SOME NOTES ON UNIVARIATE TIME SERIES ANALYSIS*

O. Basic Problem
- The basic problem involves the use of past observations of a random variable to forecast future values of that variable.
- You might enjoy reading "Forecasting–looking back and forward: Paper to celebrate the 50th anniversary of the Econometrics Institute at the Erasmus University, Rotterdam" by Clive Granger (**Journal of Econometrics**, 138 (2007), 3-13) which reflects on the past and future of forecasting methods. Professor Granger received the Nobel Prize for his work in time series and causality.

A. A, B, C's of Time Series Models
1. Stationarity
2. Non-stationary: two important examples
    a. Random walk with drift
    b. Trend stationary model
    c. Summary
    d. Dickey-Fuller Tests
3. Random walks, trends, and spurious regressions
4. Cointegration

B. Basic ARIMA Models and methods

C. Characteristics of some ARIMA Models and Identification
1. An overview and simple models
    a. AR(1)
    b. MA(1)
    c. Summary
2. AR(p)
3. MA (q)
4. ARMA (p,q)
5. ARIMA (p,d,q)

D. Diagnostic Analysis

E. Estimation

F. Forecasting

G. General comments

H.. Help: Computer Programs

**** For a "short ARIMA course" read Sections A and B, look at the figures in Section C, review Section G (1-4), then work on the Homework(next two pages).***

Homework

- **Select a time series with at least 60 observations: Search "time series data"**

- **Plot the data**

- **Explore ways to obtain a stationary time series**
    - **Plot differences**
    - **Consider taking the log of the data**

- **Perform a DF test**

- **Identify and estimate a time series model–using AC and PAC coefficients**

- **Conduct diagnostic tests of the estimated model (corrgram resid–use pac's, ac's, and Q-statistics corresponding to the estimated residuals)**

- **Write the equation of the estimated model**
  Stata reports the results of estimating :
  $$\left(1-\phi_1 L-\ldots-\phi_p L^p\right)\left(\Delta^d Y_t - const\right)=\left(1+\theta_1 L+\ldots+\theta_q L^q\right)\varepsilon_t$$

- **Determine forecasts for five future periods**

Helpful STATA Homework Commands

Data in variable named “y” and time index “t”

tsset t	Alerts to STATA to time series data. If a time index is not included in the data file, the command **gen t=_n** will generate the required variable.
scatter y t **scatter D.y t**	Plots y vs. t Plots the difference vs. t
dfuller y **dfuller D.y**	Helps determine whether the series is difference or trend stationary. The null hypothesis is difference stationary (d=1).
corrgram y, lags(#) **corrgram D.y, lags(#)**	Reports estimated ac and pac coefficients for Y and first differences. These can be used to identify the model (select p,d,q) based on the patterns of the estimated ac and pac coefficients **ac y, lags(#)** **pac y, lags(#)**
arima y, arima(p,d,q)	Estimates the identified arima model. The estimated model can be recovered from this output
predict e, resid	Stores the estimated residuals in variable *e*

corrgram e, lags(#) Reports the estimated ac and pac coefficients corresponding to the estimated errors. Correct model specification would yield statistically insignificant ac and pac coefficients. The corresponding Q-statistics would be expected to be statistically insignificant.

tsappend, add(5) Extrapolates the estimated model to obtain the next five predictions.

predict yhat Stores predicted values in the variable *yhat*

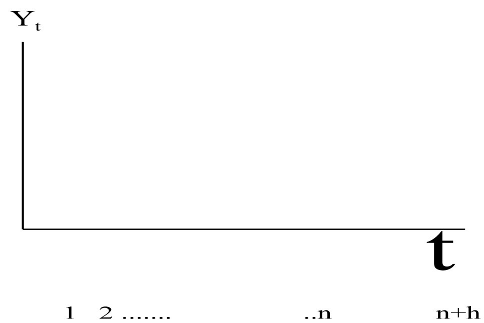
list y yhat t or Prints the indicated (if available) values of y and indicated predicted values

list y yhat t if t>N*

James B. McDonald
Brigham Young University

Univariate Time Series Analysis

0. Basic Problem: Consider the problem of, given n observations on a single variable Y (Y_1, Y_2, ..., Y_n), obtain forecasts of Y at time period n + h, denoted by $\hat{Y}_{n+h}$ or $Y_n(h)$. This might be pictorially represented as:



The data might correspond to GDP, consumer prices, foreign exchange rates, telephone calls, unemployment rates, product sales, the number of empty hospital beds in a hospital, commodity prices, or any other time series. The forecasting problem becomes that of attempting to obtain a prediction or forecast *h periods* in the future from known observations. Many techniques have been developed to extrapolate from past observations into the future. The techniques differ in structure and assumptions made, including the amount of past information used in forecasting. Some techniques assume that past (1) levels, (2) changes, or (3) percentage changes can be used to forecast the next period. Other techniques are based upon moving averages of past values and possibly allow for trends or seasonality in the underlying series. When forecasting, we would do well to remember the admonition found in 88th Section of the Doctrine and Covenants (79th verse) about studying things “which must shortly come to pass” and be very cautious about long-run forecasts.

Auto Regressive Integrated Moving Average (ARIMA) models are very general specifications which includes some of the previously mentioned forecasting approaches as special cases as well as including more general methods. These techniques will be studied in order of increasing "sophistication."

Before getting into the details of time series forecasting techniques, it is useful to provide a brief overview of potential applications of these techniques. They can be used (1) on a "stand alone" basis to predict future values of a dependent variable of interest, $Y_n(h)$, by extrapolating past trends ; (2) to predict a future value for an independent variable (X_t), $X_n(h)$ which can be substituted into an econometric model to yield

$$Y_n(h) = f(X_n(h))$$

to obtain forecasts of the dependent variable; and finally (3) these techniques can also be used to predict systematic components in the residuals (ε) ,$Y_n(h) = f(X_n(\varepsilon_n(h)$. Thus time series techniques can be used separately from the specification of an economic model or in conjunction with an estimated economic model.

A. A,B,C's of time series models

1. Stationarity

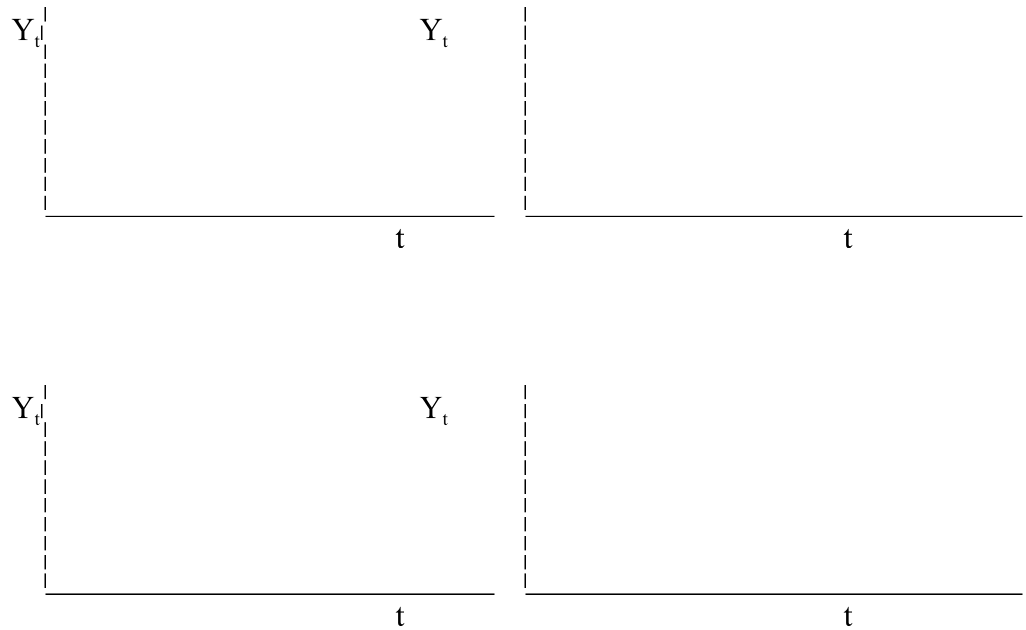
Consider a stochastic process Y_t which is defined for all integer values of t. Y_t is said to be **weakly or covariance stationary** if

- $E(Y_t) = E(Y_s) = \mu$ for all t and s (A.1a-c)
- $Var(Y_t) = Var(Y_s) = \sigma^2$ for all t and s
- $Cov(Y_t, Y_s) = \gamma_{t-s}$ depends only on *t-s*

for all t. A stronger definition of stationarity is that the joint distribution functions

$F(Y_{t+1},\ldots, Y_{t+n})$ and $F(Y_{s+1},\ldots, Y_{s+n})$ are identical for any value of t and s.

Consider the four following figures:

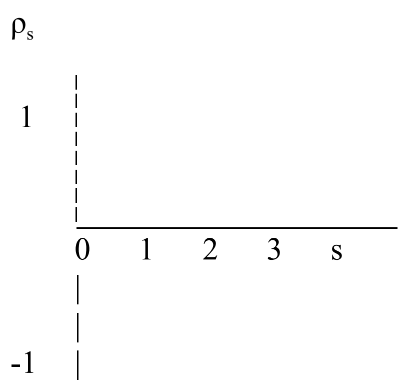


The first series would be classified as being stationary; whereas, the other three would not.

It will be useful to define the autocorrelation coefficients corresponding to Y_t in terms of the autocovariances (γ_s)

$$\rho_s = \frac{\gamma_{t-s}}{\gamma_0} \quad = \text{correlation } (Y_t, Y_s) \text{ w } \gamma_0 \text{ e } Var(Y_t) \qquad \text{(A}$$

and the plot of ρ_s against s is referred to as the correlogram.



An example of a stationary series is the autoregressive model, AR(1), where

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t \qquad \text{(A.3 a-b)}$$

where $\varepsilon_t \sim N[0,\sigma^2]$.

From this specification, it can be shown that

$$E(Y_t) = 0$$

$$Var(Y_t) = \frac{\sigma^2}{1 - \varphi_1^2} = \gamma_0$$

$$Cov(y_t, y_{t-s}) = (\varphi_1^s)\frac{\sigma^2}{1 - \varphi_1^2} = \varphi_1^s \gamma_0 = \gamma_s$$

with autocorrelation coefficients given by

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \varphi_1^s$$

which decay exponentially for $|\varphi_1| < 1$.

