POISSON (nonlinear options)
*dependent variable* *list of independent variables* ;

**Function:**

POISSON obtains estimates of the Poisson 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 Poisson model, the variance of the dependent variable equals its mean, which is rarely the case in practice. More general models, where the variance is larger than the mean, are the Negative Binomial 1 and 2 – see the NEGBIN command.

**Usage:**

The basic POISSON 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, POISSON will print a warning message and will drop those observations. POISSON also checks that the observations on the dependent variable are integers and are not negative.

The list of independent variables on the POISSON 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:**

Nonlinear options - see the NONLINEAR section of this manual.

**Example:**

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

```
POISSON PATENTS C LRND LRND(-1) LRND(-2) DSCI SIZE;
```

**Output:**

The output of POISSON 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, overdispersion test, 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. The default standard errors are the robust/QMLE Eicker-White estimates. These are consistent even for an overdispersed model. Note:

usually the overdispersion test rejects the Poisson model, and you may wish to use the Negative Binomial model instead (although as a member of the linear exponential class, the Poisson model with Eicker-White standard errors may be more robust against misspecification even when the data are overdispersed - see Cameron and Trivedi for further information on this point).

POISSON also stores some of these results in data storage for later use. The table below lists the results available after a POISSON 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. |
| @OVERDIS | scalar | 1 | Overdispersion test. |
| %OVERDIS | scalar | 1 | P-value for overdispersion test. |
| @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:**

POISSON 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. @START can be used to provide different starting values (see NONLINEAR in this manual).

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

The overdispersion test is @NOB*@RSQ from a regression of (@RES**2 - Y)/@FIT on C, @FIT . See Cameron and Trivedi (1998), p. 78, equation (3.39).

The exponential mean function is used in the Poisson model. That is, if X are the independent variables and B are their coefficients, E(Y) = exp(X*B) . This guarantees that predicted values of Y are never negative. The variance=mean property of the Poisson model implies that the penalty for a lack of fit to large Y values is much smaller than for small Y values. So if you use LSQ to run an unweighted nonlinear regression with the same exponential mean function, you will get a better fit to large Y values than with the Poisson model.

The ML command can also be used to estimate Poisson 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.

Maddala, G. S., **Limited-dependent and Qualitative Variables in Econometrics**, Cambridge University Press, New York, 1983, pp. 51-54.