Chapter 11
ESTIMATION USING TIME SERIES DATA

The analysis of discrete time series data is central to econometrics, particularly macroeconometrics. The name TSP itself (**T**ime **S**eries **P**rocessor) recognizes this fact. All procedures in TSP are designed to operate on time series as well as on other kinds of data. This chapter describes the procedures in TSP specific to time series data, and gives some hints on working with such data.

You have already encountered the simplest of such procedures in AR1: the Almon and Shiller distributed lag variables, and the Durbin-Watson and Durbin (1970) test procedures for autocorrelation of the disturbances of a regression model. This chapter describes many more: identification, estimation, and simulation of a simple time series process using Box-Jenkins techniques, estimation of a vector (the VAR procedure), estimation of GARCH-M (Generalized Auto-Regressive Conditional Heteroskedasticity with a conditional Mean term) models, estimation with a Kalman Filter, and testing for unit roots and cointegration (COINT/UNIT).

We first review the basics of time series in TSP (such as operations on lags and leads), and then discuss methods for estimating a single time series process using the Box-Jenkins (ARIMA) methodology. This is followed by descriptions of two other models that have recently been widely used in the estimation of time series models: the GARCH-M model, which allows for conditional heteroskedasticity of the disturbances, and the Kalman Filter model, which is a form of time-varying parameter model. We then discuss the use of vector autoregressions (VARs) to estimate dynamic linear relationships among several time series variables. The last subject is unit root and cointegration testing.

11.1. Techniques for time series data

Chapter 3 introduced the basic features that make time series data easy to handle in TSP: the FREQ command and lagged variables. This section reviews these concepts and gives a bit more information on using time series in TSP.

FREQ, which specifies the frequency of time series data, enables you to specify observations using a convenient date format such as 75:4 for the 4th quarter of 1975 (rather than having to use the sequence number of the observation). Currently, TSP allows the use of annual, quarterly, monthly, weekly or undated data. A frequency is a characteristic permanently attached to each time series (and stored with the series in any databank), so that TSP can always check that you are using series that conform (are of the same frequency) to the current working FREQ and SMPL you have specified. You can convert series from one frequency to another with the CONVERT command (see Section 11.1.1).

It is easy to lead or lag a variable in TSP. The notation X(-1), X(-2), etc. means the observation on X is one, two, or more periods prior to the current one. In computing this, the current SMPL is ignored. That is, if the SMPL is 1960 to 1972, 1976 to 1990, the value of X(-1) in 1976 is that for 1975, *not* that for 1976. In this case, if you specify X(-1) and the observation for 1975 is missing, you will receive a warning for some procedures and an error for others. This can cause some confusion, since the missing observation is outside the current sample.

The notation X(+1) or X(1) denotes the observation on X one period after the current one. Missing values are handled in the same way as for lags. One feature of TSP you may find useful if you want to include many lags in a regression or other procedure is the ability to specify a variable list with lags. For example, the expression X(-1)-X(-16) means all the variables X(-1) X(-2) X(-3) ...X(-15) X(-16). See LIST in the *Reference Manual* for more on specifying variable lists.

11.1.1. Changing the frequency of a series: CONVERT

Sometimes data is organized as monthly observations but the best frequency for analysis is quarterly. Or perhaps some of your series are quarterly but one series is available only annually. The CONVERT procedure can change a series frequency to a new frequency. It uses one of several methods: averaging the data (the default); choosing the first, last, or middle observations; or summing the data. Each method is appropriate for different kinds of data. A flow variable such as investment would be summed to obtain the total investment in the new longer period, but a stock variable might be averaged or the first or last observation chosen depending on the importance of timing in your use of the variable.

To convert a sales series from monthly to quarterly, use the following sequence of commands:

```
FREQ Q ; SMPL 75:1 82:4 ;
CONVERT (SUM) SALES ;
```

SALES was a monthly series running from 75:1 to 82:12 ; We specified the sample as quarterly and gave the new range in terms of quarters so the CONVERT procedure would know that we wanted to convert from monthly (the frequency of SALES) to quarterly (the current frequency). Because it would be confusing to mix frequencies in a series, CONVERT ignores the current sample range and converts the whole series no matter what you have specified as the SMPL.

If you want to preserve the old version of the series and save the new one, use this form of CONVERT:

```
CONVERT(SUM) SALESQ = SALES ;
```

Be aware, however, that you will not be able to use the old series SALES in other procedures after the new frequency has been specified. Series with different frequencies cannot be mixed in the same procedure (except for CONVERT)..

When converting from a lower frequency, straight line interpolation can be used to compute the intervening points. Use the INTERPOL option. The other convert options (SUM, etc.) can also be used.

```
FREQ M ; SMPL 75:1 82:12 ;
CONVERT(INTERPOL) SALESM = SALES ;
```

FREQ W (weekly) can be converted to quarterly, but not to monthly, because TSP does not have a built-in calender.

11.2. Box-Jenkins (ARIMA) models

One disadvantage of using structural econometric models for forecasting is that you need to know a great deal about the variables being modeled. In particular, to obtain a simulation over several periods in the future, you need to know the values for all the exogenous variables in the model over the forecast period. For this and other reasons, some forecasters use the ARIMA (AutoRegressive Integrated Moving Average) or Box-Jenkins forecasting method. This method only uses a series' own lagged values to forecast its future values. If the series follows a stationary stochastic process without too much drift or noise, this method can work well, at least over a short term forecast. (See Section 11.6 for discussion of testing for unit roots in the process.) ARIMA models can be thought of as a sophisticated extrapolation method.