While the AR(1) model is stationary if $|\varphi_1| < 1$, many economic series are not stationary. Remember a time series is not stationary if either the mean, variance, or covariance change with time. Thus a series which increases over time or is characterized by heteroskedasticity is not stationary. Since many forecasting techniques are based upon the stationarity assumption, it is comforting to note that some non-stationary series can be transformed into stationary series. We now consider two such series.

2. Non-stationary series: two important examples

a. Consider the **random walk**, with drift:

$$Y_t = \mu + Y_{t-1} + \varepsilon_t \qquad \text{(A.4)}$$

where $\varepsilon_t \sim N[0,\sigma^2]$. By recursive substitution this model can be rewritten as

$$Y_t = \mu t + \sum_{i=0}^{t} (\varepsilon_{t-i}) \qquad \text{(A.5)}$$

Note that Y_t is not stationary because the mean of Y_t (μt) increases (linearly) with t and the variance of Y_t , about the linear trend, increases with time. **It is important to note that the first difference of Y_t , denote ΔY_t, of a random walk**

$$\mathbf{\Delta Y_t = Y_t - Y_{t-1} = \mu + \varepsilon_t} \qquad \textbf{(A.6)}$$

is stationary. In working with time series it can be useful to use lag operators (L) or backshift operators (B) defined as follows:

$$L\,Y_t = Y_{t-1} \qquad B\,Y_t = Y_{t-1} \qquad \Delta Y_t = (1-L)Y_t \text{ or } (1-B)Y_t$$

Y_t is said to be integrated of order one, I(1), when the first difference of the series is stationary. A series is said to be integrated of order d, I(d), if $\Delta^d Y_t$ is stationary.

Note that if d=1, then $\Delta^d Y = \Delta Y_t = Y_t - Y_{t-1}$ and if d=0 then $\Delta^d Y_t = \Delta^0 Y_t = Y_t$.

b. Now consider the **trend stationary** model defined by

$$Y_t = \mu + \beta t + \varepsilon_t \qquad \text{(A.7)}$$

where $\varepsilon_t \sim N[0, \sigma^2]$. Like the random walk with drift, the trend stationary model has a mean ($\mu + \beta t$) which increases linearly with t; however, the variance of Y_t, about its trend, is a constant σ^2 which is in contrast to the random walk with drift whose variance increases with t. The first difference of a trend stationary model

$$Z_t = Y_t - Y_{t-1} = \left(\mu + \beta t + \varepsilon_t\right) - \left(\mu + \beta(t-1) + \varepsilon_{t-1}\right) =$$

has a constant mean (β) and variance 2 σ^2, but the errors involve a moving average $\left(\varepsilon_t - \varepsilon_{t-1}\right)$ of error terms. The correlation between Z_t and Z_{t-1} is -1/2 and correlation between Z_t and Z_{t-s} for s > 1 is zero.

c. The optimal estimation procedures for these two series are different. The best way to estimate the random walk with drift is to take first differences and then estimate the unknown parameters associated with the differenced series. The best way to work with the trend stationary series is to estimate a polynomial in "t" and then analyze the residuals. Thus, the two non stationary series are treated in different ways. The difference between these two important series can be summarized as in the following table.

Series type/ Approach	Differences: $\Delta^d Y_t$	Detrend [regress y on a polynomial time trend]
Random walk with drift • $Y_t = \mu + Y_{t-1} + \varepsilon_t$ • $Y_t = \mu t + \sum_{i=0}^{t}(\varepsilon_{t-i})$ • Behavior: Increasing mean and variance • The impact of innovations (ε_j) persist.	• Optimal	• Not optimal

Trend stationary • $Y_t = \mu + \beta t + \varepsilon_t$ • Behavior: increasing mean with constant variance • The impact of innovations (ε_i) pass.	• Not optimal	• OLS optimal • Standard statistical tests are problematical

If a series appears to increase exponentially, the previous approaches could be applied to the logarithms of the data.

The implications of the two different models can be important in some applications. A practical problem is that it may be difficult to differentiate between trend and difference stationary models and, hence, to know the most efficient estimation technique to use. One approach to discriminating between the two models is to consider the following "nesting" regression model:

$$Y_t - Y_{t-1} = \mu(1-\gamma) + \beta\gamma + \beta(1-\gamma)t + (\gamma-1)Y_{t-1} + \varepsilon_t \qquad \text{(A.9)}$$

What does this nested regression model simplify to with r

(1) $\gamma = 0$? T_ _ _ _ S_ _ _ _ _ _ _ _ _ _ (fill in the blanks)

(2) $\gamma = 1$? R_ _ _ _ _ W_ _ _ or DS (fill in the blanks)

There are a series of tests known as ***Dickey-Fuller tests*** which explore the null hypothesis, H_o: $\gamma = 1$, which are based on the regression model:

$$\mathbf{Y_t - Y_{t-1} = \alpha_0 + [}\alpha_1 \mathbf{=(\gamma-1)]Y_{t-1} + [}\beta(1-\gamma)\mathbf{]\, t} + \varepsilon_t \qquad \text{(A.10)}$$

The hypothesis $\gamma = 1$ implies that the coefficients of the variables t and $\mathbf{Y_{t-1}}$ equal 0. Standard t-tests are not appropriate, but appropriate tables have been constructed using Monte Carlo methods. If the coefficient of $\mathbf{Y_{t-1}}$, α_1 **, is negative, the evidence favors trend stationarity.**

In some applications there is concern that the $\boldsymbol{\varepsilon_t}$ will be characterized by autocorrelation. In these cases, the **augmented Dickey_Fuller test** is based on estimating the equation

$$\mathbf{Y_t - Y_{t-1} = \alpha_0 + [\alpha_1 = (\gamma-1)]Y_{t-1} + [\alpha_2 =} \beta(1-\gamma) \qquad \Sigma_{j=1}^{p-1} \phi_j \Delta Y_{t-j} \qquad + \boldsymbol{\varepsilon_t}$$

where the term involving the summation has been added to pick up the impact of autocorrelation. The test for differencing is performed by testing the null hypothesis

H_O: $\boldsymbol{\alpha_1}$ = **(γ-1) = 0** (with or without a time trend in (A.10))

This and other related tests can be performed using STATA (version 9 and 10) with the commands:

dfuller y	
dfuller y, noconst	**supresses the constant term in the regression**
dfuller y, trend	**includes a trend term in the regression**
dfuller y, lags(#)	**includes # lagged differences**

Notes:

(1) Tests of $\alpha_1 = 0$, null hypothesis of a unit root, are one tailed tests. The hypothesis of difference stationarity is rejected if the estimated test statistic is less than the

reported critical value. For example, for series without time trends, Asymptotic critical values for a unit root t - test [with no time trend] are given by

Significance level	1%	2.5%	5%	10%
Critical value	-3.43	-3.12	-2.86	-2.57

If $\frac{\hat{\alpha}_1}{s_{\hat{\alpha}_1}}$ < - 3.43, then we would reject the hypothesis of a unit root. In the case of unit roots, one approach is to work with first differences of the data. There are other approaches.

(2).Peter Phillips and others have explored the use of fractional differences where Δ^d where $0 < d < 1$ and provides an intermediate ground between working with levels (d=0) and working with differences (d =1).

3. Random Walks, trends, and Spurious Regressions

Have you ever wondered what happens if you regress one series on an unrelated series when both series grow over time? This was the question explored by Granger and Newbold in a classic paper published in the Journal of Econometrics in 1974. Their finding may not be surprising. Often, in these cases, standard statistical tests suggest statistical significance when, in fact, there is no relationship. They concluded that in such cases much larger t-statistics would be needed than suggested by traditional t-tables. Granger and Newbold find a critical value of 11.2 more appropriate than 1.96. Their study was based on Monte Carlo simulations. Peter Phillips worked with a more general model and used analytical methods, rather than simulation studies, and recommends the use of a critical value given by $\left(n^{.5}\right)\left(t_{critical\_value}\right)$.

The bottom line of all of this, is that we may want to consider fitting relationships to appropriately differenced or detrended data. However, there are alternatives.

4. Cointegration

An interesting problem arises if Y and X are integrated of different orders. This makes it impossible for the error term , $\varepsilon_t = Y_t - X_t \beta$, to be stationary. Greene gives a very abbreviated treatment of this important issue.

B. Basic ARIMA Models/methods

1. Basic models and special cases

Autoregressive-integrated moving average (ARIMA) models are an important general class of stochastic models which have been widely adopted as models for time series. These models include several of the models in the previous section as special cases and are extremely versatile in terms of their statistical properties. An ARIMA model with parameters (p,d,q) is defined as follows:

$$Y^*_t - \varphi_1 Y^*_{t-1} - \ldots - \varphi_p Y^*_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}, \tag{B.1}$$ [1]

---

[1] **Some authors use θ_i rather tha θ_i as coefficients on the right hand side. These notes were written using $-\theta_i$. Stata uses θ_i e 's and reports corresponding estimates. .**

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)Y_t^* = \left(1-\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \qquad \text{, or}$$

$$\varphi(L)Y_t^* = \theta(L)\varepsilon_t$$

where

- $Y_t^* = (\Delta^d \quad (\text{`}\mu - \quad) = ((I\cdot \mu\ ^{d}Y_t - \quad)$, e.g., for d = 1 $Y_t^* = \text{`}\mu \cdot Y_{t-1} - \quad .$
- $\varepsilon_t \sim N(0,\sigma^2)$
- $\varphi(L) = 1-\varphi_1 L - \ldots -\varphi_p L^p$
- $\theta(L) = 1-\theta_1 L - \ldots -\theta_q L^q$

Note that the portion of the expression on the left hand side of the equal sign in equation (B.1) is the **A**utoregressive (AR) portion with p lags and that on the right hand side is the **M**oving average component (MA) with q lags. The d refers to the number of times the series is differenced. This model is denoted ARIMA(p,d,q) and includes many useful models as special cases.

Some special cases include

(1) ARIMA (1,0,0) = AR(1)

$$Y_t - \varphi_1 Y_{t-1} = \varepsilon_t$$

This is a common form used to model autocorrelation in regression models.

