QF 604 ECONOMETRICS OF FINANCIAL MARKETS

LECTURE 3

Professor Lim Kian Guan

SMU Master of Science in Quantitative Finance 2023-2024 T2

LECTURE OUTLINE

01 Law of Large Numbers

- Theorem [Weak Law of Large Numbers, WLLN] Let $\{X_i\}_{i=1,2,\ldots}$ be a sequence of uncorrelated identical RVs each with mean μ and variance σ^2. Then, for any $\epsilon > 0$,

$$P \downarrow \left(\left| \frac{X_1+X_2+\cdots+X_n}{n} - \mu \right| \geq \epsilon \right) 0, \; as\, n \uparrow \infty$$

Proof: $E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \mu$ and $Var\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{\sigma^2}{n}$

From Chebyshev's inequality $\mathrm{P}(|\mathrm{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$ (see **Proof** in page 124), therefore,

$$P\left(\left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq \epsilon \right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Thus, as $n \uparrow \infty$, for any given $\epsilon > 0$, no matter how small, the probability that the sample mean deviates from μ can be made ever closer to zero.

- The WLLN says that the sample mean in the uncorrelated identical variables case (and obviously also for the special case of i.i.d.) converges in probability to the population mean μ.

01 Law of Large Numbers

- RVs X above do not need to be identical. The same WLLN result can be obtained if the RVs X have constant finite mean and constant finite variance and are uncorrelated (but not exactly identical).
- Stochastic process with constant mean and variance is weakly stationary (whether zero correlation or not)
- The WLLN may hold for weakly stationary process in general, i.e. allowing some non-zero covariance, as long as there is some ergodicity, i.e. correlations disappear in longer term.
- The WLLN gives us some degree of confidence in estimating population parameters such as mean or variance using sample mean or sample variance respectively when the sample size is large and the underlying RV is weakly stationary.

01 Central Limit Theorem

■ **Theorem**:

Suppose $\{X_i\}_{i=1,2,\ldots}$ is a sequence of i.i.d. random variables, each having mean μ and variance σ^2. Then, the $RV = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ converges in distribution to the standard normal RV as n $\uparrow \infty$.

Note that the numerator $\sum_{i=1}^{n} X_i - n\mu$ has zero mean, and the denominator is standard deviation of the numerator.

See **Proof** in pages 131 – 132.

01 Stock Return Rates

- Several empirically observed characteristics of stock price:

  (a) $P_t > 0$, i.e. prices must be strictly positive.

  (b) Return rates derived from stock prices over time are normally distributed when measured over a sufficiently long interval e.g. a month.

  (c) Returns could display a small trend or drift, i.e. increases or decreases over time.

  (d) The ex-ante anticipated variance of return rate increases with the holding period.

- Price P_t is observed at end of the period *t*, e.g., end of the trading day t. Period *t* is (*t* - 1,*t*]. Typically we say P_t is observed at t.

- Return P_{t+1}/P_t in period $t+1$measures the price relative over horizon (*t*, *t* +1]. Discrete holding period return rate ($\frac{P_{t+1}}{P_t} - 1$) or simple return rate ∈[-1,+oo) cannot have a normal distribution.

- $\ln(\frac{P_{t+1}}{P_t}) \equiv \tilde{r}_{t+1}$ is continuously compounded return rate. It has support (-∞, +∞). Can have a normal distribution. Suppose $\tilde{r}_{t+1}$ is normally distributed. Given P_t , $\ln(P_{t+1}) = \ln(P_t) + r_{t+1}$ is normally distributed. Hence P_{t+1} is lognormally distributed. We examine if P_t is lognormally distributed.

01 Stock Return Rates

- Show how r_{t+1} can reasonably be assumed as being normally distributed, over a longer time interval. Consider a short time interval $\Delta = 1/T$, such that $\ln(P_{t+\Delta}/P_t) = r_{t+\Delta}$ has mean $\mu\Delta = \mu/T$, and variance $\sigma^2\Delta = \sigma^2/T$. $\mu \neq 0$ satisfies (c).

  Aggregating the returns,

$$\ln\left(\frac{P_{t+\Delta}}{P_t}\right) + \ln\left(\frac{P_{t+2\Delta}}{P_{t+\Delta}}\right) + \ln\left(\frac{P_{t+3\Delta}}{P_{t+2\Delta}}\right) + \cdots + \ln\left(\frac{P_{t+T\Delta}}{P_{t+(T-1)\Delta}}\right) = \ln\left(\frac{P_{t+T\Delta}}{P_t}\right)$$

- The right-hand side: the continuously compounded return $\ln(P_{t+1}/P_t) \equiv r_{t+1}$ over the longer period $(t, t+1]$ has length is made up of $T = 1/\Delta$ number of periods.

- The left-hand side: invoke the Central Limit Theorem. Assume $r_{t+\Delta}$ are i.i.d. For large T, LHS is $N(T\mu\Delta, T\sigma^2\Delta)$ or $N(\mu, \sigma^2)$ since $T\Delta = 1$.

  Hence, $r_{t+1} \sim N(\mu, \sigma^2)$ which satisfies (b).

- Returns $r_{t+1}, r_{t+2}, r_{t+3}, \ldots, r_{t+k}$ are independent. Then, $Var\left(\sum_{j=1}^{k} r_{t+j}\right) = k\sigma^2$. Thus, ex-ante variance of return increases with holding period $(t, t+k]$. This satisfies characteristic (d).

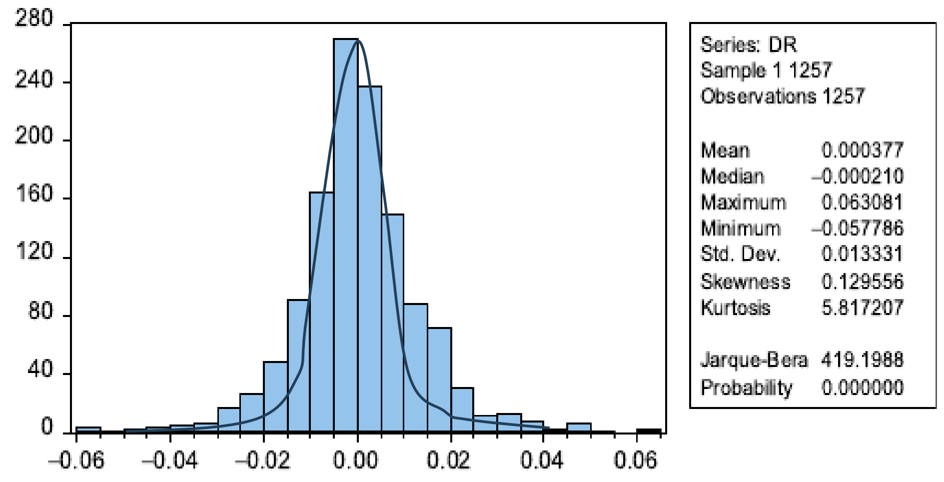
- Hence positive P_t is lognormally distributed.

