Time Series Note

Lingbo Shen

May 22, 2015

Contents

1 Modelling Ecomomic Time Series

1.1 Data Generation Processes

Let $\boldsymbol{x}_t$ be an $m \times 1$ vector of economic variables generated at time t. Such variables are typically inter0related both contemporaneously and across time. The collection $\{\boldsymbol{x}_t, -\infty < t < \infty\}$ is called a vector-valued random sequence.

The data generation process (DGP) is represented by the conditional density

$$D_t\left(\boldsymbol{x}_t \mid \mathcal{X}_{t-1}\right)$$

where $\mathcal{X}_{t-1} = \sigma\left(\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \boldsymbol{x}_{t-3}, \cdots\right)$. This is a shorthand for the σ-field representing knowledge of the past history of the system. Notice that

Remark 1.

D_t is allowed to depend on time, because the data are not assumed to be stationary.

1.2 DGPs and Models

A dynamic econometric model is a family of functions of the data which are intended to mimic aspects of the DGP, either D_t itself or functions derived from D_t such as moments. Formally, a model is a family of functions

$$\{M\left(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \cdots, \boldsymbol{d}_t; \boldsymbol{\psi}\right), \boldsymbol{\psi} \in \Psi\}, \Psi \subset \mathbb{R}^p$$

In particular, let M_D be a model of complete DGP.

- Parameters $\boldsymbol{\psi}$: p in number; parameters are constants that are common to every t.

- Parameter space: Ψ denotes the set of admissible parameter values.

- The vector $\boldsymbol{d}_t$ represents variables, treated as non-stochastic, which are intended to capture the changes in the DGP over time.

The relationship between the DGP and the model is a difficult issue. The axiom of correct specification is the assumption that there exists a model element that is identical to the corresponding function of the DGP. M_D is correctly specified if there exists $\boldsymbol{\psi}_0 \in \Psi$ such that

$$M_D\left(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \cdots, \boldsymbol{d}_t; \boldsymbol{\psi}_0\right) \quad = \quad D_t\left(\boldsymbol{x}_t \mid \mathcal{X}_{t-1}\right)$$

In general, correct specification in practical modelling exercises is an implausible assumption.

1.3 Sequence Properties

1.3.1 Stationarity

A random sequence $\{\boldsymbol{x}_t\}$ is said to be *stationary in the wide sense/covariance-stationary* if the mean, the variance and the sequence of jth-order autocovariances for $j > 0$ are all independent of t. It is said to be *stationary in the strict sense* if for every $k > 0$, the joint distributions of all collections $(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}, \boldsymbol{x}_{t+2}, \cdots, \boldsymbol{x}_{t+k})$ do not depend in any way on t.

1.3.2 Mixing

In a mixing sequence the realization of the sequence at time t contains no information about the realization at either $t - j$ or $t + j$, when j is sufficiently large. The mixing property ensures that points in the sequence appear randomly sampled when they are far enough apart.

Remark 2. *Stationarity and mixing are quite distinct properties.*

1.3.3 Ergodicity

A stationary sequence having the property that a random event involving every member of the sequence always has probability either 0 or 1 is called *ergodic*.

Theorem 1. *If $\{x_t\}$ is a stationary ergodic sequence, and $E(x_1)$ exists, $\bar{x}_n \to E(x_1)$ with probability 1.*

2 Difference Equation

2.1 The Difference Operator

First differences

$$\Delta y_t \equiv y_t - y_{t-1}$$

Second differences

$$
\begin{aligned}
\Delta^2 y_t &\equiv \Delta(\Delta y_t) \\
&\equiv \Delta y_t - \Delta y_{t-1} \\
&\equiv (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\
&\equiv y_t - 2y_{t-1} + y_{t-2}
\end{aligned}
$$

2.2 Difference Equations and Solutions

- n-th order linear difference equation (with constant coefficient)

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + X_t$$

where X_t is called forcing process. It includes stochastic terms, time trends, and other variables, but not constants and not y, nor lagged values of y.

- difference form

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + X_t$$

$$y_t = a_0 + a_1 y_{t-1} + \sum_{i=2}^{n} a_i y_{t-i} + X_t$$

$$y_t - y_{t-1} = a_0 + a_1 y_{t-1} - y_{t-1} + \sum_{i=2}^{n} a_i y_{t-i} + X_t$$

$$\Delta y_t = a_0 + (a_1 - 1)y_{t-1} + \sum_{i=2}^{n} a_i y_{t-i} + X_t$$

- A solution of a D.E shows y_t equal to a function of the x and t, plus perhaps some initial conditions for y, but not lagged values of y. Solutions are not unique.

2.2.1 Iteration

Consider following first-order linear difference equation:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t, \quad y_0 = y_0$$

We know

$$
\begin{aligned}
y_1 &= a_0 + a_1 y_0 + \varepsilon_1 \\
y_2 &= a_0 + a_1 y_1 + \varepsilon_2 \\
&= a_0 + a_1(a_0 + a_1 y_0 + \varepsilon_1) + \varepsilon_2 \\
&= a_0 + a_1 a_0 + a_1^2 y_0 + a_1 \varepsilon_1 + \varepsilon_2 \\
y_3 &= a_0 + a_1 y_2 + \varepsilon_3 \\
&= a_0 + a_1(a_0 + a_1 a_0 + a_1^2 y_0 + a_1 \varepsilon_1 + \varepsilon_2) + \varepsilon_3 \\
&= a_0 + a_1 a_0 + a_1^2 a_0 + a_1^3 y_0 + a_1^2 \varepsilon_1 + a_1 \varepsilon_2 + \varepsilon_3 \\
&\vdots \\
y_t &= a_0(1 + a_1 + a_1^2 + \cdots + a_1^{t-1}) + a_1^t y_0 + a_1^{t-1} \varepsilon_1 + a_1^{t-2} \varepsilon_2 + \cdots + \varepsilon_t \\
&= a_0 \sum_{i=0}^{t-1}(a_1)^i + a_1^t y_0 + \sum_{i=0}^{t-1}(a_1)^i \varepsilon_{t-i}
\end{aligned}
$$

2.3   Lag Operators

The lag operator L is defined

$$
L^i y_t \equiv y_{t-i}
$$

Lag operator has following properties

1. $L(\beta y_t) = \beta \cdot L y_t$

2. $L(x_t + y_t) = L x_t + L y_t$

3. $(L^i + L^j) y_t = L^i y_t + L^j y_t$

4. $L^i L^j y_t = L^{i+j} y_t = y_{t-i-j}$

5. $L^{-i} y_t = y_{t+i}$

6. For $|a| < 1$

$$
\sum_{i=0}^{+\infty}(a^i L^i) y_t = \frac{y_t}{1 - aL}
$$

7. For $|a| > 1$

$$
\sum_{i=0}^{+\infty}(a^{-i} L^{-i}) y_t = \frac{-aL y_t}{1 - aL}
$$

2.3.1   First-Order Difference Equations

Consider the first-order difference equation below

$$y_t = \phi y_{t-1} + w_t$$

which can be written using the lag operator as

$$y_t = \phi L y_t + w_t$$
$$y_t - \phi L y_t = w_t$$
$$(1 - \phi L)y_t = w_t$$
$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L)y_t = (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)w_t$$

The compound operator on the left-hand side is

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L)$$
$$= (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)$$
$$- (\phi L + \phi^2 L^2 + \cdots + \phi^t L^t + \phi^{t+1} L^{t+1})$$
$$= 1 - \phi^{t+1} L^{t+1}$$

Then we have

$$(1 - \phi^{t+1} L^{t+1})y_t = (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)w_t$$
$$y_t - \phi^{t+1} y_{-1} = w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \cdots + \phi^t w_0$$
$$y_t = \phi^{t+1} y_{-1} + w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \cdots + \phi^t w_0$$

We have known that

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L)y_t = y_t - \phi^{t+1} y_{-1}$$

if $|\phi| < 1$ and y_{-1} is finite, $\phi^{t+1} y_{-1}$ will become negligible as t becomes large

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L)y_t \cong y_t$$

A sequence $\{y_t\}_{t=-\infty}^{\infty}$ is bounded if there exists a finite number $\bar{y}$ such that $|y_t| < \bar{y}$ for all t. When $|\phi| < 1$ and when we are considering applying an operator to a bounded sequence, we have

$$(1 - \phi L)^{-1} = \lim_{t \to \infty}(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)$$

So we have

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L)y_t \; \cong \; y_t$$
$$(1 - \phi L)^{-1}(1 - \phi L)y_t \; = \; y_t$$

By this definition, we have

$$(1 - \phi L)y_t \; = \; w_t$$
$$y_t \; = \; (1 - \phi L)^{-1}w_t$$
$$= \; w_t + \phi w_{-1} + \phi^2 w_{t-2} + \phi^3 w_{t-3} + \cdots$$

2.3.2   Second-Order Difference Equations

Consider the second-order difference equation

$$y_t \; = \; \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t$$
$$(1 - \phi_1 L - \phi_2 L^2)y_t \; = \; w_t$$

The operator in the left-hand side contains a second-order polynomial in the lag operator L. Suppose we have

$$(1 - \phi_1 L - \phi_2 L^2) \; = \; (1 - \lambda_1 L)(1 - \lambda_2 L)$$
$$= \; (1 - [\lambda_1 + \lambda_2]L + \lambda_1 \lambda_2 L^2)$$

Given values for ϕ_1 and ϕ_2, we seek numbers λ_1 and λ_2 such that

$$\begin{cases} \lambda_1 + \lambda_2 = \phi_1 \\ \lambda_1 \lambda_2 = -\phi_2 \end{cases}$$

Two methods to consider this problem.

