_____

NEGBIN (MODEL=*1 or 2,* nonlinear options)
dependent variable   list of independent variables ;

_____

Function:

NEGBIN obtains estimates of the Negative Binomial model, where the dependent variable takes on only nonnegative integer count values and its expectation is an exponential linear function of the independent variables. In the Negative Binomial model, the variance of the dependent variable is larger than the mean, in contrast to the Poisson model, where the variance equals the mean (see the POISSON procedure).

Usage:

The basic NEGBIN statement is like the OLSQ statement: first list the dependent variable and then the independent variables. If you wish to have an intercept term in the regression (usually recommended), include the special variable C or CONSTANT in your list of independent variables. You may have as many independent variables as you like subject to the overall limits on the number of arguments per statement and the amount of working space, as well as the number of data observations you have available.

The observations over which the regression is computed are determined by the current sample. If any of the observations have missing values within the current sample, NEGBIN will print a warning message and will drop those observations. NEGBIN also checks that the observations on the dependent variable are integers and are not negative.

The list of independent variables on the NEGBIN command may include variables with explicit lags and leads as well as PDL (Polynomial Distributed Lag) variables. These distributed lag variables are a way to reduce the number of free coefficients when entering a large number of lagged variables in a regression by imposing smoothness on the coefficients. See the PDL section for a description of how to specify such variables.

Options:

MODEL = type of variance function. For MODEL=1, the variance is proportional to the mean:

$$VAR = MEAN*(1+alpha).$$

For the default MODEL=2, the variance is a quadratic function of the mean:

$$VAR = MEAN + alpha*MEAN**2.$$

In both cases, the parameter alpha is restricted to be non-negative. alpha=0 is the same as Poisson.

Nonlinear options - see the NONLINEAR section of this manual.

Examples:

Negative Binomial 2 regression of patents on lags of log(R&D), science sector dummy, and firm size:

    NEGBIN PATENTS C LRND LRND(-1) LRND(-2) DSCI SIZE;

NEGBIN

Negative Binomial 1 regression for the same model:

    NEGBIN(MODEL=1) PATENTS C LRND LRND(-1) LRND(-2) DSCI SIZE;

Output:

The output of NEGBIN begins with an equation title and frequency counts for the lowest 10 values of the dependent variable. Starting values and diagnostic output from the iterations will be printed. Final convergence status is printed. This is followed by the number of observations, mean and standard deviation of the dependent variable, sum of squared residuals, correlation type R-squared, likelihood ratio test for zero slopes, log likelihood, and a table of right hand side variable names, estimated coefficients, standard errors and associated t-statistics.

NEGBIN also stores some of these results in data storage for later use. The table below lists the results available after a NEGBIN command.

Name	Type	Length	Variable Description
@LHV	list	1	Name of dependent variable.
@RNMS	list	#vars	Names of right hand side variables.
@IFCONV	scalar	1	1 if convergence achieved, 0 otherwise.
@YMEAN	scalar	1	Mean of the dependent variable.
@SDEV	scalar	1	Standard deviation of the dependent variable.
@NOB	scalar	1	Number of observations.
@HIST	vector	#yvalues	Frequency counts for each dependent variable value.
@HISTVAL	vector	#yvalues	Each dependent variable value.
@SSR	scalar	1	Sum of squared residuals.
@RSQ	scalar	1	correlation type R-squared.
@LR	scalar	1	Likelihood ratio test for zero slope coefficients.
%LR	scalar	1	P-value for likelihood ratio test.
@LOGL	scalar	1	Log of likelihood function.
@SBIC	scalar	1	Schwarz Bayesian Information Criterion.
@NCOEF	scalar	1	Number of rhs variables (#vars).
@NCID	scalar	1	Number of identified coefficients.
@COEF	vector	#vars	Coefficient estimates.
@SES	vector	#vars	Standard errors.
@T	vector	#vars	T-statistics.
%T	vector	#vars	p-values for T-statistics.
@GRAD	vector	#vars	Gradient of log L at convergence.
@VCOV	matrix	#vars* #vars	Variance-covariance of estimated coefficients.
@FIT	series	#obs	Fitted probabilities.
@RES	series	#obs	Residuals.

If the regression includes a PDL variable, the following will also be stored:

@SLAG	scalar	1	Sum of the lag coefficients.
@MLAG	scalar	1	Mean lag coefficient (number of time periods).
@LAGF	vector	#lags	Estimated lag coefficients, after "unscrambling".

Method:

NEGBIN uses analytic first and second derivatives to obtain maximum likelihood estimates via the Newton-Raphson algorithm. This algorithm usually converges fairly quickly. TSP uses zeros for starting parameter values, except for the constant term and alpha. @START can be used to provide different starting values (see NONLINEAR in this manual).

Multicollinearity of the independent variables is handled with generalized inverses, like the regression procedures in TSP.

The exponential mean function is used in the NEGBIN model. That is, if X are the RHS variables and B are their coefficients, $E(Y) = \exp(X*B)$. This guarantees that predicted values of Y are never negative.

The ML command can also be used to estimate Negative Binomial models, including panel data models with fixed and random effects. See the **TSP Users's Guide** for simple examples, and see our web page for the panel examples.

References:

Cameron, A. Colin, and Pravid K. Trivedi, **Regression Analysis of Count Data**, Cambridge University Press, New York, 1998.

Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches, "*Econometric Models for Count Data with an Application to the Patents - R&D Relationship*", **Econometrica** 52, 1984, pp. 908-938.