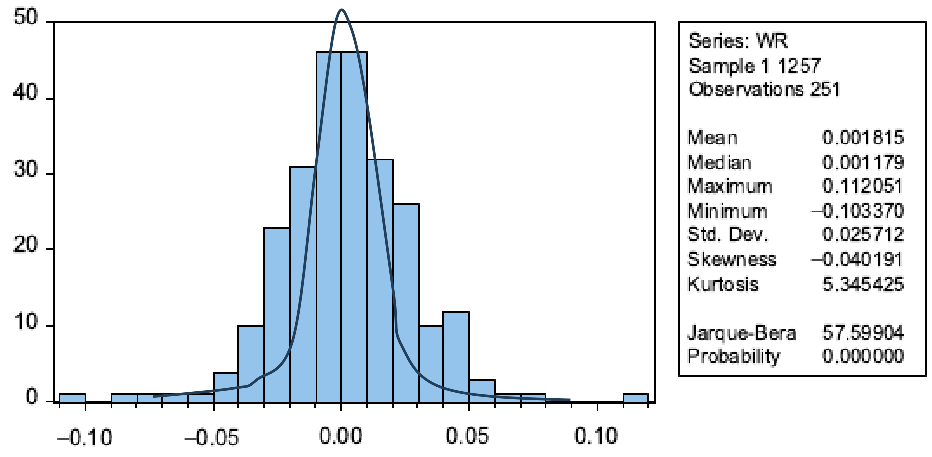
01 Test of Normality

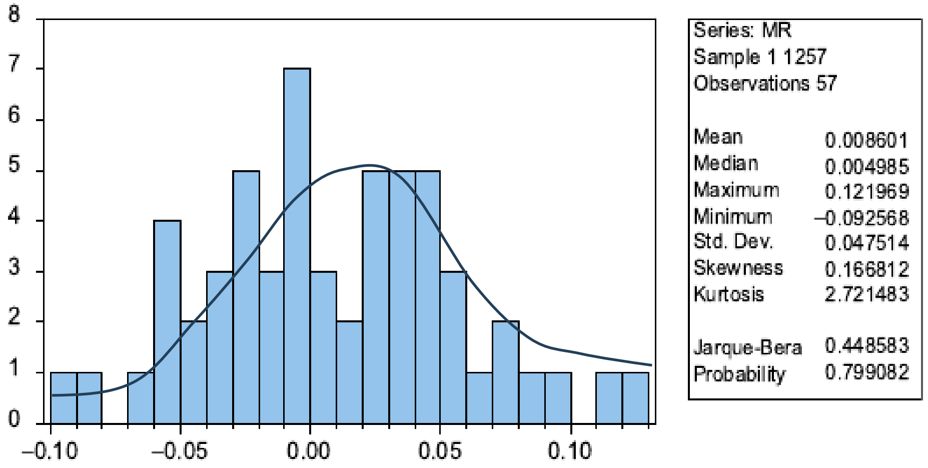
The three series of daily, weekly, and monthly continuously compounded return rates of American Express Company from 1/3/2003 to 12/31/2007 are tabulated in histograms. Descriptive statistics of these distributions: mean, standard deviation, skewness, and kurtosis are reported. The Jarque-Bera (JB) asymptotic test statistic of normality is

$$n\left[(\hat{\gamma}^2/6)+\left((\hat{k}-3)^2/24\right)\right]\sim\chi_2^2$$

where $\hat{\gamma}$ is the sample skewness estimate of $\{r_i\}$and $\hat{k}$ is the sample kurtosis estimate of standardized$\{r_i\}$.







01 Covariance Stationarity

- Time series is a (discrete) sequence of observed or realized data from a sample path of a stochastic process.
  When modeling, treat it as a sequence of r.v.'s. When doing estimation, it is a series of data.
- Deal with stationary time series in order that we can make use of the law of large numbers and central limit theorem for estimation and for testing.
  When a time series is non-stationary, it is common to transform it to a stationary series.
- Two main forms of stationarity of time series: strong form or weak form.
  Strong stationarity: finite-dimensional distribution of the time series of observations does not change with shifts in time.
  Weak stationarity: first and second moments of the stochastic process do not change with shifts in time.
- Important example of a strongly stationary process is the i.i.d. process. Strong stationarity implies weak stationarity if first and second moments are finite. An example of an exception is i.i.d. Cauchy distributed r.v.'s that have no finite first and second moments.

01 Covariance Stationarity

- Weakly stationary process $\{x_t : t \in \mathbb{Z}\}$ is also called a covariance-stationary process. In this process, $E(x_t) = \mu, \forall t,\ Var(x_t) = \sigma^2, \forall t$, and $Cov(x_t, x_s) = Cov(x_{t+h}, x_{s+h}), \forall h > 0$
- A basic building block of stationary processes is a white noise. A white noise process $\{u_t\}$ is one where each u_t has zero mean, constant variance $\sigma_u^2 < \infty$, and in addition has zero serial correlation (weak-form white noise). Special case of weakly or covariance-stationary process (constant mean, variance, and covariance under linear shift).
- Stronger version of white noise: $u_t \sim$ i.i.d.
- The Wold Theorem states that any covariance-stationary process can be constructed from (weak-form) white noises.

01 Covariance Stationarity

Theorem: (Wold Decomposition Theorem)

Any covariance-stationary time series $\{Y_t: t \in \mathbb{Z}\}$ can be represented in the form

$$Y_t = \sum_{j=0}^{\infty} \psi_j u_{t-j} + \eta_t$$

where $\psi_0 = 1, \sum_{j=1}^{\infty} \psi_j^2 < \infty$, u_t is white noise and has zero correlation with Y_{t-j} (for $j > 0$), and η_t is deterministic.

See **Proof** in page 140-141.

02 Time Series Models-Lag Operators

- Use of lag (or backward-shift) and forward-shift operators are convenient in time series models.
  A lag operator L is an operation on a time series that shifts the entire series one step back in time.
  For example, $LY_t = Y_{t-1}$
- A Lag operator has properties similar to multiplication:
  Commutative: $L(aY_t) = aL(Y_t)$
  Distributive over addition: $L(Y_t + X_t) = LY_t + LX_t$
  Operating on constant c: $Lc = c$
  Operator exists in the limit: $\lim_{j\to\infty}\left(1 + \phi L + \phi^2 L^2 + \cdots + \phi^j L^j\right) = (1 - \phi L)^{-1}$
- For any covariance-stationary zero mean time series, Wold Decomposition representation
  $x_t = \sum_{j=0}^{\infty} \psi_j u_{t-j} = \psi(L) u_t$ where $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$,
  has infinite number of parameters ψ_j's to be estimated.
  Not feasible for statistical work with finite sample. Principle of parsimony advocates using time series models that have fewer parameters.
- Wold Decomposition can be approximated by the ratio of two finite-lag polynomials: $\psi(L) \approx \frac{\theta(L)}{\phi(L)}$

02 Time Series Models-ARMA(p,q)

- Suppose a covariance-stationary process is modeled by

$$Y_t = \frac{\theta(L)}{\phi(L)} u_t \quad \text{or} \quad \phi(L) Y_t = \theta(L) u_t$$

where u is white noise ~ $(0, \sigma_u^2)$ or $WN(0, \sigma_u^2)$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \quad \text{and} \quad \theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

- Here process $\{Y_t\}$ is autoregressive with order p. u_t is moving average noise with order q. Jointly $\{Y_t\}$ is ARMA(p, q)
- Ensure that $\{Y_t\}$ process involving arbitrary $\theta(L)$ and $\phi(L)$ must be stationary
- $\theta(L)u_t$ on RHS which is a MA(q) process is covariance-stationary if q is finite since $E[\theta(L)u_t] = 0$ and $\text{Var}[\theta(L)u_t] = \sigma_u^2\left(1 + \sum_{j=1}^{q} \theta_j^2\right)$

02 Time Series Models-ARMA(p,q)

- For $Y_t = \frac{\theta(L)}{\phi(L)} u_t$, we have to check that the LHS $\phi(L)Y_t$ is indeed covariance-stationary. This is also a check if any AR(p) process, for finite p, is covariance-stationary.
- Let $\phi(L)Y_t = v_t$ where $v_t \sim WN(0, \sigma_v^2)$ $\left(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p\right)Y_t = v_t$
- Replacing the L operator with variable z, $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ is called the characteristic equation. The p number of solution z^* are called the characteristic roots.
- If any root modulus $|z| = 1$ (we use $|z|$ since some roots may be complex), the process is non-stationary.

For example, for AR(1), if $1 - \phi_1 z = 0$ has a solution $z = 1$, then $\phi_1 = 1/z = 1 \Rightarrow$ process is $Y_t = Y_{t-1} + v_t \Rightarrow$ var(Y_t) increases with t, and Y_t is non-stationary.
But if $|z| > 1$ (i.e. characteristic root is outside the unit circle), then $\phi_1 < 1 \Rightarrow$ $Y_t = \phi_1 Y_{t-1} + v_t$ and $Var(Y_t) = \sigma_v^2/(1 - \phi_1^2) < \infty$. Hence Y_t is covariance-stationary.

02 Economic and Financial Modeling

In economic and financial modelling, typically a low order Y_t ~ ARMA(p,q) process is used.

- Autoregressive order one, AR(1), process:

$$Y_t = \theta + \lambda Y_{t-1} + u_t\ , \qquad \lambda \neq 0$$

where Y_t, depends on or autoregresses on its lag value, Y_{t-1}.

- Moving average order one, MA(1), process:

$$Y_t = \theta + u_t + \alpha u_{t-1}, \qquad \alpha \neq 0$$

where the residual is made of a moving average of two white noises u_t and u_{t-1}.

