Contents

Chapter 3

Univariate ($ARIMA$) Time Series Models

One might question why univariate processes are worth studying in the first place since they appear to be so restricted. A famous theorem developed by Wold (1938) provides such a foundation.

3.1 Wold Decomposition Theorem

We present but do not prove the famous **Wold Decomposition Theorem**

- If y_t is weakly stationary mean zero asymptotically independent (memory of process is not infinite), then we can express y_t as an infinite moving average $MA(\infty)$:

$$y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} + \kappa_t \tag{3.1}$$

$$\epsilon_t = y_t - \hat{E}[y_t \mid y_{t-1}, y_{t-2,} \ldots] \tag{3.2}$$

  where ϵ_t is white noise and represents the forecasting error based on the linear projection of past information, κ_t is called the *linearly deterministic component* which can be predicted arbitrarily well from a linear function of past y_t (trends, seasonal dummies and so on)

$$\kappa_t = \hat{E}[\kappa_t \mid y_{t-1}, y_{t-2,} \ldots]$$

- The $\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ is called the linear indeterministic component.
- If $\kappa_t \equiv 0$, the process is called *purely linearly indeterministic*.
- This provides the motivation often appealed to in applied work for expressing a covariance stationary process as a infinite moving average.

- Under certain assumptions we may express the infinite polynomial as the ratio of 2 finite-order polynomial:

$$\sum_{i=1}^{\infty} L^i \psi_i = \frac{1 + \theta_1 L + \ldots + \theta_q L^q}{1 - \phi_1 L - \ldots - \phi_p L^p} \tag{3.3}$$

First we shall present the general $ARMA$ model and then specialize it to analyze various important cases. In this part, we deal only with *weakly stationary processes.* Discussion of nonstationary time series will be taken up at the end.

ARMA(p, q)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \tag{3.4}$$

To ensure that this general process is *stationary* we need to impose certain *restrictions* on the $\phi_1 \ldots \phi_p$.
Using the polynomials in the lag operators we may rewrite (3.4) as

$$\phi_p(L) y_t = \theta_q(L) \epsilon_t \tag{3.5}$$

where we have expressed:

$$\phi_p(L) = 1 - \phi_1 L - \ldots - \phi_p L^p \quad \text{and} \quad \theta_q(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$$

For the moment, we assume that there is no **drift** term. That is, the unconditional mean of y_t is zero, $E[\, y_t \,] = 0$.

3.2 Autoregressive Processes

Set all the $\theta_1 = \theta_2 = \ldots = \theta_q = 0$.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t \qquad t = 1 \ldots T \tag{3.6}$$

Without suitable restrictions on the $\phi' s$, we can not guarantee stationarity.

3.2.1 The $AR(1)$ Process

$$y_t = \phi y_{t-1} + \epsilon_t \qquad t = 1, \ldots T \tag{3.7}$$

we may rewrite this as:

$$\phi(L) y_t = \epsilon_t, \quad \text{where} \quad \phi(L) = 1 - \phi L \tag{3.8}$$

or in the moving average representation as:

$$y_t = \phi(L)^{-1} \epsilon_t = \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \ldots \tag{3.9}$$

This is an infinite moving average that we saw before:

$$y_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \tag{3.10}$$

Under the iid assumptions of ϵ_t, the variance of y_t

$$\gamma(0) = V[y_t] = E\left[\left(\sum \phi^j \epsilon_{t-j}\right)^2\right] = \sum \phi^{2j} E[\epsilon_{t-j}^2] = \sigma^2 \sum \phi^{2j} \tag{3.11}$$

- For there to be a finite variance $\Rightarrow |\phi| < 1$. With $|\phi| < 1$ we can write

$$\text{the } V[y_t] = \frac{\sigma^2}{(1-\phi^2)} \quad \text{since} \quad \sum_{j=0} \phi^{2j} = 1 + \phi^2 + \phi^4 + \ldots = \frac{1}{1-\phi^2}$$

- So far we have established for the $AR(1)$ model that it has a mean of zero and under the stationary condition $|\phi| < 1$ there is a finite variance given by

$$\gamma(0) = \sigma^2/(1-\phi^2) \tag{3.12}$$

- Equation (3.10) shows quite clearly the nature of the dependency on past ϵ_{t-h}. As j increases, the weight attached to it falls (ϕ^j) since $|\phi| < 1$.

- The autocovariance function is:

$$\gamma(h) = E[y_t y_{t-h}] = E\left[\left(\phi^h y_{t-h} + \sum_0^{h-1} \phi^j \epsilon_{t-j}\right) y_{t-h}\right] \tag{3.13}$$

$$\text{as we write } y_t = \phi^h y_{t-h} + \sum_0^{h-1} \phi^j \epsilon_{t-j}.$$

- Notice that the $\epsilon_t \ldots \epsilon_{t-h+1}$ are all *uncorrelated* with y_{t-h} , and hence we may express 3.13:

$$\gamma(h) = \phi^h E[y_{t-h}^2] = \phi^h \gamma(0) = \phi^h \frac{\sigma^2}{(1-\phi^2)} \tag{3.14}$$

  The second equality follows from stationarity.

- Notice that the autocovariances depend only on h (the distance between y_t and y_{t-h}) confirming the weak stationarity condition.

- A more convenient way of obtaining the autocovariance, together with deriving the autocorrelation function is to *assume stationarity* ($|\phi| < 1$) and then multiply $y_t = \phi y_{t-1} + \epsilon_t$ by y_{t-h} and take expectations to give:

$$E[y_t y_{t-h}] = \phi E[y_{t-1} y_{t-h}] + E[\epsilon_t y_{t-h}] \qquad h = 0, 1, 2, \ldots \tag{3.15}$$

- We know that weak stationarity $\Rightarrow E[y_{t-1}y_{t-h}] = E[y_t y_{t-h+1}] = \gamma(h-1)$ and the last term in 3.15 is zero so that :

$$\gamma(h) = \phi\gamma(h-1), \tag{3.16}$$

which also shows the autocorrelation function (see 3.14):

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h.$$

- For positive ϕ the autocorrelation function shows a smooth exponential decay and for negative ϕ the decay is also exponential but oscillates between negative and positive values (harmonic decay).
- A stationary process is always expected to return to its uncondtional mean!

