Applied Econometrics for Macro and Finance

Basic Time Series Modelling

Shiu-Sheng Chen

Department of Economics
National Taiwan University

Spring 2023

Part I

Time Series Basics

Time Series Basics

Definition

A time series (stochastic process) is a collection of random variables indexed by time (i.e., sequentially ordered over time).

$$\{Y_t : \ t \in \mathbb{T}\}$$

where

- $Y_t \in \mathbb{S}$, $\mathbb{S}$ is called the state space, which could be discrete or continuous.
- $\mathbb{T}$ is called the index set, which could also be discrete or continuous

$$\mathbb{T} = \{0, 1, 2, \ldots\} \text{ (discrete)}, \quad \mathbb{T} = [0, \infty) \text{ (continuous)}$$

Time Series Basics

- In macroeconomic time series, it is commonly considered a discrete but infinite index set:

$$\mathbb{T} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

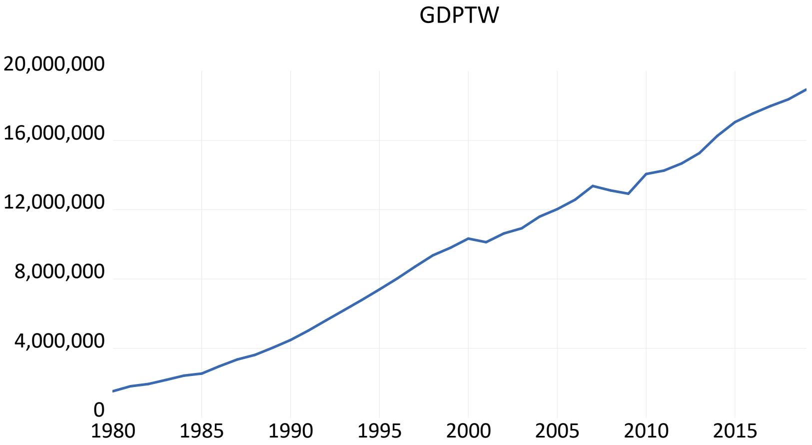
- That is, the stochastic process we consider is

$$\{\ldots, Y_1, Y_2, \ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots\} = \{Y_t\}_{t=-\infty}^{\infty}$$

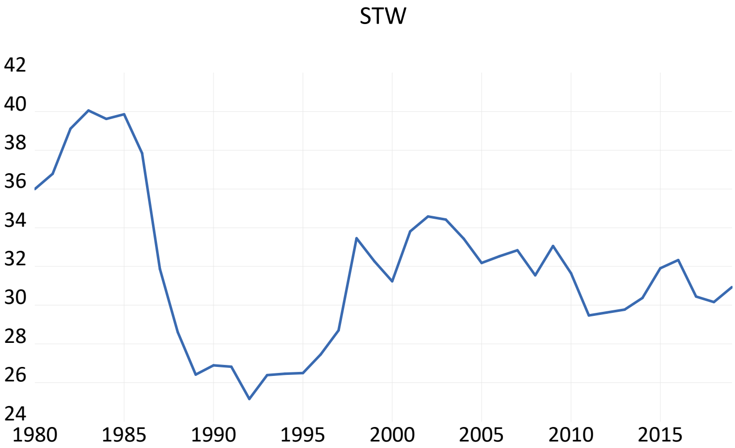
- Observed time series of length T

$$\{Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_t\} = \{y_t\}_{t=1}^{T}$$