- Method 1

$$(1 - \phi_1 z - \phi_2 z^2) \; = \; (1 - \lambda_1 z)(1 - \lambda_2 z)$$

when $z = \lambda_1^{-1}$ or $z = \lambda_2^{-1}$, the right hand side is equal to 0. When

$$\begin{cases} z_1 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \\ \\ z_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \end{cases}$$

the left-hand side is equal to 0 as well. So if we let

$$
\begin{cases}
\lambda_1^{-1} = z_1 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \\[3em]
\lambda_2^{-1} = z_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}
\end{cases}
$$

both sides are equal to 0.

- Method 2

  It is easy to find that λ_1 and λ_2 are roots of equation

$$
\lambda^2 - \phi_1\lambda - \phi_2 = 0
$$

so

$$
\begin{cases}
\lambda_1^{-1} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} \\[3em]
\lambda_2^{-1} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}
\end{cases}
$$

When we have λ_1 and λ_2 by the above two methods, we have

$$
(1 - \lambda_1 L)(1 - \lambda_2 L)y_t = w_t
$$
$$
y_t = (1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1}w_t
$$

If $\lambda_1 \neq \lambda_2$, we define

$$
(\lambda_1 - \lambda_2)^{-1}\left(\frac{\lambda_1}{1 - \lambda_1 L} - \frac{\lambda_2}{1 - \lambda_2 L}\right)
$$
$$
= (\lambda_1 - \lambda_2)^{-1}\frac{\lambda_1(1 - \lambda_2 L) - \lambda_2(1 - \lambda_1 L)}{(1 - \lambda_1 L)(1 - \lambda_2 L)}
$$
$$
= (\lambda_1 - \lambda_2)^{-1}\frac{\lambda_1 - \lambda_2}{(1 - \lambda_1 L)(1 - \lambda_2 L)}
$$
$$
= \frac{1}{(1 - \lambda_1 L)(1 - \lambda_2 L)}
$$

So we have

$$
\begin{aligned}
y_t &= (1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1} w_t \\
&= (\lambda_1 - \lambda_2)^{-1}\left(\frac{\lambda_1}{1 - \lambda_1 L} - \frac{\lambda_2}{1 - \lambda_2 L}\right) w_t \\
&= \frac{\lambda_1}{\lambda_1 - \lambda_2}\frac{1}{1 - \lambda_1 L}w_t - \frac{\lambda_2}{\lambda_1 - \lambda_2}\frac{1}{1 - \lambda_2 L}w_t \\
&= \frac{\lambda_1}{\lambda_1 - \lambda_2}(1 + \lambda_1 L + \lambda_1^2 L^2 + \lambda_1^3 L^3 + \cdots)w_t \\
&\quad - \frac{\lambda_2}{\lambda_1 - \lambda_2}(1 + \lambda_2 L + \lambda_2^2 L^2 + \lambda_2^3 L^3 + \cdots)w_t \\
&= \frac{\lambda_1}{\lambda_1 - \lambda_2}(w_t + \lambda_1 w_{t-1} + \lambda_1^2 w_{t-2} + \lambda_1^3 w_{t-3} + \cdots) \\
&\quad - \frac{\lambda_2}{\lambda_1 - \lambda_2}(w_t + \lambda_2 w_{t-1} + \lambda_2^2 w_{t-2} + \lambda_2^3 w_{t-3} + \cdots) \\
&= (c_1 + c_2)w_t + (c_1\lambda_1 + c_2\lambda_2)w_{t-1} + (c_1\lambda_1^2 + c_2\lambda_2^2)w_{t-2} + \cdots
\end{aligned}
$$

where $c_1 = \lambda_1/(\lambda_1 - \lambda_2)$ and $c_2 = -\lambda_2/(\lambda_1 - \lambda_2)$.

2.3.3 pth-Order Difference Equations

3 ARMA

The autoregressive moving average (ARMA) model

$$
y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}
$$

3.1 Stationarity

A stochastic process is *covariance stationary* or *weakly stationary* if for all t and s

$$
\begin{aligned}
E(y_t) &= E(y_{t-s}) = \mu \\
E\left[(y_t - \mu)^2\right] &= E\left[(y_{t-s} - \mu)^2\right] = \sigma_y^2 = \gamma_0 \\
E\left[(y_t - \mu)(y_{t-s} - \mu)\right] &= E\left[(y_{t-j} - \mu)(y_{t-j-s} - \mu)\right] = \gamma_s
\end{aligned}
$$

If a process is covariance stationary, the covariance between y_t and y_{t-s} depends only on s, the length of time separating the observations. It follows that for a covariance stationary process, γ_s and γ_{-s} would represent the same magnitude.

For a covariance stationary series, we can define the autocorrelation between y_t and y_{t-s}

$$
\rho_s = \frac{\gamma_s}{\gamma_0}
$$

where γ_0 is the variance of y_t

3.2   Ergodicity

Imagine a battery of I computers generating sequences $\{y_t^{(1)}\}_{t=-\infty}^{\infty}$, $\{y_t^{(2)}\}_{t=-\infty}^{\infty}$, $\ldots$, $\{y_t^{(I)}\}_{t=-\infty}^{\infty}$ and consider selecting the observation associated with date t from each sequence:

$$\{y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(I)}\}$$

This would be described as a sample of I realizations of the random variable Y_t. The expectation of the tth observation of a time series refers to the mean of the probability distribution

$$E(Y_t) \quad = \quad \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t) dy_t$$

We might view this as the probability limit of the ensemble average

$$E(Y_t) \quad = \quad \operatorname*{plim}_{I \to \infty} \frac{1}{I} \sum_{i=1}^{I} Y_t^{(i)}$$

The above expectations of a time series in terms of ensemble averages may seem a bit contrived. Usually we have a single realization of size T from the process

$$\{y_1^{(1)}, y_2^{(1)}, \ldots, y_T^{(1)}\}$$

From these observations we would calculate the sample mean $\bar{y}$, which is a time average

$$\bar{y} \quad = \quad \frac{1}{T} \sum_{t=1}^{T} y_t^{(1)}$$

A covariance stationary process is said to be *ergodic for the mean* if $\bar{y}$ converges in probability to $E(Y_t)$ as $T \to \infty$.

3.3   Moving Average Processes

3.3.1   The First-Order Moving Average Process

Let $\{\varepsilon_t\}$ be white noise and consider the process

$$y_t \quad = \quad \mu + \varepsilon_t + \theta \varepsilon_{t-1}$$

where μ and θ could be any constants. This time series is called a first-order moving average process, denoted $MA(1)$.

Expectation   The expectation of y_t is

$$E(y_t) \quad = \quad E(\mu + \varepsilon_t + \theta \varepsilon_{t-1}) = \mu + E(\varepsilon) + \theta E(\varepsilon_{t-1}) = \mu$$

Variance   The variance of y_t is

$$
\begin{aligned}
E(y_t - \mu)^2 &= E(\varepsilon_t + \theta \varepsilon_{t-1})^2 \\
&= E\left(\varepsilon_t^2 + 2\theta \varepsilon_t \varepsilon_{t-1} + \theta^2 \varepsilon_{t-1}^2\right) \\
&= \sigma^2 + 0 + \theta^2 \sigma^2 \\
&= (1 + \theta^2)\sigma^2
\end{aligned}
$$

Autocovariance   The first autocovariance of y_t is

$$
\begin{aligned}
E(y_t - \mu)(y_{t-1} - \mu) &= E(\varepsilon_t + \theta \varepsilon_{t-1})(\varepsilon_{t-1} + \theta \varepsilon_{t-2}) \\
&= E\left(\varepsilon_t \varepsilon_{t-1} + \theta \varepsilon_{t-1}^2 + \theta \varepsilon_t \varepsilon_{t-2} + \theta^2 \varepsilon_{t-1} \varepsilon_{t-2}\right) \\
&= 0 + \theta \sigma^2 + 0 + 0 \\
&= \theta^2 \sigma^2
\end{aligned}
$$

Higher autocovariances are all zero. For all $j > 1$

$$
E(y_t - \mu)(y_{t-j} - \mu) \;=\; 0
$$

Autocorrelation   The jth autocorrelation of a covariance stationary process is $\rho_j = \frac{\gamma_j}{\gamma_0}$

$$
\begin{aligned}
\rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{\theta^2 \sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta^2}{1 + \theta^2} \\
\rho_j &= 0, \quad j > 1
\end{aligned}
$$