In this section, we discuss how TSP obtains univariate ARIMA forecasts. We strongly recommend that users who are interested in this technique refer to one of the standard references for further details. The basic reference on the subject is Box and Jenkins (1976). Two other elementary books that focus on this method are Nelson (1973) and Van Daele (1983). The Pindyck and Rubinfeld text is also recommended; Section 3 is entirely devoted to time series models.

Box-Jenkins forecasting is traditionally divided into three parts: identification (determining the form of the time series process the variable follows), estimation (estimating the parameters of the process), and forecasting (extrapolating the process beyond the estimation period using the estimated parameters). In TSP these three steps correspond to the procedures BJIDENT, BJEST, and BJFRCST. Because these procedures have most of the same options, and share their values, once you have specified a set of options for BJIDENT, they are automatically assumed for any subsequent Box-Jenkins procedure, unless explicitly changed.

11.2.1. Identification: BJIDENT

Suppose you have a product's monthly sales data from 1978 through the middle of 1987, and you assume that whatever time series process generated the data is stationary (at least after differencing). To look at some characteristics of this process, you can use a BJIDENT command to display the autocorrelations and partial autocorrelations of the series:

```
BJIDENT (NDIFF=1,NSDIFF=1,NLAG=12,NLAGP=12) MSALES ;
```

Since you suspect that the original series may not be stationary, the NDIFF option specifies that the correlations be computed for the first differenced series, as well as for the original series. Similarly, the NSDIFF option specifies that seasonal differencing is to be performed. BJIDENT determines the periodicity of the seasonal factor from the frequency of your current sample; in this case the FREQ is monthly, and observations 12 periods apart would be differenced. If you have not specified a FREQ as part of your SMPL, you will have to supply the seasonal span as the NSPAN option on the BJIDENT command.

NLAG and NLAGP options specify the number of autocorrelations and partial autocorrelations to be computed, that is, the length of the lag over which they are to be computed. The default value is 20.

BJIDENT's output consists of plots of the series, both differenced and undifferenced, followed by a printout of autocorrelations and partial autocorrelations, their standard errors, and Q-statistics (Ljung-Box statistics, see Harvey (1993), p. 212) for the hypothesis that all autocorrelations of higher order are zero. Correlations are also plotted in what is called a correlogram and a partial correlogram. These can be quite useful in trying to determine the form of the process. Consult Box and Jenkins or Nelson for examples of the correlograms associated with various time series processes. The autocorrelations, partial autocorrelations, and inverse autocorrelations are also stored under the names @AC, @PAC, and @IAC.

The output of BJIDENT is given in **Example 11.1**. In our example, the series appears to be nonstationary in levels but stationary in first differences. There does not appear to be a seasonal effect; in fact, spurious correlations were introduced at the eleventh lag by seasonal differencing. The correlogram for the first differenced series suggested that the data be fit by a first order moving average error process, since there were no significant autocorrelations after the first.

11.2.2. Estimation: BJEST

Using the results of BJIDENT as a guide to the specification of the model, you can then estimate the parameters of this model with BJEST. For our example, the appropriate command might be

```
BJEST (CUMPLOT,NBACK=5,NMA=1,NSDIFF=0,NDIFF=1) MSALES ;
```

This specifies a first-order moving average model on first differenced data, with no seasonal components. Note that you do not need to specify any options that remain the same from a previous BJ command; in this case we did not need to set NDIFF to 1 if this command followed the BJIDENT command given above.

The NBACK option specifies the number of backforecasted residuals that should be used to start the process and generate the initial conditions; since we are using a moving average process of order one here, only the first will be nonzero, so NBACK can be set to a small number. In the case of autoregressive or mixed processes, you should use a much larger value. Other options control the behavior of the iteration process and the appearance of the output; consult the *Reference Manual* to learn more about them.

In BJEST's output, the usual TSP estimation results are shown for the residuals from the fitted model, along with a table of parameter estimates (only one in this case, the moving average parameter θ, followed by some statistics concerning the variable in question). This is followed by a plot of the cumulated periodogram of the residuals so you can see how well the white noise assumption of the model is satisfied. In our example, both the periodogram and the Q-statistics for serial correlation suggest that the residuals are not white noise, although they are close.

11.2.3. Forecasting: BJFRCST

Immediately following a Box-Jenkins estimation, you can perform a forecast based on the estimated time series model using the BJFRCST command. This forecast will be computed starting in any time period or range of time periods you specify; it will go forward for the number of periods given by the NHORIZ option. Confidence bounds will also be computed for the forecast. An example for the series we modeled above is:

```
BJFRCST (NHORIZ=10,ORGBEG=87:6,ORGEND=87:7) MSALES ;
```

This computes two 10-month forecasts, starting in June and July of l987. Unless requested not to, BJFRCST will also plot the forecasts and their standard error bounds.

The output from this BJFRCST is shown on the following pages in **Example 11.1.** The final model for the variable MSALES is

$$(1\text{-}L)\ MSALES_t = (1\text{-}.78L)\ e_t$$

as shown, and this model is used to generate the forecasts, using the actual value in June l987 as an initial condition for MSALES(-1). Note the 95% confidence bounds that are printed out. As may be expected, the information contributed by this model, roughly a random walk with a large measurement error, is very small, and the standard errors are correspondingly large.