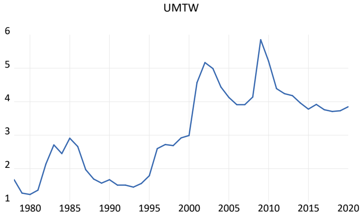
Examples

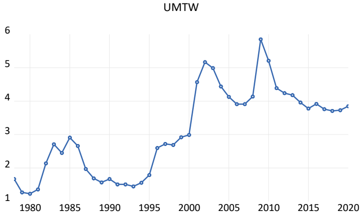

GDPTW
20,000,000
16,000,000
12,000,000
8,000,000
4,000,000
0
1980
1985
1990
1995
2000
2005
2010
2015

Examples

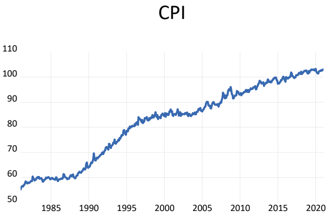

STW
42
40
38
36
34
32
30
28
26
24
1980
1985
1990
1995
2000
2005
2010
2015

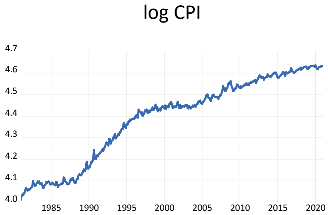
Examples

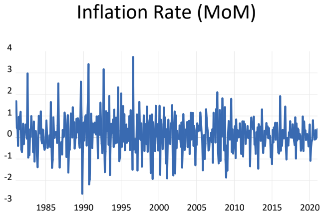

UMTW
6
5
4
3
2
1
1980
1985
1990
1995
2000
2005
2010
2015
2020

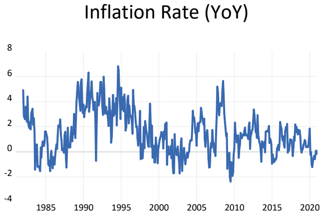

UMTW
6
5
4
3
2
1
1980
1985
1990
1995
2000
2005
2010
2015
2020

Examples: Data Manipulation

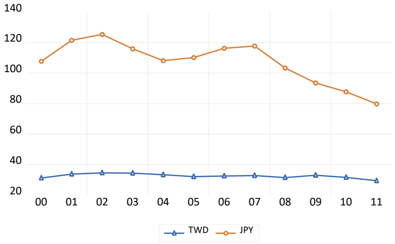


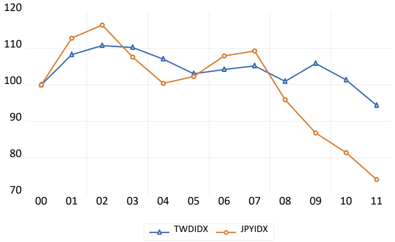






Examples: Index





Time Series

- What makes a time series different?
  - Serially correlated
  - Many of the issues that distinguish time series from cross-section econometrics concern the modeling of the dependence relationship.

Section 2

Moments and Stationarity

Moment Functions

Definition (Mean and Variance)

$$E(Y_t) = \int_{Y_t} z f_{Y_t}(z) dz = \mu_t$$

$$Var(Y_t) = \int_{Y_t} (z - \mu_t)^2 f_{Y_t}(z) dz = \sigma_t^2$$

- Where $f_{Y_t}(z)$ is the density function at time t.

Autocovariance

Definition (Autocovariance)

$$\begin{aligned}\gamma(t,s) = Cov(Y_t, Y_s) &= E[Y_s - E(Y_s)][Y_t - E(Y_t)] \\ &= \int_{Y_t}\int_{Y_s}(z-\mu_t)(x-\mu_s)f_{Y_t,Y_s}(z,x)dxdz\end{aligned}$$

- Where $f_{Y_t,Y_s}(z,x)$ is the joint density function at time t.

Autocorrelation

Definition (Autocorrelation)

$$\rho(t,s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t)Var(Y_s)}}$$

- If $s = t - 1$,
  - $\gamma(t, t-1)$ is the first order autocovariance.
  - $\rho(t, t-1)$ is the first order autocorrelation.
- If $s = t - k$,
  - $\gamma(t, t-k)$ is the k-th order autocovariance.
  - $\rho(t, t-k)$ is the k-th order autocorrelation.