3.3.2   The qth-Order Moving Average Process

A qth-order moving average process, denoted $MA(q)$, is characterized by

$$
y_t \;=\; \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}
$$

Expectation   The expectation of y_t is

$$
E(y_t) \;=\; E(\mu) + E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + \theta_2 E(\varepsilon_{t-2}) + \cdots + \theta_q E(\varepsilon_{t-q}) = \mu
$$

Variance   The variance of y_t is

$$
\begin{aligned}
E(y_t - \mu)^2 &= E(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q})^2 \\
&= E(\varepsilon_t)^2 + E(\theta_1 \varepsilon_{t-1})^2 + E(\theta_2 \varepsilon_{t-2})^2 + \cdots + E(\theta_q \varepsilon_{t-q})^2 \\
&= \sigma^2 + \theta_1^2 \sigma^2 + \theta_2^2 \sigma^2 + \cdot + \theta_q^2 \sigma^2 \\
&= (1 + \theta_1^2 + \theta_2^2 + \cdot + \theta_q^2)\sigma^2
\end{aligned}
$$

Autocovariance   The autocovariance of y_t is

$$
\begin{aligned}
E(y_t - \mu)(y_{t-j} - \mu) &= E(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}) \\
&\quad (\varepsilon_{t-j} + \theta_1\varepsilon_{t-j-1} + \theta_2\varepsilon_{t-j-2} + \cdots + \theta_q\varepsilon_{t-j-q}) \\
&= E\left(\theta_j\varepsilon_{t-j}^2 + \theta_{j+1}\theta_1\varepsilon_{t-j-1}^2 + \theta_{j+2}\theta_2\varepsilon_{t-j-2}^2 + \cdots + \theta_q\theta_{q-j}\varepsilon_{t-q}^2\right) \\
&= \left(\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q\theta_{q-j}\right)\sigma^2, \quad j = 1, 2, \ldots, q
\end{aligned}
$$

For all $j > q$

$$
E(y_t - \mu)(y_{t-j} - \mu) = 0
$$

3.3.3   The Infinite-Order Moving Average Process

Consider the process when $q \to \infty$

$$
y_t = \mu + \sum_{j=0}^{\infty} \psi_j\varepsilon_{t-j} = \mu + \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots
$$

This could be described as an $MA(\infty)$ process.

The $MA(\infty)$ process is covariance stationary if it is square summable

$$
\sum_{j=0}^{\infty} \psi_j^2 < \infty
$$

It is often to work with a slightly stronger condition called absolutely summable

$$
\sum_{j=0}^{\infty} |\psi_j| < \infty
$$

Expectation   The mean of an $MA(\infty)$ process with absolutely summable is

$$
E(y_t) = \lim_{T\infty} E(\mu + \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots + \psi_T\varepsilon_{t-T}) = \mu
$$

Autocovariance   The autocovariance of an $MA(\infty)$ process with absolutely summable is

$$
\begin{aligned}
\gamma_0 &= E(y_t - \mu)^2 \\
&= \lim_{T\infty} E(\psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots + \psi_T\varepsilon_{t-T})^2 \\
&= \lim_{T\to\infty} (\psi_0^2 + \psi_1^2 + \psi_2^2 + \cdots + \psi_T^2)\sigma^2 \\
\gamma_j &= E(y_t - \mu)(y_{t-j} - \mu) \\
&= (\psi_j\psi_0 + \psi_{j+1}\psi_1 + \psi_{j+2}\psi_2 + \psi_{j+3}\psi_3 + \cdots)\sigma^2
\end{aligned}
$$

3.4 Autoregressive Processes

3.4.1 The First-Order Autoregressive Process

A first-order autoregressive, denoted $AR(1)$, satisfies the following difference equation

$$y_t \;=\; c + \phi y_{t-1} + \varepsilon_t$$

When $|\phi| < 1$, this process is covariance stationary. The $AR(1)$ process can be rewritten as $MA(\infty)$ process as follows

$$
\begin{aligned}
y_t \;&=\; c + \phi y_{t-1} + \varepsilon_t \\
&=\; c + \varepsilon_t + \phi(c + \varepsilon_{t-1}) + \phi^2 y_{t-2} \\
&=\; c + \varepsilon_t + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + \phi^3 y_{t-3} \\
&=\; c + \varepsilon_t + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + \phi^3(c + \varepsilon_{t-3}) + \phi^4 y_{t-4} \\
&=\; c + \varepsilon_t + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + \phi^3(c + \varepsilon_{t-3}) + \cdots \\
&=\; (c + \phi c + \phi^2 c + \phi^3 c + \cdots) + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \cdots \\
&=\; \frac{c}{1-\phi} + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \cdots
\end{aligned}
$$

We can derive the expectation and autocovariance of $AR(1)$ by the above corresponding $MA(\infty)$ process. We also can derive them by assuming $AR(1)$ process is covariance stationary.

Expectation Taking expectations both sides

$$
\begin{aligned}
E(y_t) \;&=\; c + \phi E(y_{t-1}) + E(\varepsilon_t) \\
\mu \;&=\; c + \phi \mu \\
\mu \;&=\; \frac{c}{1-\phi}
\end{aligned}
$$

Autocovariance

3.4.2 The qth-Order Autoregressive Process

3.5 The Autocorrelation Function

Autocorrelation function (ACF) and the partial autocorrelation function (PACF) are useful to determine the type of time series data.

For AR(1) model

- Method 1

$$
\begin{aligned}
y_t &= a_0 + a_1 y_{t-1} + \varepsilon_t \quad \textit{assume stationary} \\
E(y_t) &= a_0 + a_1 E(y_{t-1}) + E(\varepsilon_t) \\
\Rightarrow \mu &= a_0 + a_1 \mu \\
\Rightarrow \mu &= \frac{a_0}{1 - a_1} \\
Var(y_t) &= a_1^2 Var(y_{t-1}) + Var(\varepsilon_t) \\
\Rightarrow \gamma_0 &= a_1^2 \gamma_0 + \sigma^2 \\
\Rightarrow \gamma_0 &= \frac{\sigma^2}{1 - a_1^2} \\
Cov(y_t, y_{t-s}) &= Cov(a_0 + a_1 y_{t-1} + \varepsilon_t, a_0 + a_1 y_{t-s-1} + \varepsilon_{t-s}) \\
&= Cov(a_1 y_{t-1} + \varepsilon_t, a_1 y_{t-s-1} + \varepsilon_{t-s}) \\
\Rightarrow \gamma_s &= a_1^2 \gamma_s + a_1^s \sigma^2 \\
\Rightarrow \gamma_s &= \frac{a_1^s \sigma^2}{1 - a_1^2}
\end{aligned}
$$

So the autocorrelation function for AR(1)

$$
\begin{aligned}
\rho_s &= \frac{\gamma_s}{\gamma_0} \\
&= a_1^s
\end{aligned}
$$

- Method 2

If the process started at time zero

$$
y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}
$$

Take the expectation of y_t and y_{t+s}

$$
E(y_t) = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0
$$

$$
E(y_{t+s}) = a_0 \sum_{i=0}^{t+s-1} a_1^i + a_1^{t+s} y_0
$$

If $|a_1| < 1$, as $t \to \infty$

$$
\lim_{t \to \infty} y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}
$$

$$Var(y_t) \quad = \quad E\left[(y_t - \mu)^2\right] = E\left[(\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \cdots)^2\right]$$

$$= \quad \sigma^2\left[(1 + a_1^2 + a_1^4 + \cdots)\right] = \frac{\sigma^2}{1 - a_1^2}$$

$$Cov(y_t, y_{t-s}) \quad = \quad E\left[(y_t - \mu)(y_{t-s} - \mu)\right]$$

$$= \quad E\left[(\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \cdots)(\varepsilon_{t-s} + a_1\varepsilon_{t-s-1} + a_1^2\varepsilon_{t-s-2} + \cdots)\right]$$

$$= \quad E\left(a_1^s\varepsilon_{t-s}^2 + a_1^{s+2}\varepsilon_{t-s-1}^2 + a_1^{s+4}\varepsilon_{t-s-2}^2 + \cdots\right)$$

$$= \quad \sigma^2 a_1^s(1 + a_1^2 + a_1^4 + \cdots)$$

$$= \quad \frac{\sigma^2 a_1^s}{1 - a_1^2}$$

3.5.1 The Autocorrelation Function of an AR(2) Process

We assume that $a_0 = 0$, which implies that $E(y_t) = 0$. Adding or subtracting any constant from a variable does not change its variance, covariance, correlation coefficient, etc.

