_____

PANEL (TOTAL, BYID, BETWEEN, WITHIN, VARCOMP, MEAN, REG,
PRINT, SILENT, TERSE,
ID=*id series*, TIME=*time series*, T=*number of time periods (if balanced)*,
ROBUST, VBET=*between variance*, VWITH=*within variance*, VSMALL*)
dependent variable   list of independent variables ;

_____

Function:

PANEL obtains estimates of linear regression models for panel data (several observations or time periods for each individual). The data may be unbalanced (different number of observations per individual). PANEL can also compute means by group and perform F tests between groups.

Usage:

The basic PANEL statement is like the OLSQ statement: first list the dependent variable and then the independent variables. C is optional; an intercept term is central to these models and will be added if it is not present. 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 models are computed are determined by the current sample. PANEL treats missing values, lags, and leads correctly. That is, lags and leads are applied only within an individual.

Your data must be set up with all the time periods for each individual together. Additionally, you must specify when the observations for one individual end and data for the next individual begins. The default method is to provide a series named @ID which takes on different values for each individual. If your data are balanced (the same number of time periods for every individual), the T= option can be used. Other options are also available (see below). If the data are not in this order, the SORT command can be used to reorder them; you could also sort the data by year and then individual if you wish to do variance components in the time dimension. Usually it is best to use the FREQ(PANEL) command at the top of your run to specify such ID variables, internal frequency and starting date, etc. Then these options will be used for all PANEL, AR1, GENR, etc. commands within the run.

The models you wish to estimate are specified in the options list. The default is to estimate all the models except MEAN and BYID. For the VARCOMP (random effects) model, there are additional options that specify how to compute the variance components. Small- or large sample formulas may be used, or the user can supply the values directly. If negative variances are computed using the small sample method, the method switches over to the large sample formulas, which always result in positive values. PANEL also computes a Hausman test for correlated effects by comparing the WITHIN (fixed effects) and VARCOMP(random effects) estimators.

Options:

BETWEEN/NOBETWEEN selects the "between" estimator -- a regression on the means for each individual.

BYID/**NOBYID** does a separate regression for each individual, and computes F tests for equality with the TOTAL and WITHIN estimators.

ID= the name of a series which takes on a different value for each individual. The default is @ID; alternatives are the T= and TIME= options.

PANEL

MEAN/**NOMEAN** causes the means for each individual to be printed in a table. This can be used in conjunction with the NOREG option to print means only (to suppress all the default regression models). These individual means are stored in the NI x (1+NX) matrix @MEAN, where the first column is the dependent variable.

PRINT**/NOPRINT** prints @COEFI in conjunction with BYID, and prints @FIXED for within.

REG/NOREG is used with the MEAN option above. To suppress some regression models, but print others, use the individual options -- NOBETW to suppress the BETWEEN output, etc.

ROBUST**/NOROBUST** calculates heteroskedasticity-robust standard errors (HCTYPE=1; see OLSQ) for the WITHIN coefficients. If this option is used, the Hausman test comparing WITHIN and VARCOMP is not computed.

SILENT**/NOSILENT** turns off all of the printed output. See also **TERSE**.

T= the number of time periods for each individual (for balanced data only). For unbalanced data, use the ID= option.

TERSE**/NOTERSE** turns off most of the regression output. See also **SILENT**.

TIME= the name of a time period series which increases in value for each individual and decreases between individuals. Alternatives are the ID= and T= options. Example: TIME=YEAR. This is no longer considered sufficient for identifying individuals, since the last time period for one individual may be less than the first time period of the next individual.

TOTAL/NOTOTAL selects the "total" or "pooled" estimator -- a plain OLS regression on the whole sample.

VARCOMP/NOVARCOMP selects the "variance components" or "random effects" estimator. The method of selecting the variance components is controlled with the VBET, VSMALL, and VWITH options described below. Unbalanced data are not a problem. Variance components in the time dimension are not available.

VBET= specifies the value of the "between" variance for VARCOMP.

VSMALL/NOVSMALL selects the small sample variance components formulas for VARCOMP (as opposed to the large sample formulas). Small sample formulas are unbiased but can result in negative variances, while large sample formulas are biased but always yield positive variances. To supply your own variance values, use VBET= and VWITH=.

VWITH= specifies the value of the "within" variance for VARCOMP.

WITHIN/NOWITHIN selects the "within" or "fixed effects" estimator (different intercepts for each individual).

Examples:

Global FREQ(PANEL) command, with ID variable to identify individuals:

        FREQ(PANEL,ID=FIRM); DY=Y-Y(-1);
        PANEL DY C X X(-1);

Estimate all models (7 years per individual, balanced data), and print individual means:

        PANEL(T=7,MEAN,BYID) LRNDL5 C PATENTS LRNDL4;

Print VARCOMP output only, using @ID or FREQ(PANEL) to distinguish individuals:

        PANEL(NOTOT,NOBET,NOWITH) LRNDL5 C PATENTS LRNDL4;

Estimate all models except BYID, use large sample formulas for VARCOMP:

PANEL(NOVSMALL) LRNDL5 C PATENTS LRNDL4;

Print individual means only:

PANEL(MEAN,NOREG) LRNDL5 C PATENTS LRNDL4;

Output:

The output begins with a title and a summary of the panel structure: number of individuals (NI), number of time periods (T), and total number of observations (NOB). If the data are unbalanced, TMIN and TMAX will be printed.

For each estimator, a table of regression coefficients and their standard errors is printed, along with name of the dependent variable, the sum of squared residuals, standard error of the regression, mean and standard deviation of the dependent variable, R-squared, and adjusted R-squared. Other output varies by estimator.

MEAN prints a table of means for each individual. @MEAN (#obs*#vars) is stored, and excludes any constant term.