3.2.2 The $AR(2)$ Process

The second-order autoregressive process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \tag{3.17}$$

- It is best to consider the *deterministic* second order homogeneous difference equation ($\phi_2 \neq 0$ and $\epsilon_t = 0$); the stationarity of the system does not depend on ϵ_t given our earlier iid assumptions):

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = 0. \tag{3.18}$$

The solution of this equation depends on the characteristic equation:

$$m^2 - \phi_1 m - \phi_2 = 0 \tag{3.19}$$

- Since (3.19) is a quadratic, we can find the roots, say m_1 and m_2:

$$(m - m_1)(m - m_2) = 0. \tag{3.20}$$

Using the standard quadratic formula:

$$m_1, m_2 = \frac{\left[\phi_1 \pm (\phi_1^2 + 4\phi_2)^{\frac{1}{2}}\right]}{2}.$$

We can identify three cases:

1. Both roots are real and distinct (Fuller p.44)

$$\Rightarrow y_t = b_1 m_1^t + b_2 m_2^t \tag{3.21}$$

where $b1$ and $b2$ are determined by the initial conditions

2. Real and equal roots ($m_1 = m_2 = m$):

$$y_t = (b_1 + b_2 t)m^t \tag{3.22}$$

3. Complex Conjugate Pair:

$$y_t = b_1^* m_1^t + b_1 m_2^t \tag{3.23}$$

where b_1^* is the complex conjugate of b_1 and $m_2 = m_1^*$.

Complex Numbers

- Let c and d be real numbers

$$\begin{aligned} b_1 &= c + di \\ b_1^* &= c - di \end{aligned}$$

where $i = \sqrt{-1}$

- Modulus is

$$\sqrt{c^2 + d^2}$$

- To see that these three are solutions simply substitute in (3.18).

Stationarity of AR(2)

- Regardless of the nature of the roots **:** if the roots are less than one in equations (3.21) - (3.23) (in the complex case the modulus is less than one)
- $\Rightarrow$ a shock to ε_t in (3.17) produces a time path for y_t that will dampen out (head towards zero).
- Roots of the characteristic equation must be less than one or equivalently (and this is the way you more often hear it expressed).
- The roots in the polynomial in the lag operator:

$$\phi_2(L) = 1 - \phi_1 L - \phi_2 L^2 = 0 \tag{3.24}$$

must lie outside the unit circle (notice this is not the same expression as the **characteristic equation**)

- In terms of the coefficients of the $AR(2)$ process, for (3.24) to have roots outside the unit circle then the following must be satisfied (see diagram in Box-Jenkins):

$$\begin{aligned} &(i) \quad \phi_1 + \phi_2 < 1 \\ &(ii) \quad \phi_2 - \phi_1 < 1 \\ &(iii) \quad -1 < \phi_2 < 1 \end{aligned} \tag{3.25}$$

Autocovariance Functions for an AR(2) Process:

- The trick in calculating autocovariances for any stationary $ARMA$ process is to determine the infinite moving average representation.
- Recall that an $AR(1)$ process was expressed as an **infinite** moving average process.
- The $AR(2)$ process will also have an infinite moving average representation, although more complex:

$$y_t = \phi_2^{-1}(L)\epsilon_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \tag{3.26}$$

- We can work-out what the ψ_i are by solving:

$$y_t = \psi(L)\epsilon_t \tag{3.27}$$

  and evaluating $\phi_2(L)\psi(L) = 1$.

- That is:

$$(1 - \phi_1 L - \phi_2 L^2)(\psi_0 + \psi_1 L + \psi_2 L^2 + \ldots) = 1 \tag{3.28}$$

- Gathering like terms gives:

$$\psi_0 + (\psi_1 - \phi_1)L + (\psi_2 - \phi_1\psi_1 - \phi_2\psi_0)L^2 + (\psi_3 - \phi_1\psi_2 - \phi_2\psi_1)L^3 + \ldots = 1. \tag{3.29}$$

- Notice that the coefficients in L on the RHS of (3.29) are all zero so that:

$$\psi_1 - \phi_1 = 0$$

$$\psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2} = 0 \qquad j \geq 2. \tag{3.30}$$

  setting $\psi_0 = 1$ and $\psi_1 = \phi_1$.

- Note that

$$\psi_j < \psi_{j-1} \quad \forall\ j$$

- Notice that the difference equation (3.30) has the same form as the homogeneous difference equation (3.18).
- It follows from (3.27) that:

$$\gamma(h) = \sigma^2 \sum \psi_j \psi_{j-h} \qquad h = 0, 1 \ldots \tag{3.31}$$

Yule-Walker Equations

As we did in equation (3.16), we can express the autocovariances as linear equations in the parameters of the autoregressive model. For the $AR(2)$ model (we apply the same technique of multiplying by y_{t-h} and taking expectations under the assumption of stationarity):

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2) \qquad h = 1, 2, \ldots \tag{3.32}$$

- For $h = 0$ we have (note that $E[\epsilon_t y_t] = \sigma^2$)

$$\gamma(0) = \phi_1\gamma(-1) + \phi_2\gamma(-2) + \sigma^2 = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2$$

  as $\gamma(h) = \gamma(-h)$, an even function. Dividing by $\gamma(0)$ and rearranging delivers :

$$\gamma(0) = \frac{\sigma^2}{\{1 - \phi_1\rho(1) - \phi_2\rho(2)\}} \tag{3.33}$$

  Further :

$$\rho(1) = \frac{\phi_1}{(1-\phi_2)} \text{ and } \rho(2) = \phi_1\rho(1) + \phi_2$$

  and

$$\gamma(0) = \left(\frac{1-\phi_2}{1+\phi_2}\right)\frac{\sigma^2}{[(1-\phi_2)^2 - \phi_1^2]}.$$

  This relationship obviously will extend to higher autoregressive forms.