Using Yule-Walker equations: multiply the second-order D.E by y_{t-s} for s= $0, 1, 2, \cdots$, and take expectations

$$Ey_t y_t \quad = \quad a_1 Ey_{t-1}y_t + a_2 Ey_{t-2}y_t + E\varepsilon_t y_t$$

$$Ey_t y_{t-1} \quad = \quad a_1 Ey_{t-1}y_{t-1} + a_2 Ey_{t-2}y_{t-1} + E\varepsilon_t y_{t-1}$$

$$Ey_t y_{t-2} \quad = \quad a_1 Ey_{t-1}y_{t-2} + a_2 Ey_{t-2}y_{t-2} + E\varepsilon_t y_{t-2}$$

$$\vdots$$

$$Ey_t y_{t-s} \quad = \quad a_1 Ey_{t-1}y_{t-s} + a_2 Ey_{t-2}y_{t-s} + E\varepsilon_t y_{t-s}$$

By definition, the autocovariances of a stationary series are such

$$Ey_t y_{t-s} \quad = \quad Ey_{t-s}y_t = Ey_{t-k}y_{t-k-s} = \gamma_s$$

We also know that coefficient on ε_t is unity so that $E\varepsilon_t y_t = \sigma^2$, and $E\varepsilon_t y_{t-s} = 0$, so

$$\gamma_0 \quad = \quad a_1\gamma_1 + a_2\gamma_2 + \sigma^2$$

$$\gamma_1 \quad = \quad a_1\gamma_0 + a_2\gamma_1$$

$$\gamma_2 \quad = \quad a_1\gamma_1 + a_2\gamma_0$$

$$\vdots$$

$$\gamma_s \quad = \quad a_1\gamma_{s-1} + a_2\gamma_{s-2}$$

Now we can get the ACF

$$\rho_1 = a_1\rho_0 + a_2\rho_1$$

$$\rho_s = a_1\rho_{s-1} + a_2\rho_{s-2}$$

We know $\rho_0 = 1$, so

$$\rho_1 = a_1 + a_2\rho_1$$

$$\rho_1 = \frac{a_1}{1 - a_2}$$

3.5.2 The Autocorrelation Function of an MA(1) Process

Consider the MA(1) process $y_t = \varepsilon_t + \beta\varepsilon_{t-1}$

Applying the Yule-Walker equations

$$\gamma_0 = E(y_t y_t) = E\left[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_t + \beta\varepsilon_{t-1})\right] = (1 + \beta^2)\sigma^2$$

$$\gamma_1 = E(y_t y_{t-1}) = E\left[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-1} + \beta\varepsilon_{t-2})\right] = \beta^2\sigma^2$$

$$\gamma_2 = E(y_t y_{t-2}) = E\left[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-2} + \beta\varepsilon_{t-3})\right] = 0$$

$$\gamma_s = E(y_t y_{t-s}) = E\left[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-s} + \beta\varepsilon_{t-s-1})\right] = 0 \quad \forall t > 2$$

So the ACF of MA(1)

$$\rho_0 = 1$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\beta^2}{1 + \beta^2}$$

$$\rho_s = 0 \quad \forall t > 1$$

3.5.3 The Autocorrelation Function of an ARMA(1,1) Process

Consider the ARMA(1,1) $y_t = a_1 y_{t-1} + \varepsilon_t + \beta_1\varepsilon_{t-1}$

$$E y_t y_t = a_1 E y_{t-1} y_t + E\varepsilon_t y_t + \beta_1 E\varepsilon_{t-1} y_t \Rightarrow$$

$$\gamma_0 = a_1\gamma_1 + \sigma^2 + \beta_1(a_1 + \beta_1)\sigma^2$$

$$E y_t y_{t-1} = a_1 E y_{t-1} y_{t-1} + E\varepsilon_t y_{t-1} + \beta_1 E\varepsilon_{t-1} y_{t-1} \Rightarrow$$

$$\gamma_1 = a_1\gamma_0 + \beta_1\sigma^2$$

$$E y_t y_{t-2} = a_1 E y_{t-1} y_{t-2} + E\varepsilon_t y_{t-2} + \beta_1 E\varepsilon_{t-1} y_{t-2} \Rightarrow$$

$$\gamma_2 = a_1\gamma_1$$

$$E y_t y_{t-s} = a_1 E y_{t-1} y_{t-s} + E\varepsilon_t y_{t-s} + \beta_1 E\varepsilon_{t-1} y_{t-s} \Rightarrow$$

$$\gamma_s = a_1\gamma_{s-1}$$

Solve the equations and get

$$\gamma_0 = \frac{1 + \beta_1^2 + 2a_1\beta_1}{1 - a_1^2}\sigma^2$$

$$\gamma_1 = \frac{(1 + a_1\beta_1)(a_1 + \beta_1)}{1 - a_1^2}\sigma^2$$

And the AFC

$$\rho_0 = 1$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 + a_1\beta_1)(a_1 + \beta_1)}{1 + \beta_1^2 + 2a_1\beta_1}$$

$$\rho_s = a_1\rho_s \quad \forall t > 1$$

3.6 The Partial Autoorrelation Function

In AR(1) process, y_t and y_{t-2} are correlated even though y_{t-2} does not directly appear in the model. $\rho_2 = corr(y_t, y_{t-1}) \times corr(y_{t-1}, y_{t-2}) = \rho_1^2$. All such "indirect" correlations are present in the ACF. In contrast, the partial autocorrelation between y_t and y_{t-s} eliminates the effects of the intervening values y_{t-1} through y_{t-s+1}.

Method to find the PACF:

1. Form the series $\{y_t^*\}$, where $y_t^* \equiv y_t - \mu$

2. Form the first-order autoregression equation:

$$y_t^* = \phi_{11}y_{t-1}^* + e_i$$

$$y_t^* = \phi_{21}y_{t-1}^* + \phi_{22}y_{t-2}^* + e_i$$

where ϕ_{11} is the partial autocorrelation between y_t and y_{t-1}, ϕ_{22} is the partial autocorrelation between y_t and y_{t-2}. Repeating this process for all additional lags s yields the PACF.

3.7 Sample Autoorrelations

Given that a series is stationary, we can use the sample mean, variance and autocorrelations to estimate the parameters of the actual data-generating process.

Table 1: Properties of the ACF and PACF

Process	ACF	PACF
White-noise	$\rho_s = 0$	$\phi_{ss} = 0$
AR(1): $a_1 > 0$	$\rho_s = a_1^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$
AR(1): $a_1 < 0$	$\rho_s = a_1^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$
AR(p)	Decays toward zero.	Spikes through lag p
	Coefficients may oscillate	$\phi_{ss} = 0$ for $s \geq p$
MA(1): $\beta > 0$	$\rho_s = 0, \, for \, s \geq 2$	Oscillating decay: $\phi_{11} > 0$
MA(1): $\beta > 0$	$\rho_s = 0, \, for \, s \geq 2$	Decay: $\phi_{11} < 0$
ARMA(1, 1): $a_1 > 0$		
ARMA(1, 1): $a_1 < 0$		
ARMA(p, q)		

The estimates of μ, σ^2 and ρ:

$$\overline{y} = \frac{\sum_{t=1}^{T} y_t}{T}$$

$$\hat{\sigma^2} = \frac{\sum_{t=1}^{T}(y_t - \overline{y})^2}{T}$$

$$r_s = \frac{\sum_{t=s+1}^{T}(y_t - \overline{y})(y_{t-s} - \overline{y})}{\sum_{t=1}^{T}(y_t - \overline{y})^2}$$

Box and Pierce used the sample autocorrelations to form the statistic

$$Q = T \sum_{k=1}^{s} r_k^2$$

If the data are generated from a stationary ARMA process, Q is asymptotically $\chi^2(s)$ distribution.

Ljung and Box test

$$Q = T(T+2) \sum_{k=1}^{s} \frac{r_k^2}{T-k} \sim \chi^2(s)$$

4 Asymptotics for Dynamic Models

4.1 The Simple Autoregressive Model

4.2 Martingale Difference Processes

The adapted sequence $\{s_t, \mathcal{F}_t\}$ is called a martingale if for every t the following conditions hold:

1. $E|s_t| < \infty$

2. $E(s_t \mid \mathcal{F}_{t-1}) = s_{t-1}$ a.s.

The adapted sequence $\{x_t, \mathcal{F}_t\}$ is called a martingale difference (m.d.) if for every t the following conditions hold:

1. $E|x_t| < \infty$

2. $E(s_t \mid \mathcal{F}_{t-1}) = 0$ a.s.

Here is a fundamental property of m.d. processes.

Theorem 2. *Let $\{x_t\}$ be an m.d. sequence and let $g_{t-1} = g(x_{t-1}, x_{t-2}, \cdots ,)$ be any measurable, integrable function of the lagged values of the sequence. Then $x_t g_{t-1}$ is also an m.d., and x_t and g_{t-1} are uncorrelated.*

Proof. By law of iterated expectations, we have

$$
\begin{aligned}
E(x_t g_{t-1}) &= E(E(x_t g_{t-1} \mid \mathcal{F}_{t-1})) \\
&= E(g_{t-1} E(x_t \mid \mathcal{F}_{t-1})) \\
&= E(g_{t-1} \cdot 0) \\
&= 0
\end{aligned}
$$

It means that $x_t g_{t-1}$ is also an m.d. For uncorrelation, we show that

$$
\begin{aligned}
Cov(x_t, g_{t-1}) &= E(x_t g_{t-1}) - E(x_t)E(g_{t-1}) \\
&= E(x_t g_{t-1}) - E(E(x_t \mid \mathcal{F}_{t-1})E(g_{t-1})) \\
&= 0 - 0 \\
&= 0
\end{aligned}
$$

$\square$

In particular, putting $g_{t-1} = x_{t-j}$ for any $j > 0$, the theorem implies

$$Cov(x_t, x_{t-j}) \quad = \quad 0$$

The m.d. property implies uncorrelatedness of the sequence, although it is a weaker property than independence.

