help for rcal

SJ3-4: st0049, st0050, st0051)

Regression Calibration

rcal depvar [indepvars] (label:varlist ) [(label:varlist ) ... (label:varlist )] [if exp] [in range] [, bootstrap brep(#) ltolerance(#) iterate(#) family(familyname) link(linkname) message(#) naive robust suuinit(...) ignore suu btrim(#) saving(filename) replace seed(#) scale(x2 dev #)]

where familyname is one of

gaussian | igaussian | binomial [varnameN | #N] | poisson |

and linkname is one of

identity | log | logit | probit | cloglog | opower # | power # |

and label:varlist describes a variable measured with error. The label is for the unknown measurement error covariate (label cannot be the same as an existing variable in the data set). varlist is a list of variables with the replicate measurements for the unknown label covariate (see comments for restrictions).

by ... : may be used with rcal; see help by.

rcal No predict is implemented.

Description

rcal fits generalized linear models for measurement error data using IRLS (maximum quasi-likelihood) and is similar in syntax to simex. This command is implemented by Stata's plug-in mechanism. rcal allows one or more (see comments) covariates measured with errors and uses regression calibration to estimate the missing covariates. It will allow replicate data or a user specified measurement error covariance matrix. It implements a very fast internal bootstrap (different from the regular Stata bootstrap command).

Options

bootstrap specifies that the bootstrap estimate of variance be used.

brep(#) specifies the number of bootstrap samples to consider in forming the bootstrap estimate of variance. The default is brep(199).

ltolerance(#) specifies the convergence criterion for the change in deviance between iterations; ltolerance(1e-6) is the default.

iterate(#) specifies the maximum number of iterations allowed in fitting the model; iterate(100) is the default. You should seldom need to specify iterate().

family(familyname) specifies the distribution of depvar; family(gaussian) is the default.

link(linkname) specifies the link function; the default is the canonical link for the family() specified.
message(#) The message or debug level from the plug-in module. The
default is message(2)).

robust specifies that the Huber/White/sandwich estimator of variance is
to be used in place of the traditional calculation. We do not support
the (cmd:cluster) option.

naive Uses the "naive" estimator of variance. That is, the variance is
not adjusted for measurement error. This option is for pedagogical
and diagnostic purposes and should not be otherwise used.

suuininit(matrixname) Specify the measurement error covariance matrix. This
is calculated from the replications in the measurement error
variables if it is not specified.

ignoreesuu If the measurement error covariance matrix is known, or if one
is willing to ignore the variation in its estimate use this option.
This may be relevant if the covariance comes from a large, careful
independent study, for which only summary statistics are available.

btrim(#) Percent boostrap trimming. The default is btrim(.02).

saving(filename ) Save the booststrap results to the specified file.

replace Replace the existing 'bootstrap results' file if it exists.

seed(#) specify the seed for the random number generator used by the
bootstrap. This enables for identical boostrap runs. This option is
generally not specified.

scale(x2|dev|#) overrides the default scale parameter. By default,
scale(1) is assumed for discrete distributions (binomial, Poisson,
negative binomial) and scale(x2) for continuous distributions
(Gaussian, gamma, inverse Gaussian).

scale(x2) specifies the scale parameter be set to the Pearson
chi-squared (or generalized chi-squared) statistic divided by the
residual degrees of freedom.

scale(dev) sets the scale parameter to the deviance divided by the
residual degrees of freedom. This provides an alternative to
scale(x2) for continuous distributions and over- or under-dispersed
discrete distributions.

scale(#) sets the scale parameter to #.

Special comments on multiple measurement error covariates

The number of replications for a covariate measured with error can vary
across observations. When two or more measurement error covariates exist,
they must all have the same number of replications across observations.

Special comments on standard errors

It can take a very long time to calculate the default and sandwich
variance estimates for large data sets. An estimated time to completion
is printed if the variance calculation will require more than 30 seconds.
It takes considerably less time to calculate the bootstrap variance
estimator for large data sets.

