BJEST

BJEST (CONSTANT, CUMPLOT, EXACTML,
NAR=*number of AR parameters*,
NBACK=*number of back-forecasted residuals* ,NDIFF=*degree of differencing*,
NLAG=*number of autocorrelations* ,NMA=*number of MA parameters*,
NSAR=*number of seasonal AR params*, NSDIFF=*degree of seasonal differencing,*
NSMA=*number of seasonal MA parameters*,
NSPAN=*span of seasonal*, PLOT, RELAX, ROOTS, START, nonlinear options)
[*series name*] [START *parameter name parameter value* .....]
[FIX *parameter name parameter value* .....]
[ZERO *parameter names* .....] [ZFIX *parameter names* .....] ;

Function:

BJEST estimates the parameters of an ARIMA (**A**uto**R**egressive **I**ntegrated **M**oving **A**verage) univariate time series model by the method of conditional or exact maximum likelihood. The technical details of the method used are described in Box and Jenkins, Chapter 7. TSP uses the notation of Box and Jenkins to describe the time series model.

Usage:

To estimate an ARIMA model, use the BJEST command followed by the options you want in parentheses and then the name of the series and specification of the starting values, if you want to override the default starting values.

The general form of the model which is estimated is the following:

$$\Gamma(B)\Phi(B)w_t = \Delta(B)\Theta(B)a_t + \text{constant}$$

w_t is the input series after ordinary and seasonal differencing, and a_t is the underlying "white noise" process. B is the "backshift" or lag operator:

$$Bw_t = w_{t-1} \qquad [= Lw_t]$$

and $\Gamma(B)$, $\Phi(B)$, $\Delta(B)$, and $\Theta(B)$ are all polynomials in B. The order of these polynomials is specified by the options NSAR, NAR, NSMA, and NMA respectively. $\Gamma(B)$ and $\Delta(B)$ are the seasonal AR and MA polynomials and $\Phi(B)$ and $\Theta(B)$ are the ordinary AR and MA polynomials. All the polynomials are of the form

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \ldots$$ (Note that the coefficients all have minus signs in front of them.)

The BJEST procedure uses conditional sum of squares estimation to find the best values of the coefficients of these polynomials, consistent with the unobserved variable a_t being independently identically distributed. The options specify the exact model which is to be estimated and can also be used to control the iteration process.

As with the other iterative estimation techniques in TSP, it is helpful to specify reasonable starting values for the parameters. Unlike the other nonlinear TSP procedures, the starting values for BJEST can be specified directly in the command, not with a PARAM statement, since the parameters have certain fixed predetermined names in the Box-Jenkins notation.

There are several ways to specify starting values in BJEST. The easiest way is to let TSP "guess" reasonable starting values for the parameters. TSP will choose starting values, for the non-seasonal parameters only, based on the autocorrelations of the time series. You can suppress this feature with NOSTART. If no other information is supplied

(see below), TSP will use zero as the starting value for all the parameters. In practice, this generally leads to slower convergence of the parameter estimates. You can also supply starting values in an @START vector.

You may also specify the starting values for any or all of the parameters yourself. For a model with more than one or two parameters, it may be difficult to choose starting values, since the individual parameters often do not possess a simple interpretation. However, you may wish to specify the starting values of at least some of the parameters based, perhaps, on previous estimates. In this case any parameters for which no starting value is supplied will be given the default value (TSP's guess if the START option is on, or zero for NOSTART, or the @START value).

User-supplied starting values are given after the name of the series on the command. Follow the series name with the keyword START and then a series of pairs of parameter names and starting values. The names available are:

AR(lag) or PHI(lag)	an ordinary autoregressive parameter.
MA(lag) or THETA(lag)	an ordinary moving average parameter.
SAR(lag) or GAMMA(lag)	a seasonal autoregressive parameter.
SMA(lag) or DELTA(lag)	a seasonal moving average parameter.

The lag in parentheses should be an actual number giving the position of the coefficient in the lag polynomial, i.e., AR(1) means the coefficient which multiplies the series lagged once (AR(0) is always unity).

Sometimes it is useful to fix a parameter at a certain value while others are being estimated. This may be done to limit the number of parameters being estimated, or to incorporate prior information about a parameter value, or to isolate an estimation problem which is specific to one or a few parameters. The keyword FIX followed by pairs of parameter names and values can be used to achieve this. FIX and START can be used on the same BJEST command to give starting values to some parameters and hold others fixed. The ZERO keyword is used to override the automatic starting values for some parameter(s) with zero(s). The ZFIX keyword fixes some parameter(s) to zero.