Given an arbitrary sequence $\{y_t\}$ satisfying $E|y_t| < \infty$ and $\sigma(y_t, y_{t-1}, \cdots) \subset \mathcal{F}_t$, an m.d. sequence can always be generated as the centred sequence $\{x_t\}$ where

$$x_t \quad = \quad y_t - E(y_t \mid \mathcal{F}_{t-1})$$

We consider the law of large number and central limit theorem for m.d. processes.

Theorem 3. *Let $\{x_t\}$ be an m.d. sequence. Then $\operatorname{plim} \bar{x}_n = 0$ if either of the following condition hold.*

1. *The sequence is strictly stationary and $E|x_t| < \infty$*

2. *$E|x_t|^{1+\delta} < \infty, \forall t$*

The following is central limit theorem for m.d.

Theorem 4. *Let $\{x_t, \mathcal{F}_t\}$ be a m.d. sequence with $E(x_t^2) = \sigma_t^2$ and let $\bar{\sigma}_n^2 = n^{-1} \sum_{t=1}^{n} \sigma_t^2$. If*

1. *$\operatorname{plim} n^{-1} \sum_{t=1}^{n} \left(x_t^2 - \sigma_t^2\right) = 0$*

2. *either*

   (a) *the sequence is strictly stationary*

   (b)

5 Trends and Unit Roots

5.1 The Random Walk Model

Consider the following special case AR(1) process:

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

If y_0 is a given initial condition, its solution is

$$y_t \quad = \quad y_0 + \sum_{i=1}^{t} \varepsilon_i$$

Take expectation

$$Ey_t = E\left(y_0 + \sum_{i=1}^{t} \varepsilon_i\right) = y_0$$

Taking expectation and variance

$$E_t y_{t+1} = E_t\left(y_t + \varepsilon_{t+1}\right) = y_t$$

$$E_t y_{t+s} = E_t\left(y_t + \sum_{i=1}^{s} \varepsilon_{t+i}\right) = y_t$$

$$Var(y_t) = Var\left(\sum_{i=1}^{t} \varepsilon_i\right) = t\sigma^2$$

$$Var(y_{t-s}) = Var\left(\sum_{i=1}^{t-s} \varepsilon_i\right) = (t-s)\sigma^2$$

$$E\left[(y_t - y_0)(y_{t-s} - y_0)\right] = E\left[(\sum_{i=1}^{t} \varepsilon_i)(\sum_{i=1}^{t+s} \varepsilon_i)\right]$$

$$= E\left[(\sum_{i=1}^{t-s} \varepsilon_i^2)\right]$$

$$= (t-s)\sigma^2$$

The correlation coefficient ρ_s is

$$\rho_s = \frac{(t-s)\sigma^2}{\sqrt{(t-s)\sigma^2} \times \sqrt{t\sigma^2}}$$

$$= \sqrt{\frac{t-s}{t}}$$

When t is big relative to s, the ρ_s are close to unity and decay very slowly.

The Random Walk plus Drift Model   Adding a constant term a_0:

$$y_t = y_{t-1} + a_0 + \varepsilon_t$$

Giving the initial condition y_0, its solution is

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i$$

The behavior of y_t is governed by two nonstationary components: a linear deterministic trend and the stochastic trend.

5.2 Function Spaces

- x: an element of C, that is, any continuous curve traversing the unit interval, be denoted.

- Coordinates of x: $x(r) \in \mathbb{R}$ is the unique values of x at points $r \in [0,1]$ are called the coordinates of x.

For two members of C, $x \in C$ and $y \in C$, we need to say how close together they are. Technically, C must be assigned a metric. For example, we can define **Euclidean metric** for any pair of real numbers x and y as $d_E(x,y) = |x - y|$. The pair $(\mathbb{R}, d_E)$ is known as the **Euclidean space**.

We also can define a metric called **uniform metric** as

$$d_U(x,y) \quad = \quad \sup_{0 \le r \le 1} |x(r) - y(r)|$$

This is just the *largest vertical separation* between the pair of functions over the interval. (C, d_U) is a metric space.

5.3 Brownian Motion

A Brownian motion B is a real random function on the unit interval, with the following properties:

1. $B \in C$ with probability 1.

2. $B(0) = 0$ with probability 1.

3. for any set of subintervals defined by arbitrary $0 \le r_1 < r_2 < \cdots < r_k \le 1$, the increments $B(r_1, B(r_2) - B(r_1, \cdots, B(r_k) - B(r_{k-1})$ are independent.

4. $B(t) - B(s) \sim N(0, t - s)$ for $0 \le s < t \le 1$.

5.4 The Functional Central Limit Theorem

We construct a variable $X_T(r)$ from the sample mean of the first rth fraction of the observations, $r \in [0,1]$, defined by

$$X_T(r) \quad \equiv \quad \frac{1}{T} \sum_{t=1}^{[Tr]} u_t$$

5.5   Dickey-Fuller Tests

Subtracting y_{t-1} from each side of the equation $y_t = a_1 y_{t-1} + \varepsilon_t$, we get $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$, where $\gamma = a_1 - 1$. Testing the hypothesis $a_1 = 1$ is equivalent to testing $\gamma = 0$.

Dickey and Fuller consider three different regression equations

$$\text{random walk model}$$
$$\Delta y_t \;=\; \gamma y_{t-1} + \varepsilon_t$$
$$\text{add a drift}$$
$$\Delta y_t \;=\; a_0 + \gamma y_{t-1} + \varepsilon_t$$
$$\text{add a drift and linear time trend}$$
$$\Delta y_t \;=\; a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$$

Run the OLS and get the estimated value of γ and associated standard error of these three models. However, the critical values of the t-statistics do depend on whether a drift and/or time trend is included in **regression models**. Note that the appropriate critical values depend on **sample size**. For any given level of significance, the critical values of the t-statistic decrease as sample size increases.

5.6   Augmented Dicker-Fuller test

Consider the pth-order autoregressive process:

$$y_t \;=\; a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p} + \varepsilon_t$$

jj Add and subtract $a_p y_{t-p+1}$

$$y_t \;=\; a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p+1} + a_p y_{t-p} - a_p y_{t-p+1} + \varepsilon_t$$
$$\;=\; a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} - a_p \Delta y_{t-p+1} + \varepsilon_t$$

Add and subtract $(a_{p-1} + a_p) y_{t-p+2}$

$$y_t \;=\; a_0 + \cdots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} - (a_{p-1} + a_p) y_{t-p+2} - a_p \Delta y_{t-p+1} + \varepsilon_t$$
$$\;=\; a_0 + \cdots + (a_{p-2} + a_{p-1} + a_p) y_{t-p+2} - (a_{p-1} + a_p) \Delta y_{t-p+2} - a_p \Delta y_{t-p+1} + \varepsilon_t$$

Continuing in this fashion, we get

$$\Delta y_t \quad = \quad a_0 + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t$$

$$where \quad \gamma \quad = \quad -\left(1 - \sum_{i=1}^{p} a_i\right)$$

$$\beta_i \quad = \quad \sum_{j=i}^{p} a_j$$

We can use the same Dickey-Fuller statistics which depends on the regression models and sample size.