Stationarity

- Loosing speaking, a time series $\{Y_t\}$ is said to be stationary if it has statistical properties similar to those of the time-shifted series $\{Y_{t+h}\}$ for each integer h.
- Two important concepts of stationarity
  - Weak-Stationary (Covariance-Stationary)
  - Strict-Stationary

Weak-Stationary

Definition (Weak-Stationary)

A time series $\{Y_t\}$ is said to be weak-stationary if

(a) $E(Y_t) = \mu < \infty \quad \forall \ t$

(b) $Var(Y_t) < \infty \quad \forall \ t$

(c) $\gamma(t, t-j) = Cov(Y_t, Y_{t-j}) = E\left(Y_t - \mu\right)\left(Y_{t-j} - \mu\right) = \gamma(j) < \infty \ \ \forall t, j$

In other words, a weak-stationary time series has the following features:

- constant mean (independent of t)
- finite variation
- j-th autocovariance function only depends on j and not depends on time, t

Strict-Stationary

Definition (Strict-Stationary)

A time series $\{Y_t\}$ is said to be strict-stationary if for any set of time $(t_1, t_2, \ldots, t_m)$, and for all k,

$$(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})' \overset{d}{=} (Y_{t_1-k}, Y_{t_2-k}, \ldots, Y_{t_m-k})'$$

- In other words, strict stationarity requires that the joint distribution of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})$ is invariant under time shift.
- For instance,

$$(Y_2, Y_8)' \overset{d}{=} (Y_{11}, Y_{17})'$$

Why is Stationarity Important?

- For most series we have only one observation at each point in time. Therefore, if μ_t is not the same every period, we have only one observation to estimate it.
- If all observations share the same parameter $\mu_t = \mu$, we can potentially use $\{Y_t\}_{t=1}^{T}$ to estimate the parameter.
  - That is, use the sample mean $\frac{1}{T}\sum_{t=1}^{T} Y_t$ to estimate μ.
- Remarks:
  - If we label a process as "stationary," you should interpret it as meaning "strictly stationary" in this course.
  - If a sequence is strictly stationary and if the variance and covariances are finite, then the sequence is weakly stationary.
  - Any i.i.d. process is strictly stationary.

Functions of A Strictly Stationarity Process

Theorem

If Y_t is strictly stationary, and

$$X_t = f(Y_t, Y_{t-1}, \ldots)$$

is a well-defined random variable, then X_t is strictly stationary.

Ergodicity

Definition (Ergodicity)

A strictly stationary process $\{Y_t\}$ is ergodic if

$$\gamma(k) \longrightarrow 0 \text{ as } k \longrightarrow \infty.$$

- Note that this is a sufficient condition
- An idea of asymptotic independence or weak dependence
- A strictly stationary and ergodic process is also called ergodic stationary
- Ergodicity is assumed to hold in our lectures (and in fact, in many textbooks).

Functions of Ergodic Stationarity

Theorem

If Y_t is strictly stationary and ergodic, and

$$X_t = f(Y_t, Y_{t-1}, \ldots)$$

is a well-defined random variable, then X_t is strictly stationary and ergodic.

Ergodic Theorem

Theorem

If Y_t is strictly stationary and ergodic, then

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{p} E(Y_t).$$

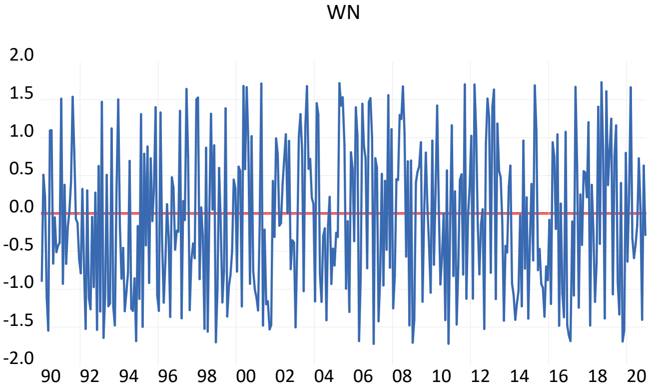
$$\frac{1}{T}\sum_{t=1}^{T} f(Y_t) \xrightarrow{p} E[f(Y_t)]$$