BJEST can also be used to check roots of polynomials for stationarity/invertibility, without doing any estimation. Just supply the coefficient values in @START, use NAR=p and/or NMA=q, and do not supply a dependent variable. The PRINT option will print the roots, or the @ARSTAT and @MAINV variables can be used.

Options:

CONSTANT/**NOCONS** specifies whether a constant term is to be included in the model.

CUMPLOT/NOCUMPLO specifies whether a cumulative periodogram of the residuals is to be plotted. The number of computations required for this plot goes up with the square of the number of observations, so that it may be better to forego this option if the number of observations is large.

EXACTML/**NOEXACT** specifies exact (versus conditional) maximum likelihood estimation. EXACTML is recommended for models with a unit root in the MA polynomial. This option always imposes stationarity of the AR polynomial(s) and invertibility MA polynomial(s).

NAR= the number of autoregressive parameters in the model. The default* is zero.

NBACK= the number of back-forecasted residuals to be calculated. The default* is 100.

NDIFF= the degree of differencing to be applied to the series. The default* is zero.

NLAG= the number of autocorrelations/Q-statistics to calculate. The default* is 20.

NMA= the number of moving average parameters to be estimated. The default* is zero.

NSAR= the number of seasonal autoregressive parameters to be estimated. The default* is zero.

NSDIFF= the degree of seasonal differencing to be applied to the series. The default* is zero (no differencing).

NSMA= the number of seasonal moving average parameters to be estimated. The default* is zero.

NSPAN= the span (number of periods) of the seasonal cycle, i.e., for quarterly data, NSPAN should be 4. The default is the current frequency (that is, 1 for annual, 4 for quarterly, 12 for monthly).

PLOT/NOPLOT specifies whether the residuals are to be plotted.

RELAX/**NORELAX** relaxes the stationarity and invertibility constraints on the lag polynomials. This cannot be used with EXACTML.

ROOTS/**NOROOTS** prints the polynomial roots, without the other output produced by PRINT.

START/NOSTART specifies whether the procedure should supply its own starting values for the parameters.

Nonlinear options control iteration and printing. They are explained in the NONLINEAR section of this manual.

* For all the Box-Jenkins procedures (BJIDENT, BJEST, and BJFRCST), TSP remembers the options from the previous Box-Jenkins command (except for nonlinear options), so that you only need to specify the ones you want to change.

Examples:

This example estimates a simple ARMA(1,1) model with no seasonal component; no plots are produced of the results.

```
BJEST (NAR=1,NMA=1,NOPLOT,NOCUMPLO) AR9MA5 ;
```

This example uses the Nelson (1973) auto sales data; a logarithmic transformation of the series is made before estimation.

```
GENR LOGAUTO = LOG(AUTOSALE) ;
BJEST (NDIF=1,NSDIFF=1,NMA=2,NSMA=1,NSPAN=12, NBACK=15)  LOGAUTO
   START THETA(1) 0.12 THETA(2) 0.20 DELTA(1) 0.82 ;
```

Note that NBACK is specified as 15 since the backcasted residuals fall to exactly zero after NSPAN+NSMA+NMA periods in a pure moving average model.

The next example estimates a third order autoregressive process with one parameter fixed:

```
BJEST GNP START PHI(2) -0.5 PHI(3) 0.1 FIX PHI(1) 0.9 ;
```

The model being estimated is

$$GNP_t - 0.9*GNP_{t-1} - \phi_2*GNP_{t-2} - \phi_3*GNP_{t-3} = a_t$$

Exact ML estimation

```
MMAKE @START .1 .1 .1;
BJEST(NAR=2,NMA=1,NDIFF=1,EXACTML) Y;
```

Output:

The output of BJEST begins with a printout of the starting values, followed by an iteration log with one line per iteration giving the value of the objective function and the convergence criterion. If the PRINT option is on, the convert values of the options and the exact time series process being used are printed. In the iteration log, parameter values and changes

are printed for each iteration when PRINT is on.

When convergence of the iterative process has been achieved or the maximum number of iterations reached, a message to that effect is printed, and the final results are displayed. These include the conventional statistics on the model: the standard error of a_t, the R-squared and F-statistic for the hypothesis that all the parameters are zero. The parameter estimates and their standard errors are shown in the usual regression output.

