_____

AR1 (OBJFN= EXACTML or GLS,
METHOD=CORC or HILU or ML or MLGRID,
RMIN=*minimum rho value*, RMAX=*maximum rho value*,
RSTART=*start value for rho*, RSTEP=*step value for rho*, TSCS,
INST=(*list of instrumental variables*), FAIR, nonlinear options)
depvarname list of independent variables ;

_____

Function:

AR1 obtains estimates of a regression equation whose errors are serially correlated. These estimates are efficient if the disturbances in the equation follow an autoregressive process of order one. The estimates may be obtained using one of two different objective functions: exact maximum likelihood (which imposes stationarity by constraining the serial correlation coefficient to be between -1 and 1 and keeps the first observation for estimation), or by GLS, which drops the first observation.

Usage:

To obtain estimates of a regression equation which are corrected for first order serial correlation, use the AR1 command as you would an OLSQ command. PDL (polynomial distributed lag) variables may be included in an AR1 statement. See the PDL section for a further description of how to specify these variables. TSP automatically deletes observations with missing values for one or more variables before estimation.

Options:

FAIR/NOFAIR specifies whether the lagged dependent and independent variables are to be added to the instrument list automatically when doing instrumental variable estimation combined with a serial correlation correction.

INST= *list of instrumental variables*. This list should include any exogenous variables that are in the equation such as the constant or time trend, as well as any other variables you wish to use as instruments. After any instruments are added by the FAIR option, there must be at least as many instruments as the number of estimated coefficients (the number of independent variables in the equation, plus one for rho). OBJFN=GLS is implied; the actual labeled objective function is E'PZ*E. See the Examples for a way to reproduce the AR1 estimates with FORM and LSQ.

Fair once argued that the lagged dependent and independent variables must be in the instrument list to obtain consistent estimates when doing instrumental variable estimation with a serial correlation correction. TSP adds them automatically if you use the FAIR option (the default); if you want to specify a different list of instruments, you must suppress this feature with a NOFAIR option.

Fair retracted his claim in 1984; it has since been disproved by Buse (1989), but the alternative instruments for consistency involve pseudo-differencing with the estimated rho (Theil's G2SLS), which is tedious to perform by hand. Buse also showed that the asymptotically most efficient estimator in this case (S2SLS) includes the lagged excluded exogenous variables as well, but he cautions that in small samples this may quickly exhaust the degrees of freedom. G2SLS estimation requires fewer instruments than S2SLS, and it will be provided in a future version of TSP.

METHOD=ML or MLGRID or CORC or HILU was formerly used to specify the estimation algorithm. This is now specified by the OBJFN= option. METHOD=ML or MLGRID imply OBJFN=EXACTML, while METHOD=CORC or HILU imply OBJFN=GLS. METHOD=ML formerly used the Beach and McKinnon algorithm, while METHOD=CORC used the Cochrane-Orcutt algorithm. Now iterations are done using the Newton-Raphson alogrithm (HITER=N in the nonlinear options) which is quadratically convergent (about the same speed as Beach-MacKinnon, but

AR1

much faster and more accurate than Cochrane-Orcutt). METHOD=HILU refers to Hildreth-Lu, a simple grid search method.

OBJFN=EXACTML or GLS specifies the objective function. EXACTML retains the first observation and includes the Jacobian term log(1-rho**2), which guarantees stationarity. GLS drops the first observation and does not impose stationarity. It is the same as nonlinear least squares on a rho-differenced equation, and can also be described as "conditional ML" (conditional on the initial residual).

   EXACTML is the usual default, but if there is a lagged dependent variable on the RHS, GLS becomes the default, because EXACTML has a small-sample bias in this case.

   GLS uses an initial grid search to locate starting values and potential multiple local optima. It is well known that multiple local optima can occur for GLS, especially when there are lagged dependent variables. Multiple optima are noted in the output if they are detected. AR1 then iterates efficiently to locate an accurate global optimum. EXACTML normally skips the grid search, because no cases of multiple local optima are known when the Jacobian is included. METHOD=MLGRID will turn on this grid search.