(2) ARIMA (p,0,0) = AR(p)

$$Y_t - \varphi_1 Y_{t-1} \ldots -\varphi_p Y_{t-p} = \varepsilon_t$$

(3) ARIMA (0,0,1) = MA(1)

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

(4) ARIMA (0,0,q) = MA(q)

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots -\theta_q \varepsilon_{t-q}$$

(5) ARIMA (p, o, q) = ARMA (p,q)

(6) ARIMA(0, 0, 0) White noise
$Y_t = \varepsilon_t$

(7) ARIMA(0,1,0) Random Walk
$$Y_t - Y_{t-1} = \varepsilon_t$$

Other special cases include exponential smoothing ARIMA (0,1,1) and the Holt-Winters nonseasonal model corresponding to an ARIMA (0,2,2).

2. Stationarity and invertibility

An ARIMA(p,d,q) models is said to be ***invertible*** if it can be expressed as an ARI(d, ∞) moc $\varepsilon_t = \frac{\phi(L)}{\theta(L)}\Delta^d(Y_t - \mu)$. Similarly, an ARIMA(p,d,q) models is ***stationary*** if it can be expressed as an IMA(d, ∞) moc $\Delta^d(Y_t - \mu) = \frac{\theta(L)}{\phi(L)}\varepsilon_t$

IMA representation is referred to as the **Wold decomposition**. Necessary and sufficient conditions for invertibility/stationarity require that each of the zeros of the respective polynomials, $\theta(z)$ and $\phi(z)$, have modulus greater than one. Certain coefficient restrictions insure invertibility/stability.

3. Basic steps in application of ARIMA models

Traditional applications of ARIMA models as forecasting tools involves the following four steps: (1) identification, (2) estimation, (3) diagnostics, and (4) forecasting.

(1) IDENTIFICATION--determine values for p, d, and q. In other words the appropriate number of autoregressive lags (p) and moving average lags (q) as well as the determination degree of differencing needs to be determined.

"d" is selected to be the number of times that the series must be differenced to obtain a stationary series.

Probably the most common method of "identifying" or selecting p and q is by analyzing the behavior of the estimated <u>autocorrelation coefficients</u> (ρ_i ‘s) and the <u>partial autocorrelation coefficients</u> φ_{ii} (related to the φ_i and will be defined shortly).

Different ARIMA models will be associated with different behaviorial patterns for the true autocorrelation and partial autocorrelation coefficients. These patterns depend upon the values for p and q as well as the corresponding numerical values of φ_i, and θ_i in the model. See the charts in section C.

A summary and preview: the patterns of the autocorrelation and partial autocorrelation coefficients for an AR(p) and MA(q) "appear" as follows

	pac patterns	**ac patterns**
AR(p) :	**p "spikes", then cut off**	**decline**
MA(q):	**decline**	**q "spikes", then cut off**
ARMA(p,q)	**decline**	**decline**

Thus, an inspection of the autocorrelation coefficients and partial autocorrelation coefficients can help identify an ARIMA model (p, d, and q) much like fingerprints or DNA can be used to "identify" a suspect. Values for p and q will be selected so that patterns of the <u>corresponding autocorrelation</u> and <u>partial autocorrelation</u> coefficients will be mimic patterns of the observed behavior of the estimated autocorrelation and partial autocorrelation coefficients.

The behavior of the autocorrelation coefficients and partial correlation coefficients corresponding to different values of p and q can be derived mathematically or using a computer program like (on 588 Blackboard)

THEORYTS

THEORYTS generates **<u>true</u>** population auto and partial autocorrelation coefficients corresponding to arbitrary values for p, q, and user provided values for φ_i, φ_i and can be used to provide a useful **tutorial** in the identification process. Sample outputs are included in these notes.

Many programs are available which will **estimate** autocorrelation and partial autocorrelation coefficients corresponding to a series of data.

The following **STATA** command will report the indicated number (#lags) of estimated auto and partial autocorrelation coefficients:

corrgram y, lags(#lags)

Now what are the partial autocorrelation coefficients?
The partial autocorrelation coefficients are denoted by φ_{ii} and are useful in determining the order of the autoregressive process.

φ_{ii} is equal to the coefficient φ_i if the model has an ith order autoregressive component, AR(i). For example,

$\varphi_{11} = \varphi_1$ in an AR(1)
$\varphi_{22} = \varphi_2$ in an AR(2)
$\varphi_{33} = \varphi_3$ in an AR(3)
.
$\varphi_{pp} = \varphi_p$ in an AR(p)

Unfortunately, in applications we do not know the true values of φ_{ii} which must be estimated from the data . The partial autocorrelation coefficients can can be estimated using (1) OLS to estimate AR(1), AR(2), AR(3), ..., AR(#lags) models or (2) Yule Walker equations which will be discussed later.

The estimated partial autocorrelation coefficients provide a tool in determining the value of *p*, the number of autoregressive lags in the model. For example, if the model is an AR(2), we would expect to find $\hat{\phi}_{11}$ ε $\hat{\phi}_{22}$ to be statistically significant, with $\hat{\phi}_{33}$ being statistically insignificant. More on this later.

Alternative and complimentary approaches to the identification process involve the use of the spectral density function or the specification of an objective function which is optimized over p and q and φ and θ. The Akaike information criterion AIC = -2 ℓn (likelihood) +2(p+q) is an example of this procedure. AIC is then minimized over p and q as well as the coefficients in the ARIMA specification.

After the model has been identified (values for p,d,q selected), the second step involves estimating the specified model.

(2) Estimation--Given values for p and q, nonlinear estimation techniques can be employed to estimate $\sigma^2 = \mathrm{Var}(\varepsilon_t)$, $\varphi_1,...,\varphi_p$, $\theta_1,...,\theta_q$. Conditional maximum likelihood or nonlinear least square estimators can be obtained. Some of the associated details are discussed in section E.

The STATA estimation command for estimating anARIMA(p,d,q) for the variable Y is given by

arima y, arima(p,d, q)

The estimation routine is based upon an equivalent AR(∞) model representations of the errors ε_t:

$$\varepsilon_t = \Delta^d\left(Y_t - \mu\right) - \phi_1\Delta^d\left(Y_{t-1} - \mu\right) ... - \phi_p\Delta^d\left(Y_{t-p} - \mu\right)$$

$\theta_q\, \varepsilon_{t-q}$ or

$$= \theta^{-1}(L)\varphi(L)\,\Delta^{d}\,(Y_t - \mu)$$

In either representation ε_t depends upon the φ_i's and θ_i's (autoregressive and moving average parameters). The associated sum of squared errors is given by

$$SSE(\varphi,\theta) = \sum \varepsilon_t^2$$

$$= \sum_t \{\Delta^d (Y_t - \mu) - \phi_1 \Delta^d (Y_{t-1} - \mu) \ldots - \phi_p \Delta^d (Y_{t-p} - \mu) + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}\}^2$$

$$= \sum \{\theta^{-1}(L)\,\varphi(L)\Delta^{d}(Y_t - \mu) \qquad \}^2 .$$

This formulation gives MLE for normally distributed error terms. Alternative formulations based on different probability density functions can be employed.

(3) Diagnostics--Given that the hypothesized model has been estimated, tests are performed to check the validity of the conjectured model.

- Given estimated values for the φ's and θ's , we can obtain estimated residuals from:

$$\varepsilon_t = \theta^{-1}(L)\varphi(L)\,\Delta^{d}\,(Y_t - \mu)$$

- The estimated residuals (ε_t) can then be tested for the existence of underlying patterns. This can be done by checking for patterns in the autocorrelation and partial autocorrelation coefficients associated with the estimated errors. In a correctly specified model, the estimated residuals should be white noise without statistically significant autocorrelation coefficients or partial autocorrelation coefficients. A "Q statistic" provides the basis for a statistical test of the hypothesis that the autocorrelation coefficents of the residuals are zero. Using the Box-Pierce Q-statistic

$$Q = T\Sigma_{k=1}^{p} r_k^2 \sim \chi^2(p)$$

or the Box–Ljung Q statistic

$$Q = T\,(T+2)\Sigma_{k=1}^{p} \frac{r_k^2}{T-k} \sim \chi^2(p)$$

- This analysis could also be performed by using the ARIMA command on the residuals and investigating the associated patterns of the auto and partial autocorrelation coefficients.

- Extra AR or MA parameters may be included and then test for the statistical significance of the extra terms using a "t-type" statistic or likelihood ratio test.

(4) Forecasting

The forecasts can be thought of as being generated from the equivalent MA(∞) representation or Wold decomposition, if it exists, corresponding to the identified and estimated ARIMA model. This form is given by:

$$\Delta^d (Y_t - \mu) = \frac{\Theta(L)}{\Phi(L)} \varepsilon_t$$

$$= \Gamma(L)\, \varepsilon_t = \varepsilon_t + \Gamma_1\, \varepsilon_{t-1} + \Gamma_2\, \varepsilon_{t-2} + \ldots$$

To illustrate how the estimation can be performed consider the STATA forecasting commands illustrated below:

Consider the Australian unemployment data from http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/. Let the unemployment rate be denoted by "y" and create a variable t which is the observation number. There are 211 observations in the time series.

```
generate t = _n                //this makes t = to the obs number
tsset t                        // sets t as the time variable
arima y, arima(1,0,1)          //this runs arima on the 211 observations
tsappend, add(9)               //this adds 9 obs to the data
predict yhat                   //this predicts y for all 220 observations
list y yhat t                  //this lists the 220 values for y, yhat and t
```

Alternatively, to estimate the model on a subset of the data (the first 200 observations) and to obtain forecasts on a holdout sample (the eleven observations not used in estimation), we could use the commands

```
arima y, arima(1,0,1), if t<201
predict yhat
list y yhat  t
```

The previous material has attempted to give a brief overview of alternative models, their characteristics and identification, estimation, model diagnostics, and forecasting procedures. Each of these issues will be considered in additional detail in subsequent sections. We turn to a more thorough analysis of a number of simple ARIMA(p,d,q) models–including an investigation of the behavior of the associated autocorrelation coefficients considered.

C. Characteristics of some ARIMA Models and Identification -- more detail

1. An overview and introduction–more details and some illustrations.

In this section we consider the behavior of the autocorrelation (ac) coefficients and partial autocorrelation (pac) coefficients for AR(1), MA(1), AR(p), MA(q), and ARMA(p,0,q) models along with a brief discussion of the Yule-Walker equations which can be used to provide estimates of the pac coefficents. In each case we will will review the basic mathematics illustrate the underlying patterns of corresponding autocorrelation and partial autocorrelation coefficients. The reader may only want to review the AR(1) and MA(1) discussion to get an idea of the underlying mechanics.

