Univariate Time Series Models

Ercan Karadas

New York University
Department of Economics

Spring, 2018

Table of Contents

Table of Contents

Introduction

- We wish to explain the dynamic behavior of a single economic variable.
- That is, we wish to model the (stochastic) process that generates a time series of data on Y_t, say.
- Thus we investigate how current values depend upon past values (directly or indirectly).
- Little economic theory.
- Powerful for producing forecasts and their uncertainty.
- In the next lecture, we extend this to multiple variables (and consider their dynamic interrelationships).

- In the univariate case a series is modeled only in terms of its own past values and some disturbance. The general expression is

$$Y_t = f\left(Y_{t-1}, Y_{t-2}, \ldots, \varepsilon_t, \varepsilon_{t-1}, \ldots\right)$$

- To make this expression operational one must specify three things:
  - the functional form $f(\cdot)$
  - the number of lags
  - a structure for the disturbance term.
- Then the specification imposes restrictions on the times series properties of the process that generates Y_t
- If, for example, one specified a linear function with one lag and a white noise disturbance, the result would be the first-order, autoregressive $AR(1)$, process,

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$$

- White noise series satisfy the following conditions

$$\begin{aligned} E(\varepsilon_t) &= 0 & &\text{for all } t \\ E(\varepsilon_t^2) &= \sigma^2 & &\text{for all } t \\ E(\varepsilon_t \varepsilon_s) &= 0 & &\text{for all } t \neq s \end{aligned}$$

- Adding the assumption of normality to these conditions gives

$$\varepsilon_t \text{ are iid } N(0, \sigma^2)$$

  which reads, 'The ε_t are independently and identically distributed normal variables with zero mean and variance σ^2'.

- We return to the $AR(1)$ relationship specified before

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$$

- This equation shows that Y_t is determined as a function of δ, θ, Y_0 and the current and previous disturbances:

$$\begin{aligned} Y_1 &= \delta + \theta Y_0 + \varepsilon_1 \\ Y_2 &= \delta + \theta\left(\delta + \theta Y_0 + \varepsilon_1\right) + \varepsilon_2 \\ &= \delta(1+\theta) + \theta^2 Y_0 + (\varepsilon_2 + \theta\varepsilon_1) \end{aligned}$$

- Proceeding in this fashion gives the general equation

$$Y_t = \delta\left(1 + \theta + \theta^2 + \cdots + \theta^t\right) + \theta^t Y_0 + (\varepsilon_t + \theta\varepsilon_{t-1} + \theta^2\varepsilon_{t-2} + \cdots)$$

- Now multiply across this equation successively by $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}$ and take expectations to obtain

$$\begin{aligned} E(Y_t\varepsilon_t) &= \sigma^2 \\ E(Y_t\varepsilon_{t-1}) &= \theta\sigma^2 \\ E(Y_t\varepsilon_{t-2}) &= \theta^2\sigma^2 \end{aligned}$$

Thus Y_t is correlated with the current and all previous disturbances but is uncorrelated with all future disturbances. Then it follows that Y_{t-1} is uncorrelated with the current disturbance ε_t and all future disturbances.

- Assuming that the process started a very long time ago, we rewrite
$$Y_t = \delta\left(1 + \theta + \theta^2 + \cdots\right) + (\varepsilon_t + \theta\varepsilon_{t-1} + \theta^2\varepsilon_{t-2} + \cdots)$$
- The stochastic properties of the Y series are determined by the stochastic properties of the ε series.
- Taking expectations of both sides of the above equation
$$E(Y_t) = \delta\left(1 + \theta + \theta^2 + \cdots\right)$$
- This expectation only exists if the infinite geometric series on the right-hand side has a limit. The necessary and sufficient condition is
$$|\theta| < 1$$
- The expectation is then
$$E(Y_t) = \frac{\delta}{1 - \theta} \equiv \mu$$
so the Y series has a constant unconditional mean μ at all poins.

- To determine the variance we can now write

$$Y_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1} + \theta^2\varepsilon_{t-2} + \cdots$$

- Squaring both sides and taking expectations:

$$\begin{aligned} V(Y_t) &= E\left[(Y_t - \mu)^2\right] \\ &= E\left[\varepsilon_t^2 + \theta^2\varepsilon_{t-1}^2 + \cdots + 2\theta\varepsilon_t\varepsilon_{t-1} + 2\theta\varepsilon_t\varepsilon_{t-2} + \cdots\right] \\ &= \frac{\sigma^2}{1-\theta^2} \equiv \sigma_y^2 \end{aligned}$$

  Thus the Y series has a constant unconditional variance, independent of time.

- We will often work with demeaned form of the stochastic process

$$y_t = Y_t - \mu$$

  Note that $E(y_t) = 0$ and $V(y_t) = V(Y_t)$.

- A new concept is that of **autocovariance**, which is the covariance of Y with a lagged value of itself.
- The kth order autocovariance is defined as

$$\gamma_k \equiv cov(Y_t, Y_{t-k})$$

- The we can compute the first-lag and second lag autocovariances as:

$$\begin{aligned}
\gamma_0 &= \sigma_y^2 \qquad \text{(by definition)} \\
\gamma_1 &\equiv cov(Y_t, Y_{t-1}) \\
&= E\left[(Y_t - \mu)(Y_{t-1} - \mu)\right] \\
&= \theta\sigma_y^2 \\
\gamma_2 &\equiv cov(Y_t, Y_{t-2}) \\
&= E\left[(Y_t - \mu)(Y_{t-2} - \mu)\right] \\
&= \theta^2\sigma_y^2
\end{aligned}$$

In general,

$$\gamma_k = \theta^k\sigma_y^2, \quad k = 0, 1, 2, \ldots$$

Note that the autocovariances thus depend only on the lag length and are independent t.