- Equation (3.32) is called the **Yule-Walker equations**.

- Another (and perhaps more common) way to express the Yule-Walker equations is in terms of the autocorrelation coefficients (dividing (3.32) by $\gamma(0)$):

$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2) \qquad h = 1, 2 \ldots \tag{3.34}$$

We could "solve" these equations with $h = 1$ and 2 for the $\phi's$:

$$\begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} \tag{3.35}$$

Equation (3.35) may, in fact, serve as a method of estimating the ϕ's. Again this may be extended to higher orders (see Harvey p.123 or Box and Jenkins pp. 55)

3.2.3 The $AR(p)$ process

We shall briefly consider the general autoregressive p process. Once $p > 2$, the analytical calculations become tedious and the intuition behind the time path of y_t in terms of the autoregressive parameters becomes somewhat lost:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t \tag{3.36}$$

The strictly autoregressive models (with no MA components) have always been popular in the applied time series literature since they are so simple to estimate (under the assumptions that we have employed this is simply OLS)

For stationarity, we require the roots of the characteristic equation:

$$m^p - \phi_1 m^{p-1} - \ldots - \phi_p = 0 \tag{3.37}$$

to be less than one in absolute value, or equivalently, the roots in the polynomial in the lag operator:

$$1 - \phi_1 L - \ldots - \phi_p L^p = 0 \tag{3.38}$$

lie outside the unit circle. (Note the difference between (3.37) and (3.38) is that we set $m = L^{-1}$ and then the whole equation is multiplied by L^p). The $AR(p)$ process yields an expression for the autocorrelation coefficients:

$$\rho(h) = \phi_1 \rho(h-1) + \ldots + \phi_p \rho(h-p) \qquad h = 1, 2, \ldots$$

- In Stata, we can use the vector autoregression modeling commands to check stability of the *estimated* coefficients of any AR model (see root.do program)

3.3 Moving Average Error Processes

3.3.1 $MA(q)$ Processes

We commence the discussion with the most general $MA(q)$ process:

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \tag{3.39}$$

A finite moving average process (given our assumptions on ϵ_t) is always stationary $\Rightarrow$ without placing any restrictions on the parameters θ
We may wish to express the MA process in the AR form:

$$y_t = \theta_q(L)\epsilon_t \quad \text{where} \quad \theta_q(L) = 1 + \theta_1 L + \ldots + \theta_q L^q \tag{3.40}$$

which we re-express as:

$$\theta_q^{-1}(L) y_t = \epsilon_t \tag{3.41}$$

- To invert this polynomial in the lag operator we will have to impose certain conditions similar to the stationary conditions for the AR process. When these conditions are satisfied we say the MA process is **invertible**.

- Consider two seemingly different $MA(1)$ processes for y_t :

$$y_t = \epsilon_t + \theta \epsilon_{t-1} \quad \theta < 1, \quad \epsilon_t \sim iid(0, \sigma^2) \tag{3.42}$$

and

$$y_t = \tilde{\epsilon}_t + \frac{1}{\theta}\tilde{\epsilon}_{t-1} \quad \tilde{\epsilon}_t \sim iid(0, \theta^2\sigma^2) \tag{3.43}$$

Both representations have the same autocovariance:

$$\gamma(0) = (1+\theta^2)\sigma^2 \quad \text{and } \gamma(1) = \theta\sigma^2 \tag{3.44}$$

- The process (3.42) is the invertible representation and (3.43) is non-invertible.
- **Every invertible MA process has a non-invertible representation that gives the same autocovariance function**
- In terms of the invertible MA (3.42) we can write the error ϵ_t in terms of **lags**

$$\epsilon_t = y_t - \theta y_{t-1} + \theta^2 y_{t-2} + \ldots \tag{3.45}$$

- However, (3.43) has an error term $\tilde{\epsilon}_t$ that is in terms of **leads**

$$\begin{aligned}
\tilde{\epsilon}_t &= \theta y_{t+1} - \theta^2 y_{t+2} + \cdots \quad \text{and solve "forward"} \qquad (3.46)\\
(1+\tilde{\theta}L)^{-1} &= \frac{-\tilde{\theta}^{-1}L^{-1}}{1-\tilde{\theta}^{-1}L^{-1}} = -\tilde{\theta}L^{-1}\left[1 + \tilde{\theta}^{-1}L^{-1} + \tilde{\theta}^{-2}L^{-2} + \ldots\right]\\
\text{with } \tilde{\theta} &= \frac{1}{\theta} > 1 \quad \text{since this gives}\\
(1+\tilde{\theta}L)^{-1} \times (1+\tilde{\theta}L) &= 1 \quad \text{as the identity operator}
\end{aligned}$$

- Clearly it is more sensible to define an error in terms of its current and past history rather than its current and all future values.
- Many estimation algorithms also have a problem with noninvertible processes especially STATA.

Conditions for Invertibility

For the general $MA(q)$ process (3.40) the conditions for invertibility are the roots from the characteristic equation:

$$m^q + \theta_1 m^{q-1} + \ldots + \theta_q = 0 \tag{3.47}$$

must be less than one in absolute value.
Another example of a stationary but not invertible process consider:

$$y_t = \epsilon_t + \epsilon_{t-1} \tag{3.48}$$

3.3.2 $MA(1)$

$$y_t = \epsilon_t + \theta\epsilon_{t-1} \tag{3.49}$$

We can show that:

$$\rho(1) = \frac{\theta}{1+\theta^2} \tag{3.50}$$

- Notice that the maximum correlation coefficient is ± 0.5 corresponding to $\theta = \pm 1$.
- A common argument against $MA(1)$ representations is that their maximum correlation is 0.5 and in this sense may be less useful in modeling dependencies than the AR processes.

