logistic regression model; see [R] clogit. Like the multinomial logistic regression model, the conditional logistic regression model also makes the IIA assumption. Consider an example, introduced in [R] asclogit, in which the demand for American, Japanese, and European cars is modeled in terms of the number of local dealers of the respective brands and of some individual attributes incorporated in interaction with the nationality of cars. We want to perform a Hausman-type test for IIA comparing the decision between all nationalities with the decision between non-American cars. The following code fragment demonstrates how to conduct a Hausman test for IIA via suest in this case.

. clogit choice japan europe maleJap maleEur incJap incEur dealer, group(id)
. estimates store allcars
. clogit choice japan maleJap incJap dealer if car!=1 , group(id)
. estimates store foreign
. suest allcars foreign
. test [allcars_choice=foreign_choice], common

Testing proportionality

The applications of `suest` that we have discussed so far concern Hausman-type tests for mis-specification. To test such a hypothesis, we compared two estimators that have the same probability limit if the hypothesis holds true, but otherwise have different limits. We may also want to compare the coefficients of models (estimators) for other substantive reasons. Although we most often want to test whether coefficients differ between models or estimators, we may occasionally want to test other constraints (see Hausman and Ruud [1987]).

Example 2

In this example, using simulated labor market data for siblings, we consider two dependent variables, income (`inc`) and whether a person was promoted in the last year (`promo`). We apply familiar economic arguments regarding human capital, according to which employees have a higher income and a higher probability of being promoted, by having more human capital. Human capital is acquired through formal education (`edu`) and on-the-job training experience (`exp`). We study whether income and promotion are “two sides of the same coin”, that is, whether they reflect a common latent variable, “human capital”. Accordingly, we want to compare the effects of different aspects of human capital on different outcome variables.

We estimate fairly simple labor market equations. The income model is estimated with `regress`, and the estimation results are stored under the name `Inc`.

. use http://www.stata-press.com/data/r13/income
. regress inc edu exp male
Source       SS       df    MS           Number of obs =  277
             F(  3,  273) = 42.34
Model     2058.44672     3  686.148908     Prob > F = 0.0000
Residual   4424.05183  273  16.2053181     R-squared = 0.3175
          ___________________________________________________________
        Adj R-squared = 0.3175
          ___________________________________________________________
Total     6482.49855  276  23.4873136     Root MSE = 4.0256

        inc           Coef.     Std. Err.     t     P>|t|    [95% Conf. Interval]
          ______________    __________    _________    _____    _________
edu      2.213707      .243247     9.10     0.000    1.734828     2.692585
exp      1.47293       .231044     6.38     0.000    1.018076     1.927785
male     .5381153      .4949466     1.09     0.278   -.4362822     1.512513
_cons    1.255497      .3115808     4.03     0.000    .642091     1.868904
. est store Inc

Being sibling data, the observations are clustered on family of origin, `famid`. In the estimation of the regression parameters, we did not specify a `vce(cluster famid)` option to adjust standard errors for clustering on family (`famid`). Thus the standard errors reported by `regress` are potentially flawed. This problem will, however, be corrected by specifying a `vce(cluster clustvar)` option with `suest`.

Next we estimate the promotion equation with `probit` and again store the results under an appropriate name.
The coefficients in the income and promotion equations definitely seem to be different. However, because the scales of the two variables are different, we would not expect the coefficients to be equal. The correct hypothesis here is that the proportionality of the coefficients of the two models, apart from the constant, are equal. This formulation would still reflect that the relative effects of the different aspects of human capital do not differ between the dependent variables. We can obtain a nonlinear Wald test for the hypothesis of proportionality by using the `testnl` command on the combined estimation results of the two estimators. Thus we first have to form the combined estimation results. At this point, we specify the `vce(cluster famid)` option to adjust for the clustering of observations on `famid`.

```
.suest Inc Promo, vce(cluster famid)
```