- Autoregressive Moving Average order, ARMA(1,1), process:

$$Y_t = \theta + \lambda Y_{t-1} + u_t + \alpha u_{t-1}, \qquad \lambda \neq 0, \alpha \neq 0$$

where Y_t autoregresses on its first lag and the residual is also a moving average.

02 Time Series Models-Autoregressive Process

AR(1) process:

$$Y_t = \theta + \lambda Y_{t-1} + u_t \,, \qquad (\lambda \neq 0), \;\; t = -T, \dots, T, \qquad (5.5)$$

where $\{u_t\}$ is i.i.d. with zero mean. $E(u_t) = 0$ and $Var(u_t) = \sigma_u^2$ for every t.

Since the process holds for $t = -T, \dots, T$, the process is equivalent to a system of equations as follows.

$$\begin{gathered} Y_T = \theta + \lambda Y_{T-1} + u_T \\ Y_{T-1} = \theta + \lambda Y_{T-2} + u_{T-1} \\ \vdots \\ Y_1 = \theta + \lambda Y_0 + u_1 \\ \vdots \\ Y_{-T+1} = \theta + \lambda Y_{-T} + u_{-T+1} \end{gathered}$$

These equations are stochastic, not deterministic, as each equation contains a random variable u_t that is not observable.

02 Time Series Models-Autoregressive Process

By repeated substitution for Y_t in this AR(1) process, Eq.(5.5) implies

$$Y_t = \theta + \lambda(\theta + \lambda Y_{t-2} + u_{t-1}) + u_t$$

$$\text{or } Y_t = (1+\lambda)\theta + \lambda^2 Y_{t-2} + (u_t + \lambda u_{t-1})$$

$$= \cdots\cdots$$

$$= (1 + \lambda + \lambda^2 + \lambda^3 + \cdots)\theta + (u_t + \lambda u_{t-1} + \lambda^2 u_{t-2} + \cdots)$$

For each t,

$$E(Y_t) = (1 + \lambda + \lambda^2 + \lambda^3 + \cdots)\theta = \frac{\theta}{1-\lambda} \text{ provided } |\lambda| < 1.$$

$$Var(Y_t) = Var\left(u_t + \lambda u_{t-1} + \cdots + \lambda^k u_{t-k} + \cdots\right)$$

$$= \sigma_u^2(1 + \lambda^2 + \lambda^4 + \cdots)$$

$$= \frac{\sigma_u^2}{1-\lambda^2} \text{ provided } |\lambda| < 1$$

Otherwise, if $|\lambda| \geq 1$, finite mean, variance do not exist

02 Time Series Models-Autoregressive Process

Autocovariance of Y_t and Y_{t-1},

$$Cov(Y_t, Y_{t-1}) = Cov(\theta + \lambda Y_{t-1} + u_t, Y_{t-1})$$
$$= \lambda \frac{\sigma_u^2}{1-\lambda^2} = \gamma(1)$$

Recall: $Var(Y_t) = \gamma(0) = \frac{\sigma_u^2}{1-\lambda^2}$

$$Corr(Y_t, Y_{t-1}) = \frac{\gamma(1)}{\gamma(0)} = \rho(1) = \lambda$$

Also, autocorrelation coefficient lag k, $Corr(Y_t, Y_{t-k}) = \lambda^k \equiv \rho(k)$.

AR(1) Y_t process is covariance-stationary with constant mean = $\frac{\theta}{1-\lambda}$, constant variance = $\frac{\sigma_u^2}{1-\lambda^2}$, and autocorrelation lag k, $\rho(k) = \lambda^{|k|}$, a function of k only, provided $|\lambda| < 1$.

02 Numerical Example

$$Y_t = 2.5 + 0.5Y_{t-1} + u_t \qquad (5.6)$$

Assume Y_{t-1} and u_t are stationary normally distributed, not correlated. Given $E(u_t) = 0$, and $Var(u_t) = 3$. Stationarity $\Rightarrow E(Y_t) = E(Y_{t-1}) = u_Y$, and $Var(Y_t) = Var(Y_{t-1}) = \sigma_Y^2$.

- Take unconditional expectation on Eq.(5.6),

$$E(Y_t) = 2.5 + 0.5E(Y_{t-1}) + E(u_t)$$

$$u_Y = 2.5 + 0.5\, u_Y$$

then $u_Y = \frac{2.5}{(1-0.5)} = 5.$

- Take unconditional variance on Eq.(5.6),

$$Var(Y_t) = 0.5^2\, Var(Y_{t-1}) + Var(u_t)$$

$$\sigma_Y^2 = 0.25\, \sigma_Y^2 + 3$$

then $\sigma_Y^2 = \frac{3}{1-0.25} = 4.$

02 Numerical Example

- The first-order autocovariance (or autocovariance at lag 1)

$$Cov(Y_t, Y_{t-1}) = Cov(2.5 + 0.5Y_{t-1} + u_t, Y_{t-1})$$
$$= 0.5\, Cov(Y_{t-1}, Y_{t-1}) = 0.5 \times 4 = 2$$

Since Y_t is stationary, $Cov(Y_{t+k}, Y_{t+k+1}) = Cov(Y_{t+k}, Y_{t+k-1}) = \gamma(1) = 2$ for any k.

- First-order autocorrelation (autocorrelation at lag 1):

$$Corr(Y_{t+k}, Y_{t+k+1}) = Corr(Y_{t+k}, Y_{t+k-1}) = \rho(1) = \frac{\gamma(1)}{\sigma_Y^2} = 0.5$$

- Second-order and higher order autocovariance:

$$Cov\left(Y_t, Y_{t-j}\right) = Cov\left(2.5 + 0.5Y_{t-1} + u_t, Y_{t-j}\right) = \gamma(j) \neq 0 \;\; for\; j > 1$$

02 Time Series Models-Moving Average Process

MA(1) process: $$Y_t = \theta + u_t + \alpha u_{t-1} \quad (\alpha \neq 0), t = -T, \dots, T \tag{5.7}$$

where $\{u_t\}$ is i.i.d. with zero mean, and variance σ_u^2

$$E(Y_t) = \theta$$
$$Var(Y_t) = \sigma_u^2(1 + \alpha^2)$$
$$Cov(Y_t, Y_{t-1}) = \alpha\sigma_u^2$$
$$Corr(Y_t, Y_{t-1}) = \frac{\alpha}{1 + \alpha^2}$$
$$Corr(Y_t, Y_{t-k}) = 0 \quad for\ k > 1$$

MA(1) Y_t is covariance-stationary with constant mean=θ, constant variance = $\sigma_u^2(1 + \alpha^2)$, and autocorrelation lag k, a function of k only:

$$\rho(k) = \begin{cases} \frac{\alpha}{1+\alpha^2}, & k = 1 \\ 0, & k > 1 \end{cases}$$

02 Time Series Models-ARMA (1,1) Process

ARMA(1,1) process:

$$Y_t = \theta + \lambda Y_{t-1} + u_t + \alpha u_{t-1}, (\lambda \neq 0, \alpha \neq 0\ ), t = -T, \dots, T,$$

where $\{u_t\}$ is i.i.d. with zero mean, and variance σ_u^2.

Then, $Y_t = \theta + \lambda(\theta + \lambda Y_{t-2} + u_{t-1} + \alpha u_{t-2}) + u_t + \alpha u_{t-1}$

$$= (1+\lambda)\theta + \lambda^2 Y_{t-2} + (u_t + \lambda u_{t-1}) + \alpha(u_{t-1} + \lambda u_{t-2})$$

$$= (1 + \lambda + \lambda^2 + \lambda^3 + \cdots)\theta + u_t + (\lambda + \alpha)u_{t-1} + (\lambda+\alpha)\lambda u_{t-2} + (\lambda+\alpha)\lambda^2 u_{t-3} + \cdots$$

02 Time Series Models-ARMA (1,1) Process

For each t, $E(Y_t) = \frac{\theta}{1-\lambda}$ provided $|\lambda| < 1$

$$Var(Y_t) = \sigma_u^2 \left[1 + \frac{(\lambda+\alpha)^2}{1-\lambda^2}\right] \text{ provided } |\lambda| < 1$$

$$Cov(Y_t, Y_{t-1}) = \lambda Var(Y_{t-1}) + \alpha\sigma_u^2$$

$$Cov(Y_t, Y_{t-k}) = \lambda^k Var(Y_{t-k}) + \alpha\lambda^{k-1}\sigma_u^2 \quad \text{for } k \geq 1$$

ARMA(1,1) Y_t is covariance-stationary with constant mean = $\frac{\theta}{1-\lambda}$, constant variance = $\sigma_u^2\left[1 + \frac{(\lambda+\alpha)^2}{1-\lambda^2}\right] = \sigma_y^2$, and autocovariance lag k, a function of k only, provided $|\lambda| < 1$.