- It is common to standardize autocovariances by dividing through the variance.
- Dividing the covariances by the variance gives the set of **autocorrelation coefficients**, also known as **serial correlation coefficients**, which we will designate by
$$\rho_k \equiv \frac{\gamma_k}{\gamma_0}, \quad k = 0, 1, 2, \ldots$$
- We can immediately write

$$\begin{aligned} \rho_0 &= 1 \\ \rho_1 &= \theta \\ \rho_2 &= \theta^2 \\ \vdots &= \vdots \\ \rho_k &= \theta^k \end{aligned}$$

- Plotting the autocorrelation coefficients against the lag lengths gives the **correlogram** of the series.
- To summarize, when $|\theta| < 1$ the mean, variance, and covariances of the Y series are constants, independent of time. The Y series is then said to be **weakly** or **covariance stationary**.

Stationarity

- A stochastic process is **weakly stationary** (or covariance stationary) if

$$
\begin{aligned}
E\{Y_t\} &= \mu < \infty \\
V\{Y_t\} &= E\{(Y_t - \mu)^2\} = \gamma_0 < \infty \\
\text{cov}\{Y_t, Y_{t-k}\} &= E\{(Y_t - \mu)(Y_{t-k} - \mu)\} = \gamma_k, \quad k = 1, 2, 3, \ldots
\end{aligned}
$$

- A stochastic process is **strictly stationary** if its properties are not affected by a shift along the time axis (do not depend upon t).

Lag operator

- We are going to study generalizations of $AR(1)$ process and some other stochastic processes but first we will introduce some technical tools to make the calculations easier
- It is convenient to use the lag operator, denoted L, defined as

$$
\begin{aligned}
L^0 y_t &= y_t \\
L y_t &= y_{t-1} \\
L^k y_t &= L^{k-1}(L y_t) = L^{k-1} y_{t-1}
\end{aligned}
$$

- For example using the lag operator we can express the first difference operator Δ as

$$
\begin{aligned}
(1-L)y_t &= y_t - y_{t-1} = \Delta y_t \\
L(1-L)y_t &= y_{t-1} - y_{t-2} = \Delta y_{t-1}
\end{aligned}
$$

- In many algebraic manipulations the lag operator may be treated as a scalar.
- One of the most important operations is taking the inverse of an expression in L.

Lag operator

- For example, let
$$\theta(L) = 1 - \theta L$$
denote a first-order polynomial in L. Now consider the multiplication
$$(1-\theta L)\left(1+\theta L+\theta^2 L^2+\cdots+\theta^p L^p\right) = 1-\theta^{p+1}L^{p+1}$$
- As $p \to \infty$, $\theta^p L^p \to 0$, provided $|\theta| < 1$. We may then write the reciprocal, or inverse of $\theta(L)$ as
$$\theta^{-1}(L) = \frac{1}{(1-\theta L)} = 1+\theta L+\theta^2 L^2+\theta^3 L^3+\cdots$$

Table of Contents

ARMA processes

- A p-th order autoregressive process or $AR(p)$ as

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + \varepsilon_t$$

- We define a q-th order moving average process or $MA(q)$ as

$$y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q}$$

- An autoregressive moving average or $ARMA(p, q)$ process as

$$y_t = \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q}$$

- Define the following **lag polynomials**

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_p L^p$$
$$\alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_q L^q$$

- These definitions allow us to write ARMA models compactly as

$$\begin{aligned} \theta(L) y_t &= \varepsilon_t & & AR(p) \\ y_t &= \alpha(L)\varepsilon_t & & MA(q) \\ \theta(L) y_t &= \alpha(L)\varepsilon_t & & ARMA(p, q) \end{aligned}$$

AR(1) Process

- We return to the $AR(1)$
$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$$
where ε is a white noise process.
- For AR(1) we will rederive the properties that we have seen above, but using the lag operator.
- Using lag operator we can write AR(1) as
$$(1 - \theta L) Y_t = \delta + \varepsilon_t$$
which gives
$$\begin{aligned} Y_t &= \frac{1}{(1-\theta L)}(\delta + \varepsilon_t) = \left(1 + \theta L + \theta^2 L^2 + \cdots\right)(\delta + \varepsilon_t) \\ &= \delta\left(1 + \theta + \theta^2 + \cdots\right) + \left(\varepsilon_t + \theta \varepsilon_{t-1} + \theta^2 \varepsilon_{t-2} + \cdots\right) \end{aligned}$$
- Provided $|\theta| < 1$ the expectation is then
$$E(Y_t) = \frac{\delta}{1-\theta} \equiv \mu$$
so the Y series has a constant unconditional mean μ at all poins.

- We have seen that the variance is

$$V(Y_t) = \frac{\sigma^2}{1-\theta^2} \equiv \sigma_y^2$$

- This variance can be derived in an alternative fashion that also facilitates the derivation of autocovariances. It is possible to rewrite AR(1) as

$$y_t = \theta y_{t-1} + \varepsilon_t$$

where $y_t = Y_t - \mu$.

- Squaring both sides and taking expectations:

$$E(y_t^2) = \theta^2 E(y_{t-1}^2) + E(\varepsilon_t^2) + 2\theta E(y_{t-1}\varepsilon_t)$$

The last term on the right-hand side vanishes, since y_{t-1} depends only on $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ and ε_t is uncorrelated with all previous values by the white noise assumption.

- When θ satisfies the stationarity condition $|\theta| < 1$,

$$\sigma_y^2 = E(y_t^2) = E(y_{t-1}^2) = \cdots$$

and the previous equation becomes

$$\sigma_y^2 = \theta^2 \sigma_y^2 + \sigma^2$$

solving for σ_y^2 gives the expression we obtain before.

AR(2) Process

- The AR(2) process is defined as
$$Y_t = \delta + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \varepsilon_t$$
where ε is a white noise process.
- By assuming stationarity, the unconditional mean is
$$\mu = \frac{\delta}{1 - \theta_1 - \theta_2}$$
- Then AR(2) in deviation form
$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t$$
- If we multiply both sides by y_t and take expectations,
$$\gamma_0 = \theta_1 \gamma_1 + \theta_2 \gamma_2 + \sigma^2$$
where used $E(y_t \varepsilon_t) = \sigma^2$, which can obtained from the previous equation.