If you do a forecast based on some previously estimated time series process, you can also supply the parameters of the model directly to BJFRCST. Consult the *Reference Manual* for further details on the notation (which is that of Box and Jenkins). An example for a simple moving average process is:

```
BJFRCST (NMA=1, NSDIFF=0, NDIFF=1, NHORIZ=12, ORGBEG=84:1, ORGEND=84:3)
   SALES S 4.021 THETA(1) -.802 ;
```

11.3. Auto-Regressive Conditional Heteroskedasticity

The ARCH model originated by Engle, and its many elaborations, are widely used in econometrics to estimate models of time series processes where the variance of the disturbance is dependent in a simple way on the behavior of the preceding observations, but the conditional mean of the disturbance is still equal to zero. ARCH processes appear to describe well many observed macroeconomic data series, such as exchange rates and stock market returns.

ARCH will estimate not only the basic ARCH model, but also more complicated models which allow the variance of the disturbance to follow both an autoregressive and a moving average process and to be conditional on other series. This latter feature makes ARCH suitable for heteroskedastic data of any kind, not just time series. In addition, the conditional mean of the dependent variable can depend on the standard deviation of the variance of the disturbance (the full GARCH-M model). These models are more fully discussed in Engle (1982, for ARCH), Bollerslev (1986, for GARCH and its identification), and McCurdy and Morgan (1988, for GARCH-M). A fuller discussion of models TSP can estimate is given in the *Reference Manual* (under the ARCH command).

Here is an example of simple ARCH model estimation of an index of stock prices, where there are no independent variables, and the variance of the disturbance is assumed to follow an autoregressive process of order 3:

```
ARCH(NAR=3) RM C ;
```

It estimates the following model by maximum likelihood (denoting the stock price series RM by y_t):

$$y_t = \eta_0 + \epsilon_t$$
$$\epsilon_t \sim N(0,h_t)$$
$$h_t = \alpha_0 + \alpha_1 \epsilon^2_{t-1} + \alpha_2 \epsilon^2_{t-2} + \alpha_3 \epsilon^2_{t-3}$$

The coefficients estimated by this command are η_0, the average stock return, and α_0, α_1, α_2, α_3, the parameters of the ARCH process. These coefficients are labeled ALPHA0, ALPHA1, etc. in the output. Because the estimation of ARCH models can be prone to numerical problems arising when estimated variances are negative, zero, or very large, ALPHA0 is constrained to be nonnegative, and the other ALPHAs are bounded between zero and one. See the *Reference Manual* for details on the methods used to achieve these constraints.

ARCH models can also have regressors; that is, the conditional mean of the dependent variable can be a function of other variables, just as in ordinary linear regression. For example, consider a market beta model where the return on a stock (R) is a function of the return on the market (RM), and the variance of its return follows an ARCH process:

```
     Current sample:  1978:7 to 1987:6

     ======================
     Box-Jenkins procedures
     Procedure   BJIDENT
     ======================

                              Autocorrelations
                              ================

                                          0    12 0
                                Series (1-B) (1-B  )   MSALES
                                               Mean =   6.3533574
                                     Standard Error =  0.25485519

                     Lags
Autocorrelations           0.754     0.613     0.538     0.543
Standard Errors    1- 4   0.962E-01 0.141     0.163     0.179
Q-statistics               63.1      105.      138.      172.
Autocorrelations           0.546     0.515     0.505     0.446
Standard Errors    5- 8   0.194     0.208     0.219     0.230
Q-statistics               206.      237.      267.      290.
Autocorrelations           0.367     0.360     0.417     0.568
Standard Errors    9-12   0.237     0.243     0.248     0.254
Q-statistics               307.      322.      344.      384.

                                          1    12 0
                                Series (1-B) (1-B  )   MSALES
                                               Mean =  0.70060792E-02
                                     Standard Error =  0.17201304

                     Lags
Autocorrelations          -0.242    -0.129    -0.142    -0.230E-01
Standard Errors    1- 4   0.967E-01 0.102     0.104     0.105
Q-statistics               6.45      8.30      10.6      10.6
Autocorrelations           0.579E-01-0.403E-01 0.934E-01 0.146E-01
Standard Errors    5- 8   0.106     0.106     0.106     0.107
Q-statistics               11.0      11.2      12.2      12.2
Autocorrelations          -0.167    -0.117    -0.135     0.648
Standard Errors    9-12   0.107     0.109     0.110     0.112
Q-statistics               15.5      17.2      19.4      70.9

                          Partial Autocorrelations
                          ========================

                                          0    12 0
                                Series (1-B) (1-B  )   MSALES
                  Standard Error of Autocorrelations =  0.96225045E-01

                     Lags
Partial Autocorrs  1- 4   0.754     0.103     0.107     0.203
Partial Autocorrs  5- 8   0.118     0.288E-01 0.997E-01-0.652E-01
Partial Autocorrs  9-12  -0.109     0.794E-01 0.160     0.382

                                          1    12 0
                                Series (1-B) (1-B  )   MSALES
                  Standard Error of Autocorrelations =  0.96673649E-01

                     Lags
Partial Autocorrs  1- 4  -0.242    -0.199    -0.252    -0.198
Partial Autocorrs  5- 8  -0.106    -0.163    -0.181E-01 0.489E-02
Partial Autocorrs  9-12  -0.193    -0.287    -0.511     0.345

                                                    0    12 0
```