02 Changing Conditional Means

Consider the AR(1) process: $Y_{t+1} = \theta + \lambda Y_t + u_{t+1} (\lambda \neq 0)$

- At *t*, information Y_t is already known, so: $E(Y_{t+1}|Y_t) = \theta + \lambda Y_t + E(u_{t+1}|Y_t)$. Conditional mean at *t* is $\theta + \lambda Y_t$. It changes with time.
- Different from the constant unconditional mean $\frac{\theta}{1-\lambda}$.
- Conditional variance at *t* is $Var(Y_{t+1}|Y_t) = Var(u_{t+1}|Y_t) = \sigma_u^2$, which is a constant no matter what Y_t is.
- This conditional variance, however, is smaller than the unconditional variance $\frac{\sigma_u^2}{1-\lambda^2}$ $(|\lambda| < 1)$. This is because the unconditional variance includes variance $\lambda^2 \sigma_y^2$ on the regressor since Y_t is not assumed to be known yet.

Consider an MA(1) process: $Y_{t+1} = \theta + u_{t+1} + \alpha u_t \ (\alpha \neq 0)$

- Conditional mean at *t* is $E(Y_{t+1}|u_t) = \theta + \alpha u_t \neq \theta$
- Conditional variance at *t* is $Var(Y_{t+1}|u_t) = \sigma_u^2 < \sigma_u^2(1 + \alpha^2)$
- Likewise for MA(1) covariance stationary process, conditional mean changes at each *t*, but conditional variance is constant and is smaller than the unconditional variance.

03 Sample Autocorrelation Function

Given a time series and knowing it is from a stationary process

- Next step is to identify the statistical time series model or its generating process
- AR and MA models produce different autocorrelation function (ACF) $\rho(k)$, $k > 0$
- We can find the sample autocorrelation function $r(k)$ and use this to try to distinguish between an AR or MA

Given a sample $\{Y_t\}$, $t = 1,2,3,\ldots,T$

- Sample autocovariance lag k is: $c(k) = \frac{1}{T}\sum_{t=1}^{T-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})$ for $k = 0,1,2,3,\ldots, p$. As a rule of thumb, $p < T/4$ given a sample size T.
- $c(k)$ is consistent, $\lim_{T\uparrow\infty} c(k) = \gamma(k)$.
- Sample autocorrelation lag k is: $r(k) = \frac{c(k)}{c(0)} = \frac{\sum_{t=1}^{T-k}(Y_t-\bar{Y})(Y_{t+k}-\bar{Y})}{\sum_{t=1}^{T}(Y_t-\bar{Y})^2}$ for $k = 0,1,2,3,\ldots, p$.
- $r(k)$ is consistent, $\lim_{T\uparrow\infty} r(k) = \rho(k)$.

03 Sample Autocorrelation Function

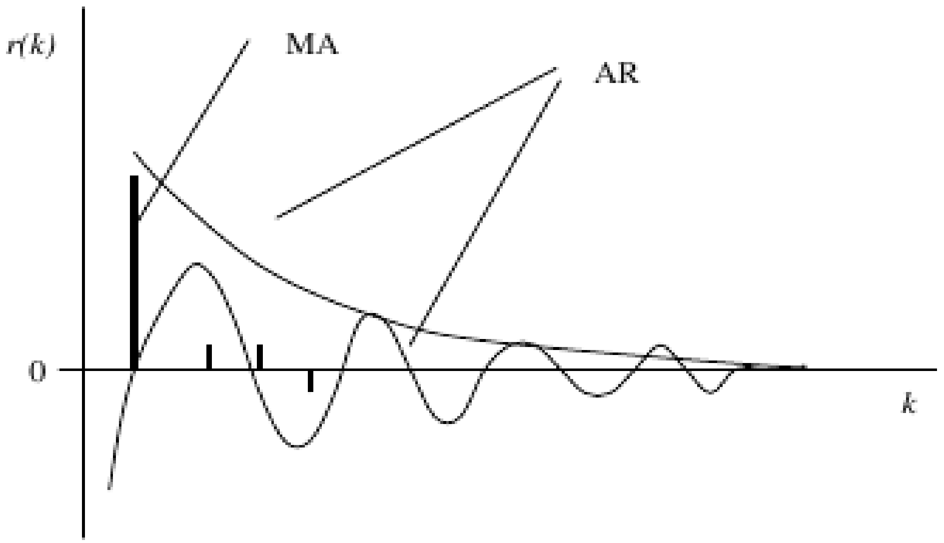


Figure 5.4: Sample Autocorrelation Functions of AR and MA Processes

Based on the AR(1) $Y_t = \theta + \lambda Y_{t-1} + u_t$ $(\lambda \neq 0)$ and the sample autocorrelation measure

$$r(k) = \frac{c(k)}{c(0)} = \frac{\sum_{t=1}^{T-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{T}(Y_t - \bar{Y})^2}$$

$\text{Var}(r(k)) \approx \frac{1}{T}\left[\frac{(1+\lambda^2)(1-\lambda^{2k})}{1-\lambda^2} - 2k\lambda^{2k}\right]$. For k=1, $\text{Var}(r(1)) \approx \frac{1}{T}[1-\lambda^2]$

03 Test of Zero Correlations for AR(1)

- For the AR(1) process, autocorrelation coefficient lag k, $\rho(k) = \lambda^k$, provided $|\lambda| < 1$.
- Suppose we test the null hypothesis $H_0: \rho(k) = 0$ for all $k > 0$. This is essentially a test of the null hypothesis $H_0: \lambda = 0$. $\mathrm{Var}(r(k)) \approx \frac{1}{T}$ for all $k > 0$.
- Under H_0, Eq.(5.5) becomes $Y_t = \theta + u_t$, where $\{u_t\}$ is i.i.d. Hence, $\{Y_t\}$ is i.i.d.

  Asymptotically as sample size T increases to $+\infty$,

$$\begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(m) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{bmatrix} \frac{1}{T} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{T} \end{bmatrix} \right)$$

- As T becomes large, the MVN distribution is approached. To test that the j^{th} autocorrelation is zero, evaluate $z_j \sim N(0, 1)$ statistic as follows: $z_j = \frac{r(j)-0}{\sqrt{1/T}}$. Reject $H_0: \rho(j) = 0$ at 95% significance level if the $|z|$ value exceeds 1.96 for a two-tailed test.

03 Identification in Box and Jenkins Approach

- AR processes have sample ACF $r(k)$ that decay to zero slowly. MA(q) processes have sample ACF $r(k)$ that are zero for $k > q$.
- The autocorrelogram (graph of $r(k)$) generated by simulated AR(3) and MA(1) shows that we reject $H_0: \rho(k) = 0$ for all $k > 1$ at the 95% significance level for the AR(3) process.
- For MA(q) processes where $\rho(k) = 0$ for $k > q$, $\mathrm{Var}(r(k)) \approx \frac{1}{T}\left[1 + 2\sum_{j=1}^{q} \rho(j)^2\right]$.
- For MA(1), Var(r(1)) may be estimated by $\frac{1}{T}[1 + 2r(1)^2] > 1/T$. In the autocorrelogram, MA(1) is identified.
- The standard error used in most statistical programs is $1/\sqrt{T}$. This standard error is reasonably accurate for AR(1) and for MA(q) processes when q is small.