2. Autoregressive model of order 1 [ARIMA (1,0,0) or AR(1)]

$$Y_t - \varphi_1 Y_{t-1} = \varepsilon_t \tag{C.1}$$

where $\varepsilon_t \sim N[0,\sigma^2]$. Through recursive substitution, (C.1) can be rewritten as

$$Y_t = \frac{\varepsilon_t}{1-\phi_1 L} = \sum_{i=0}^{\infty} \phi_1^i L^i \varepsilon_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$$

From this specification, it can be shown that

$$E(y_t) = 0 \tag{C.2}$$

$$\mathrm{Var}(Y_t) = \frac{\sigma^2}{1-\varphi_1^2} = \gamma_0$$

$$\mathrm{Cov}(y_t, y_{t-s}) = (\varphi_1^s)\frac{\sigma^2}{1-\varphi_1^2} = \varphi_1^s \gamma_0 = \gamma_s$$

<u>details</u> Details associated with these results:

- $E(Y_t) =$ $E(\varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_1^2 \varepsilon_{t-2} + \ldots) = 0$
- $\mathrm{Var}(Y_t) = \mathrm{Var}(\sum \varphi_1^i \varepsilon_{t-i})$

  $= \sum \varphi_1^{2i} \mathrm{Var}(\varepsilon_{t-i})$

$$= \sigma^2 \sum \varphi_1^{2i} \quad = \sigma^2/(1 - \varphi_1^2)$$

- $$\gamma_s = E(Y_t Y_{t-s}) = E(\varepsilon_t + \varphi_1 \varepsilon_{t-1} + ... + \varphi_1^s \varepsilon_{t-s} + ...)$$

$$\bullet(\varepsilon_{t-s} + \varphi_1 \varepsilon_{t-s-1} + ...)$$

$$= \varphi_1^s E(\varepsilon_{t-s} + \varphi_1 \varepsilon_{t-s-1} + ...)^2$$

$$= \varphi_1^s \mathrm{Var}(y_{t-s})$$

$$\frac{\varphi_1^s \sigma^2}{1 - \varphi_1^2}$$

$$\rho_s = \gamma_s/\gamma_0 = \varphi_1^s$$

The autocorrelation coefficients

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \varphi_1^s$$ decay exponentially.

Note $\varphi_1 = \rho_1$

The partial autocorrelation coefficients denoted (φ_{ii}) can be shown to be $\varphi_{11} = \rho_1$ and $\varphi_{ii} = 0$ i>1.

Some sample output from the THEORYTS program illustrates these patterns. The reader might experiment with some other values of φ_i. The first example corresponds to a positive value of φ_i and the second illustrates the impact of a negative value of φ_i

. **Example 1: φ_i =.8** DOS mode: THEORYTS

```
ENTER THE NUMBER OF AUTOREGRESSIVE PARAMETERS:  1

ENTER THE NUMBER OF MOVING AVERAGE PARAMETERS:  0

ENTER THE NUMBER OF AUTOCORRELATION COEFFICIENTS:  15

ENTER THE VALUE OF PHI(1):  .8

    ****AUTO CORRELATION COEFFICIENTS****

LAGS                                          VALUES
15                      I*                    0.035
14                      I*                    0.044
13                      I*                    0.055
12                      I*                    0.069
11                      I *                   0.086
10                      I *                   0.107
9                       I  *                  0.134
8                       I  *                  0.168
7                       I   *                 0.210
 6                      I    *                0.262
 5                      I     *               0.328
 4                      I       *              0.410
 3                      I             *       0.514
 2                      I                    *     0.640                    *
 1                      I                                 0.800

______________________________________________________________
 -1                      0                         +1

          ENTER THE NUMBER OF PARTIALS:  15

    * * * PARTIAL AUTOCORRELATION COEFFICIENTS * * *

LAGS
                                    VALUES
15                     *                      -0.000
14                     *                      -0.000
13                    *                        0.000
12                    *                        0.000
11                   *                        -0.000
10                   *                         0.000
 9                   *                        -0.000
 8                   *                         0.000
 7                  *                         -0.000
 6                  *                          0.000
5                   *                         -0.000
 4                  *                          0.000
 3                  *                         -0.000
 2                  *                         -0.000
 1                            *                0.800
______________________________________________________________

-1                         0                       +1
```

Example 2: φ_i =-.8

```
ENTER 1 FOR NEW PROCESS.  2 FOR THE SAME
PROCESS:  1
ENTER THE NUMBER OF AUTOREGRESSIVE PARAMETERS:
ENTER THE NUMBER OF MOVING AVERAGE PARAMETERS:  0
ENTER THE NUMBER OF AUTOCORRELATION COEFFICIENTS:  15
ENTER THE VALUE OF PHI(1):  .-8
 * * * * AUTO CORRELATION COEFFICIENTS * * * *
LAGS
                           VALUE
15                          *I                     -0.035
14                           I*                     0.044
13                          *I                     -0.055
12                           I*                     0.069
11                      *    I                     -0.086
10                           I *                    0.107
 9                     *     I                     -0.134
8                            I  *                   0.168
7                      *     I                     -0.210
6                            I     *                0.262
5                  *         I                     -0.320
4                            I        *             0.410
3                 *          I                     -0.512
 2                           I            *         0.640
 1    *                      I                     -0.800
_____________________________________________________________
  -1                        0                  1

ENTER THE NUMBERS OF PARTIALS  15

* * * PARTIAL AUTOCORRELATION COEFFICIENTS * * *

LAGS                                              VALUES
15                                     *           -0.000
14                                     *           -0.000
13                                     *           -0.000
12                                     *           -0.000
11                                     *           -0.000
10                                     *           -0.000
 9                                     *           -0.000
 8                                     *           -0.000
 7                                     *           -0.000
 6                                     *           -0.000
 5                                     *           -0.000
 4                                     *           -0.000
 3                                     *           -0.000
 2                                     *           -0.000
 1                  *                  I           -0.800
_____________________________________________________________
```

-1 0 +1

Notes

(1) In each case of the previous two cases, the autocorrelation coefficients decline geometrically as φ_1^s. The partial autocorrelation coefficients are all equal to zero except for φ_{11} (the first) which is equal to φ_1.

(2) It will also be instructive to note that an AR(1) with $|\varphi_1|<1$ can be written as an infinite moving average MA(∞). This is the reason that the autocorrelation coefficients decline geometrically.

(3) If the process is AR(p), then the first p partial autocorrelation coefficients may be nonzero and all others zero. The corresponding autocorrelation coefficients decline to zero.

3. Moving average of order 1, ARIMA (0,0,1) or MA(1)

The MA(1) model is defined by

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} = (1-\theta_1 L)\,\varepsilon_t = \theta(L)\varepsilon_t$$

The MA(1) model is invertible (can be written as an AR(∞)) if $|\theta_1| < 1$. In particular,

$$\varepsilon_t = \theta^{-1}(L)Y_t = (1-\theta_1 L)^{-1}\, Y_t$$

$$\sum_{i=0}^{\infty} \theta^i L^i Y_t$$

$$\sum_{i=0}^{\infty} \theta^i Y_{t-i}$$

or

$$Y_t + \theta_1\, Y_{t-1} + \theta_1^2\, Y_{t-2} + \ldots = \varepsilon_t$$

which is an AR(∞). Looking at the form of the coefficients of the lagged Y_t's, suggests that the partial autocorrelation coefficients will decline geometrically. This is in fact what happens for a MA(1).

In order to determine the behavior of the autocorrelation coeffients, we derive expressions for ρ_s. From the form for a moving average model we obtain the following

- $$E(Y_t) = E(\varepsilon_t - \theta_1\,\varepsilon_{t-1}) = E(\varepsilon_t) - \theta_1\, E(\varepsilon_{t-1})$$
  $$= 0$$
- $$\mathrm{Var}\,(Y_t) = \mathrm{Var}\,(\varepsilon_t - \theta_1\,\varepsilon_{t-1})$$
  $$= \mathrm{Var}\,(\varepsilon_t) + \theta_1^2\,\mathrm{Var}\,(\varepsilon_{t-1})$$
  $$= \sigma^2 + \theta_1^2\,\sigma^2$$
  $$= \sigma^2\,(1+\theta_1^2)$$

- $\text{Cov}(Y_t\, Y_{t-1}) = E(Y_t Y_{t-1})$

$$= E(\varepsilon_t - \theta_1\varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1\varepsilon_{t-2})$$

$$= -\theta_1 E(\varepsilon_{t-1})^2$$

$$= -\theta_1\sigma^2$$

- $\text{Cov}(Y_t\, Y_{t-2}) = E(\varepsilon_t - \theta_1\varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1\varepsilon_{t-3})$

$$= 0$$

- $\text{Cov}(Y_t\, Y_{t-s}) = E(\varepsilon_t - \theta_1\varepsilon_{t-1})(\varepsilon_{t-s} - \theta_1\varepsilon_{t-s-1})$

$$= 0$$

Therefore

$$\rho_s = \begin{cases} -\dfrac{\theta}{1 + \theta_1^2} & s = 1 \\ 0 & s = 2,3,... \end{cases}$$

These results demonstrates that a moving average process of order 1 (MA (1)) only has a "memory" of one period. Similarly a moving average process of order q (MA(q)) only has a "memory" of q periods. In other words for a MA(q) model $\rho_s = 0$ for s>q. The impact of Y_t on the Y's will completely die out after q periods.

The following computer printout illustrates typical behavior of the autocorrelation and partial autocorrelation coefficients corresponding to a MA(1) with $\theta_1 = .9$. Note that there is only one nonzero autocorrelation coefficient and the partial autocorrelation coefficients decline geometrically. Computational details will be reviewed following the graphs.

Consider the following example of an MA(1) with $\theta_1 = .9$. The autocorrelations coefficients can be shown to be:

$$\rho_i = 0 \qquad \text{for } i > 2$$

$$\rho_i = \frac{-\theta_1}{1 + \theta_1^2}$$

$$= \frac{-.9}{1 + .81} = -.497$$

The following figures illustrate the corresponding ac and pac coefficients.