1. The $MA(1)$ process has a memory of 1 period:

$$\gamma(1) = \theta\sigma^2$$

$$\gamma(h) = 0 \quad \text{for} \quad h > 1$$

2. In general an $MA(q)$ process has a memory of q periods

$$\gamma(h) = 0 \text{ if } h > q$$

3. MA processes are harder to estimate than the AR ones.

3.4 Mixed Process $ARMA$

Box-Jenkins time series analysis is most typically identified with the mixed types of models based upon the belief that a low-order polynomial in p and q in (3.3) can describe most series well:
For example the $ARMA(1,1)$

$$y_t = \phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1} \tag{3.51}$$

Notes

1. Whether a process is stationary or not depends solely on the autoregressive component. In general, we will require the roots of the polynomial in the lag operator $\{\phi(L)\}$ to be outside the unit circle.

2. Invertibility will center on the roots of the MA process $\theta(L)$.

3. We may rewrite (3.51) as

$$(1 - \phi L)y_t = (1 + \theta L)\epsilon_t \tag{3.52}$$

which leads to:

$$y_t = \frac{\epsilon_t}{1 - \phi L} + \frac{\theta\epsilon_{t-1}}{1 - \phi L} \tag{3.53}$$

If $\mid \phi \mid \leq 1$ we may write this as a infinite geometric series:

$$y_t = \epsilon_t + \sum_{j=1}^{\infty}(\theta\phi^{j-1} + \phi^j)\epsilon_{t-j} \tag{3.54}$$

giving the infinite moving average representation. It is quite clear from (3.54) that for finite variances and covariances requires the restriction $\mid \phi \mid < 1$.

4. The general ARMA(p, q) process is given in (3.4) and is repeated here:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \ldots + \theta_q\epsilon_{t-q}$$

We have seen how the MA representation may be obtained for the $AR(p)$ process. Essentially the same trick is employed for the mixed process: Solve for the $\psi(L)$(relate coefficients in the same powers of L) in the expression:

$$\phi(L)\psi(L) = \theta(L) \tag{3.55}$$

5. The mixed processes are much more difficult to identify from just observing y_t (in the Box-Jenkins sense) since we do not obtain as sharp and as informative results. For example, let us take the $ARMA(1, 1)$ in (3.51): multiplying by y_{t-h} and taking expectations yields the Yule-Walker equations:

$$\gamma(h) = \phi\gamma(h - 1) + E\left[\,\epsilon_t y_{t-h}\,\right] + \theta E\left[\,\epsilon_{t-1}y_{t-h}\,\right] \tag{3.56}$$

Since the memory of the moving average part is one for $h > 1$ these terms are zero

$$\text{if } h = 1 \Rightarrow E\left[\,\epsilon_{t-1}y_{t-1}\,\right] = E\left[\,\epsilon_{t-1}(\phi y_{t-2} + \epsilon_{t-1} + \theta\epsilon_{t-2})\,\right] = \sigma^2$$

with the first term being zero: if $h = 0$ and

$$\begin{aligned} E\left[\varepsilon_t y_t\right] &= \sigma^2 \\ E\left[\varepsilon_{t-1}y_t\right] &= \phi\sigma^2 + \theta\sigma^2 \end{aligned}$$

6. The autocovariance function is:

$$\begin{aligned} \gamma(0) &= \phi\ \gamma(1) + \sigma^2 + \theta\phi\sigma^2 + \theta^2\sigma^2 \\ \gamma(1) &= \phi\ \gamma(0) + \theta\sigma^2 \\ \gamma(h) &= \phi\ \gamma(h - 1) \qquad h = 2, 3, \ldots \end{aligned}$$

Solving for $\gamma(0)$ in the above:

$$\gamma(0) = \frac{1+\theta^2+2\phi\theta}{1-\phi^2}\sigma^2 \tag{3.57}$$

$$\gamma(1) = \frac{(1+\phi\theta)(\phi+\theta)}{1-\phi^2}\sigma^2 \tag{3.58}$$

$$\tag{3.59}$$

7. This leads to the autocorrelation coefficients:

$$\rho(1) = \frac{(1+\phi\theta)(\phi+\theta)}{1+\theta^2+2\phi\theta} \tag{3.60}$$

$$\rho(h) = \phi\ \rho(h-1) \qquad h = 2, 3, \ldots \tag{3.61}$$

Note for $h > 1$, the behaviour of the autocorrelation function is governed solely by the first order AR process. This is the same as the $AR(1)$ process studied earlier. This is **not** equivalent to saying that the MA part is making no contribution since $\rho(1)$ involves the MA parameters which in turn feeds into the difference equation (3.61).

8. The properties of higher order ARMA (p, q) processes may be derived and studied analogously. The general pattern is that the first q autocorrelations depend explicitly upon the moving average and the autocorrelation parameters. Higher order autocorrelations are given by p^{th} difference equation of the form

$$\rho(h) = \phi_1\ \rho(h-1) + \ldots + \phi_p\ \rho(h-p) \qquad h = q, q+1, \ldots \tag{3.62}$$

3.4.1 Parameter Redundancy and Common Factors

If the AR and MA polynomials in (3.4) have a root which is the same then they are said to have a **common factor**. The model is **over-parametrized**, since a model with identical properties can be constructed by reducing p and q by one. Common factors such as this may make models not identifiable and may lead to computational problems.

- Box-Jenkins provide a long discussion on the concept of **parsimony** and why it is desirable to model the phenomenon under study with as few as parameters as possible.