Remarks

The allowed link functions are
<table>
<thead>
<tr>
<th>Link function</th>
<th>glm option</th>
</tr>
</thead>
<tbody>
<tr>
<td>identity</td>
<td>link(identity)</td>
</tr>
<tr>
<td>log</td>
<td>link(log)</td>
</tr>
<tr>
<td>logit</td>
<td>link(logit)</td>
</tr>
<tr>
<td>probit</td>
<td>link(probit)</td>
</tr>
<tr>
<td>complementary log-log</td>
<td>link(cloglog)</td>
</tr>
<tr>
<td>odds power</td>
<td>link(opower #)</td>
</tr>
<tr>
<td>power</td>
<td>link(power #)</td>
</tr>
<tr>
<td>negative binomial</td>
<td>link(nbinomial)</td>
</tr>
<tr>
<td>log-log</td>
<td>link(loglog)</td>
</tr>
<tr>
<td>log-compliment</td>
<td>link(logc)</td>
</tr>
</tbody>
</table>

The allowed distribution families are

<table>
<thead>
<tr>
<th>Family</th>
<th>glm option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian(normal)</td>
<td>family(gaussian)</td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>family(igaussian)</td>
</tr>
<tr>
<td>Bernoulli/binomial</td>
<td>family(binomial)</td>
</tr>
<tr>
<td>Poisson</td>
<td>family(poission)</td>
</tr>
<tr>
<td>Negative binomial</td>
<td>family(nbinomial)</td>
</tr>
<tr>
<td>Gamma</td>
<td>family(gamma)</td>
</tr>
</tbody>
</table>

Reasonable combinations of family() and link() are

<table>
<thead>
<tr>
<th></th>
<th>id</th>
<th>log</th>
<th>logit</th>
<th>probit</th>
<th>clog</th>
<th>pow</th>
<th>opower</th>
<th>nbinomial</th>
<th>loglog</th>
<th>logc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inv. Gau.</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>binomial</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Poisson</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>neg. bin.</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>gamma</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Nonstandard combinations other than those marked out above are allowed, but the user is responsible for seeing that the data fit the combination and for the interpretation of the results.

If you specify family() but not link(), you obtain the canonical link for the family:

<table>
<thead>
<tr>
<th>family()</th>
<th>default link()</th>
</tr>
</thead>
<tbody>
<tr>
<td>family(gaussian)</td>
<td>link(identity)</td>
</tr>
<tr>
<td>family(igaussian)</td>
<td>link(power -2)</td>
</tr>
<tr>
<td>family(binomial)</td>
<td>link(logit)</td>
</tr>
<tr>
<td>family(poission)</td>
<td>link(log)</td>
</tr>
<tr>
<td>family(nbinomial)</td>
<td>link(log)</td>
</tr>
<tr>
<td>family(gamma)</td>
<td>link(power -1)</td>
</tr>
</tbody>
</table>

Examples

```
. * generate some data
. set obs 1000
. gen x1 = uniform()
. gen x2 = uniform()
. gen x3 = uniform()
. gen err = invnorm(uniform())
. gen y = 1+2*x1+3*x2+4*x3+err

. * estimate with x3 known
. qvf y x1 x2 x3, bootstrap
```
. * simulate measurement error covariate
  . gen a1 = x3 + .3*invnorm(uniform())
  . gen a2 = x3 + .3*invnorm(uniform())

. * estimate x1, x2 & w3 using regression calibration
  . rcal (y=x1 x2) (w3: a1 a2), bstrap
  . rcal (y=x1 x2) (w3: a1 a2), bstrap saving("rcalboot.txt") replace
  . eret list

. * display and use a covariance error matrix
  . mat list e(suu)
  . mat suu = (.1)
  . rcal (y=x1 x2) (w3: a1 a2), bstrap suuinit(suu)

Also see

Online: help for qvf, simex