Simultaneous results for Inc, Promo

```
Number of obs = 277
(Std. Err. adjusted for 135 clusters in famid)

<table>
<thead>
<tr>
<th>Robust</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[95% Conf. Interval]</td>
</tr>
<tr>
<td>Robust</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inc_mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edu</td>
<td>2.213707</td>
<td>.2483907</td>
<td>8.91</td>
<td>0.000</td>
</tr>
<tr>
<td>exp</td>
<td>1.47293</td>
<td>.190583</td>
<td>7.79</td>
<td>0.000</td>
</tr>
<tr>
<td>male</td>
<td>0.5381153</td>
<td>.4979227</td>
<td>1.08</td>
<td>0.280</td>
</tr>
<tr>
<td>_cons</td>
<td>1.255497</td>
<td>.3374977</td>
<td>3.72</td>
<td>0.000</td>
</tr>
<tr>
<td>Inc_lnvar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>2.785339</td>
<td>.079597</td>
<td>34.99</td>
<td>0.000</td>
</tr>
<tr>
<td>Promo_promo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edu</td>
<td>0.4593002</td>
<td>.0898537</td>
<td>5.11</td>
<td>0.000</td>
</tr>
<tr>
<td>exp</td>
<td>0.3593023</td>
<td>.0805774</td>
<td>4.46</td>
<td>0.000</td>
</tr>
<tr>
<td>male</td>
<td>0.2079983</td>
<td>.1656413</td>
<td>1.26</td>
<td>0.209</td>
</tr>
<tr>
<td>_cons</td>
<td>0.464622</td>
<td>.1088166</td>
<td>4.27</td>
<td>0.000</td>
</tr>
</tbody>
</table>
```
The standard errors reported by `suest` are identical to those reported by the respective estimation commands when invoked with the `vce(cluster famid)` option. We are now ready to test for proportionality:

$$
H_0 : \frac{\beta_{\text{Income, edu}}}{\beta_{\text{Promotion, edu}}} = \frac{\beta_{\text{Income, exp}}}{\beta_{\text{Promotion, exp}}} = \frac{\beta_{\text{Income, male}}}{\beta_{\text{Promotion, male}}}
$$

It is straightforward to translate this into syntax suitable for `testnl`, recalling that the coefficient of variable $v$ in equation $eq$ is denoted by $[eq]v$.

```
. testnl [Inc_mean]edu/[Promo_promo]edu = [Inc_mean]exp/[Promo_promo]exp = [Inc_mean]male/[Promo_promo]male
   (1) [Inc_mean]edu/[Promo_promo]edu = [Inc_mean]exp/[Promo_promo]exp
   (2) [Inc_mean]edu/[Promo_promo]edu = [Inc_mean]male/[Promo_promo]male
   chi2(2) = 0.61
   Prob > chi2 = 0.7385
```

From the evidence, we fail to reject the hypotheses that the coefficients of the income and promotion equations are proportional. Thus it is not unreasonable to assume that income and promotion can be explained by the same latent variable, “labor market success”.

A disadvantage of the nonlinear Wald test is that it is not invariant with respect to representation: a Wald test for a mathematically equivalent formulation of the nonlinear constraint usually leads to a different test result. An equivalent formulation of the proportionality hypothesis is

$$
H_0 : \frac{\beta_{\text{Income, edu}}}{\beta_{\text{Promotion, exp}}} = \frac{\beta_{\text{Promotion, edu}}}{\beta_{\text{Income, exp}}} \quad \text{and} \quad \frac{\beta_{\text{Income, edu}}}{\beta_{\text{Promotion, male}}} = \frac{\beta_{\text{Promotion, edu}}}{\beta_{\text{Income, male}}}
$$

This formulation is “more linear” in the coefficients. The asymptotic $\chi^2$ distribution of the nonlinear Wald statistic can be expected to be more accurate for this representation.

```
. testnl ([Inc_mean]edu*[Promo_promo]exp = [Inc_mean]exp*[Promo_promo]edu)
   > ([Inc_mean]edu*[Promo_promo]male = [Inc_mean]male*[Promo_promo]edu)
   (1) [Inc_mean]edu*[Promo_promo]exp = [Inc_mean]exp*[Promo_promo]edu
   (2) [Inc_mean]edu*[Promo_promo]male = [Inc_mean]male*[Promo_promo]edu
   chi2(2) = 0.46
   Prob > chi2 = 0.7936
```

Here the two representations lead to similar test statistics and $p$-values. As before, we fail to reject the hypothesis of proportionality of the coefficients of the two models.