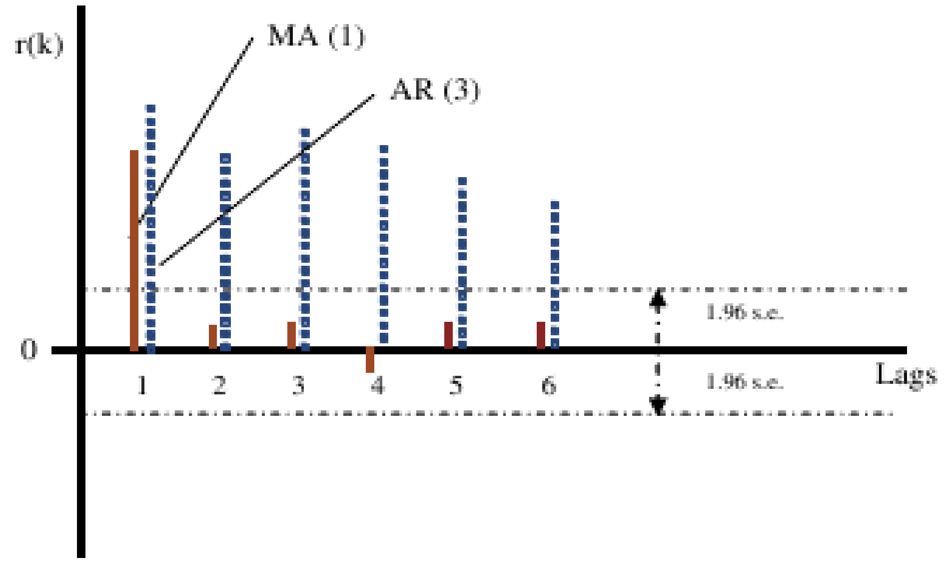


Figure 5.5: Identification Using Sample Autocorrelogram

Ref: George Box and G.M. Jenkins, 1976, “Time Series Analysis: Forcasting and Control” Prentice-Hall.

03 Ljung-Box Test

To test if all autocorrelations are simultaneously zero, $H_0: \rho(1) = \rho(2) = \cdots = \rho(m) = 0$ (it is common practice to set m at 6, 12, 18, provided $m < T/4$ as a rule of thumb), we can apply the **Box and Pierce (1970) Q-statistic**:

$$\hat{Q}_m = T\sum_{k=1}^{m}[r(k)]^2 = \sum_{k=1}^{m}\left[\sqrt{T}r(k)\right]^2 = \sum_{k=1}^{m} z_k^2 \sim \chi_m^2$$

This is an asymptotic test statistic. The **Ljung and Box (1978) test statistic** provides for approximate finite sample correction to the above asymptotic test statistic:

$$\hat{Q}'_m = T(T+2)\sum_{k=1}^{m}\frac{[r(k)]^2}{T-k} \sim \chi_m^2$$

This Ljung-Box test is appropriate in situations when the null hypothesis is a white noise or approximately white noise such as stock return rate.

03 MA Process as Infinite Order AR Process

MA processes can sometimes be represented by infinite order AR processes.
As an example, consider MA(1):

$$Y_t = \theta + u_t + \alpha u_{t-1} \ (\alpha \neq 0), \text{ or } \ Y_t = \theta + (1 + \alpha B) u_t$$

So, $(1+\alpha B)^{-1} Y_t = (1+\alpha B)^{-1}\theta + u_t$.

Note that $(1+x)^{-1} = 1 - x + x^2 - x^3 + x^4 - \cdots$ for $|x| < 1$. Let constant $c = (1+\alpha B)^{-1}\theta = \theta/(1+\alpha)$. Then, $(1 - \alpha B + \alpha^2 B^2 - \alpha^3 B^3 + \alpha^4 B^4 - \cdots) Y_t = c + u_t$,

$$Y_t - \alpha Y_{t-1} + \alpha^2 Y_{t-2} - \alpha^3 Y_{t-3} + \alpha^4 Y_{t-4} - \cdots = c + u_t.$$

Thus, $Y_t = c + \alpha Y_{t-1} - \alpha^2 Y_{t-2} + \alpha^3 Y_{t-3} - \alpha^4 Y_{t-4} + \cdots + u_t$, which is an infinite order AR process.

This AR(∞) process is not a proper representation that allows infinitely past numbers of Y_{t-k} 's to forecast a **finite $\boldsymbol{Y_t}$ unless it is stationary with finite mean and variance**. If it is not stationary, Y_t may increase by too much, based on an infinite number of explanations of past Y_{t-k}'s.

03 Invertible MA Process

- In the MA(1): $Y_t = \theta + u_t + \alpha u_{t-1}$ $(\alpha \neq 0)$ represented by AR(∞):
  $Y_t = c + \alpha Y_{t-1} - \alpha^2 Y_{t-2} + \alpha^3 Y_{t-3} - \alpha^4 Y_{t-4} + \cdots + u_t$
  This covariance-stationary provided $|\alpha| < 1$.

- If a stationary MA(q) process can be equivalently represented as a stationary AR(∞) process, i.e., convergent, then the MA(q) process is said to be **invertible**.

- Invertibility of an MA(q) process to stationary AR(∞) allows expression of current Y_t and future Y_{t+k} in terms of past Y_{t-k}, $k > 0$. This could facilitate interpretations of past impact and forecasts respectively.

- Although all finite order MA(q) processes are stationary, not all are invertible. For example, $Y_t = u_t - 0.3u_{t-1}$ is invertible, but $Y_t = u_t - 1.3u_{t-1}$ is not.

- MA(1): $Y_t = \theta + u_t + \alpha u_{t-1}$ $(\alpha \neq 0)$ and $Y_t = \theta + u_t + \frac{1}{\alpha} u_{t-1}$ $(\alpha \neq 0)$ both are stationary and have same lag-one autocorrelation $\rho(1) = \frac{\alpha}{1+\alpha^2}$, and $\rho(k) = 0$ for $k > 1$, only the MA(1) with $\alpha < 1$ is invertible.

03 Correlations of AR(p) Processes

Stationary AP(p), $u_t \sim$ i.i.d.

$$Y_t = \theta + \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + \cdots + \lambda_p Y_{t-p} + u_t$$

Taking expectations: $\mu = \theta + \lambda_1 \mu + \lambda_2 \mu + \cdots + \lambda_p \mu$ where $\text{E}(Y_t) = \mu$

Then $Y_t - \mu = \lambda_1 (Y_{t-1} - \mu) + \lambda_2 (Y_{t-2} - \mu) + \cdots + \lambda_p (Y_{t-p} - \mu) + u_t$

$$Cov(Y_{t-1} - \mu, Y_t - \mu) = \lambda_1 Var(Y_{t-1}) + \lambda_2 Cov(Y_{t-1} - \mu, Y_{t-2} - \mu) + \lambda_3 Cov(Y_{t-1} - \mu, Y_{t-3} - \mu)$$
$$+ \cdots + \lambda_p Cov(Y_{t-1} - \mu, Y_{t-p} - \mu)$$

Or, $\gamma(1) = \lambda_1 \gamma(0) + \lambda_2 \gamma(1) + \lambda_3 \gamma(2) + \cdots + \lambda_p \gamma(p-1)$

Dividing by $\gamma(0)$: $\rho(1) = \lambda_1 + \lambda_2 \rho(1) + \lambda_3 \rho(2) + \cdots + \lambda_p \rho(p-1)$

03 Yule-Walker Equations

In general,

$$\rho(1) = \lambda_1\rho(0) + \lambda_2\rho(1) + \lambda_3\rho(2) + \lambda_4\rho(3) + \cdots + \lambda_{p-1}\rho(p-2) + \lambda_p\rho(p-1)$$

$$\rho(2) = \lambda_1\rho(1) + \lambda_2\rho(0) + \lambda_3\rho(1) + \lambda_4\rho(2) + \cdots + \lambda_{p-1}\rho(p-3) + \lambda_p\rho(p-2)$$

$$\rho(3) = \lambda_1\rho(2) + \lambda_2\rho(1) + \lambda_3\rho(0) + \lambda_4\rho(1) + \cdots + \lambda_{p-1}\rho(p-4) + \lambda_p\rho(p-3)$$

$$\vdots$$

$$\rho(p) = \lambda_1\rho(p-1) + \lambda_2\rho(p-2) + \lambda_3\rho(p-3) + \cdots + \lambda_{p-1}\rho(1) + \lambda_p\rho(0)$$

where $\rho(0) \equiv 1$.

These equations derived from AR(p) are called the Yule-Walker equations.

03 Yule-Walker Equations

If we replace the $\rho(k)$ by sample $r(k)$ as approximates, then the p Yule-Walker equations can be solved as follows for the parameters estimates $\hat{\lambda}_k$'s.

$$R = \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{pmatrix} \qquad \Phi = \begin{bmatrix} 1 & & r(p-1) \\ r(1) & \cdots & r(p-2) \\ r(2) & & r(p-3) \\ \vdots & \ddots & \vdots \\ r(p-1) & \cdots & 1 \end{bmatrix} \qquad \Lambda = \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \vdots \\ \hat{\lambda}_p \end{pmatrix}$$

And $R = \Phi\Lambda$. Therefore, $\Lambda = \Phi^{-1}R$.

