Chapter 4

Analysis Facilities

MODLER is designed primarily to permit and support the creation, maintenance and use of econometric models that commonly involve multiple equations. Among their particular properties, such models ordinarily represent an attempt to establish the relationships between economic variables over time, rather than expressing the time patterns of the behavior of single economic variables. However, preliminary to the attempted specification of these relationships, it is both useful and informative to start by initially considering the observed time series behavior of the individual variables. Among other considerations, these properties can affect the characteristics and appropriate form of these relationships.

There are at least two aspects of individual variables' observed behavior that need to be considered. One of these is that the values of these variables generally originate as empirical measurements, which variously take the form of primary observations, compilations, or constructions that each are made with reference to a specific conceptual definition. Primary, usually microeconomic observations, such as the price of a loaf of bread, can conceptually represent, for example, either the list price or a transacted price. Similarly, the measured income of individuals can represent money income or else a more involved concept that may include in-kind payments and even imputations or certain exclusions. In turn, aggregate variables, macroeconomic variables specifically, are usually built up from primary measurements, but often involving not simply adding up or cumulating, but rather an explicit constructive process, even in the case of variables commonly thought of as simple compilations. Of course, variables representing abstract concepts, such as inflation or the cost-of-living or wealth, will inevitably take the form of constructions. In all cases, the measured values of economic variables must therefore be regarded, from the outset, simply as estimates, estimates that are usually subject to both errors of observation and errors of construction. Consequently, there is always an initial question whether any particular variable, considered on its own as a set of observations, adequately measures the concept it ostensibly represents. Subject to this judgment, there are then a number of questions that can be asked about a variable's conceptual characteristics, apart from any errors of measurement, construction, or compilation. Actually, in practice, the situation is yet more involved, for the values that are observed are also subject to transmission errors, misprints, and other such transcription errors.

In order to address each of these issues, MODLER is designed to support the evaluation of the properties of time series variables in several ways. The analytical facilities offered by MODLER range from those related to display, such as plots of variables over time, to specific tests of their properties as time series. Of these, the plotting facilities deserve careful attention initially, for they provide a potentially very informative first stage analytic facility. Although seemingly prosaic, plots of series, both in level form and in transformation, can provide vital, sometimes invaluable, preliminary information about the revealed properties of time series. To consider this matter properly, it is helpful to begin by defining the potential errors of measurement quite broadly: whenever the observations on a given time series are first considered, errors potentially associated with a particular observation should be interpreted to include those from *any* source, among these even misprints or those of transcription. It should be borne in mind that even government statistical agencies, trade organizations and other purveyors of economic measurements do sometimes inadvertently publish the wrong numbers.

From the perspective of an end user, these are all, broadly speaking, measurement errors because at first there is no way to distinguish them simply by looking at a print out or screen display, although outright "misprints" and other transcription errors can often be fairly evident, particularly when they take the form of a "serious" error. These errors can be spotted relatively easily since, almost by definition, they are not related to the actual measurement properties of a given time series. Moreover, it is when they are large, relative to other types of "measurement" error, that such errors are clearly of greatest concern—that is when they take the form of "outlier" values. A tiny transcription error will in any case often be lost in the general measurement noise, but a "serious" error can distort the subsequent analysis.

Because of human visual acuity, one of the particular benefits to be obtained from simply plotting a series against time is that outlier values become immediately obvious, even in comparison to tabular displays of that series. Or if not obvious from looking at a time plot of the level values of a series, because the error may not in this context be "sufficiently" large, it may nevertheless become evident when the series is plotted in either rate of change or difference form; in this transformed context a transcription error may change the normal pattern of the graphical display. Of course, recognizing that a value is an outlier, and then in response giving that outlier its proper interpretation are not the same thing. Obviously, the existence of an outlier per se does not necessarily signify a misprint or transcription error, because it may have a particular economic meaning or cause. Nevertheless, the identification of such values in a particular case does signal potential measurement error, or at the very least raise a question as to what it means. The possibility of measurement error may also put you on your guard not to automatically ascribe a property to a time series that would otherwise be compounded in an after-the-fact statistical result.