Example: $\theta_1 = .9$

ENTER THE NUMBER OF AR PARAMETERS: 0

ENTER THE NUMBER OF MA PARAMETERS: 1

ENTER THE NUMBER OF AC COEFFICIENTS: 15

ENTER THE VALUE OF THETA   (1): .9

```
* * * * AUTO CORRELATION COEFFICIENTS * * * *

LAGS                                                      VALUES
15                                       *                0.000
14                                       *                0.000
13                                       *                0.000
12                                       *                0.000
11                                       *                0.000
10                                       *                0.000
 9                                       *                0.000
 8                                       *                0.000
 7                                       *                0.000
 6                                       *                0.000
 5                                       *                0.000
 4                                       *                0.000
 3                                       *                0.000
 2                                       *                0.000
 1                              *        I                -0.497
________________________________________________________________
      -1                                 0             +1
```

ENTER THE NUMBER OF PAC COEFFICIENTS (15)

```
* * * PARTIAL AUTOCORRELATION COEFFICIENTS * * *

LAGS                                                      VALUES
15                                      * I               -0.041
14                                      * I               -0.045
13                                      * I               -0.051
12                                      * I               -0.057
11                                      * I               -0.065
10                                      * I               -0.073
 9                                  *     I               -0.084
 8                                  *     I               -0.096
 7                                  *     I               -0.112
 6                                 *     I                -0.131
 5                                 *     I                -0.156
 4                             *         I                -0.191
 3                         *             I                -0.243
 2                        *              I                -0.329
 1                   *                   I                -0.497
________________________________________________________________
      -1                                 0             +1
```

Important note:
Sampling variation and the possible presence of autoregressive and moving average components makes the identification process more difficult.

We now turn to an analysis of higher order AR and mixed processes.

4. Autoregressive model of order p, ARIMA (p,0,0) or AR(p)

$$Y_t - \varphi_1 Y_{t-1} - \ldots - \varphi_p Y_{t-p} = \varepsilon_t \tag{C.3}$$

$$\varphi(B) y_t = \varepsilon_t \tag{C.4}$$

(a) <u>An AR(p) will be stationary and can be written as a MA(∞) if the roots of $\varphi(z)$ are greater than one in absolute value</u>.

<u>Details</u>: Factoring the polynomical

$$\varphi(z) = 1 - \varphi_1 z - \varphi_2 z^2 \ldots \varphi_p z^p$$

$$= (1-\tilde{\varphi}_1 z)(1-\tilde{\varphi}_2 z) \ldots (1-\tilde{\varphi}_p z)$$

The roots of $\varphi(z)$ are then given by

$$(1-\tilde{\varphi}_i z) = 0 \, , \; z = 1/\tilde{\varphi}_i .$$

Now consider

$$y_t = \{1/\varphi(L)\}\varepsilon_t$$

$$= \left(\frac{1}{1-\tilde{\varphi}_1 L}\right) \cdots \left(\frac{1}{1-\tilde{\varphi}_p L}\right) \varepsilon_t$$

$$= \left(\sum_{i=0}^{\infty} \tilde{\varphi}_1^i L^i\right) \cdots \left(\sum_{j=0}^{\infty} \tilde{\varphi}_p^j L^j\right) \varepsilon_t$$

$$= \{1 + \tilde{\varphi}_1 \beta + \tilde{\varphi}_1^2 \beta^2 + \ldots\} .. \{1 + \tilde{\varphi}_p L + \tilde{\varphi}_p^2 L^2 \ldots\} \varepsilon_t$$

$$= \{1 + (\tilde{\varphi}_1 + \tilde{\varphi}_2 + \ldots + \tilde{\varphi}_p)L$$

$$+ (\tilde{\varphi}_1^2 + \tilde{\varphi}_2^2 + ... + \tilde{\varphi}_p^2 + \tilde{\varphi}_1\tilde{\varphi}_2 + ... + \tilde{\varphi}_1\tilde{\varphi}_p$$

$$+ ... + \tilde{\varphi}_{p-1}\tilde{\varphi}_p) L^2\} \varepsilon_t + ...$$

which is valid if the $|\tilde{\varphi}|<1$, i.e., the roots of $\varphi(z)$ are greater than one in absolute value.

(b) $E(Y_t) = 0$ if the series (C.3) is stationary.

(c) Yule Walker equations (an alternative approach to evaluating the pac's)

The relationship between the autocorrelation coefficients and φ_i's in (C.3) is given by

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & & \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{pmatrix}$$

or

$$\rho_k = \sum_{j=1}^{p} \varphi_j \rho_{k-j} \qquad \text{for } k > 0$$

(C.5) is referred to as the system of Yule Walker equations. The square matrix on the right hand side of (C.5) is a Toeplitz matrix in the ρ_i's. The φ_i's can then be expressed in terms of the ρ_i's and the ρ_i's can be expressed in terms of the φ_i's. For example, for

p=1: $\rho_1 = \varphi_1$

p=2:
$$\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}$$

$$\varphi_1 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \qquad \rho_1 = \frac{\varphi_1}{1 - \rho_2}$$

$$\varphi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \qquad \rho_2 = \frac{\varphi_2(1 - \varphi_2) + \varphi_1^2}{1 - \rho_2}$$

Derivation of the Yule Walker Equations: Multiply (C.3) by y_{t-k} and take the expected value

$$E(y_t\, y_{t-k}) - \varphi_1 E(y_{t-1}\, y_{t-k}) \ldots\ \varphi_p\, E(y_{t-p}\, y_{t-k})$$
$$= E(\varepsilon_t\, y_{t-k})$$

or

$$\gamma_k - \varphi_1 \gamma_{k-1} \cdots\ \varphi_p\, \gamma_{k-p} = 0$$

Dividing by $\gamma_0 = Var(y_t)$ yields

$$\rho_k - \varphi_1\, \rho_{k-1} - \ldots\ \varphi_p \rho_{k-p} = 0.$$

(d) Partial Autocorrelation Coefficients

If different values of p are selected and the "last coefficient" φ_p obtained for each p, using the Yule Walker equations, these coefficients are referred to as partial autocorrelation coefficients and are useful in determining the order of the autoregressive process. This is analogous to deciding on how many terms to include in a multiple regression. For an autoregressive process of order p, the first p partial autocorrelations will be nonzero and higher order partial autocorrelation coefficients will equal zero. The partial autocorrelation coefficients are denoted by φ_{ii}.

- As an example for p = 1, the Yule Walker equations are

  $$\rho_1 = \varphi_1 = \varphi_{11}$$

- For p = 2, the Yule Walker equations are

$$\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix};$$

therefore, using Cramer's Rule to solve for φ_2 yields

$$\varphi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

- For p = 3, the Yule Walker equations are

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix}.$$

The corresponding third partial autocorrelation coefficient is

$$\varphi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

- More generally

$$\varphi_{ii} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & & \vdots \\ \rho_{i-1} & \rho_{i-2} & & \rho_i \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-1} \\ \rho_1 & 1 & \cdots & \rho_{i-2} \\ \vdots & \vdots & & \vdots \\ \rho_{i-1} & \rho_{i-2} & \cdots & 1 \end{vmatrix}}$$

<u>Note</u>:

If the <u>actual value</u> of p is 1, then

$$\varphi_{11} = \rho_1$$

$$\varphi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\rho_1^2 - \rho_1^2}{1 - \rho_1^2} = 0$$

since $\rho_i = \varphi_1^i$ for a first order autoregressive process. Therefore,

$$\varphi_{ii} = 0 \quad i \geq 2.$$

The determination of an appropriate estimate for p becomes a statistical question. The ρ_k can be estimated by

$$\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) / \Sigma(y_t - \bar{y})^2$$

The Yule-Walker equations can then be used to estimate the corresponding partial autocorrelation coefficients. The associated asymptotic standard errors were shown to be

$$s_{\hat{\varphi}_{kk}} = \frac{1}{\sqrt{T}} \qquad k \geq + 1$$

by Quenouille. If one assumes that n (number of observations used in fitting) is large enough for $\hat{\varphi}_{kk}$ to be approximately normally distributed, we have a procedure which can be used in determining a reasonable value of p.

Consider the following AR(2) example $(\varphi_1, \varphi_2) = (.2, .7)$ using the Yule Walker

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2}$$

$$\rho_0 = 1$$

$$\rho_1 = \frac{\varphi_1}{1 - \varphi_2} = \frac{02}{1 - .7} = \frac{2}{3}$$

$$= .667$$

$$\rho_2 = \frac{\varphi_2(1 - \varphi_2) + \varphi_1^2}{1 - \varphi}$$

$$= \frac{(.7)(.3) + .04}{1 - .7}$$

$$= \frac{.25}{.30} = .833$$

$$\varphi_{11} = \rho_1 = .667$$

$$\varphi_{22} = \frac{\rho_2 - \rho_2^2}{1 - \rho_1^2} = \frac{\left(\frac{5}{6}\right) - \left(\frac{2}{3}\right)^2}{1 - \left(\frac{2}{3}\right)^2} \qquad \frac{21}{30} = .7 \qquad =$$

$$\varphi_{33} = 0$$

Example: $(\varphi_1, \varphi_2) = (.2, .7)$

```
ENTER THE NUMBER OF AUTOREGRESSIVE PARAMETERS: 2

ENTER THE NUMBER OF MOVING AVERAGE PARAMETERS: 0

ENTER THE NUMBER OF AUTOCORRELATION COEFFICIENTS: 15

ENTER THE VALUE OF PHI(1): .2

ENTER THE VALUE OF PHI(2): .7

* * * * AUTO CORRELATION COEFFICIENTS * * * *

LAGS                                         VALUES
15                    I         *             0.343
14                    I         *             0.368
13                    I          *            0.384
12                    I          *            0.416
11                    I           *           0.430
10                    I            *          0.471
9                     I             *         0.480
8                     I              *        0.536
7                     I              *        0.533
6                     I               *       0.614
5                     I              *        0.585
4                     I                 *     0.710
3                     I               *       0.633
2                     I                   *   0.833
1                     I              *        0.667

      -1              0                         +1

ENTER THE NUMBER OF PARTIALS     15

 * * * PARTIAL AUTOCORRELATION COEFFICIENTS * * *

LAGS                                          VALUES
15                                            -0.000
14                                             0.000
13                                             0.000
12                    *                        0.000
11                    *                        0.000
10                    *                       -0.000
9                     *                        0.000
8                     *                       -0.000
7                     *                       -0.000
6                     *                        0.000
5                     *                       -0.000
4                     *                        0.000
3                     *                       -0.000
2                     I          *             0.700
1                     I         *              0.667

      -1              0                         +1
```

5. q^{th} order moving average model, ARIMA (0,0,q) or MA(q)

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots -\theta_q\varepsilon_{t-q} \tag{C.8}$$

$$= \theta(L)\,\varepsilon_t$$

In order for the process defined by (C.8) to be invertible, the roots of θ(L) must have modulus greater than one,

$$\theta(z) = \prod_{i=1}^{q} (1 - \theta_i z) = 0$$

Hint: $\theta^{-1}(L) = \prod_{j=1}^{q} (1 - \tilde{\theta}_j L)^{-1} = \prod_{j=1}^{q} \sum_{i=0}^{\infty} (\tilde{\theta}_j^i L^i)$

all j.