- Example of Overparameterization

(Harvey) Suppose we have the following ARMA (2, 1) process:

$$y_t = 0.2y_{t-1} + 0.15y_{t-2} + \epsilon_t + 0.3\epsilon_{t-1}$$

Note that the AR polynomial can be factored into two parts:

$$(1 - 0.2L - .15L^2) = (1 - 0.5L)(1 + 0.3L)$$

$$y_t = \phi^{-1}(L)\theta(L)\epsilon_t = \frac{(1 + 0.3L)}{(1 - 0.5L)(1 + 0.3L)}\epsilon_t$$

which may be rewritten as the simple $AR(1)$ process:

•

$$y_t = 0.5y_{t-1} + \epsilon_t.$$

- Hence it will be important to check for redundancies or common factor restrictions to simplify our models.
- Serial correlation in common regression equations can be thought of as imposing common factor restrictions
- As we shall see, STATA estimates the full likelihood using a Kalman filter for updating which can often result in some pretty strange looking estimates.
- The problem is due to common factors, typically with negative roots on the moving average terms:

$$\begin{aligned} y_t &= \epsilon_t \text{ Truth} \\ (1 - \phi L)y_t &= (1 + \theta L)\epsilon_t \text{ Fit} \end{aligned}$$

with the fitted $ARMA(1, 1)$ STATA will find $\phi \approx 1$ and $\theta \approx -1$

3.4.2 Aggregation

In macro economics aggregation is often a culprit for a lot of ills. In times series analysis aggregation can lead to some strange time dependencies in the aggregate data.

- Consider two series y_t ARMA(p_1, q_1) and x_t ARMA(p_2, q_2), sum aggregation leads to $z_t = y_t + x_t$ with a ARMA(p_3, q_3) with $p_3 \leq p_1 + p_2$ and

$$q_3 \leq \max\{p_1 + q_2, p_2 + q_1\} \tag{3.63}$$

- Some other general results on aggregation:

$$\begin{aligned} AR(p) + \text{white noise} &= ARMA(p, p) \\ AR(p_1) + AR(p_2) &= ARMA(p_1 + p_2, \max[p_1, p_2]) \\ MA(q_1) + MA(q_2) &= \text{ARMA}(0, \max(q_1, q_2)) \\ ARMA(p, q) + \text{ white noise} &= ARMA(p, \max[p, q]) \\ AR(p) + MA(q) &= ARMA(p, p + q) \end{aligned}$$

3.4.3 Multivariate Processes

There is nothing stopping us from considering vector processes such as vector ARMA models or vector autoregressive models (VAR). The latter processes have found a special place in macro econometrics and causation studies. We will discuss these again (Chapter 7) once we have firmly established the ground rules for univariate representations.

3.5 Identification: Autocorrelation (AC) and Partial Autocorrelation Functions (PAC)

It is often easy to lose sight of the fact that when writing these kinds of models out that we only observe the y_t, $t = 1, \ldots T$. We need practical schemes to identify a candidate model for the order and kind of model that in some way provides a useful starting point. Fitting all factorial combinations is neither feasible nor desirable. The most popular tools for doing this are the autocorrelation and partial autocorrelation function. These two calculations not only assist in identifying a canddiate model(s) but can serve as a check at the end of the modelling to make sure that you have not over-fitted the model(i.e. built too large a time series model)

3.5.1 Autocorrelation Functions (ACF) or Correlogram

The autocorrelation function is simply the autocorrelation graphed against the lags. Stationary series have a decaying autocorrelation structure We know that if a process is pure MA (there is no AR component) that the autocorrelation function is zero for all lags that exceed the order of the moving average ($h > q$). On the other hand, for mixed or AR processes, the autocorrelation function is infinite (albeit dying out for stationary process). Typically in practice we will not know the order (be it MA, AR or ARMA) and will have to **identify** it.

- The standard error for the correlogram (ACF) is usually approximated by

$$V[\hat{\rho}(k)]^{\frac{1}{2}} = 1/\sqrt{T}$$

  for testing whether $\rho(k)$ is equal to zero at any lag k.

- Stata uses the Bartlett formula for calculating the variance of $\hat{\rho}(k)$

$$V[\hat{\rho}(k)] = \begin{cases} \frac{1}{T} \text{ if } k = 1 \\ \frac{1}{T}\left(1 + 2\sum_{i=1}^{k-1} \hat{\rho}(i)\right) \text{ if } k > 1 \end{cases}$$

- This is used to determine how fast the correlogram is dying in k (for some $k = 1, \ldots,$)

$$\hat{\rho}(k) \times \sqrt{V[\hat{\rho}(k)]} \sim N(0,1) \text{ asymptotically under } H_0$$

or in graphs confidence intervals can be drawn for some pre-specified confidence level.

- Choice of k is often some function of the total data size T or some multiple of the data frequency (say quarterly data 12 or 16)

3.5.2 Partial Autocorrelation Functions (PAC)

The **partial autocorrelation function** is a function that is **zero for lags that are greater than the order of the AR part**. That is, there will be, say p non-zero parts to the partial autocorrelation coefficients.

Recall the Yule-Walker Representation for the $AR(2)$

$$\rho(h) = \phi_1\ \rho(h-1) + \phi_2\ \rho(h-2) \qquad h = 1, 2 \ldots \tag{3.64}$$

which extends to $AR(p)$.

Denote ϕ_{kj} as the j^{th} coefficient in an autoregressive process of order k, so that ϕ_{kk} is the last coefficient:

$$\rho(j) = \phi_{k1}\ \rho(j-1) + \ldots + \phi_{k(k-1)}\ \rho(j-k+1) + \phi_{kk}\ \rho(j-k) \qquad j = 1, 2, \ldots, k \tag{3.65}$$

and the graph of these last coefficients ϕ_{kk} are called the partial autocorrelation
This leads to another representation of the Yule-Walker equations:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \rho_{k-1} & \rho_{k-2} & \cdots & & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \rho_k \end{bmatrix} \tag{3.66}$$

or in matrix form:

$$\mathbf{P_k \phi_k = \rho_k} \tag{3.67}$$

We may solve in terms of the ϕ_{kk} as before for $k = 1, 2, \ldots$. Note that we are considering different AR models since we vary k.
(see Box-Jenkins $p64$):

$$\begin{array}{lcl} \phi_{11} & = & \rho_1 \\ \phi_{22} & = & \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} & = & \text{see Box and Jenkins } p64 \end{array} \tag{3.68}$$