---

**Testing cross-model hypotheses**

**Example 3**

In this example, we demonstrate how some cross-model hypotheses can be tested using the facilities already available in most estimation commands. This demonstration will explain the intricate relationship between the cluster adjustment of the robust estimator of variance and the `suest` command. It will also be made clear that a new facility is required to perform more general cross-model testing.
We want to test whether the effect of \(x_1\) on the binary variable \(y_1\) is the same as the effect of \(x_2\) on the binary \(y_2\); see Clogg, Petkova, and Haritou (1995). In this setting, \(x_1\) may equal \(x_2\), and \(y_1\) may equal \(y_2\). We assume that logistic regression models can be used to model the responses, and for simplicity, we ignore further predictor variables in these models. If the two logit models are fit on independent samples so that the estimators are (stochastically) independent, a Wald test for \(\hat{b}[x_1] = \hat{b}[x_2]\) rejects the null hypothesis if

\[
\frac{\hat{b}(x_1) - \hat{b}(x_2)}{\hat{\sigma}^2(\hat{b}(x_1)) + \hat{\sigma}^2(\hat{b}(x_2))^{1/2}}
\]

is larger than the appropriate \(\chi^2_1\) threshold. If the models are fit on the same sample (or on dependent samples), so that the estimators are stochastically dependent, the above test that ignores the covariance between the estimators is not appropriate.

It is instructive to see how this problem can be tackled by “stacking” data. In the stacked format, we doubled the number of observations. The dependent variable is \(y_1\) in the first half of the data and is \(y_2\) in the second half of the data. The predictor variable \(z_1\) is set to \(x_1\) in the first half of the expanded data and to 0 in the rest. Similarly, \(z_2\) is 0 in the first half and \(x_2\) in the second half. The following diagram illustrates the transformation, in the terminology of the \texttt{reshape} command, from wide to long format.

\[
\begin{pmatrix}
  id & y_1 & y_2 & x_1 & x_2 \\
  1 & y_{11} & y_{21} & x_{11} & x_{21} \\
  2 & y_{12} & y_{22} & x_{12} & x_{22} \\
  3 & y_{13} & y_{23} & x_{13} & x_{23}
\end{pmatrix}
\implies
\begin{pmatrix}
  id & y & z_1 & z_2 & \text{model} \\
  1 & y_{11} & x_{11} & 0 & 1 \\
  2 & y_{12} & x_{12} & 0 & 1 \\
  3 & y_{13} & x_{13} & 0 & 1 \\
  1 & y_{21} & 0 & x_{21} & 2 \\
  2 & y_{22} & 0 & x_{22} & 2 \\
  3 & y_{23} & 0 & x_{23} & 2
\end{pmatrix}
\]

The observations in the long format data organization are clustered on the original subjects and are identified with the identifier \(id\). The clustering on \(id\) has to be accounted for when fitting a simultaneous model. The simplest way to deal with clustering is to use the cluster adjustment of the robust or sandwich estimator; see \texttt{P} \texttt{robust}. The data manipulation can be accomplished easily with the \texttt{stack} command; see \texttt{D} \texttt{stack}. Subsequently, we fit a simultaneous logit model and perform a Wald test for the hypothesis that the coefficients of \(z_1\) and \(z_2\) are the same. A full setup to obtain the cross-model Wald test could then be as follows:

\begin{verbatim}
. generate zero = 0 // a variable that is always 0
. generate one = 1 // a variable that is always 1
. generate two = 2 // a variable that is always 2
. stack id y1 x1 zero one id y2 zero x2 two, into(id y z1 z2 model)
. generate model2 = (model==2)
. logit y model2 z1 z2, vce(cluster id)
. test _b[z1] = _b[z2]
\end{verbatim}

The coefficient of \(z_1\) represents the effect of \(x_1\) on \(y_1\), and similarly, \(z_2\) for the effect of \(x_2\) on \(y_2\). The variable \texttt{model2} is a dummy for the “second model”, which is included to allow the intercept in the second model to differ from that in the first model. The estimates of the coefficient of \(z_1\) and its standard error in the combined model are the same as the estimates of the coefficient of \(z_1\) and its standard error if we fit the model on the unstacked data.

\begin{verbatim}
. logit y1 x1, vce(robust)
\end{verbatim}
The `vce(cluster clustvar)` option specified with the `logit` command for the stacked data ensures that the covariances of \(_b[z1]\) and \(_b[z2]\) are indeed estimated. This estimation ensures that the Wald test for the equality of the coefficients is correct. If we had not specified the `vce(cluster clustvar)` option, the (co)variance matrix of the coefficients would have been block-diagonal; that is, the covariances of \(_b[z1]\) and \(_b[z2]\) would have been 0. Then `test` would have effectively used the invalid formula for the Wald test for two independent samples.