The roots of θ(z) are equal to $1/\tilde{\theta}_j$.

The autocovariances and autocorrelations can be evaluated by considering

$$\gamma_k = E(y_t\, y_{t-k})$$

$$\gamma_k = E[\varepsilon_t - \theta_1\varepsilon_{t-1} - \dots -\theta_q\varepsilon_{t-q})(\varepsilon_{t-k} - \theta_1\varepsilon_{t-k-1} - \dots - \theta_q\,\varepsilon_{t-k-q})] \tag{C.9}$$

$$\gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2.$$

$$\gamma_k = (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_q\theta_{q-k})\sigma^2 \tag{C.10}$$

for k = 1, 2, ... , q

= 0 for k > q.

$$\rho_k = \frac{-\theta_k + \theta_{k+1}\theta_1 + \dots + \theta_q\theta_{q-k}}{1 + \theta_1^2 + \dots + \theta_q^2}$$

k=1,2,...,q (C.11)

= 0 for k > q

From (C.10) we see that the autocorrelation function of a MA(q) has a "cut-off" at lag q. We might say that a MA(q) has a memory of Length q.

Bartlet's approximation for the standard error of estimators of ρ_k is useful in determining an estimate of q. For any process for which the autocorrelation's ρ_k are zero for k > q, Bartlet's approximation is given by

$$s^2_{\hat{\rho}_k} = \left(\frac{1}{T}\right)\left\{1 + 2\sum_{i=1}^{q} \rho_i^2\right\} \text{ for } k>q$$

The reader should be reminded that the appropriateness of the convention of using the limiting normal density with (C.7) or (C.12) for purposes of assessing statistical significance is questionable for small samples.

Consider the following example:

The auto and partial autocorrelation coefficients corresponding to

MA(2): $y_t = \varepsilon_t - .5\varepsilon_{t-1} - .3\varepsilon_{t-2}$

are given in the next figure.

$$\rho_1 = \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2} = -.26 \quad \text{and}$$

$$\rho_2 = \frac{-\theta_1}{1 + \theta_1^2 + \theta_2^2} = -.224$$

Example of an MA(2) with $(\theta_1 = .5, \theta_2 = .3)$

```
ENTER THE NUMBER OF AUTOREGRESSIVE PARAMETERS: 0
ENTER THE NUMBER OF MOVING AVERAGE PARAMETERS: 2
ENTER THE NUMBER OF AUTOCORRELATION COEFFICIENTS:  15
ENTER THE VALUE OF THETA(1):  .5
ENTER THE VALUE OF THETA(2):  .3
    * * * AUTO CORRELATION COEFFICIENTS * * *
LAGS                                                      VALUES
15                                      *                 0.000
14                                      *                 0.000
13                                      *                 0.000
12                                      *                 0.000
11                                      *                 0.000
10                                      *                 0.000
 9                                      *                 0.000
 8                                      *                 0.000
 7                                      *                 0.000
 6                                      *                 0.000
 5                                      *                 0.000
 4                                      *                 0.000
 3                                      *                 0.000
 2                               *      I                 -0.224
 1                          *           I                 -0.261
__________________________________________________________________
              -1                        0             +1


ENTER THE NUMBER OF PARTIALS:  15
* * * PARTIAL AUTOCORRELATION COEFFICIENTS * * *
LAGS                             VALUES

15                                      *I                -0.023
14                                     * I                -0.027
13                                     * I                -0.032
12                                     * I                -0.037
11                                     * I                -0.044
10                                     * I                -0.052
 9                                     * I                -0.062
 8                                     * I                -0.074
 7                                *     I                 -0.088
 6                                *     I                 -0.107
 5                               *      I                 -0.127
 4                               *      I                 -0.165
 3                          *          I                  -0.189
 2                        *             I                 -0.313
 1                          *           I                 -0.261
__________________________________________________________
       -1                               0           +1
```

6. Autoregressive Moving Average Processes, ARIMA (p,0,q), ARMA (p,q)

$$Y_t - \varphi 1 Y_{t-1} - \ldots - \varphi_p Y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \tag{C.13}$$

$$\varphi(L)Y_t = \varphi(L)\varepsilon_t$$

The process defined by (C.13) will be stationary if the roots of $\varphi(L)$ have modulus greater than one and will be invertable if the roots of $\theta(L)$ have modulus greater than one.

Given that the conditions of stationarity and invertibility are satisfied, we note that an ARMA (p,q) process can be expressed as

$$\text{AR}(\infty): \ \theta^{-1}(L)\varphi(L)Y_t = \varepsilon_t \tag{C.14}$$

$$\text{MA}(\infty): \ Y_t = \varphi^{-1}(L)\theta(L)\varepsilon_t \tag{C.15}$$

Multiplying (C.13) by y_{t-k} and taking expected values we see that

$$\rho_k = \varphi_1\rho_{k-1} + \ldots + \varphi_p\rho_{k-p} + \rho_k(Y,u) - \theta_1 \rho_{k-1}(Y,u) \tag{C.16}$$

$$-\theta_q\rho_{k-q}(y,u)$$

where $\rho_i(Y,u) = E(Y_{t-i}, u_t) = 0 \ \ i > 0$.

(C.16) simplifies to

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2} + \ldots + \varphi_p \rho_{k-p} \text{ for } k \geq q + 1$$

The first q autocorrelation coefficients will depend upon the moving average parameters θ_i as well as the autoregressive parameters φ_j. Therefore, the autocorrelation coefficients will exhibit an irregular pattern at lags 1 through g, then tail off according to (C.16).

The process (C.13) is equivalent to an AR(∞); hence, the partial autocorrelation coefficients tail off eventually in the same manner as with a pure autoregressive process.

These results will assist in the determination of p, q. The following tables are taken from Nelson (p. 89) and Box and Jenkins (176,7) and provide useful summary information to assist in the determination of p and q. It should be noted that the autocorrelation coefficients and partial autocorrelations provide the basis for this determination. Recall that the asymptotic standard errors of $\hat{\rho}_k$ and $\hat{\varphi}_{kk}$ are given by

$$s_{\hat{\rho}_k} = \left(\frac{1}{\sqrt{T}}\right)\left\{1 + 2(\rho_1^2 + \ldots + \rho_q^2)\right\}^{1/2}$$

$$s_{\hat{\varphi}_{kk}} = \frac{1}{\sqrt{T}} \quad \text{for } k > p.$$

Table 1. Characteristic Behavior of Autocorrelations and Partial Autocorrelations for Three Classes of Processes.

Class of processes	Autocorrelations	Partial autocorrelations
Moving average(q)	Spikes at lags 1 through q, then cut off	Tail off
Autoregressive (p)	Tail off according to $\rho_j = \varphi_1\rho_{j-1} + \ldots + \varphi_p\rho_{j-p}$	Spikes at lags 1 through p, then cut off
Mixed auto-regressive-moving average	Irregular pattern at lags 1 through q,then tail off according to $\rho_j = \varphi_1\rho_{j-1} + \ldots + \varphi_p\rho_{j-p}$	

Table 2. Behavior of the autocorrelation functions for the dth difference of an ARIMA process of order (p,d,q). (Table A and Charts B, C, and D are included at the end of this volume to facilitate the calculation of approximate estimates of the parameters for first-order moving average, second-order autoregressive, second-order moving average, and for the mixed (ARMA) (1,1) process.)

Order	(1,d,0)	(0,d,1)
Behavior of ρ_k	decays exponentially	only ρ_1 nonzero
Behavior of ϕ_{kk}	only φ_{11} nonzero	decay exponentially
Preliminary estimates	$\varphi_1 = \rho_1$	$\rho_1 = \frac{\theta_1}{1 + \theta^2}$
Admissible region	$-1 < \varphi_1 < 1$	$-1 < \theta_1 < 1$

Order	(2,d,0)	(0,d,2)
Behavior of ρ_k	mixture of exponentials or damped sine wave	only ρ_1 and ρ_2 nonzero
Behavior of φ_{kk}	only φ_{11} and φ_{22} are nonzero	dominated by mixture of exponentials or damped sine wave
Preliminary estimates	$\varphi_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2}$	$\rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2\ \theta_2^2}$
	$\varphi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho^2}$	$\rho_2 = \frac{\theta_2}{1 + \theta_1^2\ \theta_2^2}$
	$-1 < \varphi_2 < 1$	$-1 < \theta_2 < 1$
Admissible region	$\varphi_2 + \varphi_1 < 1$	$\theta_2 + \theta_1 < 1$
	$\varphi_2 - \varphi_1 < 1$	$\theta_2 - \theta_1 < 1$

Order	(1,d,1)
Behavior of ρ_k	decays exponentially from first lag
Behavior of φ_{kk}	dominated by exponential decay from first lag
Preliminary estimates from	$\rho_1 = \frac{(1-\theta_1\varphi_1)(\varphi_1-\theta_1)}{1 + \theta_1^2 - 2\varphi_1\theta_1}$ $\quad \rho_2 \quad \rho_1\varphi_1$
Admissible region	$-1 < \varphi_1 < 1 \qquad -1 < \theta_1 < 1$

(5) ARIMA (p,d,q) The foregoing discussion has assumed that the underlying series is stationary. If this is not the case, for any of several reasons, the previous approaches are not strictly appropriate. For example if a time series didn't have a fixed mean, then the series wouldn't be stationary. It will frequently be the case that $Y_t - Y_{t-1} = (I-L)Y_t$, or $(I-L)^2Y_t$, or $(I-L)^dY_t$, will be stationary. If such a value of d can be determined then the previous techniques can be applied to the "differenced" series, $(I-L)^dY_t$. For example if Y_t is basically distributed with constant variance about a linear (quadratic) trend, then $(I-L)Y_t$, $(I-L)^2Y_t$, will be stationary. If Y_t shows evidence of trend stationarity, then Y could be regressed on a polynomial in "t" and the previously described techniques can be applied to the resulting residuals. Sometimes nonlinear transformation on Y_t , such as ℓnY_t, may facilitate the search for a stationary process.