Notes On Partial Autocorrelation

1. The quantity ϕ_{kk}, regarded as a function of the lag k, is called the **partial autocorrelation function**.

2. Each of the coefficients $\phi_{11}, \phi_{22}, \ldots, \phi_{kk}$ are functions only of the $1, 2, \ldots, k$ correlations respectively. Suppose that the true process is p :

$$\begin{aligned} \phi_{kk} &\neq 0 \ \forall\ k \leq p \\ \phi_{kk} &= 0 \ \forall\ \mathbf{k > p} \end{aligned} \tag{3.69}$$

   To see this, note that for an AR(1) process from (3.68) and recall

$$\rho_2 = \rho_1^2 \Rightarrow \phi_{22} = 0. \tag{3.70}$$

3. Time series programs typically let the user select what values (there are of course defaults) but choice is often dictated by data frequency (12 or 24)

4. **Estimation of the Partial Autocorrelation Function**

   (a) Estimate by ordinary least squares an autoregressive process of orders $1, 2, 3, \ldots$, successively taking the coefficients $\hat{\phi}_{11}, \hat{\phi}_{22}, \hat{\phi}_{33}, \ldots$or

   (b) Substitute the estimated (sample) autocorrelation coefficients $\hat{\rho}(j)$ into (3.67) and 'solve' for the $\hat{\phi}_{kk}$

5. **Testing the order of the process: Partial autocorrelation function**

$$\begin{aligned} H_o &: \text{process is of order } p \\ H_1 &: \text{process is of order } (p+1) \text{ or higher} \end{aligned}$$

   It can be shown under this null the $\hat{\phi}_{kk}$ are approximately independently distributed for $k \geq p + 1$.

6. Test null by testing the significance of the $\hat{\phi}_{kk}$ for $k \geq p + 1$

$$Z = \frac{\hat{\phi}_{kk}}{\left\{ Var\left[\hat{\phi}_{kk}\right] \right\}^{\frac{1}{2}}} \sim N(0,1) \text{ asymptotically under } H_0 \tag{3.71}$$

   The variance on partial autocorrelations functions (under H_0):

$$Var\left[\hat{\phi}_{kk}\right] \cong \frac{1}{T} \qquad\qquad k \geq p+1$$

   where T is the number of observations used in the fitting . Hence the test statistic is

$$\hat{\phi}_{kk} \times \sqrt{T} \sim N(0,1) \text{ asymptotically under } H_0 \tag{3.72}$$

7. The autocorrelation function and the estimated partial autocorrelation function are two ways of getting at the same problem – namely determining the order of the process. This often serves as the initial phase to finding an appropriate time series model.

8. We will look at how to estimate consistently $ARIMA$ models in Chapter 4.

3.5.3 A Portmanteau Test for Serial Correlation: Box-Pearce Q Statistic

- Usuallly applied after some $ARMA$ model is fitted and the residulas want to be checked for autocorrelation
- The basic idea is to examine the correlations in the data and determine whether these are significantly different from zero.
- $$\begin{aligned} H_0 &: \text{errors are random} \\ H_1 &: \text{errors are not} \end{aligned}$$

  Notice that the alternative is extremely general and that there are a number of alternatives that could give rise to a rejection of the null hypothesis (for instance, structural breaks or time-dependent variances: ARCH).
- Hence these kinds of general tests are called **portmanteau-tests.**
- A common test (called the Box-Pearce Q statistic) is:

$$Q = T\sum_{h=1}^{M} r^2(h) \sim \chi^2_M \text{ (under } H_0) \tag{3.73}$$

  where M is the number of autocorrelations (based on the data or residuals once a model has been estimated) calculated.
- This is selected by the researcher (although common choices are $M = 12, 24, 36$).
- The issue in the choice of M is power.
- If too high a choice is made and the error is really, say and $MA(1)$, then we will introduce a number of correlations that are zero and lose power.
- On the other hand, too small an M might overlook some correlations at higher h, particularly if there are seasonal effects.
- Note that for this test, we have taken advantage of the fact that the $r(h)$ under the null is standardized normal and independent.
- As indicated in (3.73), the test is asymptotic and unfortunately the small sample properties are poor.
- A **modified** Box-Pearce statistic (due to Ljung and Box) is:

$$Q^* = T(T+2)\sum_{h=1}^{M}(T-h)^{-1}r^2(h) \sim \chi^2_M \text{ (under } H_0) \tag{3.74}$$

  where the idea is $(T-h)/T(T+2)$ gives a better approximation to the variance than does $1/T$.

- In fact, these tests be motivated as an Lagrange Multiplier (LM) test of an $AR(M)$ or $MA(M)$ against white noise.
- The alternative hypothesis is quite general and does not distinguish among AR, MA, or $ARMA$ alternatives (see Harvey).
- STATA produces the Q statistics in corrgram command (as well as ACF, PAC) and the Box Pierce Statistic from the command *wntestq*
- If the ARMA(p, q) model is first estimated and the residuals are tested for randomness. In this case, the rule is to choose the degrees of freedom to be $M - p - q$ for the χ^2.
- Both statistics (3.73) and (3.74) tend to favour larger models.

3.6 Seasonality and a General Formulation of a Time Series Model

3.6.1 Seasonality

Many economic time series data (weekly, monthly and quarterly data) have pronounced seasonal patterns. Seasonal effects occur, and therefore must be modeled. The modeling strategy is divided between **permanent** and **transitory** effects. The permanent effects (regular and happen in the same time year to year) are either **deterministic** (often captured by the addition of seasonal dummies or other kinds of variables) or **stochastic** (introduce stochastic components like a local trend). The transitory effects are modeled as **stationary time series processes**. In this set up, observations of the same seasons are correlated from year to year (time dependent) but the correlations are decreasing as the years get further apart (i.e. are stationary).