- For instance,

$$\hat{\gamma}(j) = \frac{1}{T-j}\sum_{t=j+1}^{T} (Y_t - \bar{Y})(Y_{t-j} - \bar{Y}) \xrightarrow{p} \gamma(j)$$

Example

- The most simple stationary time series is the independent Gaussian white noise process $Y_t \sim^{i.i.d.} N(0, \sigma^2)$



Two Slightly More General Processes

- Independent white noise

$$Y_t \sim^{i.i.d.} (0, \sigma^2)$$

- White noise

$$Y_t \sim WN(0, \sigma^2)$$

where $E[Y_t Y_{t-s}] = 0 \quad \forall \, s, t$

Part II

Some Useful Linear Processes

Some Useful Time Series Models

- Moving Average (MA) Models
- Autoregressive (AR) Models
- Autoregressive and Moving Average (ARMA) Models

MA(q) Process

- A very important class of covariance-stationary processes, called linear processes, can be created by taking a moving average of a white noise process.
- The moving average process of order q

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}, \quad e_t \sim^{i.i.d.} (0, \sigma^2)$$

- The ACF of a MA(q)
  - The first q-th order autocorrelations of a MA(q) are non-zero, the autocorrelations above q are zero.

Example: MA(1) Process

- MA(1)

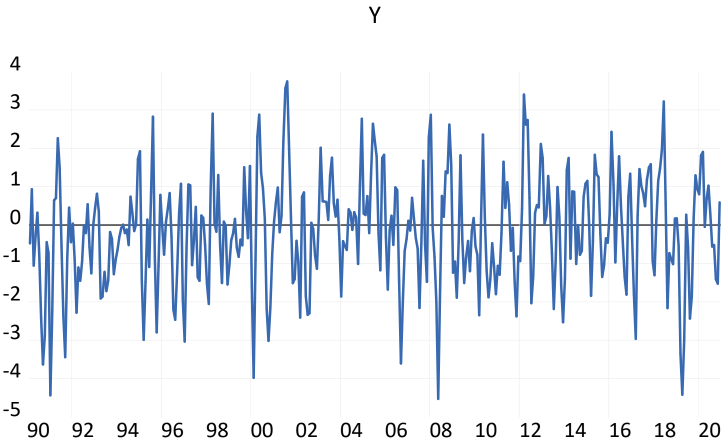
$$Y_t = e_t + \theta_1 e_{t-1}, \quad e_t \sim^{i.i.d.} (0, \sigma^2)$$

- The ACF of MA(1) process

$$\begin{aligned}
\gamma(0) &= (1 + \theta_1^2)\sigma^2 \\
\gamma(1) &= \theta_1 \sigma^2 \\
\gamma(j) &= 0, \quad \text{for} \quad j > 1
\end{aligned}$$

MA(1) Process

- $\theta_1 = 0.95$



Lag Operator

- The presentation of time series models is simplified using lag operator notation:

$$L^j y_t = y_{t-j}$$

- We can write the equation as

$$Y_t = \theta(L) e_t$$

where $\theta(L)$ is a q-th order polynomial in L

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

Infinite-order MA Process: MA(∞)

- Given a MA(∞):

$$Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots$$

  - Note that this infinite sum makes sense only if the partial sum $\sum_{j=0}^{n} \psi_j e_{t-j}$ converges (to a random variable in mean square) as $n \to \infty$.