AR(2) Process

- Similarly, multiplying by y_{t-1}, y_{t-2} and taking expectations,

$$\gamma_1 = \theta_1\gamma_0 + \theta_2\gamma_1$$
$$\gamma_2 = \theta_1\gamma_1 + \theta_2\gamma_0$$

- Now, substituting these into the previous expression for the variance

$$\gamma_0 = \frac{(1-\theta_2)\sigma^2}{(1+\theta_2)(1-\theta_1-\theta_2)(1+\theta_1-\theta_2)}$$

- Under stationarity this variance must be a constant, positive number. Requiring each term in parentheses to be positive gives

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 - \theta_1 < 1$$
$$|\theta_2| < 1$$

  These are the stationarity conditions for the AR(2) process.

- The relations between autocovariances may be restated in terms of autocorrelation coefficients, namely,

$$\rho_1 = \theta_1 + \theta_2\rho_1$$
$$\rho_2 = \theta_1\rho_1 + \theta_2\rho_0$$

AR(2) Process

- The relations between autocovariances may be restated in terms of autocorrelation coefficients, namely,

$$\rho_1 = \theta_1 + \theta_2 \rho_1$$
$$\rho_2 = \theta_1 \rho_1 + \theta_2 \rho_0$$

  These are the **Yule-Walker** equations for the AR(2) process. Solving for the first two autocorrelation coefficients gives

$$\rho_1 = \frac{\theta_1}{1 - \theta_2}, \qquad \rho_2 = \frac{\theta_1^2}{1 - \theta_2} + \theta_2$$

- The acf for the AR(2) process is

$$\rho_k = \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2}, \qquad k = 3, 4, \ldots$$

- This is a second-order difference equation with the first two values given above. Moreover. the coefficients of the difference equation are those of the AR(2) process.
- Thus the stationarity conditions ensure that the acf dies out as the lag increases. The acf will be a damped exponential or, if the roots of the acf are complex, a damped sine wave.

AR(2) Process

- An alternative and enlightening view of the AR(2) process is obtained by recasting it in lag operator notation:

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

  where $\theta(L) = 1 - \theta_1 L - \theta_2 L^2$.
- Express this quadratic as the product of two factors

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L)$$

- The connection between the λ and the θ parameters is

$$\lambda_1 + \lambda_2 = \theta_1 \text{ and } \lambda_1 \lambda_2 = -\theta_2$$

- The λ's may be seen as the roots of

$$\lambda^2 - \theta_1 \lambda - \theta_2 = 0$$

  which is the **characteristic equation** of the AR(s).
- Its roots are

$$\lambda_{1,2} = \frac{\theta_1 \pm \sqrt{\theta_1^2 + 4\theta_2}}{2}$$

AR(2) Process

- The inverse $\theta^{-1}(L)$ may be written

$$\theta^{-1}(L) = \frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)} = \frac{c}{1-\lambda_1 L} + \frac{d}{1-\lambda_2 L}$$

  where $c = -\lambda_1/(\lambda_2 - \lambda_1)$ and $d = \lambda_2/(\lambda_2 - \lambda_1)$

- Then

$$y_t = \theta^{-1}(L)\varepsilon_t = \frac{c}{1-\lambda_1 L}\varepsilon_t + \frac{d}{1-\lambda_2 L}\varepsilon_t$$

- From the results on the AR(1) process, stationarity of the AR(2) process requires

$$|\lambda_1| < 1 \text{ and } |\lambda_2| < 1$$

  Restating these conditions in terms of the θ parameters gives the stationarity conditions already derived before.

AR(2) Process

- The AR(2) case allows the possibility of a pair of complex roots, which will occur if $\theta_1^2 + 4\theta_2 < 0$. The roots may then be written as

$$\lambda_{1,2} = h \mp vi$$

  where $h = \theta_1/2$ and $v = (1/2)\sqrt{-(\theta_1^2 + 4\theta_2)}$ are real numbers.
- The autocorrelation coefficients will now display sine wave fluctuations, which will dampen toward zero provided the complex roots have *moduli* less than one. The absolute value or modulus of each complex root is

$$|\lambda_j| = \sqrt{h^2 + v^2} = -\theta_2, \qquad j = 1, 2$$

  which gives $0 < -\theta_2 < 1$ as the condition for the correlogram to be a damped sine wave.
- For real or complex roots the stationarity condition is that the moduli of the roots should be less than one, or that the roots lie within the unit circle.

Characteristic Equation

- In some books, characteristic equation is defined as

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2$$

(Note that the tranformation $z = 1/\lambda$ converts $\theta(z) = 0$ to $\lambda^2 - \theta_1\lambda - \theta_2 = 0$)

- The roots of $\theta(z)$ are the values of z that solve the equation

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z) = 0$$

- The roots are obviously

$$z_j = \frac{1}{\lambda_j}, \qquad j = 1, 2$$

so that if λ's lie within the unit circle, z's lie outside the unit circle.

- The stationarity condition is commonly stated in the literature as the roots of the relevant polynomial in the lag operator lying outside the unit circle.

MA(q) processes

- Recall that we defined a q-th order moving average process or $MA(q)$ as

$$y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q},$$

  where y_t is in deviation form and ε_t is a white noise process.