Because MODLER's plotting, as well as its table making facilities, have previously

been considered in Chapter 2, the particular characteristics of these display facilities will not be explored further here. However, it should be evident already that what can be learned initially from simply plotting a series and various of its transformations can be beneficial. Furthermore, although MODLER's analysis facilities might be considered here in terms of specific computational offerings, they should not be treated simply as a given set of tools to be applied in rote fashion.   Instead, it can be argued that the process of investigation, in true Sherlock Holmes fashion, should utilize and weigh all the evidence that can be obtained. Therefore, in the subsequent discussion, it will be assumed that the analysis facilities that are the specific focus of that discussion are being used only after a preliminary analysis that might have involved simple plots and other methods of visual examination.

   Representing an attempt to present them collectively, a number of the specific time series analysis facilities available to MODLER users are listed together on the Central Control screen under the menu heading **Analysis**. For your convenience, this list is shown below in Figure 4-1.   The items have been subdivided into two main categories, **Evaluate Single Series** and **Multivariate Analysis**.

   As you can see, the first set includes the choices **Display Autocorrelation Matrix**, **Plot Autocorrelation Function**, **Unit Root Statistics**, **Growth Rate Triangles**, and **Trend Statistics**. The second includes **Correlation/Covariance** and **Regression**. Certain of these techniques are somewhat ad hoc in nature; others are based upon a formal logic. They will be considered in the order they are listed.
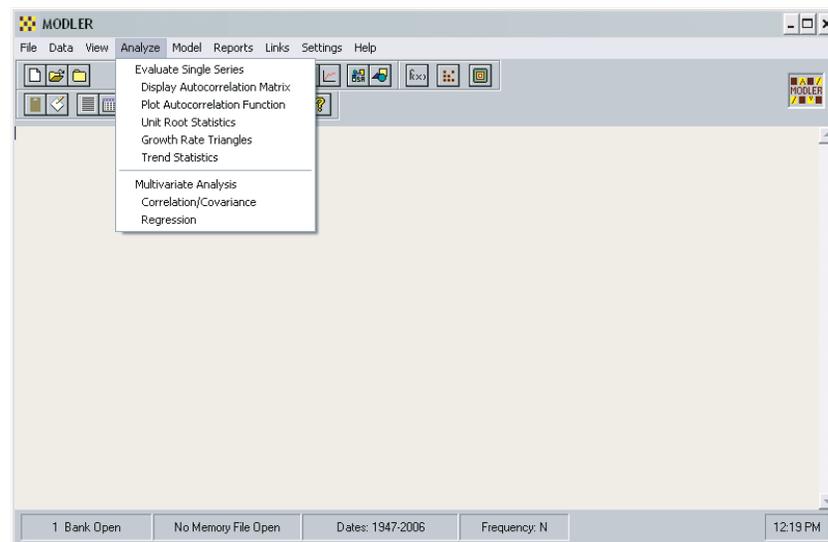

Figure 4-1. Analysis Options

Evaluation of a Single Series

  The first three of the single series analysis options, **Display Autocorrelation Matrix**, **Plot Autocorrelation Function**, and **Unit Root Statistics**, are based upon standard time series analysis constructs that are related and these can each be formally presented.   It is useful to begin with the concept of an autoregressive process and to proceed from there.

Autoregressive Processes

  Following the early treatment by Box and Jenkins (1970) and that more recently in an econometric textbook context by Johnston and DiNardo (1997, pp. 204-214), a univariate autoregressive process (AR) of order p can be characterized by:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... + \alpha_k x_{t-p} + u_t \qquad (4.1)$$

where the α_i (i=0,1,2,...p) are scalar constants. If we then define a Lag operator L, such that:

$$Lx_t = x_{t-1}$$
$$L^p x_t = x_{t-p}$$

and, for later use, the operator Δ:

$$\Delta x_t = x_t - x_{t-1} = (1 - L)x_t$$

it is possible to define the polynomial operator:

$$\phi(L) = 1 - \alpha_1 L + \alpha_2 L^2 + ... + \alpha_p L^p$$

which allows us to respecify expression (4.1) as:

$$\phi(L) x_t = u_t \qquad (4.2)$$

and then to define the transformation:

$$x_t = \phi^{-1} (L) u_t$$

This particular transformation illustrates that the autoregressive series x_t can be expressed alternatively as a weighted sum of the u_t.   A further generalization is:

$$\phi(L) x_t = \theta(L)u_t$$

which is of course a mixed autoregressive-moving average (ARMA) process and obviously implies a different assumption about the properties of the disturbances. Notice in particular that, in general, the weights of this moving average of the u_t need not either sum to 1 or be positive.