Example 11.1: Box-Jenkins Procedures

```
      Autocorrelation Function of the series (1-B) (1-B  )   MSALES

        -1.00     -0.33     0.33      1.00
        |-+---------+----0----+---------+-|
1       |             +   |   +       R   | 0.75360
2       |            +    |    +    R     | 0.61262
3       |           +     |     +  R      | 0.53768
4       |           +     |     +  R      | 0.54307
5       |          +      |      + R      | 0.54622
6       |          +      |      + R      | 0.51510
7       |          +      |      + R      | 0.50496
8       |         +       |       R       | 0.44582
9       |         +       |      R+       | 0.36709
10      |         +       |     R +       | 0.36003
11      |         +       |      R+       | 0.41700
12      |         +       |       + R     | 0.56821
        |-+---------+----0----+---------+-|
        -1.00     -0.33     0.33      1.00


                                                      0    12 0
  Partial Autocorrelation Function of the series (1-B) (1-B  )   MSALES

        -1.00     -0.33     0.33      1.00
        |-+---------+----0----+---------+-|
1       |             +   |   +       R   | 0.75360
2       |             +   |  R+           | 0.10346
3       |             +   |  R+           | 0.10716
4       |             +   |   R           | 0.20348
5       |             +   |  R+           | 0.11780
6       |             +  R|   +           | 0.028766
7       |             +   |R  +           | 0.099730
8       |             + R |   +           | -0.065198
9       |             +R  |   +           | -0.10883
10      |             +   |R  +           | 0.079371
11      |             +   |  R+           | 0.16025
12      |             +   |   +  R        | 0.38193
        |-+---------+----0----+---------+-|
        -1.00     -0.33     0.33      1.00
                                                 1    12 0


      Autocorrelation Function of the series (1-B) (1-B  )   MSALES

        -1.00     -0.33     0.33      1.00
        |-+---------+----0----+---------+-|
1       |           R+    |   +           | -0.24219
2       |            +R   |   +           | -0.12906
3       |            +R   |   +           | -0.14156
4       |            +  R  +           | -0.022954
5       |            +   |R  +           | 0.057947
6       |            + R |   +           | -0.040291
7       |            +   |R  +           | 0.093379
8       |            +  R  +           | 0.014561
9       |            +R   |   +           | -0.16665
10      |            +R   |   +           | -0.11693
11      |            +R   |   +           | -0.13541
12      |            +    |   +       R   | 0.64762
        |-+---------+----0----+---------+-|
        -1.00     -0.33     0.33      1.00
                                                      1    12 0
  Partial Autocorrelation Function of the series (1-B) (1-B  )   MSALES
```

. . .output omitted . . .

Example 11.1: (Continued, p. 2)

```
     ======================
     Box-Jenkins procedures
     Procedure   BJEST
     ======================

     Working space used: 1339
                               STARTING VALUES

                      THETA1
     VALUE          0.24219

F=  -41.496     FNEW=  -42.823     ISQZ=  0 STEP=  1.0000     CRIT=  1.5934
F=  -44.053     FNEW=  -44.053     ISQZ=  0 STEP=  1.0000     CRIT= 0.97485E-04
F=  -44.053     FNEW=  -44.053     ISQZ=  0 STEP=  1.0000     CRIT= 0.58730E-04

CONVERGENCE ACHIEVED AFTER  20 ITERATIONS
   40 FUNCTION EVALUATIONS.

                       Results of Box-Jenkins Estimation
                       =================================

                    Statistics Based on Differenced Series
                    ======================================

     Dependent variable: MSALES

    Current sample:  1978:7 to 1987:6
Number of observations:  107

       Mean of dep. var. = .700608E-02
  Std. dev. of dep. var. = .172013
Sum of squared residuals = 2.74978
   Variance of residuals = .025941
Std. error of regression = .161063
              R-squared = .168253
      Adjusted R-squared = .168253
           LM het. test = 2.16667 [.141]
         Durbin-Watson = 1.57358
        Log likelihood = 44.0535

                        Standard
Parameter  Estimate       Error       t-statistic   P-value
THETA1     .640457      .074674       8.57668       [.000]

     Standard Errors computed from   quadratic form of analytic
     first derivatives (Gauss)
                      Autocorrelations of the Residuals

                      1             2             3
     Autocorr       0.19782      -0.12400      -0.23976
     Q-stat         4.30575       6.01375      12.46029
     P-value          .           0.014195     0.0019692


                      4             5             6
     Autocorr      -0.13902     -0.033966     -0.044417
     Q-stat        14.64857      14.78048      15.00830
     P-value      0.0021430     0.0051788      0.010327


                      7             8             9
     Autocorr       0.11677      -0.10021      -0.26080
     Q-stat        15.02421      16.20722      24.30189
     P-value       0.020069      0.023289     0.0020393


                     10            11            12
     Autocorr      -0.17599      0.092015       0.62773
     Q-stat        28.02603      29.05466      77.43045
     P-value     0.00094436     0.0012209   4.62274D-12
```