- Straightforward to characterize the properties of y_t:

$$\begin{aligned}
E[y_t] &= 0 \\
\gamma_0 &= E\left[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q})^2\right] \\
&= \left(1 + \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_q^2\right)\sigma^2 \\
\gamma_1 &= E\left[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q})(\varepsilon_{t-1} + \alpha_1 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q-1})\right] \\
&= (\alpha_1 + \alpha_2\alpha_1 + \cdots + \alpha_q\alpha_{q-1})\,\sigma^2 \\
\gamma_2 &= (\alpha_2 + \alpha_3\alpha_1 + \cdots + \alpha_q\alpha_{q-2})\,\sigma^2 \\
\vdots &= \vdots \\
\gamma_q &= \alpha_q\sigma^2 \\
\gamma_k &= 0, \qquad k > q
\end{aligned}$$

- Then autocorrelation function, ACF, can be computed as

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\alpha_k + \alpha_{k+1}\alpha_1 + \cdots + \alpha_q \alpha_{q-k}}{1 + \alpha_1^2 + \cdots + \alpha_q^2} & k \leq q \\ 0 & k > q \end{cases}$$

- Note that MA(q) process is stationary
- As y_t is a linear combination of stationary terms: the mean and variance are constant and the autocovariances, γ_k, depend on k but not on t

- Similar results hold for the infinite horizon moving average process, MA(∞).
- The MA(∞) process is stationary if the variance is bounded:

$$\gamma_0 = \left(1 + \alpha_1^2 + \alpha_2^2 + \cdots\right)\sigma^2 < \infty$$

- This requires

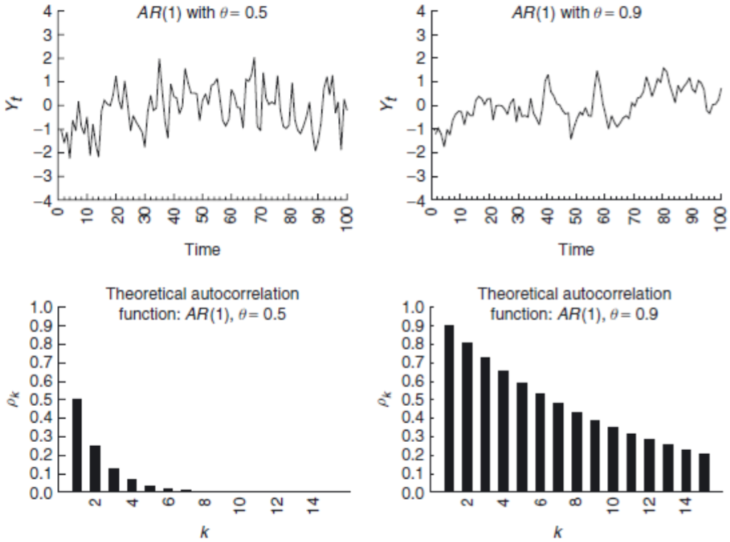
$$\sum_{i=1}^{\infty} \alpha_i^2 < \infty$$

- The moving average coefficients measure the dynamic impact of a shock to the process

$$\frac{\partial y_t}{\partial \varepsilon_t} = 1, \quad \frac{\partial y_t}{\partial \varepsilon_{t-1}} = \alpha_1, \quad \frac{\partial y_t}{\partial \varepsilon_{t-2}} = \alpha_2, \quad \ldots$$

- These partial derivatives are known as **impulse responses**
- Therefore stationarity implies that the impulse response function dies out.

AR(1) Data Series and Autocorrelation Functions



MA(1) Data Series and Autocorrelation Functions

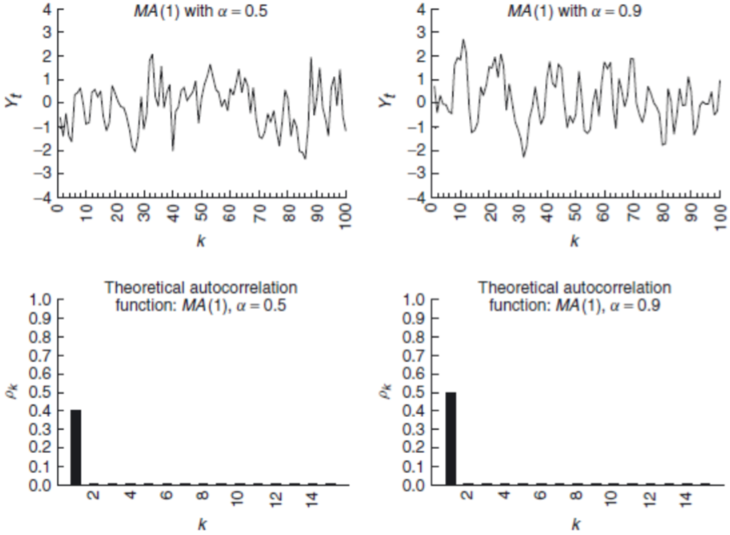


Table of Contents

Stationarity of ARMA(p,q)

- Recall the ARMA models written in terms of lag operator

$$\begin{aligned} \theta(L)y_t &= \varepsilon_t & \qquad AR(p) \\ y_t &= \alpha(L)\varepsilon_t & \qquad MA(q) \\ \theta(L)y_t &= \alpha(L)\varepsilon_t & \qquad ARMA(p,q) \end{aligned}$$

- If the AR lag polynomial is invertible, we can write $ARMA(p,q)$ as

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

  that is a moving average model of infinite order and it's stationary.
- Therefore the $ARMA(p,q)$ model $\theta(L)y_t = \alpha(L)\varepsilon_t$ is stationary if and only if AR polynomial $\theta(L)$ is invertible which in return requires the solutions $z_1, \ldots, z_p$ to $\theta(z) = 0$ to be larger than one (in absolute value).
- For example, $1 - \theta L$ is invertible if the solution to $1 - \theta z = 0$ is larger than 1 (or smaller than -1), i.e. if $|\theta| < 1$.