03 Yule-Walker Equations

The other parameters can be estimated as follows.

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} Y_t , \qquad \hat{\theta} = \hat{\mu}(1 - \hat{\lambda}_1 - \hat{\lambda}_2 - \cdots - \hat{\lambda}_p)$$

The estimate of the variance of u_t can be obtained from

$$\frac{1}{T-p-1}\sum_{t=1}^{T}\left(Y_t - \hat{\theta} - \hat{\lambda}_1 Y_{t-1} - \cdots - \hat{\lambda}_p Y_{t-p}\right)^2$$

It is important to note that the identification of the appropriate process must be done before estimation of the parameters is possible since the approach to the estimation requires knowledge of the relevant process.

03 Yule-Walker Equations

Important to note: $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p$ exist deterministically provided the underlying process is AR(p).

- Suppose true process is AR(1): $Y_t - \mu = \lambda_1(Y_{t-1} - \mu) + u_t$

$$Cov(Y_{t-1} - \mu, Y_t - \mu) = \lambda_1 Var(Y_{t-1}) \Rightarrow \rho(1) = \lambda_1\rho(0) = \lambda_1$$

$$Cov(Y_{t-2} - \mu, Y_t - \mu) = \lambda_1 Cov(Y_{t-2} - \mu, Y_{t-1} - \mu) \Rightarrow \rho(2) = \lambda_1\rho(1) = \lambda_1^2$$

Any estimation of λ_2 based on r(1), r(2) will be close to zero as λ_2 = 0.

- Suppose true process is AR(2): $Y_t - \mu = \lambda_1(Y_{t-1} - \mu) + \lambda_2(Y_{t-2} - \mu) + u_t$

$$Cov(Y_{t-1} - \mu, Y_t - \mu) = \lambda_1 Var(Y_{t-1}) + \lambda_2 Cov(Y_{t-1} - \mu, Y_{t-2} - \mu) \Rightarrow \rho(1) = \lambda_1\rho(0) + \lambda_2\rho(1)$$

$$Cov(Y_{t-2} - \mu, Y_t - \mu) = \lambda_1 Cov(Y_{t-2} - \mu, Y_{t-1} - \mu) + \lambda_2 Var(Y_{t-2}) \Rightarrow \rho(2) = \lambda_1\rho(1) + \lambda_2\rho(0)$$

$$Cov(Y_{t-3} - \mu, Y_t - \mu) = \lambda_1 Cov(Y_{t-3} - \mu, Y_{t-1} - \mu) + \lambda_2 Cov(Y_{t-3} - \mu, Y_{t-2} - \mu)$$

$$\Rightarrow \rho(3) = \lambda_1\rho(2) + \lambda_2\rho(1)$$

Any estimation of λ_3 based on r(1), r(2), r(3) will be close to zero as λ_3 = 0.

03 Partial Autocorrelation Function

For AR(p), $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p$ are not zeros, but $\hat{\lambda}_{p+1}$ is approximately zero based on p+1 Yule-walker equations.

For AR(p > 1), $\hat{\lambda}_1$ based on assuming AR(1) and 1 Yule-Walker equation is not zero.

For AR(p > 2), $\hat{\lambda}_1, \hat{\lambda}_2$ based on assuming AR(2) and 2 Yule-Walker equations are not zero.

For AR(p > 3), $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$ based on assuming AR(3) and 3 Yule-Walker equations are not zero.

……………

and so on.

This implies if $\hat{\lambda}_1 \neq 0$, p > 1. If $\hat{\lambda}_2 \neq 0$, p > 2. If $\hat{\lambda}_3 \neq 0$, p > 3. But suppose $\hat{\lambda}_k \approx 0$, then p = k, i.e., AR process of k^{th} order. Estimates above of $\hat{\lambda}_j$ are called partial autocorrelations of Y_t with its lag Y_{t-j} because the effects by other lags are held constant.

The sample ACF allows the identification respectively of either an AR or an MA processes depending on whether the sample $r(k)$ decays (reduces to 0) slowly or is clearly zero after some lag k. However, even if an AR is identified, it is still difficult to identify the order of the lag, p, since all AR(p) processes show similar decay patterns of ACF.

Complementary tool using the partial autocorrelation function (PACF) is used to identify p in the AR(p) process. The PACF also helps to confirm MA(q) processes as invertible MA(q) is AR(∞) and has slow decay in PACF.

03 Partial Autocorrelation Function

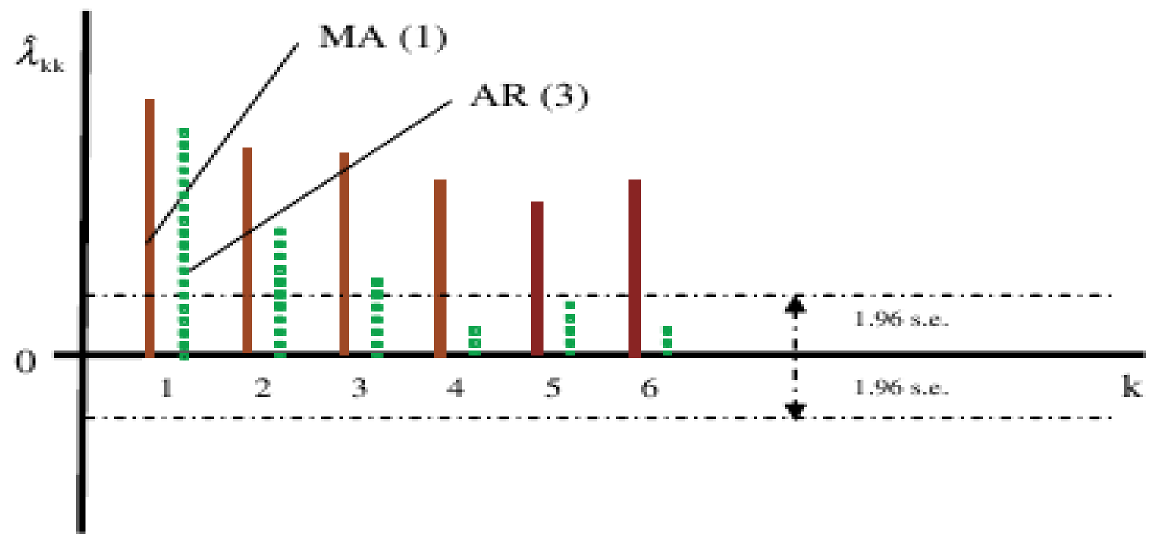


Figure 5.6: Sample Partial Autocorrelogram

If the correct order of the AR is p, then for $k > p$, $Var(\hat{\lambda}_{kk}) = T^{-1}$. Therefore, we can apply hypothesis testing to determine if for a particular k, $H_0: \lambda_{kk} = 0$, should be rejected or not by considering if statistic $\left|\frac{\hat{\lambda}_{kk}}{\sqrt{T-1}}\right| > 1.96$ at the 5% level of significance. In addition to the sample ACF shown earlier, sample PACF above shows that the AR(3) process is identified as its PACF becomes insignificant at $k = 4$.

04 Singapore's GDP Growth

Singapore's GDP from 1985 to 1995 is seen to be rising with a time trend. If we input a time trend variable T, and then run an OLS regression on an intercept and time trend, we obtain $Y = 7977.59 + 292.74 * T$, where Y is GDP and $T = 1,2, \ldots, 44$ quarters from 1985 to 1995.