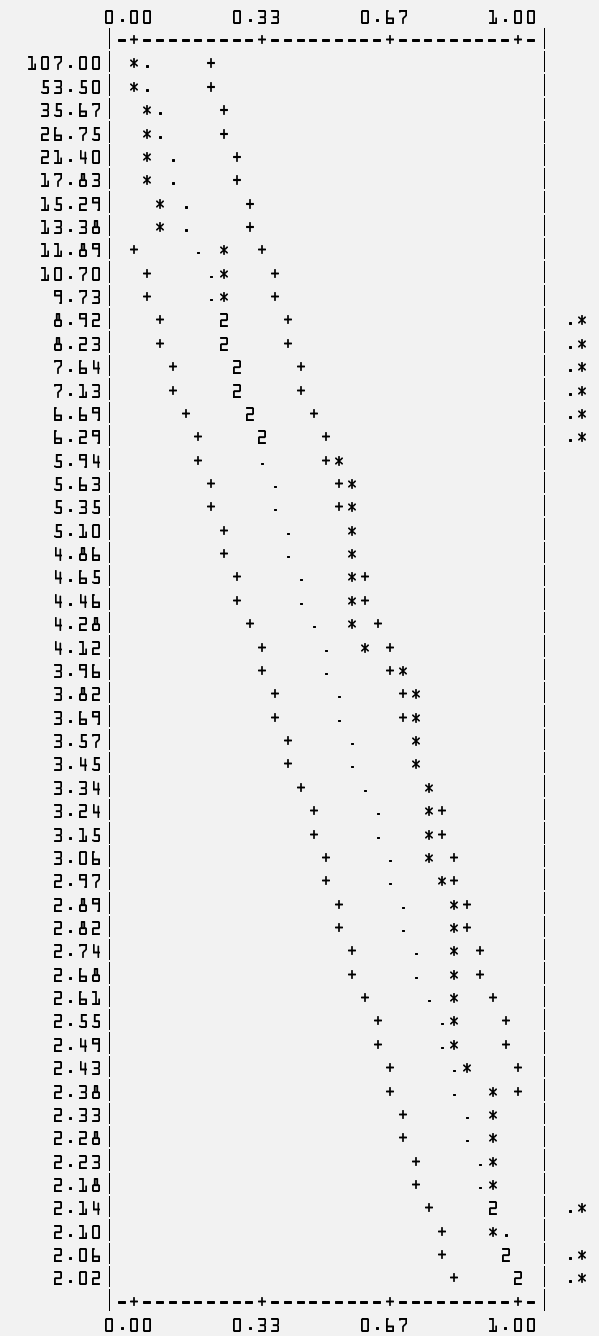
Example 11.1: (Continued, p. 3)

```
Normalized Cumulative Periodogram of Residuals

Expected CP plotted with (.)   Actual CP plotted with (*)

Band (+) marks the 10% Kolmogorov-Smirnov limits

Period
```




```
Current sample:  1978:7 to 1989:6
```

Example 11.1: (Continued, p. 4)

```
======================
Box-Jenkins procedures
Procedure   BJFRCST
======================


TIME SERIES:MSALES

STANDARD ERROR OF THE DISTURBANCE = 0.16106

THETA(B)

          1 - 0.64046 B

PHI(B)

          1 -  B


              FORECAST STANDARD ERRORS AND PSI WEIGHTS
                STD ERR            PSI
        1       0.16106       0.35954
        2       0.17116       0.35954
        3       0.18069       0.35954
        4       0.18974       0.35954
        5       0.19838       0.35954
        6       0.20666       0.35954
        7       0.21462       0.35954
        8       0.22230       0.35954
        9       0.22972       0.35954
       10       0.23690       0.35954


       Forecasts and 95% Confidence Bounds (Origin = 1987:6)

               Lowr Bnd      Forecast      Uppr Bnd
1987:6          6.69481       6.69481       6.69481
1987:7          6.30942       6.62509       6.94077
1987:8          6.28963       6.62509       6.96056
1987:9          6.27095       6.62509       6.97924
1987:10         6.25321       6.62509       6.99698
1987:11         6.23627       6.62509       7.01391
1987:12         6.22005       6.62509       7.03014
1988:1          6.20444       6.62509       7.04574
1988:2          6.18940       6.62509       7.06079
1988:3          6.17486       6.62509       7.07533
1988:4          6.16077       6.62509       7.08941


                 Actual         Error
1987:6          6.69481       0.00000
1987:7          6.42341      -0.20168
1987:8          6.28953      -0.33556
1987:9          6.50847      -0.11662
1987:10         6.66708       0.041991
1987:11         6.71113       0.086039
1987:12         6.69159       0.066494
1988:1          6.62446   -0.00062958
1988:2          6.58714     -0.037956
1988:3          6.31028      -0.31481
1988:4          6.56541     -0.059687
```

Example 11.1: (Continued, p. 5)

```
The Series is plotted with "A"
The Forecast is plotted with "F"

          MINIMUM                          MAXIMUM
        5.86005                          7.09995
          |-+------------------------------+-|
1986:6    |                             A    |
1986:7    |                 A                |
1986:8    |                     A            |
1986:9    |                         A        |
1986:10   |                        A         |
1986:11   |                       A          |
1986:12   |                      A           |
1987:1    |                        A         |
1987:2    |                      A           |
1987:3    |                A                 |
1987:4    |                       A          |
1987:5    |                      A           |
1987:6    |                        A         |
1987:7    |              *  A     F        *  |
1987:8    |             A*        F        *  |
1987:9    |             *      A  F         * |
1987:10   |             *         FA        * |
1987:11   |             *         F A       * |
1987:12   |            *          FA         *|
1988:1    |            *         AF          *|
1988:2    |           *          AF           *
1988:3    |           *  A       F            *
1988:4    |           *         A F           *
          |-+------------------------------+-|
        5.86005                          7.09995
          MINIMUM                          MAXIMUM
```

Example 11.1: (Continued, p. 6)

```
ARCH(NAR=3) R C RM ;
```

Because of ARCH's ability to estimate a model where the variance of the disturbance depends on a set of regressor variables, ARCH can also be used to estimate an ordinary weighted least squares model by ML, even when there is no dependence across the observations. For example, consider a model with size-related heteroskedasticity:

$$y_t = \eta_0 + \eta_1 x_{1t} + \eta_2 x_{2t} + \epsilon_t$$
$$\epsilon_t \sim N(0, h_t)$$
$$h_t = a_0 + a_1 g_t$$

where g_t is the size of the t^{th} observation.