Stationarity of ARMA(p,q)

- For example, $1 - \theta L$ is invertible if the solution to $1 - \theta z = 0$ is larger than 1 (or smaller than -1), i.e. if $|\theta| < 1$.
- Consider the following autoregressive processes. Are they stationary?
  - $y_t = 0.8y_{t-1} + \varepsilon_t$
  - $y_t = 1.2y_{t-1} + \varepsilon_t$
  - $y_t = 1.2y_{t-1} - 0.32y_{t-2} + \varepsilon_t$
- The first one is stationary as we discussed above ($|0.8| < 1$). In contrast, the AR(1) model in the second example has a noninvertible lag polynomial so nonstationary.
- The characteristic equation corresponding to the third example is

  $$\theta(z) = 1 - 1.2z + 0.32z^2$$

  or

  $$\theta(z) = (1 - 0.8z)(1 - 0.4z)$$

- The solutions (characteristic roots) are $1/0.8$ and $1/0.4$, which are both larger than one. Consequently the AR polynomial is invertible and the process is stationary.

Stationarity and unit roots

- As an another example consider the following ARMA(2,1) model

$$y_t = 1.2y_{t-1} - 0.2y_{t-2} + \varepsilon_t - 0.5\varepsilon_{t-1}$$

- This is nonstationary because $z = 1$ is a solution to the characteristic polynomial $1 - 1.2z + 0.2z^2 = 0$.
- Note that a quick check on the stationarity of an AR process is obtained by checking whether its coefficients of the characteristic equation add up to less than one.
- When they add up to one the process is nonstationary and has a unit root.
- So we can write the process in the example as

$$(1 - 0.2L)(1 - L)y_t = (1 - 0.5L)\varepsilon_t$$

- Let us generalize this case, and then we will come back to the example

Stationarity and unit roots

- Suppose that we can write the ARMA(p,q) process as

$$\theta^*(L)(1-L)y_t = \theta^*(L)\Delta y_t = \alpha(L)\varepsilon_t$$

  where $\theta^*(L)$ is invertible in L of order $p-1$.
- Because the roots of the AR polynomial are the solutions to $\theta^*(z)(1-z)=0$, there is one solution $z=1$, or in other words a single unit root.
- Thus this result shows that $\Delta y_t = (1-L)y_t$ can be described by a stationary ARMA model if the process for y_t has only one unit root.
- Consequently, we can eliminate the nonstationarity by transforming the series into first-differences.

Stationarity and unit roots

- Going back to the example

$$(1-0.2L)(1-L)y_t = (1-0.5L)\varepsilon_t$$

- Since $\theta^*(L) = 1 - 0.2L$ is invertible Δy_t is described by a stationary ARMA(1,1) process given by

$$\Delta y_t = 0.2\Delta y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$$

- A series with one unit root becomes stationary after first differencing
- Such a series is called integrated of order one or $I(1)$.
- In some cases, taking firstdifferences is insufficient to obtain stationarity and another differencing step is required. In this case the stationary series is given by

$$\Delta^2 y_t = \Delta(\Delta y_t) = \Delta y_t - \Delta y_{t-1}$$

- If a series must be differenced twice before it becomes stationary, then it is said to be integrated of order two, denoted $I(2)$, and it must have two unit roots.
- Accordingly, a series y_t is $I(2)$ if Δy_t is nonstationary but $\Delta^2 y_t$ is stationary.

$I(0)$ versus $I(1)$

- A stationary I(0) series:
  - Fluctuates around its mean with a finite variance that does not depend upon time,
  - Is mean-reverting: tendency to return to its mean,
  - Has limited memory: effect of a shock dies out.
  - Autocorrelations die out (fairly) rapidly with k.
- A nonstationary I(1) series:
  - Wanders widely: no finite (unconditional) mean or variance.
  - Has infinitely long memory: shocks have permanent effect
  - Estimated autocorrelations decay to zero very slowly (often almost linearly).

Table of Contents

Testing for unit roots

- First, consider the AR(1) process

$$Y_t = \delta + \theta Y_{t-1} + \varepsilon_t$$

where $\theta = 1$ corresponds to a unit root.

- The standard t-statistic for this hypothesis is given by

$$DF = \frac{\hat{\theta} - 1}{\text{se}(\hat{\theta})}$$

- But has a nonstandard distribution under the null.
- The reason for this is that the nonstationarity of the process invalidates standard results on the distribution of the OLS estimator $\hat{\theta}$
- For example, if $\theta = 1$, the variance of Y_t is not even defined
- The correct distribution of this test statistics is given by Dickey and Fuller (1979) and it is known as the **Dickey-Fuller test**.
- As a rule, the Dickey-Fuller test regression contains an intercept term.
- Its distribution is skewed to the left. Critical values are given in the following table:

1% and 5% critical values for Dickey-Fuller tests

	Without trend		With trend	
Sample size	1 %	5 %	1 %	5 %
$T = 25$	−3.75	−3.00	−4.38	−3.60
$T = 50$	−3.58	−2.93	−4.15	−3.50
$T = 100$	−3.51	−2.89	−4.04	−3.45
$T = 250$	−3.46	−2.88	−3.99	−3.43
$T = 500$	−3.44	−2.87	−3.98	−3.42
$T = \infty$	−3.43	−2.86	−3.96	−3.41

Testing for unit roots

- Usually, a slightly more convenient regression procedure is used. In this case, the model is rewritten as

$$\Delta Y_t = \delta + (\theta - 1) Y_{t-1} + \varepsilon_t$$

  from which the t-statistic for $\theta - 1 = 0$ is identical to DF above.
- If there is a unit root the model reduces to

$$\Delta Y_t = \delta + \varepsilon_t$$

  and called **difference stationary**.
- It is also possible that nonstationarity is caused by the presence of a deterministic time trend in the process, rather than by the presence of a unit root. This happens when the AR(1) model is extended to

$$Y_t = \delta + \theta Y_{t-1} + \gamma t + \varepsilon_t$$

  with $|\theta| < 1$ and $\gamma \neq 0$.
- In this case, we have a nonstationary process because of the linear trend γt.
- In this case the process Y_t is called **trend stationary**.
- In contrast to the unit root case, shocks to a trend stationary process are transitory, and their effects die out over time.