3.6.2 Seasonal Autoregressive-Moving Average Models

As an example, suppose we have quarterly data and that we have reason to believe that there is a seasonal effect at the quarters:

$$y_t = \phi_4 y_{t-4} + \epsilon_t \qquad | \phi_4 | < 1 \tag{3.75}$$

We see that this is a special case of the general $AR(4)$ model introduced earlier (since $\phi_1 = \phi_2 = \phi_3 = 0$). It should not be a surprise to find that (3.75) has similar properties to an $AR(1)$ model and that the techniques of analysis are similar.

The correlation coefficient :

$$\rho(h) = \begin{cases} \phi_4^{h/4} & h = 0, 4, 8, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{3.76}$$

Further, the closer ϕ_4 is to unity, the stronger is the seasonal pattern (the pattern is nondeterministic so long as we confine the process to be stationary).

The seasonal effects need not be restricted to the AR portion of the model and we may specify a **seasonal autoregressive moving average model:**

$$\phi^*(L^s)y_t \;=\; \theta^*(L^s)\epsilon_t \tag{3.77}$$

where $\phi^*(L^s) \;=\; 1-\phi_1^*-\ldots-\phi_P^*L^{P_s}$and $\theta^*(L^s) \;=\; 1-\theta_1^*-\ldots-\theta_Q^*L^{Q_s}$ and the value s will depend upon the frequency of the data (weekly, monthly etc.). This model (3.77) is a **pure** seasonal model since there are gaps in the autocorrelation function at non-seasonal lags. This particular model is highly unusual for economic time series. We may add the nonseasonal components in two ways (additive and the more common multiplicative both of which can be estimated in STATA).

3.6.3 Additive Seasonal Models

For the additive model, we simply augment the model by adding on the relevant seasonal components. This is the most natural way for applied people to think of seasonal ARIMA models. Thus in (3.75) we may have:

$$y_t \;=\; \phi_1 y_{t-1}+\phi_4 y_{t-4}+\epsilon_t \tag{3.78}$$

3.6.4 Multiplicative Seasonal Models (Box-Jenkins)

$$\phi^*(L^s)\phi(L)y_t \;=\; \theta^*(L^s)\theta(L)\epsilon_t \tag{3.79}$$

This way of handling seasonality fits well into the notation and manipulating the polynomials. Hence we may include an $AR(1)$ process multiplicatively on the seasonal model (3.75) to give:

$$(1-\phi_1 L)(1-\phi_4 L^4)y_t \;=\; \epsilon_t \tag{3.80}$$

so that $\phi^*(L^s)\phi(L) \;=\; (1-\phi_1 L)(1-\phi_4 L^4)$. This leads to a model of the form:

$$y_t \;=\; \phi_1 y_{t-1}+\phi_4 y_{t-4}+\phi_5 y_{t-5}+\epsilon_t \tag{3.81}$$

and the restriction $\phi_5 \;=\; -\phi_1\phi_4$. STATA imposes these constraints automatically. Without any further information, there is no reason to prefer the additive model over the multiplicative or vice versa. However, in the nonstationary environment with integrated time series, the multiplicative form is a much more natural and direct form of seasonality. Indeed this leads up to what is conventionally referred to as the Box-Jenkins time series model.

There are other kinds of additive models (Harvey p. 173) that tend to produce rather nonlinear restricted parameter space in the MA and AR parts. Such models will be discussed in the Kalman filtering representation.

3.6.5 X-12 Seasonal Adjustment

- X-12-ARIMA is a seasonal adjustment program developed at the U.S. Census Bureau and used widely ($http://www.census.gov/srd/www/x12a/$)
- The program is based on the Bureau's earlier X-11 program and the X-11-ARIMA program developed at Statistics Canada.
- There is a interactive version in Stata 12 that you can search for $sax12$
- X-12-ARIMA is an **iterative** procedure. For a monthly series with a multiplicative decomposition, X-12-ARIMA, very generally, uses the steps below to estimate the trend and seasonal components. (
  1. X-12-ARIMA estimates a rough trend-cycle.
  2. It then estimates the detrended series by dividing the original series by the trend estimate.
  3. Using the detrended data, it estimates the seasonal component using moving average filters for each month.
  4. It estimates the irregular component by dividing the seasonal component into the detrended series. X-12-ARIMA uses this irregular component to detect the extreme values.
  5. It then estimates the preliminary seasonally adjusted series by dividing the seasonal component, corrected for extreme values, from the original series.
- X-12-ARIMA repeats this process many times, getting more and more refined estimates of the trend and the seasonal factors.
- Details of the procedure can be found in the book "Seasonal Adjustment with the X-11 Method" by Ladiray and Quenneville (2001).

3.6.6 Deterministic Seasonal Models

We can always add onto a standard regression model variables that are intended to capture the seasonal effects. Perhaps the most common additive structure are the seasonal dummies (assume a simple white noise process plus seasonal effects)

$$y_t = \sum_{j=1}^{s} \gamma_j D_{jt} + \epsilon_t \tag{3.82}$$

where $D_{jt} = 1$ if season j and 0 otherwise. In this setup we interpret γ_j as the seasonal effect. Less common, but also possible, is to divide the seasons over the frequency interval $[0, 2\pi]$ and use the trigonometric functions:

$$y_t = \sum_{j} (\alpha_j \cos \omega_j t + \beta_j \sin \omega_j t) + \epsilon_t \tag{3.83}$$

where the summation $j = 1$ to $T/2$ if s is even and $j = 1$ to $(T-1)/2$ if s is odd. In this case, we have s seasonal parameters to estimate and the seasonal decomposition is in terms of the frequency. We will analyze the frequency representation of time series data later (Chapter 8).

3.6.7 Seasonal Differencing

Just as in the modeling of time series models with differencing to achieve stationarity, we can consider a class of seasonal differencing.