Given *any* time series, x_t, it is possible to define the *covariance* between any two realizations at time t versus another, t-k:

$$\gamma_k = \text{cov}(x_t, x_{t-k}) = E[(x_t - \mu)(x_{t-k} - \mu)]$$

The *autocovariance* at lag k is defined as:

$$\rho_k = \gamma_k/\gamma_0$$

where μ is the population mean value of the x_t, assuming that it meaningfully exists. The estimated covariance can be computed as:

$$g_k = \Sigma\,(x_t - \bar{x})(x_{t-k} - \bar{x})\,/\,T$$

and the estimated autocovariance as:

$$r_k = g_k/g_0$$

The population autocovariance matrix of the x_t is then defined by:

$$P_k = \begin{bmatrix} 1 & \rho_1 & . & . & . & \rho_{k-1} \\ \rho_1 & 1 & & . & . & \rho_{k-2} \\ . & . & & & . & . \\ \gamma_{k-2} & . & & 1 & & \rho_1 \\ \gamma_{k-1} & . & & & . & 1 \end{bmatrix}$$

although, being a symmetric matrix, it can be more parsimoniously represented in triangular form. In contrast, the sample matrix of course uses instead the *estimated* autocovariance coefficients, the r_k.

A generally important issue is obviously the properties of a given time series, the properties of which may or may not be those of any idealized process.   In order to fix ideas, in relation to the definitions just presented, it is useful to consider certain simple, ideal cases first.   Specifically, consider a first order AR process:

$$y_t = m + \alpha\,y_{t-1} + u_t$$

such that m and α are scalar constants and where the disturbances, u_t, constitute a

serially uncorrelated white noise process, with $E(u_t) = 0$. Then using the lag operator, L, it follows that:

$$(1 - \alpha L)y_t = m + u_t$$

Furthermore, if $\alpha < 1$, then clearly:

$$E(y_t) = m/(1-\alpha) = \mu$$

where μ is therefore a finite constant by this ratio. Consider in turn the transformation $x_t = y_t - \mu$, which immediately implies:

$$x_t = \alpha\, x_{t-1} + u_t$$

where in addition, as demonstrated by Johnston (1997, p. 208):

$$\alpha = \gamma_1/\gamma_0$$

In general:

$$\gamma_k = \alpha\gamma_{k-1} = \alpha^k \gamma_0$$

and moreover:

$$\rho_k = \gamma_k/\gamma_0$$

$$\rho_k = \alpha\rho_{k-1} = \alpha_k$$

for k=1, 2, ..., the latter expression being what is known as the population *autocorrelation function* of the series. Inasmuch as $\rho_k = \rho_{-k}$, this population function is also necessarily symmetric about zero; therefore, only the positive portion need be considered. In contrast, it can be shown (Johnston, 1997, pp. 209-210) that the autocorrelation function for a stationary AR(2) process:

$$x_t = \alpha_1\, x_{t-1} + \alpha_2\, x_{t-2} + u_t$$

is given by:

$$\rho_k = \alpha_1\, \rho_{k-1} + \alpha_\rho\, \rho_{k-2} \qquad k=3,4,...$$

where:

$$\rho_1 = \alpha_1/(1-\alpha_2) \qquad \rho_2 = \alpha_1^2 / (1-\alpha_2) + \alpha_2$$

The evident implication of these examples is that when any time series process is stationary, the autocorrelation function has a particular, quite definite form that depends on the order of the process. Furthermore, if the observed values of a given realized process can reasonably be regarded as implying these population properties, it follows therefore that its estimated autocorrelation function and the correlogram should also exhibit, within a degree of approximation, these same properties.

Considering any arbitraily chosen candidate series, the plot of its *measured* r_k against the order of the lag k is called the *sample* autocorrelation function. This function provides an estimate of the population autocorrelation function of ρ_k plotted against progressive values of the lag, under the hypothesis that these population counterparts exist. The plot display itself is normally called the (sample) *correlogram*. Notice, in particular, that the calculations are made using the realizations directly, rather than by enforcing the particular properties of a stationary series.