Testing for unit roots

- It is possible to test whether Y_t follows a random walk against the alternative that it follows the trend stationary process.
- This can be tested by running the regression

$$\Delta Y_t = \delta + (\theta - 1)Y_{t-1} + \gamma t + \varepsilon_t$$

  - The null hypothesis is corresponding to a **random walk** is

    $$H_0: \delta = \gamma = \theta - 1 = 0$$

    (random walk with drift if $\delta \neq 0$)
  - The alternative is $|\theta| < 1$, corresponding to a deterministic trend if $\gamma \neq 0$.
- Instead of testing this joint hypothesis, it is quite common to use the t-ratio corresponding to $\hat{\theta} - 1$, denoted by DF_τ, assuming that the other restrictions in the null hypotheses are satisfied.
- Although the null hypothesis is still the same as in the previous unit root test, the testing regression is different and thus we have a different distribution of the test statistic.
- The critical values for DF_τ given in the last two columns of the table above.

- If a graphical inspection of the series indicates a clear positive or negative trend, it is most appropriate to perform the Dickey-Fuller test with a trend.
- However, that unnecessarily including a time trend may result in a loss of power.

Augmented Dickey-Fuller test

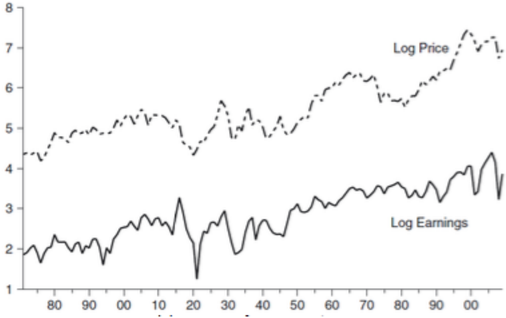
- We can generalize this to AR(p) process which is called **Augmented Dickey-Fuller test**
- The test regression contains additional lags of ΔY_t.
- To test there is single unit root in AR(p) we run the following regression

$$\Delta Y_t = \delta + \pi Y_{t-1} + c_1 \Delta Y_{t-1} + \cdots + c_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

- And the null hypothesis of a unit root corresponds to $\pi = 0$.
- The inclusion of the additional lags is done to make the error term in the test regression white noise.
- If AR(p) assumption is correct, and under the null hypothesis of a unit root, the asymptotic distributions of the DF is the same as before.
- Similarly, if the above equation contains a time trend the distribution of DF_τ statistics is again given by the last two columns in the table.
- Alternative tests:
  - Phillips - Perron test;
  - KPSS test (where the unit root is the alternative hypothesis);
  - And many more.

Illustration: stock prices and earnings

- Annual data on (log) stock price index (S&P 500) and (log) composite earnings over 1871 - 2009 (T=139).



- First, we consider the log price series.
- Not very likely to be stationary.
- Standard Dickey-Fuller test (with intercept and trend) gives:

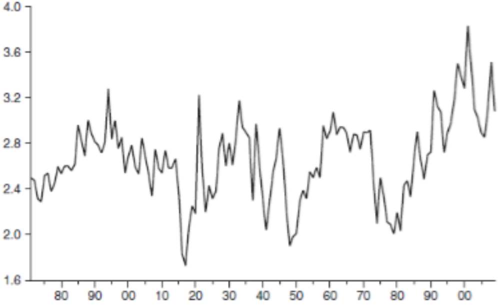
$$\Delta Y_t = \underset{(0.038)}{0.437} + \underset{(0.0074)}{0.00176}\ t - \underset{(0.0376)}{0.0984}\ Y_{t-1} + e_t$$

  resulting $DF = -2.62$.
- Augmented Dickey-Fuller tests give:

DF	ADF(1)	ADF(2)	ADF(3)	ADF(4)	ADF(5)	ADF(6)
-2.621	-2.744	-2.273	-2.618	-2.255	-2.154	-2.345

  $\rightarrow$ No rejection of unit root hypothesis.
- Results for log earnings are reasonably similar.

- To conclude, we consider log(price/earnings)



- Standard Dickey-Fuller test (with intercept and trend) produces:

$$\Delta Y_t = \underset{(0.155)}{0.685} - \underset{(0.058)}{0.255}\ Y_{t-1} + e_t$$

resulting $DF = -4.42$, corresponding to a clear rejection.

- However, putting in additional lags (ADF) or using some of the alternative tests in some cases leads to different conclusions.
- Data are apparently insufficiently informative on this issue.
- Mean reversion in the earnings/price ratio, if present, appears to be slow.

Table of Contents

Steps of ARMA modeling

How to apply and implement ARMA models in practice?

1. Specify AR and MA orders p and q;
2. Estimate parameters by (ordinary) least squares or maximum likelihood.
3. Evaluate the model by applying misspecification tests and other diagnostic measures.
4. Modify the model if necessary.
5. Use the model for description or forecasting.

The partial autocorrelation function

- While the ACF can help us to determine the order of a moving average model, for autoregressive models this is less helpful.
- To determine the order of a AR process we will use the partial autocorrelation function
- The k-th order **partial autocorrelation coefficient** is defined as the estimate for θ_k in an $AR(k)$ model

$$Y_t = \delta + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_k Y_{t-k} + \varepsilon_t$$

- Let us denote this by $\hat{\theta}_{kk}$
- When true model is $AR(p)$, but if we estimate $AR(k)$ with $k \geq p$ it holds that

$$\hat{\theta}_{kk} \longrightarrow 0 \text{ if } k > p$$

  and

$$\sqrt{T}(\hat{\theta}_{kk} - 0) \rightarrow N(0,1) \quad \text{if } k > p.$$

- Consequently, the partial autocorrelation coefficients (or the partial autocorrelation function (PACF)) can be used to determine the order of an AR process.