ARCH can estimate this model (called OLS-W by the program) with the command

```
ARCH (GT=G) Y C X1 X2 ;
```

G denotes the series that contains the size variable. The GT option is used to specify a list of series that will enter into the regression function for the variance h_t. Note that the constant (C) is automatically included in this list via the a_0 parameter, so you should not include it.

11.4. The Kalman Filter (KALMAN)

Strictly speaking, the term Kalman Filter refers to an estimation method commonly used to estimate "state-space" models, rather than the model itself. This class of models consists of two parts: the transition equation, which describes the evolution of a set of state variables, and the measurement equation, which describes how the data actually observed is generated from the state variables. Its importance in economics is partly due to its ability to model time-varying parameters in an intuitively appealing way.

In addition, the Kalman Filter estimation method is an updating method that bases the regression estimates for each time period on last period's estimates plus the data for the current time period; that is, it bases estimates only on data up to and including the current period. This makes it useful for investigating structural change in parameters or constructing forecasts based only on historical data.

As with many time series methods economists use (such as ARIMA models), State Space Models originated in the engineering literature (Kalman 1960) and were imported into economics by Rosenberg (1968) among others. A good reference is Harvey (1981); there is also a special issue of the *Annals of Economic and Social Measurement* (October 1973) on time-varying parameters. A more elementary reference is Maddala (1977), Chapter 17.

Using the notation in Harvey (1981), the model KALMAN estimates can be written in the following way (assuming a single dependent variable for simplicity):

$$y_t = Z_t\alpha_t + \xi_t$$
$$\alpha_t = T\alpha_{t-1} + \eta_t$$
$$\xi_t \sim N(0,\sigma^2)$$
$$\eta_t \sim N(0,\sigma^2 Q_t)$$
$$\alpha_0 \sim N(a_0,\sigma^2 P_0)$$

y_t is the dependent variable and there are m independent variables Z_t. The first equation is an ordinary regression equation with time-varying parameters. The second equation defines the evolution of these parameters. Note that matrices S_t and R_t in Harvey's notation have been set to identity matrices and matrix T_t to a constant matrix in this use of the Kalman Filter.

When estimating this model in TSP, you can optionally supply the matrices T (the transition matrix BTRANS), Q_t (the variance of the transition equation VTRANS), P_0 (the variance of the prior distribution on the parameter vector VBPRIOR), and the vector a_0 (the prior on the coefficients BPRIOR). If you fail to supply any of them, reasonable defaults will be used.

For example, the simplest KALMAN command looks like this:

```
KALMAN CONS C GNP ;
```

The above command estimates the regression of consumption on GNP in a recursive manner, allowing the coefficient of GNP and the constant to evolve as random walks (with a signal-to-noise ratio of 1). These coefficients are the model's "state" variables, and consumption is the measured variable. The intercept and GNP series are treated as known fixed constants. In Harvey's notation, the model is

$$y_t = Z_t\alpha_t + \xi_t$$
$$\alpha_t = \alpha_{t-1} + \eta_t$$
$$\eta_t \sim N(0,\sigma^2 I)$$

α_0 will be estimated from the first m data observations, where m is the number of coefficients in α_t.

A slight generalization of this random walk parameter model is the Cooley-Prescott (1973) adaptive regression model, which has had some success in forecasting. You can estimate this model with the command

```
KALMAN(VTRANS=SIGMAV) CONS C GNP ;
```

SIGMAV corresponds to Cooley and Prescott's Σ_v, the variance of the errors in the transition equation. In the absence of specific prior information on the form of this matrix, they suggest using a diagonal matrix whose elements represent the relative variability of the regression's different coefficients. This idea could also be used to mix time-varying parameters with constant parameters (by setting their variances to zero) in the same regression.

KALMAN always calculates recursive residuals; these residuals can also be obtained with OLSQ when the REGOPT(CALC) RECRES; command is in effect. However, KALMAN can also print the evolving coefficients of the recursive model:

```
KALMAN(NOETRANS,PRINT) CONS C GNP ;
```

The *Reference Manual* discusses the KALMAN procedure further; useful features of the KALMAN command not discussed here but described in that manual include the stochastically convergent parameter model, the use of priors in estimating Kalman Filters, and the ability to estimate several measurement equations simultaneously.

11.5. Vector Autoregressions (VAR)

A vector autoregression model is the unconstrained reduced form of a dynamic simultaneous equations model; that is, it expresses a vector of endogenous variables as linear functions of their own and each other's lagged values. Contemporaneous and lagged exogenous variables may also be included in the system. This style of simultaneous equation modeling was introduced into econometrics by Sims (1980) and is now widely used for small to medium-sized macroeconometric models, particularly for forecasting.

Estimation of an unconstrained vector autoregression is quite straightforward, even in the presence of contemporaneous correlation of the disturbances. Consider the following VAR:

$$Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + ... + GX_t + E_t$$

where Y_t is an n by 1 vector of endogenous variables, the B's are n by n matrices of coefficients, G is an n by m matrix of coefficients, X_t is an m by 1 vector of exogenous variables, and E_t is an n by 1 vector of disturbances. Because the same list of right-hand side variables (all the lags of the Y_t) appear in the n equations, this set of equations can be estimated consistently and *efficiently* by ordinary least squares; there is no need for joint estimation.