MODLER can both display the sample autocorrelation matrix for any given series and plot its sample autocorrelation function from which in principle the properties of the series can be inferred. In the first case, if you select **Analysis** and then choose **Display Autocorrelation Matrix**, the program will respond by displaying the form shown in Figure 4-2, asking you to specify a series name.
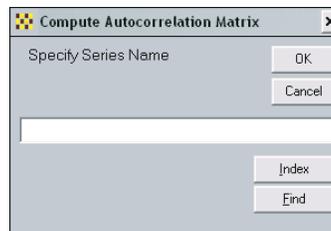

Figure 4-2. Choose Series

Once you provide an appropriate name and then either click **OK** or press the enter key, you should then see a display similar to that shown in Figure 4-3 on the next page. Here, the lower triangular portion of the matrix is displayed, reflecting its symmetric nature. Initially, only a small portion of the matrix may be shown, however, using your mouse, you will be able to explose a larger (or smaller) portion of it, the specific amount depending upon your screen resolution. Notice that you can also pan left or right and scroll up or down.

In addition, this screen displays its own menu that offer options—such as the **File** option in particular— in order to select alternative series' autocorrelation matrices for display or to print the current display or copy it to a word processor, among the choices. Under **View** you will discover several other options, including the capability to display the raw data for the series or plot the series itself. You can actually perform a good bit of related analysis without leaving this particular screen.
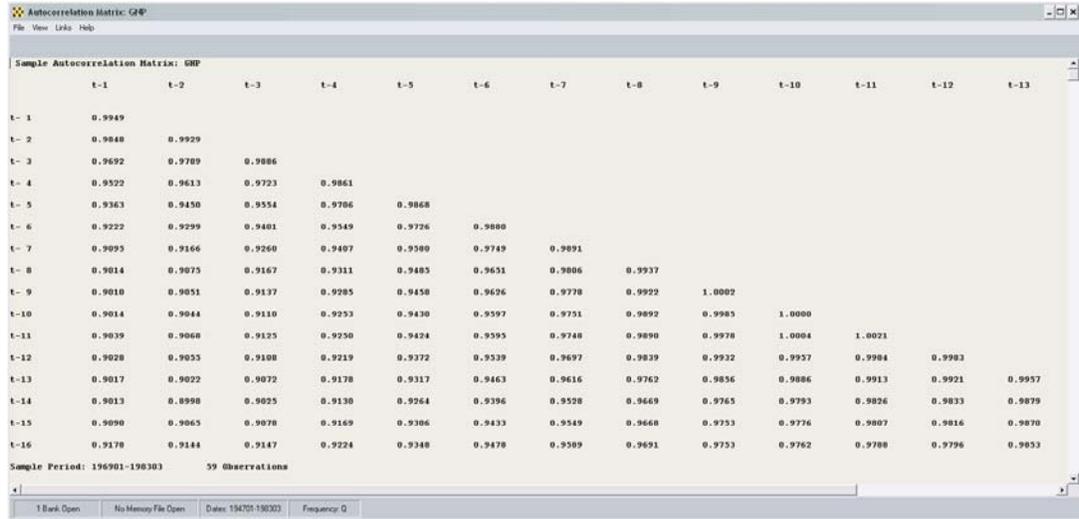
Figure 4-3. Display of Autocorrelation Matrix

A similar command sequence can used to display a series' autocorrelation function. First, select **Analysis** from the menu of the Central Control screen and then choose **Plot Autocorrelation Function**. You should then see a form quite similar to that previously shown in Figure 4-2, except that the banner line will now refer to the autocorrelation function rather than the autocorrelation matrix.   Then, once you have provided a series name and pressed the enter key or the OK button on this form, you should see a bar plot of the type shown in Figure 4-4.   Notice that the vertical axis specifies the value of the autocorrelation function and the horizontal axis the value of the corresponding lag.   In addition, the values of the autocorrelation function corresponding to each lag value are display on the screen just to the right of the plot, although this particular sub-display can be hidden by choosing to turn the **Data Table** off under the **Values** menu choice.
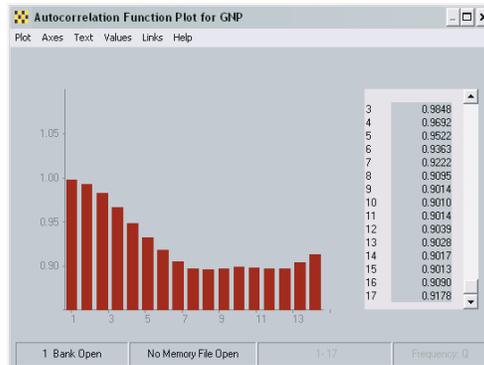
Figure 4-4.   Autocorrelation Function Display