In this example, the two logit models were fit on the same data. The same setup would apply, without modification, when the two logit models were fit on overlapping data that resulted, for instance, if the \(y\) or \(x\) variables were missing in some observations.

The `suest` command allows us to obtain the above Wald test more efficiently by avoiding the data manipulation, obviating the need to fit a model with twice the number of coefficients. The test statistic produced by the above code fragment is identical to that obtained via `suest` on the original (unstacked) data:

```
. logit y1 x1
. estimates store M1
. logit y2 x2
. estimates store M2
. suest M1 M2
. test [M1]x1=[M2]x2
```

The stacking method can be applied not only to the testing of cross-model hypotheses for logit models but also to any estimation command that supports the `vce(cluster clustvar)` option. The stacking approach clearly generalizes to stacking more than two logit or other models, testing more general linear hypotheses, and testing nonlinear cross-model hypotheses (see [R] `testnl`). In all these cases, `suest` would yield identical statistical results but at smaller costs in terms of data management, computer storage, and computer time.

Is `suest` nothing but a convenience command? No, there are two disadvantages to the stacking method, both of which are resolved via `suest`. First, if the models include ancillary parameters (in a regression model, the residual variance; in an ordinal response model, the cutpoints; and in lognormal survival-time regression, the time scale parameter), these parameters are constrained to be equal between the stacked models. In `suest`, this constraint is relaxed. Second, the stacking method does not generalize to compare different statistical models, such as a probit model and a regression model. As demonstrated in the previous section, `suest` can deal with this situation.
Stored results

`suest` stores the following in `e()`:

Scalars
- `e(N)` number of observations
- `e(N_clust)` number of clusters
- `e(rank)` rank of `e(V)`

Macros
- `e(cmd)` `suest`
- `e(eqnames#)` original names of equations of model #
- `e(names)` list of model names
- `e(wtype)` weight type
- `e(wexp)` weight expression
- `e(clustvar)` name of cluster variable
- `e(vcetype)` `vcetype` specified in `vce()`
- `e(vce)` title used to label Std. Err.
- `e(properties)` `b V`

Matrices
- `e(b)` stacked coefficient vector of the models
- `e(V)` variance–covariance matrix of the estimators

Functions
- `e(sample)` marks estimation sample

Methods and formulas

The estimation of the simultaneous (co)variance of a series of $k$ estimators is a nonstandard application of the sandwich estimator, as implemented by the command `[P] _robust`. You may want to read this manual entry before reading further.

The starting point is that we have fit $k$ different models on the same data—partially overlapping or nonoverlapping data are special cases. We want to derive the simultaneous distribution of these $k$ estimators, for instance, to test a cross-estimator hypothesis $H_0$. As in the framework of Hausman testing, $H_0$ will often be of the form that different estimators have the same probability limit under some hypothesis, while the estimators have different limits if the hypothesis is violated.

We consider (vector) estimators $\hat{\beta}_i$ to be defined as “the” solution of the estimation equations $G_i$,  
\[
G_i(b_i) = \sum_j w_{ij}u_{ij}(b_i) = 0, \quad i = 1, \ldots, k
\]

We refer to the $u_{ij}$ as the “scores”. Specifying some weights $w_{ij} = 0$ trivially accommodates for partially overlapping or even disjointed data. Under “suitable regularity conditions” (see White [1982; 1996] for details), the $\hat{\beta}_i$ are asymptotically normally distributed, with the variance estimated consistently by the sandwich estimator
\[
V_i = \text{Var}(\hat{\beta}_i) = D_i^{-1} \sum_j w_{ij}u_{ij}u_{ij}'D_i^{-1}
\]
where $D_i$ is the Jacobian of $G_i$ evaluated at $\hat{\beta}_i$. In the context of maximum likelihood estimation, $D_i$ can be estimated consistently by (minus) the Hessian of the log likelihood or by the Fisher information matrix. If the model is also well specified, the sandwiched term $(\sum_j w_{ij} u_{ij} u_{ij}'')$ converges in probability to $D_i$, so $V_i$ may be consistently estimated by $D_i^{-1}$.