BYID prints an F test vs. TOTAL (labelled F-stat for A,B=Ai,Bi), and an F test vs. WITHIN (labelled F-stat for Ai,B=Ai,Bi), in the output of the respective estimators. Only @COEFI and @SSRI are stored. Use the PRINT option to print @COEFI.

WITHIN prints an F test vs. TOTAL (labelled F-stat for A,B=Ai,B), and stores @FIXED effects vector.

VARCOMP prints the actual variance components, the method used to compute them, and the implied differencing factor (THETA). A Hausman specification test comparing VARCOMP (null hypothesis) and WITHIN is computed.

PANEL stores the standard regression results in data storage for later use using @names, but with B, I, T, V, and W appended to distinguish between the different estimators. For example, @COEFW is the within coefficients, @RESW are the within residuals, and @SSRV is the sum of squared residuals from VARCOMP. @RESB is a matrix.

Name	Type	Length	Variable Description
@LHV	list	1	Name of the dependent variable.
@RNMST/B/W/V	list	#vars	Names of right hand side variables.
@SSRT/I/B/W/V	scalar	1	Sum of squared residuals. (@SSRI=BYID, etc.)
@S2T/B/W/V	scalar	1	Variance of residuals. (@S2B=BETWEEN, etc.)
@ST/B/W/V	scalar	1	Standard error of the regression.
@YMEANT/B/W/V	scalar	1	Mean of the dependent variable.
@SDEVT/B/W/V	scalar	1	Standard deviation of the dependent variable.
@NOB	scalar	1	Number of observations.
@RSQT/B/W/V	scalar	1	R-squared.
@ARSQT/B/W/V	scalar	1	Adjusted R-squared.
@NCOEFT/B/W/V	scalar	1	Number of coefficients.
@NCIDT/B/W/V	scalar	1	Number of identified coefs (with non-zero SEs).
@LMHETT/W/V	scalar	1	LM heteroskedasticity test.
%LMHETT/W/V	scalar	1	P-value of LM heteroskedasticity test.
@DWT/W/V	scalar	1	Durbin-Watson autocorrelation test.
%DWUT/W/V	scalar	1	Upper bound on P-value of DW.
%DWLT/W/V	scalar	1	Lower bound on P-value of DW.
@HAUS	scalar	1	Hausman test value.
%HAUS	scalar	1	Hausman test p-value.
@HAUSDF	scalar	1	Hausman test degrees of freedom.
@COEFT/I/B/W/V	vector	#vars	Coefficient estimates.
@SEST/B/W/V	vector	#vars	Standard Errors.

PANEL

@TT/B/W/V	vector	#vars	T-statistics.
@FIXED	vector	#individuals	Fixed effect estimates.
@VCOVT/B/W/V	matrix	#vars*#vars	Variance-covariance of estimated coefficients.
@REST/I/B/W/V	series	#obs	Residuals=actual - fitted values of the dependent variable.

Method:

PANEL computes means for each variable by individual. These are used directly in the BETWEEN regression. WITHIN subtracts the individual means from each variable and runs a regression on this transformed data (any variables which are constant over time for every individual are not identified).

VARCOMP does a transformation similar to WITHIN. (1-SQRT(THETA)) times the mean is subtracted from each variable (including the constant term). THETA = VWITH/(VWITH + T*VBET), where T does not have to be equal for each individual. VWITH is the variance of the basic error terms (Uit), while VBET is the variance of the individual-specific error terms (Ai). The small and large sample formulas used for these variance components are:

	Small Sample	**Large Sample**
VWITH	@SSRW/(NOB-NX-NI)	@SSRW/NOB
VTOT	@SSRT/(NOB-NX-1)	(not used)
VBET	VTOT-VWITH	(@SSRT-@SSRW)/NOB

If the small sample formula produces a non-positive VBET, PANEL switches over to the large sample formulas automatically. The large sample formulas are asymptotically correct if T is (becomes) large relative to NI (not usually the case); otherwise they will be biased. Note that if THETA=1, this corresponds to VBET=0 and VARCOMP will be the same as TOTAL. If THETA=0, this corresponds to VWITH=0, and VARCOMP will be the same as WITHIN.

For each F test (described under **Output**), a P-value and an alternative critical value are printed. The critical value has a size which becomes smaller as the number of observations grows -- this is an alternative to the conventional testing procedure, which is certain to reject all point null hypotheses when sample sizes become large. It is based on a Bayesian flat prior, and computed from the formula in the Leamer reference:

$$\text{Fcrit} = \frac{T-k}{p}(T^{p/T}-1)$$

Where T = total number of observations, k = number of estimated parameters in the unrestricted model, and p = the number of restrictions.

All regressions are computed with the standard orthonormalized data matrices to insure accurate coefficients and variance estimates under possible multicollinearity (methods using moment matrices are less accurate).

The Durbin-Watson test and bounds on its P-values are computed following the Bhargava et al reference, extended to the unbalanced data case. The P-values are computed using the Farebrother/Imhof method, since there can be multiple equal eigenvalues.

References:

Bhargava, A., L. Franzini, and W. Narendanathan, "Serial Correlation and the Fixed Effects Model," **Review of Economic Studies** XLIX, 1982, pp.533-549.

Farebrother, R.W., "Algorithm AS 256", **Applied Statistics** 39, 1990. Code posted on StatLib, in Pascal.

Hsiao, Cheng, **Analysis of Panel Data**, Cambridge University Press, Cambridge, England, 1986.

Leamer, Edward E., **Specification Searches: Ad Hoc Inference with Nonexperimental Data**, Wiley, New York, 1978,

p. 114.

Maddala, G.S., **Econometrics**, McGraw-Hill, New York, 1977, pp. 326-329.

StatLib, **http://lib.stat.cmu.edu/apstat/**