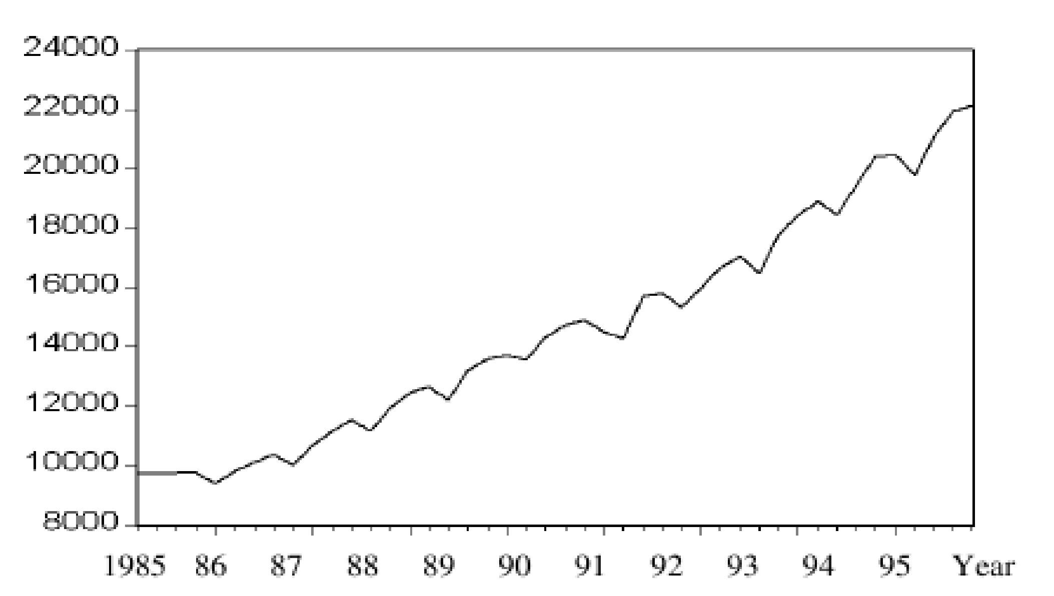
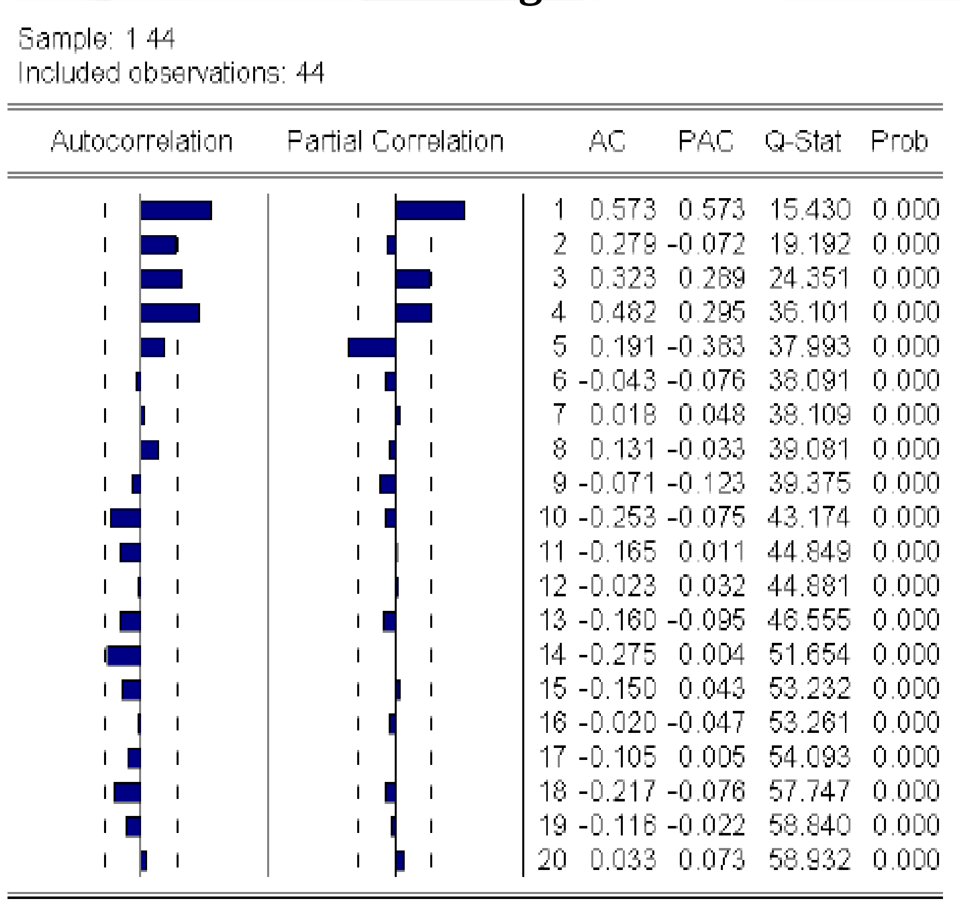


Figure 5.7: Singapore's Per Capita GDP in US$ from 1985 to 1995

We apply a constant and time trend filter to obtain the residual that is identified as AR(1)

ACF and PACF on the regression residuals



	AC	PAC	Q-Stat	Prob
1	0.573	0.573	15.430	0.000
2	0.279	-0.072	19.192	0.000
3	0.323	0.289	24.351	0.000
4	0.482	0.295	36.101	0.000
5	0.191	-0.383	37.993	0.000
6	-0.043	-0.076	38.091	0.000
7	0.018	0.048	38.109	0.000
8	0.131	-0.033	39.081	0.000
9	-0.071	-0.123	39.375	0.000
10	-0.253	-0.075	43.174	0.000
11	-0.165	0.011	44.849	0.000
12	-0.023	0.032	44.881	0.000
13	-0.160	-0.095	46.555	0.000
14	-0.275	0.004	51.654	0.000
15	-0.150	0.043	53.232	0.000
16	-0.020	-0.047	53.261	0.000
17	-0.105	0.005	54.093	0.000
18	-0.217	-0.076	57.747	0.000
19	-0.116	-0.022	58.840	0.000
20	0.033	0.073	58.932	0.000

Figure 5.8: Sample Autocorrelation and Partial Autocorrelation Statistics

04 Singapore's GDP Growth

We approximate the process as $(Y_t - 7977.59 - 292.74 * T) = 0.573 * (Y_{t-1} - 7977.59 - 292.74 * [\mathrm{T} - 1]) + u_t$, where u_t is i.i.d.

Then $Y_t = 3598 + 126T + 0.573Y_{t-1} + u_t$. We create a fitted (in-sample fit) series $\widehat{Y_t} = 3598 + 126T + 0.573Y_{t-1}$. This fitted $\widehat{Y_t}$ and the actual Y_t are plotted together.

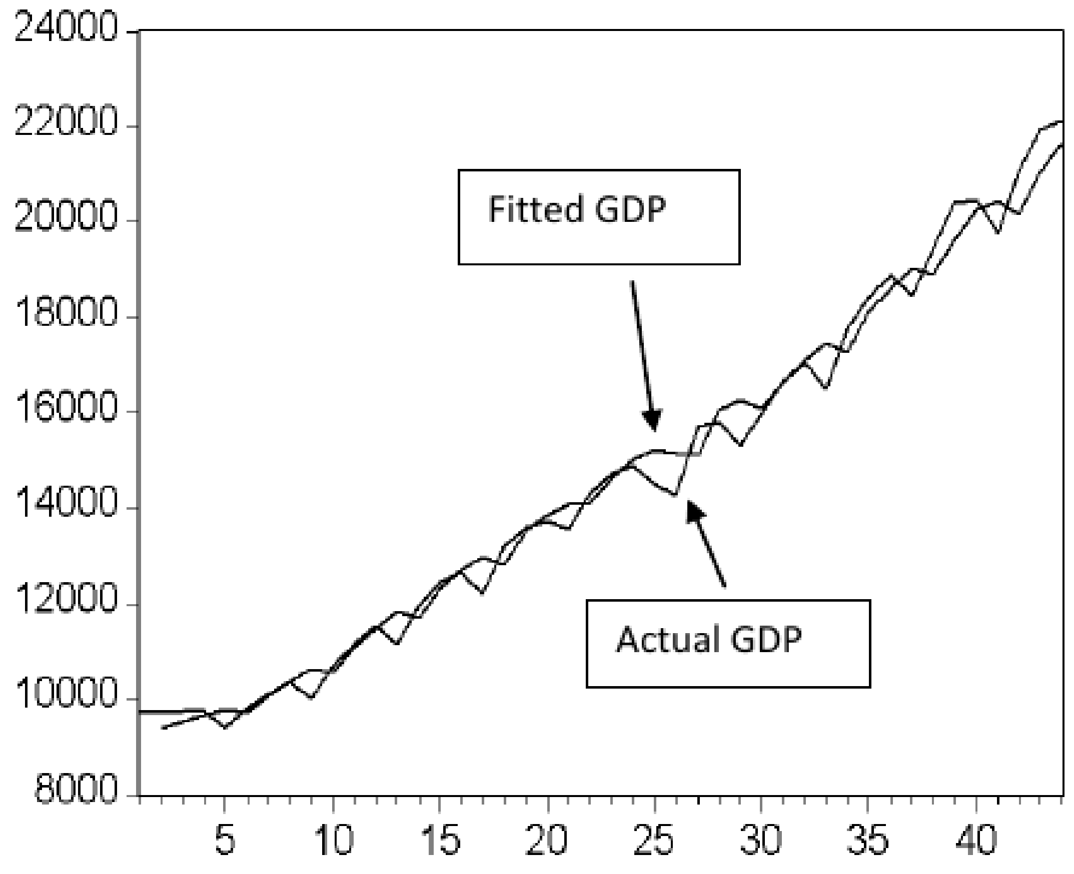


Figure 5.9: Fitted GDP Equation $\hat{Y}_t = 3598 + 126T + 0.573Y_{t-1}$, 1985–1995

04 Autoregressive Integrated Moving Average Process (ARIMA)

- Suppose for a stationary time series $\{Y_t\}$, both its ACF and PACF decay slowly or exponentially without reducing to zero, then an ARMA(p, q), p ≠ 0, q ≠ 0, model is identified. Sometimes, a time series ACF does not reduce to zero, and yet its PACF also does not reduce to zero, because it is an autoregressive integrated moving average process ARIMA(p, d, q). This means that we need to take d number of differences in $\{Y_t\}$ in order to arrive at a stationary ARMA(p, q).