$\Delta_s y_t = y_t - y_{t-s}$, so that we may model say the process as say:

$$\Delta_s y_t = \epsilon_t \tag{3.84}$$

Hence, differencing occurs at the seasonals rather than at the frequency of data measurement.

3.7 The General Multiplicative Seasonal ARIMA Model

The fact that we may require differencing to achieve stationarity (both at the frequency of data collection and at the seasonal levels) has led to a general representation of the process. Note that the differencing makes the variables stationary. All the manipulations and discussion earlier have all been in terms of the stationary variables. Now the variables are allowed to be non stationary and it is the differencing (it may take more than first differencing) which achieves the stationarity. The general model is written as:

$$\phi^*(L^s)\phi(L)\Delta^d\Delta_s^D y_t = \theta^*(L^s)\theta(L)\epsilon_t \tag{3.85}$$

where $\Delta^d = (1-L)^d$ (see Chapter 2) $\Delta_s^D = (1-L^s)^D$ and d and D are integers denoting the differencing in the non seasonals and seasonals. This is the **multiplicative seasonal ARIMA model** identified with Box and Jenkins. The model is indexed by $(p,d,q) \times (P,D,Q)$ with the latter representing the seasonal components. It is this structure that is commonly estimated in most canned time series packages. Of course, this model could be further generalized by letting the seasonal patterns be subsumed into the structure such as:

$$\phi(L)\Delta^d\Delta_s^D y_t = \theta(L)\epsilon_t \tag{3.86}$$

where the $\phi(L)$ and $\theta(L)$ contain the seasonal as well as the non seasonal components. With suitable restrictions, we may obtain (3.85) from (3.86). The multiplicative model (3.85) is capable of handling quite complex models of seasonal behaviour (see Harvey p. 180).

3.8 Overdifferencing

If a time series process is stationary, then first differencing will cause non-invertibility with the MA associated polynomial having roots on the unit circle depending on the amount of differencing. For example take a linear trend model:

$$y_t \;=\; \alpha + \beta t + \epsilon_t \qquad t \;=\; 1, \ldots, T$$

Now first difference

$$\Delta y_t \;=\; y_t - y_{t-1} \;=\; \beta + \epsilon_t - \epsilon_{t-1}$$

Now note that although the series is stationary, it is non-invertible. We say there is a singularity in the error structure. Differencing again will remove the drift (β) but the error structure will be $\epsilon_t - 2\epsilon_{t-1} - \epsilon_{t-2}$. We say that the series is overdifferenced in that the error structure which was stationary is now non invertible. The consequence of this is typically an estimation problem (the identification problem see, Chapter 4).

3.9 Identification with Non Stationary Time Series

- The basic idea is to apply the same methods of identification using the correlogram and partial autocorrelation function discussed earlier for stationary processes **once the degree of differencing is chosen**.
- Therefore, if the series in its raw form has a correlogram that does not appear to be dying out as the lag length increases or the first-order correlation coefficient is near the unit root, then differencing is indicated.
- Basically, we are searching for some level of differencing that will result in the transformed series having the properties of the stationary processes studied earlier.
- Of course, this kind of visual check for dying out of autocorrelation functions is not formal and does not control for any type-I or type-II error.

3.9.1 Trend Stationary Processes

- Often differencing is NOT appropriate and the application of a difference will induce negative autocorrelation (a negative unit root in the MA).
- Process with slow decaying $ACF's$ may not be evidence of a unit root and researcher should consider trend alternatives:

$$y_t = \alpha + \beta_1 t + \ldots \beta_m t^m + u_t \tag{3.87}$$

where u_t is weakly stationary.

- In these circumstances the error would be modelled as an ARMA process
- of course, the models can be nested (easily in STATA) and **ARMA(p, q)** plus Trend can be estimated
- $$y_t = \mu + +\beta t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \quad (3.88)$$
- Of course the trick is to determine m which could be done by significance testing on the $\beta' s$ or graphing the $ACF' s$ of $\hat{u}_t(m)$ for different choices of m until the $ACF' s$ look stationary.
- **After m is determined the investigator models $\hat{u}_t(m)$ as a $ARMA$ model**

3.9.2 Unit Roots Again

The Box-Jenkins approach did not have a formal test for considering whether a series was first (or even higher) difference stationary. Instead, the researcher by applying various diagnostic tests (discussed in Chapter 6) together with the autocorrelation and correlogram to determine whether the series is stationary or not. If it is not, then the series is first differenced and the identification procedure is reapplied. Again checks on the series are conducted to determine whether the first differenced series is stationary: if not then the series is differenced again. This procedure is to be repeated until the final series appeared to be stationary. Since the procedure rested on the judgements of the researcher, it is not possible to evaluate formally the size and power behind these techniques. Later, Chapter 10, we shall discuss formal tests for first-difference stationary processes.

Differencing will not always achieve stationarity. For example, certain trending models cannot be made stationary by any amount of differencing. (see Harvey p. 185). Over the last 20 years there a great deal of research effort has been directed at nonstationary time series. In certain data sets you may find that it is not just a simple matter of first differencing, using a simple polynomial trend, or whatever. We will return to this issue in Chapters 10 and 11.

As for seasonal differencing, the correlogram will indicate "high" correlation at seasonal lags (possibly near unity for the correlations) and these correlations at seasonal frequencies may not appear to die out. Hence seasonal differencing may be indicated. Again there are now formal tests to decide the order of seasonal differencing.

3.10 Stata commands to simulate $ARMA$ models

There are two commands available to do $ARMA$ simulations which are available on your course page.

1. *arma.ado* is a ado command that I wrote some time ago that simulates an arbitrary $ARMA(1,1)$ with a trend, constant and arbitrary variance for the innovation. While it is limited to order 1 processes, it is handy to convince yourself that you can sort out trend and nonstationary process in general. To see the syntax of this ado file type: *help arma*

2. *sim_arma* can simulate arbitrary $ARMA(p,q)$ models and with some manipulation you could also augment these as well.