RMIN= specifies the minimum value of the serial correlation parameter rho for the initial grid search (when OBJFN=GLS or METHOD=MLGRID are used). The default value is -0.9.

RMAX= specifies the maximum value of rho for the grid search methods. The default value is 1.05 for OBJFN=GLS, or .95 for METHOD=MLGRID.

RSTEP= specifies the increment to be used in the grid search over rho. The default value is 0.1, until rho=.8. Then the values .85, .9, .95 are used, plus .9999, 1.0001, and 1.05 when OBJFN=GLS. These last 3 values help to detect optima with rho > 1, which are usually not reached during iterations when rho starts below 1.

RSTART= specifies a starting value of rho for the iterative methods. Ordinarily zero is used for OBJFN=EXACTML, but faster convergence may be achieved if a value closer to the true answer is chosen. RSTART can also be used to override the default grid search for OBJFN=GLS, but multiple local optima would not be detected.

TSCS/NOTSCS specifies EXACTML estimation for time series-cross section data when the FREQ(PANEL) command is in effect (then TSCS is the default) or when SMPL gaps have been set up to separate the cross section units. (See the example below.) OBJFN=GLS is not implemented for panel data.

(Obsolete) WEIGHT= is a former AR1 option which is no longer supported. The ML or LSQ commands should be used instead to implement a weight.

Nonlinear options are described under NONLINEAR in this manual. HITER/HCOV=N (second derivatives, default) and G (first derivatives) are both available. MAXIT=0 can be used to avoid iterations; to perform just a simple grid search without the additional accuracy of iterations. AR1 uses a special default TOL=1E-6 (.000001).

Examples:

This example estimates the consumption function for the illustrative model with a serial correlation correction, first using the maximum likelihood method, and then searching over rho to verify that the likelihood is unimodal in the relevant range.[1]

   AR1 (PRINT) CONS C GNP ;
   AR1 (METHOD=MLGRID, RSTEP=0.05) CONS C GNP ;

---

[1]It is probably that this search is unnecessary for the default OBJFN=EXACTML

The next three estimations are exactly equivalent and demonstrate the FAIR option with instrumental variables:

```
SMPL 11,50;
AR1 (INST=(C,G,TIME,LM)) CONS C GNP ;
AR1 (NOFAIR,INST=(C,G,TIME,LM,GNP(-1),CONS(-1))) CONS C GNP;
FORM(NAR=1,PARAM,VARPREF=B) EQAR1 CONS C GNP;
SMPL 12,50;        ? Drop first observation, to match AR1(OBJFN=GLS)
LSQ(INST=(C,G,TIME,LM,GNP(-1),CONS(-1))) EQAR1;
```

Lagged dependent variable (default OBJFN=GLS, since EXACTML has a small sample bias):

```
AR1 CONS C GNP CONS(-1);
```

Time series-cross section with 10 years of data and 3 cross section units, and fixed effects:

```
SMPL 1,30;
FREQ(PANEL,T=10);
TREND OBS;
FIRM = 1 + INT((OBS-1)/10);
DUMMY FIRM FIRMS;   ? Create FIRM1-FIRM3 dummy variables, stored in FIRMS list
AR1 SALES FIRMS ADV POP GNP ;
```

Output:

The AR1 procedure produces quite a lot of printed output. An equation title and the chosen objective function are printed first. If the PRINT option is on, this is followed by the list of option values, the initial estimates for all the coefficients, iteration output for all coefficients, and any grid values for rho and the objective function.

The usual regression output follows, as described under the OLSQ command. The left column of regression statistics are in terms of the rho-transformed dependent variable and independent variables, and the (theoretically) uncorrelated error which results from AR1 estimation. This means that the R-squared, for example, is not comparable with one from OLS estimation on the same data. The right column of statistics is based on the original (untransformed) data (this second R-squared would be comparable with one from OLS).