6 Multivarite Model

6.1 Vector Autoregression Model

Consider the first order VAR model (**Structural equation**):

$$y_t \quad = \quad b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt}$$

$$z_t \quad = \quad b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt}$$

$$assume$$

$$\{y_t\}, \ \{z_t\} \ are \ stationary$$

$$\varepsilon_{yt} \sim W.N(0, \sigma_y^2)$$

$$\varepsilon_{zt} \sim W.N(0, \sigma_z^2)$$

$$\varepsilon_{yt}, \ \varepsilon_{zt} \ are \ uncorrelated$$

$$
\begin{aligned}
y_t + b_{12} z_t &= b_{10} + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{\mathrm{yt}} \\
b_{21} y_t + z_t &= b_{20} + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{\mathrm{zt}}
\end{aligned}
$$

$$
\begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix} \begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} b_{10} \\ b_{20} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{\mathrm{yt}} \\ \varepsilon_{\mathrm{zt}} \end{pmatrix}
$$

$$
\begin{aligned}
\mathrm{B} x_t &= \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t \\
x_t &= B^{-1}\Gamma_0 + B^{-1}\Gamma_1 x_{t-1} + B^{-1}\varepsilon_t \\
&= A_0 + A_1 x_{t-1} + e_t
\end{aligned}
$$

$$
A_0 = B^{-1}\Gamma_0 = \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix}^{-1} \begin{pmatrix} b_{10} \\ b_{20} \end{pmatrix}
$$

$$
A_1 = B^{-1}\Gamma_1 = \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}
$$

$$
e_t = B^{-1}\varepsilon_t = \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon_{\mathrm{yt}} \\ \varepsilon_{\mathrm{zt}} \end{pmatrix}
$$

Rewrite the vector form $x_t = A_0 + A_1 x_{t-1} + e_t$

$$
\begin{aligned}
y_t &= a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t} \\
z_t &= a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t}
\end{aligned}
$$

We know that

$$
\begin{aligned}
e_t &= B^{-1}\varepsilon_t \\
&= \frac{1}{1 - b_{12}b_{21}} \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{\mathrm{yt}} \\ \varepsilon_{\mathrm{zt}} \end{pmatrix} \\
&= \frac{1}{1 - b_{12}b_{21}} \begin{pmatrix} \varepsilon_{\mathrm{yt}} - b_{12}\varepsilon_{\mathrm{zt}} \\ \varepsilon_{\mathrm{zt}} - b_{21}\varepsilon_{\mathrm{yt}} \end{pmatrix}
\end{aligned}
$$

So

$$
e_{1t} = \frac{\varepsilon_{\mathrm{yt}} - b_{12}\varepsilon_{\mathrm{zt}}}{1 - b_{12}b_{21}}
$$

$$
e_{2t} = \frac{\varepsilon_{\mathrm{zt}} - b_{21}\varepsilon_{\mathrm{yt}}}{1 - b_{12}b_{21}}
$$

To derive the properties of e_t

$$E(e_{1t}) = E\left(\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right) = 0$$

$$E(e_{1t}^2) = E\left[\left(\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right)^2\right]$$

$$= \frac{\sigma_{yt}^2 + b_{12}^2\sigma_{zt}^2}{(1 - b_{12}b_{21})^2}$$

$$E(e_{1t}e_{1t-i}) = E\left[\frac{(\varepsilon_{yt} - b_{12}\varepsilon_{zt})(\varepsilon_{yt-i} - b_{12}\varepsilon_{zt-i})}{(1 - b_{12}b_{21})^2}\right] = 0$$

$$E(e_{1t}e_{2t}) = E\left[\frac{(\varepsilon_{yt} - b_{12}\varepsilon_{zt})(\varepsilon_{zt} - b_{21}\varepsilon_{yt})}{(1 - b_{12}b_{21})^2}\right]$$

$$= \frac{-\left(b_{21}\sigma_{yt}^2 + b_{12}\sigma_{zt}^2\right)}{(1 - b_{12}b_{21})^2}$$

6.1.1 Identification

Compare the number of parameters in the structural VAR with the number of parameters from the standard form VAR model.

Form	Estimate	Calculation	N
Reduced form	a_{10}, a_{20}, a_{11}, a_{12}, a_{21}, a_{22}	$var(e_{1t})$, $var(e_{2t})$, $cov(e_{1t}, e_{2t})$	9
Structural form	b_{10}, b_{20}, γ_{11}, γ_{12}, γ_{21}, γ_{22}, b_{12}, b_{21}	σ_y, σ_z	10

The primitive system contains 10 parameters, whereas the VAR estimation yields only nine parameters. Unless restrict one of the parameters, it is not possible to identify the primitive system. If exactly one parameter of the primitive system is restricted, the system is exactly identified, and if more than one parameter is restricted, the system is over-identified.

6.2 VAR

$$Y_t = AY_{t-1} + \epsilon_t = A^2 Y_{t-2} + A\epsilon_{t-1} + \epsilon_t = A^m Y_{t-m} + A^{m-1}\epsilon_{t-m+1} + .. + \epsilon_t$$

This explodes if A^m explodes if $m \to \infty$, this does not explode if it converges to 0. Thus if all eigenvalues of A are smaller than 1 in absolute value.

6.3 Vector Auto Regression: VAR

Definition VAR: for k-variate time series:

$$Y_t = c + \Gamma_1 Y_{t-1} + ... + \Gamma_p Y_{t-p} + \epsilon_t$$

Y_t : K-dim $\qquad$ c : K-dim $\qquad$ Γ_1 : K x H -dim $\qquad$ ϵ_t : K-dim

Probalistic analysis VAR(1) identical to AR(1) Model:

$$Y_t = AY_{t-1} + \epsilon_t, \qquad \epsilon_t \text{ iid}$$

Stationary solution: write as VMA(∞):

$$Y_t = A^m Y_{t-m} + \sum_{j=0}^{m-1} A^j \epsilon_{t-j} \overset{m \to \infty}{\to} \sum_{j=0}^{\infty} A^j \epsilon_{t-j} \qquad \text{if all absolute eigenvalues of A} < 1.$$

Expectation:

$$E(Y_t) = E(\sum_{j=0}^{\infty} A^j \epsilon_{t-j}) = 0 \text{ if } E(\epsilon_t) = 0$$

Variance:

$$Var(Y_t) = Var(\sum_{j=0}^{\infty} A^j \epsilon_{t-j}) = \sum_{j=0}^{\infty} Var(A^j \epsilon_{t-j}) = \sum_{j=0}^{\infty} A^j Var(\epsilon_{t-j})(A^j)^T$$

Alternative:

$$Var(Y_t) = Var(AY_{t-1} + \epsilon_t) = Var(AY_{t-1}) + Var(\epsilon_t) = AVar(Y_{t-1})A^T + Var(\epsilon_t) = AVar(Y_t)A^T + Var(\epsilon_t).$$

Last equality due to stationarity.

Slides of Sender, VAR(1):

$$Var(Y_{t+k}) = \sum_{i=1}^{k} \theta(i) \cdot \Psi \cdot \theta^T(i)$$

$$\text{with} \qquad \Psi = E(\epsilon_t \cdot \epsilon_t^T)$$

$$\text{and} \qquad \theta(k) = (I + \phi_1 + \phi_1^2 + \cdots + \phi_1^{k-1})$$

6.4 Impulse response functions

For instance: What is the effect on the, say, third component of Y_t if ϵ_{t-7} has a unit shock in its second component $\to$ look at $(A^7)_{3,2}$. A not known, but has to be estimated so use Confidence Intervals. To calculate these, use the delta-method.

7   Trends and Unit Roots

7.1   The Random Walk Model

Consider the following special case AR(1) process:

$$
\begin{aligned}
y_t &= y_{t-1} + u_t \\
\Delta y_t &= u_t
\end{aligned}
$$

If y_0 is a given initial condition, its solution is

$$
y_t = y_0 + \sum_{i=1}^{t} u_i
$$

Take expectation

$$
E y_t = E \left(y_0 + \sum_{i=1}^{t} u_i \right) = y_0
$$

Taking expectation and variance

$$
\begin{aligned}
E_t y_{t+1} &= E_t \left(y_t + \varepsilon_{t+1} \right) = y_t \\
E_t y_{t+s} &= E_t \left(y_t + \sum_{i=1}^{s} \varepsilon_{t+i} \right) = y_t \\
Var(y_t) &= Var \left(\sum_{i=1}^{t} u_i \right) = t\sigma^2 \\
Var(y_{t-s}) &= Var \left(\sum_{i=1}^{t-s} u_i \right) = (t-s)\sigma^2 \\
E\left[(y_t - y_0)(y_{t-s} - y_0) \right] &= E \left[(\sum_{i=1}^{t} u_i)(\sum_{i=1}^{t+s} u_i) \right] \\
&= E \left[(\sum_{i=1}^{t-s} u_i^2) \right] \\
&= (t-s)\sigma^2
\end{aligned}
$$

The correlation coefficient ρ_s is

$$
\begin{aligned}
\rho_s &= \frac{(t-s)\sigma^2}{\sqrt{(t-s)\sigma^2} \times \sqrt{t\sigma^2}} \\
&= \sqrt{\frac{t-s}{t}}
\end{aligned}
$$

When t is big relative to s, the ρ_s are close to unity and decay very slowly.

The Random Walk plus Drift Model   Adding a constant term a_0:

$$y_t = y_{t-1} + a_0 + u_t$$

Giving the initial condition y_0, its solution is

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} u_i$$

The behavior of y_t is governed by two nonstationary components: a linear deterministic trend and the stochastic trend.

7.2   Function Spaces

- x: an element of C, that is, any continuous curve traversing the unit interval, be denoted.

- Coordinates of x: $x(r) \in \mathbb{R}$ is the unique values of x at points $r \in [0,1]$ are called the coordinates of x.

For two members of C, $x \in C$ and $y \in C$, we need to say how close together they are. Technically, C must be assigned a metric. For example, we can define **Euclidean metric** for any pair of real numbers x and y as $d_E(x,y) = |x - y|$. The pair $(\mathbb{R}, d_E)$ is known as the **Euclidean space**.

We also can define a metric called **uniform metric** as

$$d_U(x,y) = \sup_{0 \leq r \leq 1} |x(r) - y(r)|$$

This is just the *largest vertical separation* between the pair of functions over the interval. (C, d_U) is a metric space.