Unit Root Tests

The concepts behind the unit root tests provided by MODLER are explained in almost any modern econometrics textbook, such as Greene (2004, pp. 632-646), and are only described here to the degree necessary to explain the nature of the test statistics that MODLER provides. A time series, x_t, is specifically said to obey a random walk (about a zero mean) if:

$$x_t = x_{t-1} + u_t$$

where u_t is a white noise, with a ero mean and a constant variance In contrast, a given realized series x_t might alternatively behave such that either: $^{.}$

$$x_t = \alpha^* + dT + u_t$$
or
$$x_t = \alpha^* + \gamma\, x_{t-1} + u_t$$

where T is a time trend; $\alpha*$, δ, and γ are constants (such that, possibly, $\gamma = 1$); and where, in each case, what might be observed from the series time plot is apparently strongly trended behavior. Significantly, these alternative characterizations each individually define distinct behavioral patterns. Furthermore, as alternatives, they constitute non-nested hypotheses. Notice, specifically, that if $\gamma = 1$, it is possible to solve the second (difference) equation above to yield:

$$x_t = \alpha^* + (u_0 + \ldots + u_t)$$

which implies that x_t in this case can be characterized as a variable that for all t has an increasing variance about a linear trend. The alternative form may instead express (constant) variation about a linear trend; in this case, its behavior can be termed "trend stationary," , given that in this case, as before, the u_t are themselves distributed with zero mean and a constant variance.

However, in order to evaluate these specifications as alternatives, they fortunately can each be nested within the more general specification:

$$\Delta x_t = \alpha + \beta T + (\gamma\text{-}1)\, x_{t-1} + u_t$$

such that $\alpha = \alpha*\gamma$, $\beta = \delta\gamma$. Notice that $\gamma = 1$ versus $\delta = 0$ are the conditions under which one or the other of the more specialized equation forms are "selected." The implication of $\gamma = 1$, $\delta = 0$, is that the variable behaves as a random walk with drift. In contrast, the condition $\gamma < 1$ possibly implies trend stationary behavior and $\gamma > 1$ (generally thought to be unusual for economic series) explosive behavior. Notice, in addition, that an implication of the random walk with drift specification is that

economic shocks are permanent (and cumulative) in effect, in contrast to the impermanent effect of shocks in a trend stationary context, which damp out over time.

Given that each of the specifications considered here are easily estimated using Ordinary Least Squares, at first sight it would appear that standard tests might be inferentially applied. However, the principle problem associated with unit root tests is that when the hypothesis of a unit root is true ($\gamma = 1$), the standard inferential approach, based upon the assumption that u_t is either normally distributed or asymptotically normally distributed, is no longer valid. Instead, the test statistics associated with the parameter estimators of the right-hand-side terms must be evaluated taking into account that these estimators may be biased, among other characteristics. Although the calculation of the test statistic is algorithmically the same as for the standard t-statistic, the implication of the possibility of a unit root is that the critical values against which these values should be compared differ from those found in standard t tables.   The requisite tests are therefore described as Dickey-Fuller tests, reflecting the work by Dickey and Fuller to derive this set of critical values, given the specific observational context.

However, it also needs to be recognized that the properties of the u_t have so far simply been assumed. It is actually quite possible that these disturbances could instead be autocorrelated, not independently distributed. In order to allow for the possibility of autocorrelation, another, somewhat more elaborate specification can be considered, that given by:

$$\Delta x_t = \alpha + \beta T + (\gamma\text{-}1)\, x_{t-1} + (\delta_1\, \Delta x_{t-1} + \ldots + \delta_k\, \Delta x_{t-k}) + u_t$$

This form is the basis for the Augmented Dickey-Fuller test, with the inclusion of the one or more lagged difference terms (as well as an implied redefinition of certain other parameters) having the effect of "whitening" the disturbance term of this specification, as compared to those considered before.   An additional complexity here is that the appropriate number of lagged difference terms is of course not known a priori, so that one aspect of the more elaborate testing procedure that must be performed in this case is the general need to consider and evaluate the number of such terms used in the application of the augmented test(s).