- A popular condition to make this happen is to assume that $\{\psi_j\}$ is absolutely summable:

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

MA(∞) with Absolutely Summable Coefficients

- Consider a MA(∞):

$$Y_t = \psi(L)e_t,$$

where

$$\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \cdots$$

$$e_t \sim^{i.i.d.} (0, \sigma^2)$$

and

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

- e_t i.i.d. with finite variance $\Longrightarrow$ e_t is stationary
- $\sum_{j=0}^{\infty} |\psi_j| < \infty \Longrightarrow Y_t$ converges $\Longrightarrow Y_t$ well-defined
- Hence, $\{Y_t\}$ is also stationary

MA(∞) with Absolutely Summable Coefficients

- That is, a MA(∞) process with absolutely summable coefficients is stationary
- The above results encompass MA(q) as special case that

$$\psi_j = \begin{cases} \theta_j & j = 0, 1, 2, \ldots, q \\ 0 & j > q \end{cases}$$

- Hence, a finite order MA(q) process is always stationary

Absolutely Summable Inverses of Lag Polynomials

Theorem (Hayashi (2000), P374)

Given p-th degree lag polynomial

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$

If all roots of $\phi(z) = 0$ lie outside the unit root circle, then

(a) *The lag polynomial can be inverted:*

$$\phi(L)^{-1} = \psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots$$

(b) *The inverted coefficients are absolutely summable:*

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

AR(p) model

- Consider an AR(p) model:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + e_t,$$

where

$$e_t \sim^{i.i.d.} (0, \sigma^2)$$

- In lag operator form

$$\beta(L) Y_t = e_t,$$

where

$$\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p$$

Stationary AR(p) Process

- Given an AR(p):

$$\beta(L)Y_t = e_t, \quad e_t \sim^{i.i.d.} (0, \sigma^2)$$

- If the roots of $\beta(z) = 0$ all lie outside the unit circle, then $\{Y_t\}$ has an MA(∞) representation with absolutely summable coefficients:

$$Y_t = \beta(L)^{-1} e_t = \psi(L) e_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots$$

where

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

- It follows that $\{Y_t\}$ is stationary.

Stationary AR(1) Process

- AR(1)

$$Y_t = \beta_1 Y_{t-1} + e_t,$$

- Autocovariance functions

$$\gamma(0) = \frac{\sigma^2}{1-\beta_1^2}$$

Moreover, for $k > 0$,

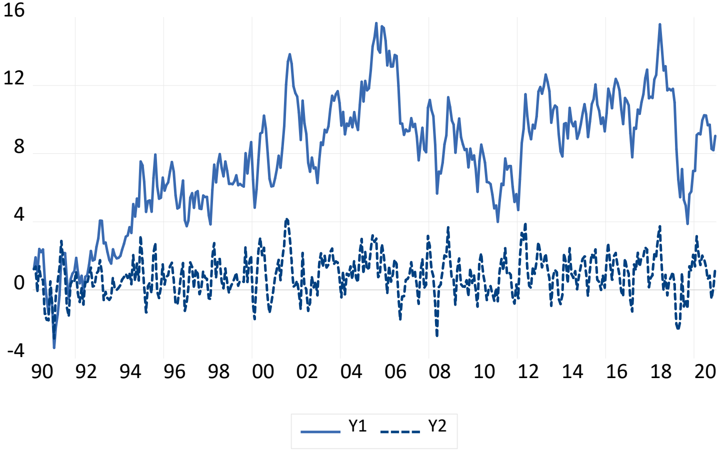
$$\gamma(k) = \beta_1 \gamma(k-1)$$

Hence,

$$\gamma(k) = \frac{\beta_1^k \sigma^2}{1-\beta_1^2}$$

AR(1) Process

- $\beta_1 = 0.95$ vs. $\beta_1 = 0.50$



Impulse Response Functions

- Suppose an AR(1) series starts out at zero. Then there is a unit shock, $\varepsilon_t = 1$ and then all shocks $\varepsilon_{t+1} = \varepsilon_{t+2} = \cdots = 0$ afterwards.