7.3   Brownian Motion

A Brownian motion B is a real random function on the unit interval, with the following properties:

1. $B \in C$ with probability 1.

2. $B(0) = 0$ with probability 1.

3. for any set of subintervals defined by arbitrary $0 \leq r_1 < r_2 < \cdots < r_k \leq 1$, the increments $B(r_1, B(r_2) - B(r_1, \cdots, B(r_k) - B(r_{k-1})$ are independent.

4. $B(t) - B(s) \sim N(0, t - s)$ for $0 \leq s < t \leq 1$.

7.4　The Functional Central Limit Theorem

We construct a variable $X_T(r)$ from the sample mean of the first rth fraction of the observations, $r \in [0, 1]$, defined by

$$X_T(r) \equiv \frac{1}{T} \sum_{t=1}^{[Tr]} u_t$$

For any given realization, $X_T(r)$ is a step function in r, with

$$X_T(r) = \begin{cases} 0 & 0 \le r < 1/T \\ \frac{u_1}{T} & 1/T \le r < 2/T \\ \frac{u_1 + u_2}{T} & 2/T \le r < 3/T \\ \vdots & \\ \frac{u_1 + u_2 + \cdots + u_{T-1}}{T} & (T-1)/T \le r < 1 \\ \frac{u_1 + u_2 + \cdots + u_{T-1} + u_T}{T} & r = 1 \end{cases}$$

Then

$$
\begin{aligned}
\sqrt{T} X_T(r) &= \sqrt{T} \frac{1}{T} \sum_{t=1}^{[Tr]} u_t \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \\
&= \frac{\sqrt{[Tr]}}{\sqrt{T}} \frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} u_t \\
&\approx_d \sqrt{r} N(0, \sigma^2) \\
&\sim N(0, r\sigma^2)
\end{aligned}
$$

The last second comes from the facts that $\frac{\sqrt{[Tr]}}{\sqrt{T}} \to \sqrt{r}$ and $\frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} u_t \approx_d N(0, \sigma^2)$.

If we consider a sample mean based on observations $[Tr_1]$ through $[Tr_2]$ for $r_2 > r_1$, we would conclude that

$$\sqrt{T} \left[X_T(r_2) - X_T(r_2) \right] / \sigma \approx N(0, r_2 - r_1)$$

It should not be surprising that the sequence of stochastic functions $\left\{ \sqrt{T} X_T(\cdot)/\sigma \right\}_{T=1}^{\infty}$ has an asymptotic probability law that is described by standard Brownian motion

$$\sqrt{T} X_T(\cdot)/\sigma \to W(\cdot)$$

This result is called *functional central limit theorem*.

Remark 3. *We should notice the difference between $X_T(\cdot)$ and $X_T(r)$. $X_T(\cdot)$ denotes a random function while $X_T(r)$ denotes the value that function assumes at date r; thus, $X_T(\cdot)$ is a function, while $X_T(r)$ is a random variable.*

Convergence in law for random functions Let $S(\cdot)$ be a continuous-time stochastic process with $S(r)$ representing its value at time date r for $r \in [0, 1]$. Suppose that for any given realization $S(\cdot)$ is a continuous function of r with probability 1.

Definition 1. *For $\left\{S_T(\cdot)\right\}_{T=1}^{\infty}$ a sequence of such continuous functions, we say that $S_T(\cdot) \overset{L}{\to} S(\cdot)$ if all of the following hold:*

1. *For any finite collection of k particular dates $0 \leq r_1 < r_2 < \cdots < r_k \leq 1$, the sequence of $k-$dimensional random vectors $\{\boldsymbol{y}_T\}_{T=1}^{\infty}$ converges in distribution to the vector $\boldsymbol{y}$, where*

$$
\boldsymbol{y}_T \equiv
\begin{bmatrix}
S_T(r_1) \\
S_T(r_2) \\
\vdots \\
S_T(r_k)
\end{bmatrix}
\qquad
\boldsymbol{y} \equiv
\begin{bmatrix}
S(r_1) \\
S(r_2) \\
\vdots \\
S(r_k)
\end{bmatrix}
$$

2. *For each $\varepsilon > 0$, the probability that $S_T(r_1)$ differs from $S_T(r_2)$ for any dates r_1 and r_2 within δ of each other goes to zero uniformly in T as $\delta \to 0$.*

3. *$P\{|S_T(0)| > \lambda\} \to 0$ uniformly in T as $\lambda \to \infty$.*

Remark 4. *This definition applies to sequences of continuous functions, though the function $X_T(r)$ is a discontinuous step function. Fortunately, the discontinuities occur at a countable set of points.*

Convergence in probability to sequences of random functions Let $\left\{S_T(\cdot)\right\}_{T=1}^{\infty}$ and $\left\{V_T(\cdot)\right\}_{T=1}^{\infty}$ denote sequences of random continuous functions with $S_T : r \in [0, 1] \to \mathbb{R}^1$ and $V_T : r \in [0, 1] \to \mathbb{R}^1$. Let the scalar Y_T represent the largest amount by which $S_T(r)$ differs from $V_T(r)$ for any r defined by

$$
Y_T \quad \equiv \quad \sup_{0 \leq r \leq 1} |S_T(r) - V_T(r)|
$$

Definition 2. *If the sequence of scalars $\left\{Y_T(\cdot)\right\}_{T=1}^{\infty}$ converges in probability to zero, then we say that the sequence of functions $S_T(\cdot)$ converges in probability to $V_T(\cdot)$. That is, the expression*

$$
S_T(\cdot) \quad \to_p \quad V_T(\cdot)
$$

is interpreted to mean that

$$\sup_{r \in [0,1]} |S_T(r) - V_T(r)| \quad \to_p \quad 0$$

Continuous mapping theorem   The result is similar to Slutsky's theorem.

Theorem 5. *The continuous mapping theorem states that if $S_T(\cdot) \xrightarrow{L} S(\cdot)$ and $g(\cdot)$ is a continuous **functional**, then $g\left(S_T(\cdot)\right) \xrightarrow{L} g\left(S(\cdot)\right)$.*

For example, consider $S_T(r) \equiv [\sqrt{T} X_T(r)]^2$ and $\sqrt{T} X_T(r) \xrightarrow{L} \sigma W(\cdot)$. It follows that

$$S_T(\cdot) \quad \xrightarrow{L} \quad \sigma^2 [W(\cdot)]^2$$

7.5   Unit Root Processes

A random walk $y_t = y_{t-1} + u_t$ where $\{u_t\}$ is an i.i.d. sequence with zero mean and variance σ^2. If $y_0 = 0$, then
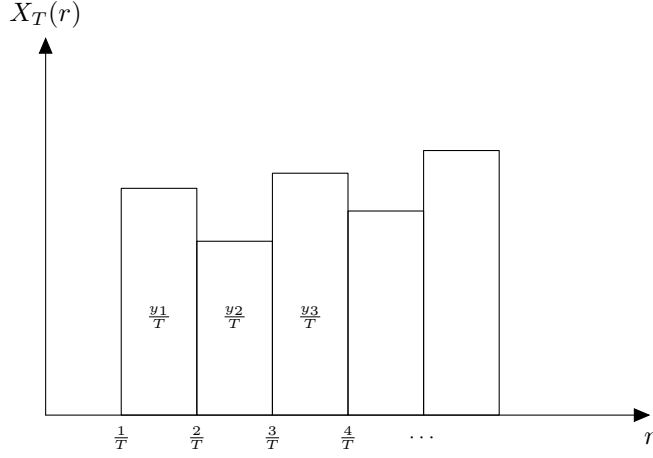
$$y_t = u_1 + u_2 + \cdots + u_t$$

The stochastic function $X_T(r)$ defined as

$$X_T(r) \quad = \quad \begin{cases} 0 & 0 \le r < 1/T \\[2mm] \frac{y_1}{T} & 1/T \le r < 2/T \\[2mm] \frac{y_2}{T} & 2/T \le r < 3/T \\[2mm] \vdots & \\[2mm] \frac{y_{T-1}}{T} & (T-1)/T \le r < 1 \\[2mm] \frac{y_T}{T} & r = 1 \end{cases}$$

The figure below plots $X_T(r)$ as a function of r. The area under this step function is the sum of T rectangles. The integral of $X_T(r)$ is thus equivalent to

$$\int_0^1 X_T(r) dr \quad = \quad \frac{y_1}{T^2} + \frac{y_2}{T^2} + \cdots + \frac{y_{T-1}}{T^2}$$

Then we have

$$\int_0^1 \sqrt{T} X_T(r) dr \;\; = \;\; \sqrt{T}\left[\frac{y_1}{T^2} + \frac{y_2}{T^2} + \cdots + \frac{y_{T-1}}{T^2}\right]$$

$$= \;\; T^{-3/2}\sum_{t=1}^{T} y_{t-1}$$

By functional central limit theorem and continuous mapping theorem that as $T \to \infty$

$$\int_0^1 \sqrt{T} X_T(r) dr \;\; \xrightarrow{L} \;\; \sigma \int_0^1 W(r) dr$$

which implies that

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} \;\; \xrightarrow{L} \;\; \sigma \int_0^1 W(r) dr \tag{1}$$