D. Diagnostic Analysis

There are several approaches which can be utilized to determine the "validity" of the estimated ARIMA models. Three of the most common involve : (1) considering more general models, (2) an analysis of estimated residuals, and (3) the Q-statistic.

1. Generalized Model. Assume that ARIMA(p,d,q) has been "identified." The researcher might estimate an ARIMA(p',d,q') where p' and/or q' are larger than p, q and then check the statistical significance of the additional coefficients. This approach has at least two limitations. The validity of the statistical inference (test statistics) is questionable for small samples and an ARIMA (p,d,q) process is uniquely determined by the autocorrelation structure up to a multiple of polynomials in L. t "type" statistics or likelihood ratio tests may be used.

2. Analysis of Estimated Residuals

The estimated residuals can be obtained from the stimated ARIMA model as follows:

$$\hat{\varepsilon}_t = \theta^{-1}(L)\ \hat{\varphi}(L)(I-L)^d y_t$$

One might consider an analysis of the behavior of autocorrelation coefficients

$$\hat{\rho}_k(\hat{\varepsilon}_t) = \sum_t \hat{\varepsilon}_t \hat{\varepsilon}_{t-k} / \sum_t \hat{\varepsilon}_t^2$$

and partial autocorrelation coefficients corresponding to the estimated residuals.

It should be mentioned that the distributional characteristics of $\hat{\varepsilon}_t$ are not necessarily exactly the same as for ε_t. If the model has been correctly specified, the estimated residuals and the associated auto and partial autocorrelation coefficients should correspond to white noise. However, if the estimated residuals appear to have an AR or MA component, then the model specification should be respecified. For example, if the $\hat{\varepsilon}_t$'s appear to be an AR(1), then one more AR component should be included in the specification for the Y_t series.

3. Q-Statistics.

Box and Pierce and Box and Ljung, respectively, define the following Q-statitics

$$Q = n \sum_{k=1}^{M} \hat{\rho}_k^2(\hat{\varepsilon}_t)$$

$$Q = T(T+2)\sum_{k=1}^{M} \frac{\hat{\rho}_k^2}{T-k}$$

which can be used to test the hypothesis that the autocorrelations coefficients are zero. The two Q-statistics are asymptotically distributed as $\chi^2(M-p-q)$ under the hypotheses of ε_t being independently and identically distributed as $N(0,\sigma^2)$. This follows because $\hat{\rho}_i \sim N[0, 1/T^{.5}]$ if the model is correctly specified. The hypothesis is rejected by large values of Q.

E. Estimation

1. Background. Once values for p and q have been determined, the coefficients in the ARIMA (p,d,q) need to be estimated in order to use the model.

$$(1-\varphi_1 L - \ldots -\varphi_p L^p)Y_t = (1-\theta_1 L - \ldots -\theta_q{}^q)\varepsilon_t \qquad \text{(E.1)}$$

or

$$\varphi(L)Y_t = \theta(L)\varepsilon_t$$

Note that ε_t can be explicitly expressed as

$$\varepsilon_t = Y_t - \varphi_1 Y_{t-1} \ldots -\varphi_p Y_{t-p} + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\, \varepsilon_{t-q} \qquad \text{(E.2)}$$

$$= \varphi(B)Y_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}$$

and also

$$= \theta^{-1}(L)\varphi(L)Y_t \qquad \text{(E.3)}$$

Note that in both representations, ε_t depends upon the φ_i's and θ_i's (autoregressive and moving average parameters). The associated sum of squared errors is given by

$$SSE(\varphi,\theta) = \sum \varepsilon_t^2 \qquad \text{(E.4)}$$

$$= \sum_t (Y_t - \varphi_1 Y_{t-1} \ldots -\varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q})^2$$

$$= \sum \{\theta^{-1}(L)\ \varphi(L)Y_t\}^2$$

Under the assumption that the ε_t's are independently and identically distributed as $N(0,\sigma^2)$ the <u>log likelihood</u> function is

$$\ell(\theta,\varphi,\sigma^2) = \ell n \frac{e^{-\Sigma\varepsilon_t^2/2\sigma^2}}{(2\pi)^{N/2}(\sigma^2)^{N/2}}$$

$$= \frac{-SSE(\varphi,\theta)}{2\sigma^2} - \frac{N}{2}\ell n(2\pi) - \frac{N}{2}\ell n(\sigma^2)$$

Hence, minimizing $SSE(\varphi,\theta)$ with respect to φ and θ is a necessary part of obtaining maximum likelihood estimators and the normality assumption.

2. Numerical optimization.

A close inspection of the first expression for SSE in (E.4) reveals that ε_t depends on the previous random disturbances and observations on the variable Y, e.g.

$$\varepsilon_1 = Y_1 - \varphi_1 Y_0 - \varphi_2 Y_{-1} \ldots -\varphi_p Y_{1-p} + \theta_1 \varepsilon_0 + \ldots + \theta_q \varepsilon_{1-q}$$

Several approaches to these problems have been proposed. This is frequently referred to as the question as to how to <u>initialize the series</u>. One approach is to replace the unobservable values of Y_t and ε_t by their expected values--in this case 0; hence

$$\varepsilon_1 = Y_1$$

$$\varepsilon_2 = Y_2 - \varphi_1 Y_1 + \varphi_1 \varepsilon_1$$

$$\varepsilon_3 = Y_3 - \varphi_1 Y_2 - \varphi_2 Y_1 + \varphi_1 \varepsilon_2 + \varphi \varepsilon_1$$

The associated sum of squared error terms is

$$\sum_{t=1}^{N} \varepsilon_t^2$$

Another approach is to start the sum at t = p + 1

$$SSE = \sum_{t=p+1}^{N} (y_t - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} \dots + \theta_q \varepsilon_{t-q})^2$$

where ε_p, ε_{p-1} , ... , ε_{p+1-q} are set equal to zero.

The minimization of (E.4) requires a nonlinear optimization routine if there is a moving average component to the series, i.e., some of the θi's are nonzero. *If there is not a moving average component, (all θi's = 0), then a regular linear regression package can be used in the estimation process. A number of studies have found that autoregressive models perform very well. Recall that an ARIMA(p,d,q) model can be expressed as an AR(∞).

Nonlinear optimization routines require that initial estimates of the parameters be provided. The Yule Walker equations are frequently used for this purpose.

$$\begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & & . \\ . & . & & . \\ \rho_{p-1} & & \cdots & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_p \end{pmatrix}$$

where ρ_i is estimated by

$$\hat{\rho}_i = \frac{\sum_t (y_t - \bar{y})(x_t - \bar{x})/n}{\sum_t (y_t - \bar{y})^2/n} = \frac{\operatorname{cov}(y_t, y_{t-i})}{\operatorname{var}(y_t)}$$

3. Nonnormal data

MLE estimates of the ARIMA model can be obtained by maximizing the log-likelihood function

$$\ell(\;) = \sum_{t=1}^{n} \ell n f\left(\Theta^{-1}(L)\Phi^{-1}(L)\Delta^d Y_t; \psi\right)$$

over the ARIMA and distributional parameters.

F. Forecasts

1. Forecasts

Let $Y_n(h)$ denote a forecast of Y_{n+h} made at time t = n. Assume that Y_t is an ARIMA(p,d,q) stochastic process, i.e.,

$$\varphi(L)Y_t = \theta(L)\varepsilon_t \tag{F.1}$$

Y_t can be expressed in terms of ε_t as

$$\begin{aligned} Y_t &= \frac{\theta(L)}{\varphi(L)}\,\varepsilon_t \\ &= \Gamma(L)\,\varepsilon_t \\ &= 1\,\varepsilon_t + \Gamma_1\varepsilon_{t-1} + \Gamma_2\,\varepsilon_{t-2} + \ldots \end{aligned} \tag{F.2}$$

From (F.2) we can express Y_{n+h} as

$$Y_{n+h} = \underbrace{\varepsilon_{n+h} + \Gamma_1\varepsilon_{n+h-1} + \ldots + \Gamma_{h-1}\varepsilon_{n+1}}_{\text{Unknown at t=n}} + \underbrace{\Gamma_h\varepsilon_n + \Gamma_{h+1}\,\varepsilon_{n-1} + \ldots}_{\text{Can be estimated at t=n}} \tag{F.3}$$

It can be shown that the optimal (minimum mean squared error) forecast of Y_{N+h} at time N is given by

$$\begin{aligned} Y_n(h) &= \Gamma_h\varepsilon_n + \Gamma_{h+1}\,\varepsilon_{n-1} + \Gamma_{h+2}\varepsilon_{n-2} + \ldots \\ &= \sum_{j=0}^{\infty} \Gamma_{h+j}\,\varepsilon_{n-j} \end{aligned} \tag{F.4}$$

A recurrence relationship can be developed from (F.3) which facilitates evaluation of forecasts.

$$\begin{aligned} Y_n(h) &= \sum_{i=1}^{p} \varphi_i Y_n(h-i) + \sum_{i=0}^{\infty} \theta_{i+h}\,\varepsilon_{n-i} \\ &= \varphi_1\,Y_n(h-1) + \ldots + \varphi_p y_n(h-p) \\ &\quad + \theta_h\varepsilon_n + \theta_{h+1}\,\varepsilon_{n-1} + \ldots + \theta_{q+h}\,\varepsilon_{n-q} \end{aligned}$$

As an example consider forecasts corresponding to an **AR(1) model.**

$$Y_n(h) = \varphi_1 Y_n(h-1)$$

$$Y_n(1) = \varphi_1 Y_n$$

$$Y_n(h) = \varphi_1^h Y_n$$

2. Forecast errors and confidence intervals.

The forecast error is defined by

$$e_N(h) = Y_{n+h} - Y_n(h) \tag{F.6}$$

$$= \sum_{i=0}^{\infty} \Gamma_i \varepsilon_{n+h-i} - \sum_{j=0}^{\infty} \Gamma_{h+j} \varepsilon_{n-j}$$

$$= \varepsilon_{n+h} + \Gamma_1 \varepsilon_{n+h-1} + \ldots + \Gamma_{h-1} \varepsilon_{n+1}$$

(see equations F.3 and F.4) , where Y_{n+h} and $Y_n(h)$, denote the actua forecast at t+n, respectively.