As in OLSQ and INST, a table of coefficient estimates is printed. RHO is always the last coefficient in the table; its inclusion guarantees that the standard errors are always consistent, even if there are lagged dependent variables on the RHS. The fitted values ($\hat{y}_t$ = @FIT) and residuals (e_t = @RES) are computed as follows:

$$\hat{y}_t = X_t b \qquad\qquad \text{for } t = 1 \quad (\text{OBJFN=EXACTML})$$
$$\hat{y}_t = X_t b + \rho*(y_{t-1} - X_{t-1}b) \qquad \text{for } t > 1$$
$$e_t = y_t - \hat{y}_t$$

AR1

AR1 also stores this regression output in data storage for later use. The table below lists the results available after an AR1 command. Note: the number of coefficients (# vars) always includes RHO (ρ).

Name	Type	Length	Variable Description
@RNMS	list	#vars	Names of right hand side variables.
@LHV	list	1	Name of the dependent variable.
@RHO	scalar	1	Serial correlation parameter at convergence.
@SSR	scalar	1	Sum of squared residuals.
@S	scalar	1	Standard error of regression.
@YMEAN	scalar	1	Mean of the transformed dependent variable.
@SDEV	scalar	1	Standard deviation of the dependent variable.
@NOB	scalar	1	Number of observations.
@DW	scalar	1	Durbin-Watson statistic.
@RSQ	scalar	1	R-squared.
@ARSQ	scalar	1	Adjusted R-squared.
@IFCONV	scalar	1	1 if convergence achieved
@LOGL	scalar	1	Log of likelihood function.
@COEF	vector	#vars	Coefficient estimates.
@SES	vector	#vars	Standard Errors.
@T	vector	#vars	t-statistics.
@VCOV	matrix	#vars*#vars	Variance-covariance of estimated coefficients.
@RES	series	#obs	Fitted residuals from model.
@FIT	series	#obs	Fitted values of dependent variable.

If the regression includes PDL variables, @SLAG, @MLAG, and @LAGF will also be stored (see OLSQ for details).

Method:

AR1 uses an initial grid search to local possible multiple local optima (when appropriate - for OBJFN=GLS), and then iterates efficiently to a global optimum with second derivatives. The likelihood function and treatment of the initial observation are described completely in Davidson and MacKinnon (1993).

References:

Beach, Charles M. and James G. MacKinnon, "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," **Econometrica 46,** 1978, pp. 51-58.

Buse, A., "Efficient Estimation of a Structural Equation with First Order Autocorrelation," **Journal of Quantitative Economics** 5, January 1989, pp. 59-72.

Cochrane, D. and G. H. Orcutt, "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," **JASA 44,** 1949, pp. 32-61.

Davidson, Russell, and James G. MacKinnon, **Estimation and Inference in Econometrics**, Oxford University Press, New York, NY, 1993, Chapter 10. (This is the best single reference)

Dufour, J-M, M. J. I. Gaudry, and T. C. Liem, "The Cochrane-Orcutt Procedure: Numerical Examples of Multiple Admissible Minima," **Economics Letters 6**, 1980, pp.43-48.

Fair, Ray C., "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," **Econometrica 38,** 1970, pp. 507-516.

Fair, Ray C., **Specification, Estimation and Analysis of Macroeconomic Models**, Harvard University Press, Cambridge, MA, 1984.

Hildreth, C. and J. Y. Lu, "Demand Relations with Autocorrelated Disturbances," **Research Bulletin 276,** Michigan State University Agricultural Experiment Station, 1960.

Judge et al, **The Theory and Practice of Econometrics,** John Wiley & Sons, New York, 1981, Chapter 5.

Maddala, G. S., **Econometrics,** McGraw Hill Book Company, New York, 1977, pp. 274-291.

Pindyck, Robert S., and Daniel L. Rubinfeld, **Econometric Models and Economic Forecasts,** McGraw Hill Book Company, New York, 1976, pp. 106-120.

Prais, S. J. and C. B. Winsten, "Trend Estimators and Serial Correlation," **Cowles Commission Discussion Paper No. 373**, Chicago, 1954.

Rao, P. and Z. Griliches, "Small Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors," **JASA 64,** 1969, pp. 253-272.