If the order of any polynomial is greater than one, its roots are shown so that you can check that they are outside the unit circle (as is required for stationarity). A table of the autocorrelations and Ljung-Box modified Q-statistics of the residuals is printed after this. If the PRINT option is on, the exact model estimated is again printed in lag notation, along with some summary statistics for the residuals. Then comes a printout of the lagged cross correlations between the differenced series w_t and the white noise (residual) series a_t.

If the PLOT option is on, a time series plot of the residuals (a_t) is printed. If the CUMPLOT option is on, a normalized cumulative periodogram is also plotted (see the CUMPLOT option above).

The following variables are stored (statistics based on differenced variable if differencing is specified):

Name	Type	Length	Variable Description
@LHV	list	1	Name of the dependent variable.
@RNMS	list	#vars	Names of right hand side parameters.
@SSR	scalar	1	Sum of squared residuals.
@S	scalar	1	Standard error of the regression.
@S2	scalar	1	Standard error squared.
@YMEAN	scalar	1	Mean of the dependent variable.
@SDEV	scalar	1	Standard deviation of the dependent variable.
@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.
@NCOEF	scalar	1	Number of coefficients.
@NCID	scalar	1	Number of identified coeficients.
@COEF	vector	#coefs	Coefficient estimates.
@SES	vector	#coefs	Standard Errors.
@T	vector	#coefs	T-statistics on coefficients.
@GRAD	vector	#coefs	Values of the gradient at convergence.
@VCOV	matrix	#coefs* #coefs	Variance-covariance of estimated coefficients.
@QSTAT	vector	NLAG	Ljung-Box modified Q-statistics
%QSTAT	vector	NLAG	P-values for Q-statistics
@ARSTAT	scalar	1	1 if AR polynomial is stationary
@MAINV	scalar	1	1 if MA polynomial is invertible
@FIT	series	#obs	Fitted values of the dependent variable.
@RES	series	#obs	Residuals=actual - fitted values of the dependent variable.

Method:

The method used by BJEST (for the default NOEXACTML) to estimate the parameters is essentially the one described by Box and Jenkins in their book. It uses a conventional nonlinear least squares algorithm with numerical derivatives. The major difference between the estimation of time series models and estimation in the traditional (nonlinear least squares) way relates to the use of "back-forecasted" residuals. The likelihood function for the ARIMA model depends on the infinite past sequence of residuals. If we estimate the time series model by simply setting the values of these past residuals to zero, their unconditional expectation, we might seriously misestimate the parameters if the initial disturbance, a_0, happens to be very different from zero.

The solution to this problem, suggested by Box and Jenkins, is to invert the representation of the time series process, i.e., write the same process as if the future outcomes were determining the past. Thus it describes the relationships which the time series will, **ex post**, exhibit. This representation of the backward process constructs back-forecasts of the disturbance series, the a_t series and then uses these calculated residuals in the likelihood function. By using a reasonable number of these backcasted residuals, the problems introduced by an unusually high positive or negative value for the first disturbance in the time series can be eliminated. The method described above is known as the "unconditional SSR" method, and it was used in TSP 4.2 and earlier. In TSP 4.3, BJEST was changed to use the "conditional SSR" method, which does not count the sum of squared backforecasted residuals in the likelihood function. In TSP 4.4, the EXACTML method was made available, which is superior to the "unconditional SSR" method.

If the process is a pure moving average process, this backcast becomes zero after a fixed number of time periods; consequently, you can set NBACK to a fairly small number in this case.

When the EXACTML option is used, no backforecasting is done; the AS 197 algorithm (Mélard 1984) is used.

References:

Box, George P. and Gwilym M. Jenkins, **Times Series Analysis: Forecasting and Control**, Holden-Day, New York, 1976.

Ljung, G.M., and Box, George P., "On a measure of lack of fit in times series models," **Biometrika 66**, 1978, pp. 297-303.

Mélard, G., "Algorithm AS 197: A Fast Algorithm for the Exact Likelihood of Autoregressive-moving Average Models," **Applied Statistics**, 1984, p.104-109. (code available on StatLib)

Nelson, Charles, **Applied Times Series Analysis for Managerial Forecasting**, Holden-Day, New York, 1973.

Pindyck, Robert S. and Daniel L. Rubinfeld, **Econometric Models and Economic Forecasts**, McGraw-Hill Book Co., New York, 1976, Chapter 15.

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