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

  - Period t, we have $Y_t = 1$, period $t+1$, we have $Y_{t+1} = \rho$, period $t+n$, we have $Y_{t+n} = \rho^n$ and so on.
  - The shock fades away gradually. How fast depends on the size of ρ.
  - The time path of Y_t after this hypothetical shock is known as the Impulse Response Function (IRF).

Impulse Response Functions

- Note that

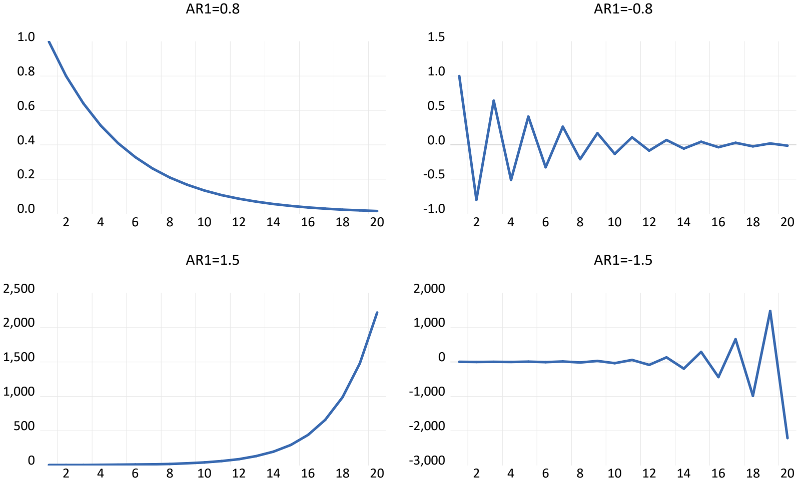
$$Y_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots + \rho^j\varepsilon_{t-j} + \rho^{j+1}Y_{t-j-1}$$

  Hence the IRF is

$$\frac{\partial Y_{t+j}}{\partial \varepsilon_t} = \rho^j$$

- Notice that when the IRF decays, $\{Y_t\}$ exhibits mean-reverting behavior. That is, $\{Y_t\}$ fluctuates about the mean value.

Example: AR(1) Impulse Responses



Impulse Response Functions

- IRF graphs are commonly used to illustrate dynamic properties of time series data.
  - The propagation mechanisms (persistence)
- The IRF decays at a geometric rate. The decay rate of the IRF is sometimes reported as a half-life, the lag j^* at which the IRF reaches 0.5.

$$\rho^{j^*} = 0.5$$

- Hence,

$$j^* = \frac{\log(0.5)}{\log(\rho)}$$

Volatility: Shocks and Propagation Mechanisms

- Consider the AR(1) model

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

- Recall that the variance of Y_t is given by

$$Var(Y_t) = \gamma(0) = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

- That is, the volatility of the series is partly due to size of shocks but also due to the strength of the propagation mechanism.

Impulse Response Functions for an AR(p) Model

- Given an AR(p) model:

$$Y_t = \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \varepsilon_t$$

- In state-space form,

$$\begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_{p-1} & \beta_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

That is,

$$\mathcal{Y}_t = A\mathcal{Y}_{t-1} + \epsilon_t$$

Impulse Response Functions for an AR(p) Model

- By backward iteration,

$$Y_{t+k} = AY_{t+k-1} + \epsilon_{t+k} = \sum_{j=0}^{\infty} A^j \epsilon_{t+k-j}$$

- Thus

$$y_{t+k} = \begin{bmatrix} 1 & 0 \cdots 0 \end{bmatrix} \sum_{j=0}^{\infty} A^j \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \varepsilon_{t+k-j}$$