Although you could use OLSQ to estimate a VAR model one equation at a time (listing all the relevant lagged variables for each estimation), TSP also provides the VAR command to automate this process. VAR also provides some regression output specific to the VAR methodology, such as impulse response functions and forecast error variance decompositions. For example, suppose you wish to estimate the vector autoregression for the six variables money, real GNP, unemployment, wages, price level, and import prices described in the Sims paper. You would use:

```
VAR(NLAGS=4) M,RGNP,U,W,P,PM ;
```

The above command produces as output the regression coefficients for the six ordinary leastsquares regressions of each dependent variable on four lags each of M, RGNP, U, W, P, and PM (24 coefficients for each equation), along with the log of the likelihood function under the multivariate normal assumption on the disturbances. In addition, the program will display (and store under the name @IMPRES) the response of each endogenous variable to Choleski-factored shocks in the given order over ten periods (the impulse response function).

Exogenous variables may be included in the vector autoregression, and calculation of the impulse response function can be modified with options. For example, the command

```
VAR(NLAGS=2,NHORIZ=5,SHOCK=UNIT) M GNP | C CONS P ;
```

specifies that M and GNP are regressed on the lagged variables M(-1), M(-2), GNP(-1), GNP(-2), the constant, CONS, and P. The impulse response function will be calculated for only five time periods and a unit shock to the endogenous variables will be used. Except for computation of the impulse response, the VAR command above is equivalent to:

```
OLSQ M   M(-1) M(-2) GNP(-1) GNP(-2) C CONS P ;
OLSQ GNP M(-1) M(-2) GNP(-1) GNP(-2) C CONS P ;
```

VAR automatically computes block exogeneity tests to see if lags of the other endogenous variables enter a given equation significantly. In the bivariate example above, these are Granger causality tests (F tests) for GNP(-1) and GNP(-2) in the M equation, and for M(-1) and M(-2) in the GNP equation.

A common use of VARs is as a base case for more restricted structural models of lag relationships of a set of endogenous variables. With TSP's ANALYZ procedure it is possible to calculate an asymptotic test statistic for a joint hypothesis

on the reduced form VAR coefficients. This hypothesis can be linear or nonlinear. For example, suppose you wish to test that the coefficients on the lagged Ms and GNPs are proportional in the two VAR equations given above. You could use the following set of commands to perform the task:

```
VAR(NLAGS=2) M GNP | C CONS P ;
UNMAKE @COEF B11-B14 G11-G13 B21-B24 G21-G23 ;      ? Give the estimated coefficients names
FRML PREQ1 B11/B21 - B12/B22 ;                      ? Proportionality constraints
FRML PREQ2 B12/B22 - B13/B23 ;
FRML PREQ3 B13/B23 - B14/B24 ;
ANALYZ PREQ1-PREQ3 ;                                ? Test with 3 degrees of freedom
```

11.6. Testing for Unit Roots and Cointegration: COINT

Simply stated, the test for a unit root in a time series y_t is the test that a regression of y_t on y_{t-1} yields a coefficient of one. This test is complicated by several features arising from the nonstationarity of y_t under the null hypothesis: 1) The ordinary t-statistic does not have the usual distribution, so you cannot use tables of t-statistics to find its p-value. 2) The correct distribution depends on nuisance parameters in the regression, in particular, whether the constant or the time trend is included.

In a well-known paper, Dickey and Fuller (1979) suggest a method for computing a test for a unit root in a time series, and present critical values for their proposed tests with and without a trend variable included. The method consists of running the stationary regression y_t-y_{t-1} on y_{t-1} either with or without a constant and time trend and testing whether the coefficient of y_{t-1} is zero. They provide the appropriate critical values for such a test in a table. Since the distribution of the resulting t-statistic generally depends on the value of the intercept in the model unless a time trend is included, most researchers choose to include both a constant and a time trend and then use the tables appropriate for that case.

Since then, a large literature on unit root tests has appeared, describing many alternative tests (some of which are variations of the above). The COINT command in TSP can compute 3 different types of unit root tests: the Dickey-Fuller (tau), Phillips-Perron (Z, "nonparametric"), and Weighted Symmetric. Each allows for various exogenous variables like time trends and seasonal dummies/trends, and each allows for a series of "augmenting" lags to control for additional serial correlation. See the references and the *Reference Manual* under COINT for further information.

For example, suppose you want to test whether the logarithm of consumption in Chapter 3's Illustrative Model has a unit root (is integrated of order 1). The following set of statements will perform the (augmented) Weighted Symmetric and Dickey-Fuller tests and print their approximate (asymptotic) P-values:

```
SMPL 49 75 ;
COINT LCONS;
```

In this example, COINT chooses 2 augmenting lags for the WS test and 10 lags for the DF; the P-values are .68 and .74 respectively, so the null hypothesis of a unit root would not be rejected at the .05 level.

Alternately, the Dickey-Fuller test could be computed "by hand"!

```
SMPL 50 75 ;
DCONS = LCONS-LCONS(-1) ;
OLSQ DCONS  LCONS(-1) C TIME;
CDF(DICKEYF) @T(1) ;
```

This example assumes that there is no further serial correlation since it does not add lagged y differences to the model. The resulting statistic was -1.84 with a corresponding asymptotic P-value of .69, so the null of a unit root is accepted at the .05 level. Note that if we had used the conventional t-table to evaluate this P-value, we might have rejected this hypothesis. Options for CDF allow you to compute the P-values without assuming the presence of a trend or constant. See the references and the *Reference Manual* for further details. Also be aware that the residuals from the Dickey-Fuller regression should be serially uncorrelated for the test to be valid, although they do not generally need to be homoskedastic (Phillips 1987). The Weighted Symmetric test is recommended over the Dickey-Fuller test, because it

has (sometimes only slightly) higher power (see Pantula, Gonzalez-Farias, and Fuller 1994). That is, the WS test is more likely to reject the unit root (null hypothesis) when it is in fact false. The Phillips-Perron test is a variant of the Dickey-Fuller which tackles the problem of additional serial correlation in the residuals by using a GMM-type method to compute a residual variance that is "robust" to autocorrelation.