The partial autocorrelation function

- Testing an AR(k-1) model versus AR(k) model implies testing the null hypothesis that $\hat{\theta}_{kk} = 0$
- Under the null hypothesis that the model is AR(k-1) the approximate standard error of $\hat{\theta}_{kk}$ is $1/\sqrt{T}$ so that $\theta_{kk} = 0$ is rejected if $|\sqrt{T}\hat{\theta}_{kk}| > 1.96$.
- This way one can look at the PACF and test for each lag whether the partial autocorrelation coefficient is zero. For a genuine AR(p) model the partial autocorrelations will be close to zero after the pth lag
- In summary, we start with an AR(k) model with k reasonably large and then if θ_{kk} is statistically insignificant then we reduce the model to AR(k-1) and so on...

ACF and PACF

- In summary, an AR(p) process is characterized by
  - An ACF that is infinite in extent (it tails off)
  - A PACF that is (close to) zero for lags larger than p.
- For an $MA(q)$ process we have:
  - An ACF that is (close to) zero for lags larger than q.
  - A PACF that is infinite in extent (it tails off).
- In the absence of any of these, a combined ARMA model may provide a parsimonious representation.
- Note that we can compute PACF from the sample data by running basically a regression. On the other hand, the computation of ACF involves (unobserved) population parameters.
- Therefore, instead of ACF we use the **sample autocorrelation function** which is just sample analog of ACF:

$$\hat{\rho}_k = \frac{\frac{1}{T-k}\sum_{t=k+1}^{T}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\frac{1}{T}\sum_{t=1}^{T}(Y_t - \bar{Y})^2}$$

- Choosing p and q
  - In general, it is recommended to try to keep the model as parsimonious as possible. That is, we prefer, for example, $AR(2)$ to $AR(3)$ and $ARMA(1,1)$ to $MA(6)$
  - Diagnostic checking: residual analysis (residuals should not exhibit serial correlation), and overfitting;
  - Out-of-sample forecasting;

Estimating ARMA models

- Autoregressive models can be estimated by ordinary least squares.
- Moving average models can be estimated by nonlinear least squares.
- Alternatively, maximum likelihood can be used if distributional assumptions are made on the error term (normality). Results are equivalent to least squares.
- Estimation is reasonably straightforward.
- The more difficult issue is to determine which model to estimate.

Residual Analysis: Ljung-Box test

- A residual analysis is usually based on the fact that the residuals e_t of an adequate model should be approximately white noise.
- A plot of the residuals can be a useful tool in checking for outliers.
- There are also formal tests to examine the estimated residual autocorrelations
- A popular one is **Ljung-Box test** which is based on the test statistic

$$Q_K = T(T+2)\sum_{k=1}^{T}\frac{1}{T-k}r_k^2 \sim \chi^2(K-p-q)$$

  where
  - r_k's are the estimated AR coefficients on the residuals e_t
  - K is a number chosen by the researcher
- The null hypothesis is that the ARMA(p,q) model is correctly specified, so the residuals exhibit no autocorrelation
- Under this null Q_K is approximately $\chi^2(K-p-q)$ (so the test makes sense only when the choice satisfy $K > p+q$)

Model selection criteria

- A model selection criterion provide a trade-off between goodness-of-fit and parsimony (# of parameters to be estimated)
- There are two popular ones
- Akaikes Information Criterion (AIC):

$$AIC = \log \hat{\sigma}^2 + 2\frac{p+q+1}{T}$$

- Schwarzs Bayesian Information Criterion (BIC):

$$BIC = \log \hat{\sigma}^2 + \frac{p+q+1}{T} \log T$$

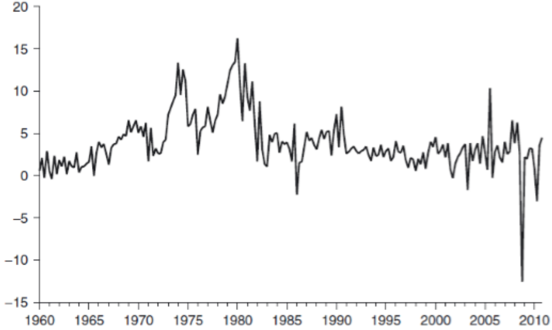
- Lower values are preferred.
- BIC has larger punishment for more parameters than AIC. (Results in more parsimonious models.)

Table of Contents

Illustration: the persistence of inflation

- Inflation is a key variable in economics
- Its persistence has received substantial attention.
- Persistence in this context means:
  - how long and how strongly does a 1% shock to inflation today affect future inflation rates?
  - How long does it take for the inflation rate to return to its previous level, if ever?
- We consider seasonally adjusted US inflation rates from 1960-2010 (T=204 quarters), based on the consumer price index (CPI)

Quarterly (Annualized) Inflation in the United States, 1960-2017

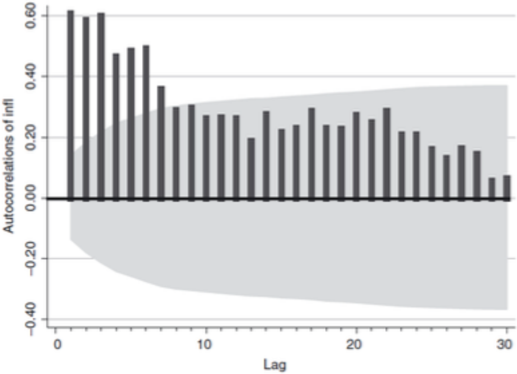


- The figure shows that inflation was relatively low in the 1960s, whereas it rose steadily in the 1970s, with peaks around 1974 and 1980.
- At the beginning of the 1980s, the Federal Reserve enforced its policy to reduce inflation rates, leading to lower and more stable inflation rates until the 1990s.
- The first decade of the new century exhibits increased variation in inflation rates, partly attributable to the recession and the high variation in commodity prices, like crude oil, in this period.

Unit Root?