The variance of the forecast error is given by

$$\mathrm{Var}(e_N(h)) = \mathrm{Var}(\varepsilon_{n+h} + \ldots + \Gamma_{h-1} \varepsilon_{n+1}) \tag{F.7}$$

$$= \mathrm{Var}(\varepsilon_{n+h}) + \ldots + \Gamma^2_{h-1} \mathrm{Var}(\varepsilon_{n+h})$$

$$= \sigma_\varepsilon^2 (1 + \Gamma_1^2 + \ldots + \Gamma_{h-1}^2)$$

$$\mathrm{Var}(e_n(h)) = \sigma_\varepsilon^2 (1 + \Gamma_1^2 + \ldots + \Gamma_{h-1}^2) \tag{F.7}$$

Note: (1) That the variance of the forecast error increases as the lead time increases

(2) σ_ε^2 can be estimated by

$$\hat{\sigma}_\varepsilon^2 = \frac{SSE}{N-p-q}$$

(3) "Asymptotic" confidence intervals are given by

$$Y_n(h) \pm Z\alpha \sqrt{\hat{\sigma}^2 (\Gamma_0^2 + \ldots + \Gamma_{h-1}^2)}$$

The forecasts will eventually converge to the "trend."

(4) The expression for the variance of the forecast error in (F.7) doesn't take account of parameter uncertainty.

G. General Comments

1. It should be mentioned that

$$\varphi(L)(I-L)^d y_t = \theta(L)\,\varepsilon_t$$

is uniquely determined by a given autocorrelation structure (given d) up to a multiple of a polynomial in L. For example,

$$Y_t - .5Y_{t-1} = \varepsilon_t \text{ and}$$
$$Y_t - .75Y_{t-1} + .125Y_{t-2} = (1-.5L)(1-.25L)Y_t = \varepsilon_t - .25\varepsilon_{t-1}$$

Are observationally equivalent. This suggests working with the simplest possible model, the Principle of Parsimony.

2. If the time series exhibits seasonal behavior [large ρ_{12} (monthly) or ρ_4 (quarterly)] then the previously developed model can be modified to incorporate this behavior into the modeling process. Seasonal components can be incorporated into an ARIMA model in several different ways.

- $\Phi(L)(1-L)^d(Y_t-\mu) = \Theta(L)\,\varepsilon_t$ $\quad \Delta = I - L \quad$ basic mode
- $\Phi(L)(1-L)^d(1-L^s)(Y_t-\mu) = \Theta(L)\,\varepsilon_t$
- $\Phi_s(L^s)\Phi(L)(1-L)^d(Y_t-\mu) = \Theta_s(L^s)\Theta(L)\varepsilon_t$

  Seasonal differencing with seasonal AR and MA components

STATA can accomodate these variations of the basic ARIMA model.

3. Just a reminder: the simple exponential smoothing is an ARIMA(0,1,1) model and the Holt-Winters nonseasonal predictor is an ARIMA (0,2,2) model.

4. A number of texts suggest that time series models based on ARIMA formulations should only be expected to be "successful" if at least 40 observations are available. For smaller samples, use other techniques–such as exponential smoothing or Holt Winters. . .

5. A complementary method of analysis is that of spectral analysis. The spectral density of a stationary series is given by

$$g(f) = 2\left\{1 + 2\sum_{k=1}^{\infty} \rho_k \operatorname{Cos} 2\pi fk\right\}$$

$0 \le f \le ½$

In practice this is estimated using "lag windows" by

$$g(f) = 2\left\{1 + 2 \sum_{k=1}^{n-1} \lambda_k \rho_k \ \mathrm{Cos}\, 2\pi f k\right\}$$

The spectral density function provides information about the cyclical behavior of a series and shows how the variance of a stochastic process is distributed between a continuous range of frequencies.

<u>Spectral Density Functions for</u>

ARMA(p,q) $\varphi(L)y_t = \theta(L)\varepsilon_t$

$$y_t = \varphi^{-1}(L)\theta(L)\varepsilon_t$$

The spectral density function is given by

$$f(u) = \sigma^2 \frac{\theta(e^{iu})\theta(e^{-iu})}{\varphi(e^{iu})\varphi(e^{-iu})}$$

$$-\pi < u < \pi$$

DeMoivre's Theorem.

$$e^{i\theta} = \cos\theta + i \sin\theta$$

$$(cos\theta + i\sin\theta)^m = e^{i\theta m} = \cos m\theta + \sin m\theta$$

<u>Model</u>	Spectral Density Function
ARMA(0,0)	σ_ε^2
AR(1)	$\sigma_\varepsilon^2/(1-2\varphi_1 \cos\omega + \sigma_1^2)$
AR(2)	$\sigma^2/[1 + \varphi_1^2 + \sigma_2^2 - 2\varphi_1(1-\varphi_2)\cos\omega - 2\varphi_2 \cos 2\omega]$
MA(1)	$\sigma^2(1 + \theta_1^2 - 2\theta_1 \cos\omega)$
MA(2)	$\sigma^2(1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1-\theta_2)\cos\omega - 2\theta_2 \cos 2\omega)$

ARMA(1,1) $$\frac{\sigma^2(1 + \theta_1^2 - 2\theta_1\cos\omega)}{(1 + \varphi_1^2 - 2\varphi_1\cos\omega)}$$

ARMA(1,2) $$\frac{\sigma^2(1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1-\theta_2)\cos\omega - 2\theta_2(\cos2\omega)}{(1 + \varphi_1^2 - 2\varphi_1\cos\omega)}$$

ARMA(2,1) $$\frac{\sigma^2(1 + \theta_1^2 - 2\theta_1\cos\omega)}{1 + \varphi_1^2 + \varphi_2^2 - 2\varphi_1(1-\varphi_2)\cos\omega}$$

ARMA (2,2) $$\frac{\sigma^2(1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1-\theta_2)\cos\omega - 2\theta_2\cos2\omega)}{1 + \varphi_1^2 + \varphi_2^2 - 2\varphi_1(1-\varphi_2)\cos\omega - 2\varphi_2\cos\omega}$$

<u>APPENDIX: COMPUTER PROGRAMS</u> (TIME SERIES ANALYSIS)

I. THEORYTS–look at ac and pac

ENTER THE NUMBER OF AUTOREGRESSIVE PARAMETERS: 2

ENTER THE NUMBER OF MOVING AVERAGE PARAMETERS: 0

ENTER THE NUMBER OF AUTOCORRELATION COEFS:25

ENTER THE NUMBER OF PARTIAL AUTO CORRELATION COEFFICIENTS: 25

ENTER THE VALUE OF PHI(1): .2

ENTER THE VALUE OF PHI(2): .7

II. RUN SHR:DATATSF–generates ARIMA series

HOW MANY SERIES DO YOU WANT? 1

HOW MANY OBSERVATIONS ARE TO BE IN EACH SERIES? 300

WHAT ARE THE AR PARAMETERS? (UP TO TWO) .9

WHAT ARE THE MA PARAMETERS? (UP TO TWO)

DO YOU WANT THE SIBYL RUNNER HEADING? N

WHAT IS THE SEASONAL AR PARAMETER? 0

WHAT IS THE SEASONAL MA PARAMETER?

HOW MANY TIMES SHOULD THE SERIES BE SUMMED? 0

HOW MANY TIMES SEASONALLY SUMMED? 0

WHAT IS THE SPAN OF SEASONALITY? 12

WHAT IS THE MEAN OF THE SERIES? 0

WHAT TYPE OF ERROR TERMS DO YOU WANT?
(N=NORMAL, L=LOG-NORMAL, P-PARETO, E=EXPONENTIAL) N

WHAT VALUE OF SIGMA DO YOU WANT? 1. **(remember the decimal)**

STOP

III. RUN SHR:BYUTSF-estimates ARIMA models with different pdf’s

* * * * BYU TIME SERIES FORECASTING PACKAGE * * * *

WOULD YOU LIKE TO TEST LEVELS OF DIFFERENCING? N (see next section, used for identification)

WHAT FORECASTING TECHNIQUE WOULD YOU LIKE TO RUN?
FOR HELP TYPE HELP

THE FOLLOWING FORECASTING TECHNIQUES ARE AVAILABLE

BOXJN	=	BOX-JENKINS MODEL
EXPS	=	SINGLE PARAMETER EXPONENTIAL SMOOTHING
WINTR	=	WINTER'S 3-PARAMETER MODEL
MARM	=	TRANSFER FUNCTIONS (MARMA MODELS)

WHICH OF THESE WOULD YOU LIKE TO RUN? BOXJN

WHAT IS THE DATA FILE NAME? SER1.DAT

HOW MANY OBSERVATIONS ARE TO BE USED? 300

IS THIS A SIBYL-RUNNER FILE? N

WOULD YOU LIKE TO TEST LEVELS OF DIFFERENCING? N

WHAT IS THE DATA FILE NAME? L.DAT

HOW MANY OBSERVATIONS ARE TO BE USED? 50

IS THIS A SIBYL-RUNNER FILE? N

DO YOU WANT A GRAPH OF THE DATA? N

HOW MANY TIMES MUST THE SERIES BE DIFFERENCED? 1

HOW MANY TIMES MUST THE SERIES BE SEASONALLY DIFFERENCED?

WHAT IS THE SPAN OF SEASONALITY?

DO YOU WANT TO SEE GRAPH OF THE NEW DATA? N

WOULD YOU LIKE TO SEE THE SPECTRUM? N

HOW MANY AUTOCORRELATION COEFFICIENTS ARE TO BE SEEN? 10

HOW MANY PARTIAL AUTO'S ARE TO BE SEEN? 10

WOULD YOU LIKE TO REPEAT THIS PROCEDURE? N