The cointegration of time series is a methodology pioneered by Engle and Granger (1987). Two or more time series are said to be *cointegrated* if a linear combination of them is I(0) (is stationary, or has all roots outside the unit circle) even though individually they are each I(1). Thus the hypothesis of cointegration consists of two parts: tests for I(1) of the individual series and I(0) of a linear combination. Usually the term cointegration testing refers only to the second part of the hypothesis; the test is performed *conditional* on the fact that each component series is I(1). Although this hypothesis sounds different from the hypothesis of a unit root, the practice of testing for cointegration is quite similar. TSP gives the P-values for the Engle-Granger versions of these tests in the CDF procedure under the DICKEYF option.

As an example, consider testing that logarithms of real consumption and real GNP from the Illustrative Model are cointegrated. It is easy to establish that each is I(1) separately (with asymptotic P-values of .69 and .66). The TSP commands to evaluate the second part of the hypothesis are the following:

```
SMPL 49 75 ;
COINT(ALLORD) LCONS LGNP;   ? this also performs the individual unit root tests at the same time
```

When LCONS is the dependent variable of the cointegrating regression, COINT chooses 2 augmenting lags, and obtains a test statistic of -1.65, which has a P-value of .89 . When LGNP is the dependent variable, 10 augmenting lags are chosen, and the test statistic and P-value are -1.29 and .95 respectively. So the null hypothesis of a unit root in the cointegrating regression cannot be rejected at the .05 level in either test. We can conclude that the linear combination of LCONS and LGNP is not I(0), so they are not cointegrated (at this significance level).

If done manually (without augmentation, and only shown for LCONS as the dependent variable) the COINT test would look like this:

```
SMPL 49 75 ;
OLSQ LCONS LGNP C TIME ;        ? the cointegrating regression
SMPL 50 75 ;
DRES = @RES-@RES(-1) ;
OLSQ DRES @RES(-1) ;            ? Engle-Granger test
CDF(DICKEYF,NVAR=2) @T ;        ? (Dickey-Fuller test on residuals from the cointegrating regression)
```

In the above example, consumption is regressed on a constant, time, and GNP to obtain the cointegrating vector, residuals are constructed, and then the first-differenced residuals are regressed on the lagged residual. Under the hypothesis of stationarity, the coefficient on this variable should be -1; the t-statistic for this hypothesis is the Engle-Granger statistic. One complication is that the actual value of the Engle-Granger statistic (although not its distribution) will be affected by the choice of left-hand variable in the first regression (consumption or GNP); COINT with the ALLORD option computes both tests.

To compute the asymptotic P-value manually for the Engle-Granger statistic, use the DICKEYF option of the CDF procedure with the NVAR option to specify the number of cointegrating variables used in computing the test statistic. TSP provides P-values for cointegrating regressions with up to 6 variables, using the response surface estimates given by MacKinnon (1990, 1994). For this example, the value of the statistic was -3.07, with a P-value equal to .23. If we had put GNP on the left-hand side of the original regression, the corresponding value of the statistic would have been -3.29, with a P-value of .16. The null hypothesis of a unit root in the cointegrating regression would not be rejected at the .05 level, so we could not conclude that these variables are cointegrated. Note that these P-values are much lower than the more correct ones given above, which include augmentation (correction for serial correlation of the residuals of the cointegrating relationship). Note also that the test can be sensitive to the estimation sample. If 50,75 is used for the cointegrating regression, the unaugmented P-values are .00097 and .029, which would lead to the conclusion that cointegration does exist.

A second type of cointegration test performed by the COINT command is the Johansen-Juselius (maximum likelihood) test. This involves testing for a particular restriction on the coefficient matrix of lagged dependent variables in a VAR.

It is actually estimated by running two VARs and obtaining the eigenvalues for a function of their joint residual covariance matrix. Then likelihood ratio tests (with finite sample corrections) are made to check for the number of cointegrating vectors of the original system. The Johansen-Juselius test is often "oversized" (i.e. a P-value of .01 may be printed when the true rejection frequency should be .10 or so), implying that "too much" cointegration (or too many cointegrating vectors) tend to be found. For example:

```
SMPL 58:2 84:3;
COINT(JOH,SEAS,NOTREND,NOEG,NOUNIT,MAXLAG=1,MINLAG=1) Y1-Y4;
```

This command computes the trace tests for the Finnish data from the Johansen-Juselius(1990) paper. Note that COINT computes the tests with the finite sample adjustment, but this doesn't affect the results much (at least in this case of autoregressive order 2). In this case, we would conclude at the .05 level that there are 2 cointegrating vectors (because the first 2 nulls are rejected and the third is accepted). The P-values are interpolated from Osterwald-Lenum(1992) Table 1.1* (because a constant term is included, and Table 1.1* is more conservative that Table 1).

eigenvalue	**null hypothesis**	**trace test w/o adj.**	**test w/ adjustment**	**P-value**
.31	r=0	76.13	70.28	.0003
.23	r≤1	37.65	34.75	.023
.073	r≤2	11.00	10.16	.43
.029	r≤3	3.11	2.87	.44