Evidently, the unit root test procedure that might be pursued in a given context will depend upon the a priori assumptions that are made, which may vary depending upon such things as the perceived nature of the time series considered. For instance, there is presumably a difference between a macroeconomic variable, such as GDP, compared to the price of the stock of some corporation or a stock price index. In the first case there may be a strong presumption that the time behavior of the variable is strongly trended, with the question being whether this represents actual trended behavior, possibly of the stationary trend variety, or something closer to a random

walk with drift. In the second case, that of a stock price, the behavior may represent a type of random walk.

Obviously, all such a priori restrictions need to be considered carefully, but because they may be justified, in practice MODLER permits you to consider any (or all) of the following Dickey-Fuller specifications:

$$\Delta x_t = \alpha + \beta T + (\gamma\text{-}1)\, x_{t\text{-}1} + u_t$$

$$x_t = \alpha^* + \gamma x_{t\text{-}1} + u_t$$

$$x_t = \gamma x_{t\text{-}1} + u_t$$

in addition to those associated with the Augmented Dickey-Fuller tests.   The first of the three above is described as involving Both Drift and Trend, the second as No Trend, Drift Only, and the third as Random Walk: No Drift, No Trend.

In order to perform unit root tests, you can begin by choosing the entry **Unit Root Tests** under the Central Control screen **Analysis** menu option.   When you have made this choice, you should next see the form shown in Figure 4-5. Simply type in the name of the series you wish to subject to tests.   For example, using the Demobank as the data source, try GNP.

Figure 4-5. Variable Selection Form

In response, MODLER will display a screen similar to that shown in Figure 4-6. As it happens, this screen presumes you to be a regular user of this facility. It provides summary results, which you will generally need to be familar with in order to make sense of.   However, if you look at the menu line on this screen, you will see, in particular, the following choices: **Summary**, **Complete**, and **Defaults**, among others.   Choosing the first of these provides you with the ability to specialize the summary results.   **Complete**, in contrast, displays in each case, the full set of regression results and statistics, thus providing you with the capability to see in detail what each line of the summary table implies. The defalults, in turn, provide you with more control over the way in which the tests are performed.
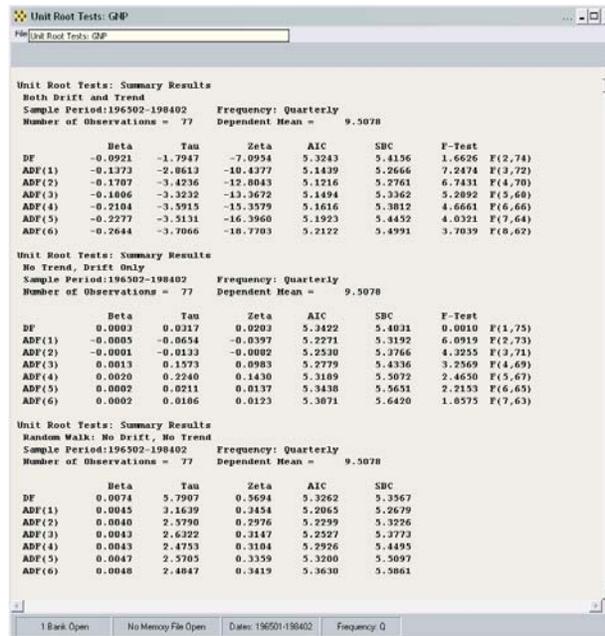
Figure 4-6. Unit Root Tests Summary Screen

Once you are looking at the screen displayed in Figure 4-6, at least for the first time, it might be useful to choose the menu item **File**, which, among other things, gives you the ability to select the variable to test: **Select Variable**.   Generally, the Unit Root Tests screen is designed to give you access, in one place, to a number of related facilities. One of these is to test progressively each member of a set of variables, one after the other, hence the **Select Variable** command.   The other **File** menu elements include the capability to capture the test display currently on the screen, sending it to an editable file, to a word processor, or to your printer, thus permitting you to preserve an audit trail of what you have done.