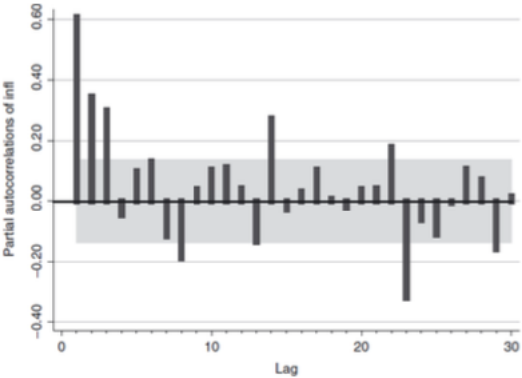
- As a first step, we test for the presence of a unit root in the quarterly inflation series using the augmented Dickey-Fuller test.
- The Augmented Dickey-Fuller test with an intercept and 2 lags takes a value of -3.078, which is marginally significant at 5%
- With four lags, the test statistic reduces to -2.764 and when more lags are added, it becomes less likely to reject the null of a unit root.
- Since the results are not too strong in either direction we can say that the inflation is either I(0) or I(1) with a high degree of persistence
- Our next look at the data involves an inspection of the sample autocorrelation and partial autocorrelation functions

Sample Autocorrelation Function of Inflation Rate



- The ACF shows that inflation is highly persistent, with the first seven autocorrelation coefficients being statistically significant from zero.

Sample Partial Autocorrelation Function of Inflation Rate



- The PACF indicates statistical significance of the first three partial autocorrelation coefficients, after which the PACF is close to zero, with an occasional peak in either direction.

Some estimated models: AR(3)

- To continue our analysis we assume that inflation is $I(0)$, but with a high degree of persistence.
- Based on the sample ACF and PACF, the first model we estimate is a AR(3) (because the PACF becomes insignificant at lag 4):

$$Y_t = \underset{(0.068)}{0.292}\, Y_{t-1} + \underset{(0.069)}{0.227}\, Y_{t-2} + \underset{(0.069)}{0.300}\, Y_{t-3} + e_t$$

AIC = 4.577; BIC = 4.642; s=2.363.

(For brevity, we do not report the estimated intercepts here and in the following regressions)

- Now we need to check whether the residuals e_t are autocorrelated or not (residual analysis). For that we compute the Ljung-Box test statistics for $K = 6$ and $K = 12$

$$Q_6 = 10.57 (p = 0.014) \text{ and } Q_{12} = 16.96 (p = 0.049)$$

- Tests suggest residual autocorrelation is still present.
- In order to accommodate this we consider two extensions: estimate AR(4) and ARMA(3,1)

AR(4) and ARMA(3,1)

- Regression results for AR(4):

$$\begin{array}{cccccc} Y_t = & 0.305\ Y_{t-1} + & 0.328\ Y_{t-2} + & 0.313 Y_{t-3} \text{ - } & 0.043 Y_{t-4} & + e_t \\ & (0.071) & (0.071) & (0.072) & (0.072) & \end{array}$$

  AIC = 4.585; BIC = 4.666; s=2.36.
  $Q_6 = 11.143$ (p=0.004) and $Q_{12} = 17.505$ (p=0.025)

- Regression results for ARMA(3,1):

$$\begin{array}{cccccc} Y_t = & 1.104\ Y_{t-1} + & 0.303\ Y_{t-2} + & 0.365 Y_{t-3} + & 0.207 e_{t-1} & + e_t \\ & (0.212) & (0.108) & (0.089) & (0.227) & \end{array}$$

  AIC = 4.585; BIC = 4.666; s=2.36
  $Q_6 = 10.882$ (p=0.004) and $Q_{12} = 17.286$ (p=0.027)

- Neither of these two extended specifications is superior to the original AR(3) model.
- Nevertheless, each of the three models estimated so far still exhibit some residual serial correlation.
- An inspection of the residual ACF and PACF suggests that including a 6th lag may be appropriate. Therefore, as a next specification, we consider an AR(6) model

AR(6)

- Regression results for AR(6):

$$
\begin{array}{lllll}
Y_t = & \underset{(0.071)}{0.297}\, Y_{t-1} + & \underset{(0.074)}{0.218}\, Y_{t-2} + & \underset{(0.077)}{0.248}\, Y_{t-3} & \\
- & \underset{(0.077)}{0.106}\, Y_{t-4} + & \underset{(0.075)}{0.062}\, Y_{t-5} + & \underset{(0.072)}{0.132}\, Y_{t-6} & + e_t
\end{array}
$$

AIC = 4.577; BIC = 4.691; s=2.347
$Q_{12} = 13.605$ (p=0.034)

- The three additional lags of Y_t in this model are individually insignificant at the 5% level.
- The AIC is slightly better than for the AR(3) model, but the BIC, which has a larger punishment for the additional parameters, favours the more parsimonious AR(3) specification.
- The Ljung-Box test marginally rejects the null hypothesis of no residual autocorrelation at the 5% level.

Reduced AR(6)

- As a final specification, we estimate an AR(6) model, but exclude the intermediate lags 4 and 5:

$$\begin{array}{lllllll} Y_t & = & 0.270 Y_{t-1} + & 0.216\ Y_{t-2} + & 0.242 Y_{t-3} + & 0.125 Y_{t-6} & + e_t \\ & & (0.068) & (0.069) & (0.075) & (0.066) & \end{array}$$

AIC $=$ 4.569; BIC $=$ 4.650; s=2.348
$Q_6 = 5.174$ (p=0.075) and $Q_{12} = 15.106$ (p=0.057)

- This specification (marginally) satisfies the Ljung-Box portmanteau tests.
- The 6th lag, however, is only marginally significant.
- The AIC favours the latter specification to the AR(3) model, whereas the BIC favours the AR(3)
- The results of our specification search are not clear-cut, and either the AR(6) or the (reduced) AR(6) model can be defended.