- For example, if $\{Y_t\}$is ARIMA(1,1,1), then $\{Y_t - Y_{t-1}\}$ is ARMA(1,1). In such a case, we have to take differences and then check the resulting ACF, PACF in order to proceed to determine ARMA.

- There is a special case ARIMA(0,1,1) that is interesting, amongst others.

$$\Delta Y_t \equiv Y_t - Y_{t-1} = u_t - \alpha u_{t-1}, \quad |\alpha| < 1.$$

Then, $(1-B)Y_t = (1-\alpha B)u_t$. So,

$$\begin{aligned}
&\frac{(1-B)}{(1-\alpha B)}Y_t = u_t \\
&\Leftrightarrow \frac{(1-\alpha B)-(1-\alpha)B}{(1-\alpha B)}Y_t = u_t \\
&\Leftrightarrow [1-(1-\alpha)(B+\alpha B^2+\alpha^2 B^3+\cdots)]Y_t = u_t \\
&\Leftrightarrow Y_t = Y^*_{t-1} + u_t, \quad Y^*_{t-1} \equiv \delta\sum_{j=1}^{\infty}(1-\delta)^{j-1}Y_{t-j}, \quad \delta = 1-\alpha \\
&\Leftrightarrow E(Y_t \mid Y_{t-1}, Y_{t-2}, \ldots,) = Y^*_{t-1} = \delta Y_{t-1} + (1-\delta)Y^*_{t-2}.
\end{aligned}$$

Forecast of next Y_t at t-1 is weighted average of Y_{t-1} and exponentially weighted moving average of past Y_{t-j}'s, for j > 1.

04 ARIMA

- Another special application of ARIMA is deseasonalisation. Suppose a time series $\{S_t\}$ is monthly sales. It is noted that every December (when Christmas and New Year comes around) sales will be higher because of an additive seasonal component X (assume this is a constant for simplicity). Otherwise, $S_t = Y_t$.
- Assume Y_t is stationary. Then, the stochastic process of sales $\{S_t\}$ will look as follows. $\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_{11}, \tilde{Y}_{12} + X, \tilde{Y}_{13}, \tilde{Y}_{14}, \ldots, \tilde{Y}_{23}, \tilde{Y}_{24} + X, \tilde{Y}_{25}, \tilde{Y}_{26}, \ldots, \tilde{Y}_{35}, \tilde{Y}_{36} + X, \tilde{Y}_{37}, \tilde{Y}_{38}, \ldots$
- This is clearly a nonstationary series even if $\{Y_t\}$ by itself is stationary. This is because the means will jump by X each December. A stationary series can be obtained from the above for purpose of forecast and analysis by performing the appropriate differencing. $(1 - B^{12})S_t = Y_t - Y_{t-12}$.
- Suppose $(1 - B^{12})S_t = u_t$, a white noise. Then, this can be notated as $(0, 1, 0)_{12}$. If $(1 - \theta B^{12})S_t = u_t$, then it is $(1, 0, 0)_{12}$. Notice that the subscript 12 denotes the power of B. If $(1 - \theta B)(1 - B^{12})S_t = u_t$, then it is $(1, 0, 0) \times (0, 1, 0)_{12}$. $(1 - B^{12})S_t = (1 - \alpha B)u_t$ is $(0, 1, 0)_{12} \times (0, 0, 1)$.

04 Modelling Inflation Rates

We define monthly inflation rate I_t as the change from month $t-1$ to month t in the natural log of the Consumer Price Index (CPI) denoted by P_t.

$$I_t = \ln\left(\frac{P_t}{P_{t-1}}\right).$$

Using U.S. data for the period 1953–1977, Fama and Gibbons (1984) reported the following sample autocorrelations in their Table 1 (shown as follows in Table 5.2)

Table 5.2. Autocorrelations of Monthly Inflation Rates and Rate Changes

	$r(1)$	$r(2)$	$r(3)$	$r(4)$	$r(5)$	$r(6)$
I_t	0.55	0.58	0.52	0.52	0.52	0.52
$I_t - I_{t-1}$	−0.53	0.11	−0.06	−0.01	−0.00	0.03
	$r(7)$	$r(8)$	$r(9)$	$r(10)$	$r(11)$	$r(12)$
I_t	0.48	0.49	0.51	0.48	0.44	0.47
$I_t - I_{t-1}$	−0.04	−0.04	0.06	0.02	−0.08	0.09

04 Modelling Inflation Rates

Given sample size $N = 299$, the standard error of $r(k)$ is approximately 0.058. Using 95% significance level or a critical region outside of 1.96 standard errors, or about 0.113, $r(k)$'s for the I_t process are all significantly greater than 0, but $r(k)$'s for $I_t - I_{t-1}$ process are all not significantly different from 0 except for $r(1)$. The autocorrelation of I_t is seen to decline particularly slowly, suggesting plausibility of an ARIMA process. The ACF of $I_t - I_{t-1}$ suggests an MA(1) process. Thus, I_t is plausibly ARIMA(0,1,1).
Using this identification,

$$I_t - I_{t-1} = u_t + \alpha u_{t-1}, \quad |\alpha| < 1. \tag{5.14}$$

From (5.14), $E_{t-1}(I_t) = I_{t-1} + \alpha u_{t-1}$ since u_t is i.i.d. Substitute this conditional forecast or expectation back into (5.14), then

$$I_t = E_{t-1}(I_t) + u_t. \tag{5.15}$$

But $E_t(I_{t+1}) = I_t + \alpha u_t$ and $E_{t-1}(I_t) = I_{t-1} + \alpha u_{t-1}$ imply that

$$\begin{aligned} E_t(I_{t+1}) - E_{t-1}(I_t) &= I_t - I_{t-1} + \alpha(u_t - u_{t-1}) \\ &= u_t + \alpha u_{t-1} + \alpha(u_t - u_{t-1}) \\ &= (1+\alpha)u_t. \end{aligned} \tag{5.16}$$

04 Modelling Inflation Rates

This indicates that the forecast follows a random walk since u_t is i.i.d. From (5.15) , $u_t = I_t - E_{t-1}(I_t)$ can also be interpreted as the unexpected inflation rate realised at t.

Suppose α is estimated as -0.8. Then, (5.14) can be written as:

$$I_t - I_{t-1} = u_t - 0.8u_{t-1}.$$

From (5.16) the variance in the change of expected inflation is $(1-0.8)^2\sigma_u^2$ or $0.04\sigma_u^2$, while the variance of the unexpected inflation is σ_u^2. Thus, the latter variance is much bigger.

This suggests that using past inflation rates in forecasting future inflation as in the time series model (5.15) above is not as efficient. Other relevant economic information could be harnessed to produce forecast with less variance in the unexpected inflation, i.e. produce forecast with less surprise. Fama and Gibbons (1984) show such approaches using the Fisher effect which says that

$$R_t = E_{t-1}(r_t) + E_{t-1}(I_t),$$

where R_t is the nominal risk-free interest rate from end of period $t-1$ to end of period t, and this is known at $t-1$, r_t is the real interest rate for the same period as the nominal rate, and I_t is the inflation rate.

Practice Exercises (not graded)

+SARIMAX.ipynb
Airline-passengers.csv