To derive the simultaneous distribution of the estimators, we consider the “stacked” estimation equation,

$$G(\beta) = \left\{ G_1(\hat{\beta}_1)' \ G_1(\hat{\beta}_2)' \ldots \ G_k(\hat{\beta}_k)' \right\}' = 0$$

Under “suitable regularity conditions” (see White [1996] for details), $\hat{\beta}$ is asymptotically jointly normally distributed. The Jacobian and scores of the simultaneous equation are easily expressed in the Jacobian and scores of the separate equations. The Jacobian of $G$,

$$D(\beta) = \left. \frac{dG(\beta)}{d\beta} \right|_{\beta=\hat{\beta}}$$

is block diagonal with blocks $D_1, \ldots, D_k$. The inverse of $D(\beta)$ is again block diagonal, with the inverses of $D_i$ on the diagonal. The scores $u$ of $G$ are simply obtained as the concatenated scores of the separate equations:

$$u_j = (u_{1j}' \ u_{2j}' \ldots \ u_{kj}')'$$

Out-of-sample (that is, where $w_{ij} = 0$) values of the score variables are defined as 0 (thus we drop the $i$ subscript from the common weight variable). The sandwich estimator for the asymptotic variance of $\hat{\beta}$ reads

$$V = \text{Var}(\beta) = D(\beta)^{-1} \left( \sum_j w_j u_j u_j' \right) D(\beta)^{-1}$$

Taking a “partitioned” look at this expression, we see that $V(\hat{\beta}_i)$ is estimated by

$$D_i^{-1} \left( \sum_j w_{ij} u_{ij} u_{ij}' \right) D_i^{-1}$$

which is, yet again, the familiar sandwich estimator for $\hat{\beta}_i$ based on the separate estimation equation $G_i$. Thus considering several estimators simultaneously in this way does not affect the estimators of the asymptotic variances of these estimators. However, as a bonus of stacking, we obtained a sandwich-type estimate of the covariance $V_{ih}$ of estimators $\hat{\beta}_i$ and $\hat{\beta}_h$,

$$V_{ih} = \text{Cov}(\hat{\beta}_i, \hat{\beta}_h) = D_i^{-1} \left( \sum_j w_{ij} u_{ij} u_{ih}' \right) D_h^{-1}$$

which is also obtained by White (1982).

This estimator for the covariance of estimators is an application of the cluster modification of the sandwich estimator proposed by Rogers (1993). Consider the stacked data format as discussed in the logit example, and assume that Stata would be able to estimate a “stacked model” in which different models apply to different observations, for example, a probit model for the first half, a regression model for the second half, and a one-to-one cluster relation between the first and second half. If there are no common parameters to both models, the score statistics of parameters for the stacked models are zero in the half of the data in which they do not occur. In Rogers’ method, we have to sum the score statistics over the observations within a cluster. This step boils down to concatenating the score statistics at the level of the cluster.
We compare the sandwich estimator of the (co)variance $V_{12}$ of two estimators with the estimator of variance $\tilde{V}_{12}$ applied in the classic Hausman test. Hausman (1978) showed that if $\hat{\beta}_1$ is consistent under $H_0$ and $\hat{\beta}_2$ is efficient under $H_0$, then asymptotically

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \text{Var}(\hat{\beta}_2)$$

and so $\text{var}(\hat{\beta}_1 - \hat{\beta}_2)$ is consistently estimated by $V_1 - V_2$.

Acknowledgment

`suest` was written by Jeroen Weesie of the Department of Sociology at Utrecht University, The Netherlands. This research is supported by grant PGS 50-370 by The Netherlands Organization for Scientific Research.

An earlier version of `suest` was published in the *Stata Technical Bulletin* (1999). The current version of `suest` is not backward compatible with the STB version because of the introduction of new ways to manage estimation results via the `estimates` command.

References


Also see

[R] estimates — Save and manipulate estimation results
[R] hausman — Hausman specification test
[R] lincom — Linear combinations of estimators
[R] nlcom — Nonlinear combinations of estimators
[R] test — Test linear hypotheses after estimation
[R] testnl — Test nonlinear hypotheses after estimation
[P] _robust — Robust variance estimates