If you next select **Complete**, you will be presented with a set of choices that permit you to display the complete regression results for any particular Dickey-Fuller test or all such tests, particular Augmented Dickey-Fuller tests or all such tests, or some combination of both Dickey-Fuller and Augmented Dickey-Fuller tests. If alternatively you select **Summary**, you will be given the same type of choices for the summary displays for these tests.   The **Default** menu element gives you the ability to change various defaults, including the sample period used.

In contrast, the **View** menu element gives you the ability to display the observations on the currently selected (test) variable or, alternatively, plots of that series, alone or together with other time series.   You can also view the **Trend Statistics** display for the series or its sample autocorrelation matrix or plot its sample autocorrelation function. In addition the View options include the capability to display the contents of any MODLER macro file or even a word processor document.

Links provides you with links to the clipboard viewer, your notes editor, and the onscreen calculator, as well as to your word processor or presentation manager or the Internet via the www.modler.com website.

You should view these facilities as simply the current offerings. Any suggestions that you may have towards the improvement of the Unit Root Tests screen will be gratefully received.

Growth Rate Triangles

A Growth Rate Triangle is simply a display of the point-to-point rates of growth of a time series between some time t and another time t-k. If, on the Central Control screen, you choose **Analysis** and then **Growth Rate Triangles**, you should next see the form shown in Figure 4-7.
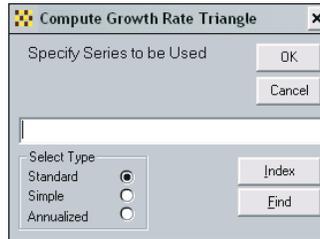
Figure 4-7. Growth Rate Triangle Series Selection

If you then specify a particular series, such as GNP, making no other changes, you should next see the display shown in Figure 4-8.   This table shows the absolute point-to-point growth rates.

In contrast, using the radio buttons in the bottom left corner of the form shown in

Figure 4-7, you were to choose **Simple** or **Annualized**, the effect would be to cause the point-to-point growth rates to be computed as

  Notice, in addition, while you are looking at the display in Figure 4-8, that the Growth Triangle screen, like the Unit Root Tests screen, provides several menu choices that permit you, on the one hand, to capture under **File** the display and send it to a file, your printer or a word processor, and, on the other, under **View** to display series observations or plot one or more series or transformations. You can also view the **Trend Statistics** display for the series or plot its sample autocorrelation function. In addition the **View** options include the capability to display the contents of any MODLER macro file or even a word processor document.

Trend Statistics

  If, on the Central Control screen, you choose **Analysis** and then **Trend Statistics**, you should next see the form shown in Figure 4-8.

Figure 4-8. Select Variable

  Specify a series name, such as GNP, and you then should next see the type of screen shown in Figure 4-9.
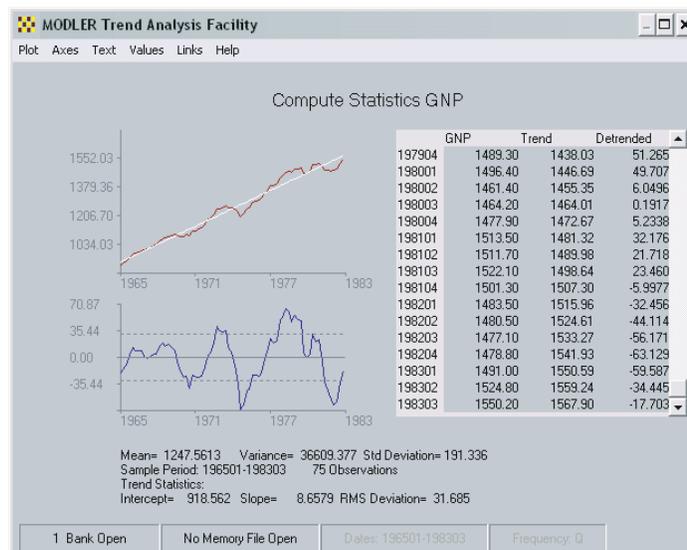
  This screen shows two plots, the topmost one being a plot of the specified series upon which is superimposed a regression trend line and the bottom one a plot of the

difference between the trend values and the values of the variable at each time point shown on the graph.   In addition, the data table to the right of these plots provides the observation values for the series, its trend, and the detrended differences. Under the bottom most plot are the trend coefficients and the computed statistics associated with the series and the detrended differences.