- For $j = k$,

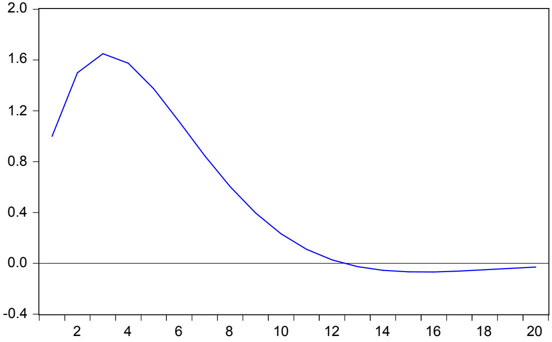
$$\frac{\partial y_{t+k}}{\partial \varepsilon_t} = \begin{bmatrix} 1 & 0 \cdots 0 \end{bmatrix} A^k \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Example: AR(2) Impulse Responses

- Given an AR(2) model

$$Y_t = 1.5Y_{t-1} - 0.6Y_{t-2} + \varepsilon_t$$

- A humped-shape IRF



General Linear Process

Let $\beta(L)$ and $\theta(L)$ denote lag polynomials, an ARMA(p,q) process is

$$\beta(L)Y_t = \theta(L)e_t$$

$$\beta(L) = 1 - \beta_1 L - \cdots - \beta_p L^p$$

$$\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$$

- Y_t is stationary if all roots of $\beta(z) = 0$ lie outside the unit circle
- Y_t is invertible if all roots of $\theta(z) = 0$ lie outside the unit circle
  - The inverse may not be absolutely summable unless the invertibility condition holds

Part III

Estimation of ARMA Models

Estimation of ARMA Models

- For ARMA(p,q) model,

$$Y_t = c + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

  we use conditional MLEs.

- For AR(p) model,

$$Y_t = c + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \varepsilon_t$$

  the conditional MLEs are equivalent to the least squares estimates.

EViews Example

```
wfcreate(wf=C4_1)  m 1990:1 2021:1
read(b2,s=C2_7) TSbookData.xls TWD JPY GBP UKCPI USCPI
genr q_GBP = log(GBP) + log(UKCPI) - log(USCPI)
equation eq_arma1.ls q_GBP q_GBP(-1) ma(1) c
equation eq_arma2.ls q_GBP ar(1) ma(1) c
freeze(rootplot) eq_arma2.arma(type=root)
```

Part IV

Integrated Processes

Long-Run Variance

Definition

For an ergodic stationary process $\{Y_t\}$, the long-run variance is defined by

$$\lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t\right) = \sum_{j=-\infty}^{\infty}\gamma(j) = \gamma(0) + 2\sum_{j=1}^{\infty}\gamma(j)$$

where $\gamma(j) = Cov(Y_t, Y_{t-j})$

- See (6.5.2) in Hayashi, pp. 401.
- Note that by the CLT for an ergodic stationary process,

$$\sqrt{T}(\bar{Y} - \mu) \xrightarrow{d} N\left(0, \sum_{j=-\infty}^{\infty}\gamma(j)\right)$$

Integreted Processes

- $I(0)$ process:
  - Integrated of order zero
  - An $I(0)$ process is a stationary process whose long-run variance is finite and positive.
- $I(d)$ process:
  - A process Y_t is said to be integrated of order d, if

$$\Delta^d Y_t \sim I(0)$$

- Why requiring positive long-run variance?

Integreted Processes

- Consider

$$Y_t = \varepsilon_t - \varepsilon_{t-1}, \quad \varepsilon_t \sim^{i.i.d.} (0, \sigma^2)$$

- We know that

$$\gamma(0) = 2\sigma^2, \ \ \gamma(1) = \gamma(-1) = -\sigma^2, \ \ \gamma(j) = 0 \ \text{ for } \ |j| > 1$$

- The long-run variance of Y_t is

$$\gamma = \sum_{j=-\infty}^{\infty} \gamma(j) = -\sigma^2 + 2\sigma^2 - \sigma^2 = 0$$

- Y_t is not $I(0)$. Without requiring $\gamma > 0$, Y_t would be $I(0)$, which implies an independent white noise ε_t is $I(1) \Rightarrow$ make no sense!
  - Differencing any stationary process will imply that the resulting process has a long-run variance equal to zero.
  - It is called overdifferencing