We can write

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} \;\; = \;\; T^{-3/2}\left[u_1 + (u_1 + u_2) + \cdots + (u_1 + u_2 + \cdots + u_{T-1})\right]$$

$$= \;\; T^{-3/2}\left[(T-1)u_1 + (T-2)u_2 + \cdots u_{T-1}\right]$$

$$= \;\; T^{-3/2}\sum_{t=1}^{T}(T-t)u_t$$

$$= \;\; T^{-1/2}\sum_{t=1}^{T} u_t - T^{-3/2}\sum_{t=1}^{T} t u_t \tag{2}$$

From (2), we can derive

$$T^{-3/2} \sum_{t=1}^{T} t u_t \quad = \quad T^{-1/2} \sum_{t=1}^{T} u_t - T^{-3/2} \sum_{t=1}^{T} y_{t-1}$$

$$\rightarrow_d \quad \sigma W(1) - \sigma \int_0^1 W(r) dr \tag{3}$$

Theorem 6. *In summary, for random walk without drift* $y_t = y_{t-1} + u_t$ *where* $y_0 = 0$ *and* $\{u_t\}$ *is an i.i.d. sequence with mean zero and variance* σ^2. *Then we have*

$$T^{-1/2} \sum_{t=1}^{T} u_t \quad \xrightarrow{L} \quad \sigma W(1) \tag{4}$$

$$T^{-1} \sum_{t=1}^{T} y_{t-1} u_t \quad \xrightarrow{L} \quad \frac{1}{2} \sigma^2 [W(1)^2 - 1] \tag{5}$$

$$T^{-3/2} \sum_{t=1}^{T} t u_t \quad \xrightarrow{L} \quad \sigma W(1) - \sigma \int_0^1 W(r) dr \tag{6}$$

$$T^{-3/2} \sum_{t=1}^{T} y_{t-1} \quad \xrightarrow{L} \quad \sigma \int_0^1 W(r) dr \tag{7}$$

$$T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \quad \xrightarrow{L} \quad \sigma^2 \int_0^1 W(r)^2 dr \tag{8}$$

$$T^{-5/2} \sum_{t=1}^{T} t y_{t-1} \quad \xrightarrow{L} \quad \sigma \int_0^1 r W(r) dr \tag{9}$$

$$T^{-3} \sum_{t=1}^{T} t y_{t-1}^2 \quad \xrightarrow{L} \quad \sigma^2 \int_0^1 r W(r)^2 dr \tag{10}$$

$$T^{-(v+1)} \sum_{t=1}^{T} t^v \quad \rightarrow \quad \frac{1}{v+1}, \quad v \in \mathbb{N} \tag{11}$$

Proof. (4) is trivial. (6) is shown in (3). (7) is shown in (1). Now we prove the remaining properties.

We define $S_T(r)$ as

$$S_T(r) \quad = \quad \left(\sqrt{T} X_T(r) \right)^2 = T[X_T(r)]^2$$

Then it can be written as a step function

$$
S_T(r) = \begin{cases} 0 & 0 \le r < 1/T \\ \frac{y_1^2}{T} & 1/T \le r < 2/T \\ \frac{y_2^2}{T} & 2/T \le r < 3/T \\ \vdots & \\ \frac{y_{T-1}^2}{T} & (T-1)/T \le r < 1 \\ \frac{y_T^2}{T} & r = 1 \end{cases}
$$

Then we have

$$
\int_0^1 S_T(r)dr = \frac{y_1^2}{T^2} + \frac{y_2^2}{T^2} + \cdots + \frac{y_{T-1}^2}{T^2} = T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{L} \sigma^2 \int_0^1 W(r)^2 dr \qquad (12)
$$

Now we show (8).

For (9), we prove as

$$
T^{-5/2}\sum_{t=1}^{T} t y_{t-1} = T^{-3/2}\sum_{t=1}^{T} \frac{t}{T} y_{t-1} \xrightarrow{L} \sigma \int_0^1 r W(r) dr
$$

for $r = t/T$.

For (10)

$$
T^{-3}\sum_{t=1}^{T} t y_{t-1}^2 = T^{-2}\sum_{t=1}^{T} (t/T) y_{t-1}^2 \xrightarrow{L} \sigma^2 \int_0^1 r W(r)^2 dr
$$

For (5)

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} y_{t-1} u_t &= \frac{1}{2}\frac{1}{T} y_T^2 - \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T} u_t^2 \\
&= \frac{1}{2} S_T(1) - \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T} u_t^2 \\
&\xrightarrow{L} \frac{1}{2}\sigma^2[W(1)]^2 - \frac{1}{2}\sigma^2 \\
&\xrightarrow{L} \frac{1}{2}\sigma^2[W(1)^2 - 1]
\end{aligned}
$$

$\square$

7.5.1   No Constant No Time Trend

Consider OLS estimation of ρ based on an AR(1) regression

$$
y_t = \rho y_{t-1} + u_t \qquad (13)
$$

where u_t is i.i.d. with mean zero and variance σ^2. OLS estimate is

$$\hat{\rho}_T = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

when true value of ρ is 1. Then we have

$$T(\hat{\rho}_T - 1) = \frac{T^{-1} \sum_{t=1}^{T} y_{t-1} u_t}{T^{-2} \sum_{t=1}^{T} y_{t-1}^2}$$

For random walk process $y_t = y_0 + u_1 + u_2 + \cdots + u_t$, apart from the initial term y_0, by (5)

$$T^{-1} \sum_{t=1}^{T} y_{t-1} u_t \xrightarrow{L} \frac{1}{2} \sigma^2 [W(1)^2 - 1]$$

and by (8)

$$T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{L} \sigma^2 \int_0^1 W(r)^2 dr$$

Under the null hypothesis that $\rho = 1$, the OLS estimate is characterized by

$$T(\hat{\rho}_T - 1) = \frac{T^{-1} \sum_{t=1}^{T} y_{t-1} u_t}{T^{-2} \sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{L} \frac{\frac{1}{2}[W(1)^2 - 1]}{\int_0^1 W(r)^2 dr}$$

$[W(1)]^2$ is a $\chi^2(1)$ variable.

7.5.2 Constant No Time Trend

The DGP is still $y_t = y_{t-1} + u_t$. Suppose now that a constant term is included in the AR(1) model that is to be estimated by OLS

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

7.5.3 No Constant No Time Trend

7.6 Dickey-Fuller Tests

7.6.1 Serial Correlation

When ε_t is serially correlated, two approaches are developed. Subtracting y_{t-1} from each side of the equation $y_t = a_1 y_{t-1} + u_t$, we get $\Delta y_t = \gamma y_{t-1} + u_t$, where $\gamma = a_1 - 1$. Testing the hypothesis $a_1 = 1$ is equivalent to testing $\gamma = 0$.

Dickey and Fuller consider three different regression equations

<div align="center">random walk model</div>

$$\Delta y_t \;=\; \gamma y_{t-1} + u_t$$

<div align="center">add a drift</div>

$$\Delta y_t \;=\; a_0 + \gamma y_{t-1} + u_t$$

<div align="center">add a drift and linear time trend</div>

$$\Delta y_t \;=\; a_0 + \gamma y_{t-1} + a_2 t + u_t$$

Run the OLS and get the estimated value of γ and associated standard error of these three models. However, the critical values of the t-statistics do depend on whether a drift and/or time trend is included in **regression models**. Note that the appropriate critical values depend on **sample size**. For any given level of significance, the critical values of the t-statistic decrease as sample size increases.

7.7 Augmented Dicker-Fuller test

Consider the pth-order autoregressive process:

$$y_t \;=\; a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p} + u_t$$

jj Add and subtract $a_p y_{t-p+1}$

$$
\begin{aligned}
y_t \;=\;& a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p+1} + a_p y_{t-p} - a_p y_{t-p+1} + u_t \\
\;=\;& a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} - a_p \Delta y_{t-p+1} + u_t
\end{aligned}
$$

Add and subtract $(a_{p-1} + a_p) y_{t-p+2}$

$$
\begin{aligned}
y_t \;=\;& a_0 + \cdots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} - (a_{p-1} + a_p) y_{t-p+2} - a_p \Delta y_{t-p+1} + u_t \\
\;=\;& a_0 + \cdots + (a_{p-2} + a_{p-1} + a_p) y_{t-p+2} - (a_{p-1} + a_p) \Delta y_{t-p+2} - a_p \Delta y_{t-p+1} + u_t
\end{aligned}
$$

Continuing in this fashion, we get

$$\Delta y_t \;=\; a_0 + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i+1} + u_t$$

$$where \;\; \gamma \;=\; -\left(1 - \sum_{i=1}^{p} a_i\right)$$

$$\beta_i \;=\; \sum_{j=i}^{p} a_j$$

We can use the same Dickey-Fuller statistics which depends on the regression models and sample size.