  Multivariate Analysis

  The two types of multivariate analysis supported in this context are the generation of covariance and correlation matrices for sets of time series and regression analysis.

Correlation/Covariance

  If, on the Central Control screen, you choose **Analysis** and then **Correlation/Co-variance**, you should next see the form shown in Figure 4-9. Depending upon
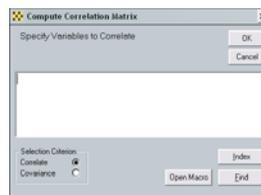

Figure 4-9. Correlation/Covariance Variable Selection

whether you prefer to display a correlation or a covariance matrix, select between these choices using the radio buttons in the bottom-left-hand corner of the form. Then simply list the variables you wish to use and click **OK**.   At this point you should see a display similar to that shown in Figure 4-10.


```
Sample Correlation Matrix

              GNP            CONS           M1$

GNP           1.0000

CONS          0.9955         1.0000

M1$           0.9603         0.9740         1.0000

Sample Period: 196501-198303       75 Observations
```
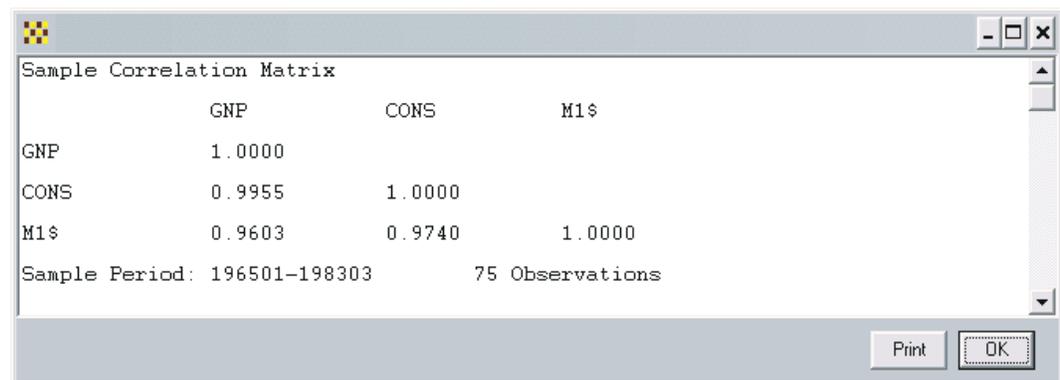
Figure 4-10. Correlation Matrix Display

  The computations will be performed for the longest common period of availability of any of the series included in the list of variables that you provided.

Regression

   The regression option exists simply because Analysis provides one of the routes to performing a regression.   At the Central Control screen, choose **Analysis** and then **Regression**, you should next see the form shown in Figure 4-11.
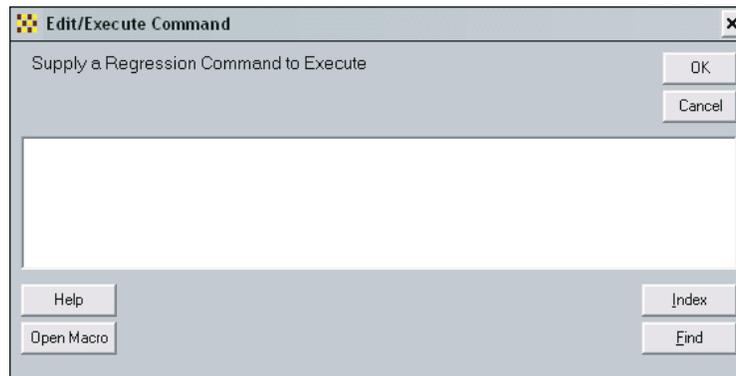
   The **Help** button provides a description of the format of a regression command, as well as providing related details.   However, otherwise, information about this Edit/Execute Command form is provide in Chapter 5.