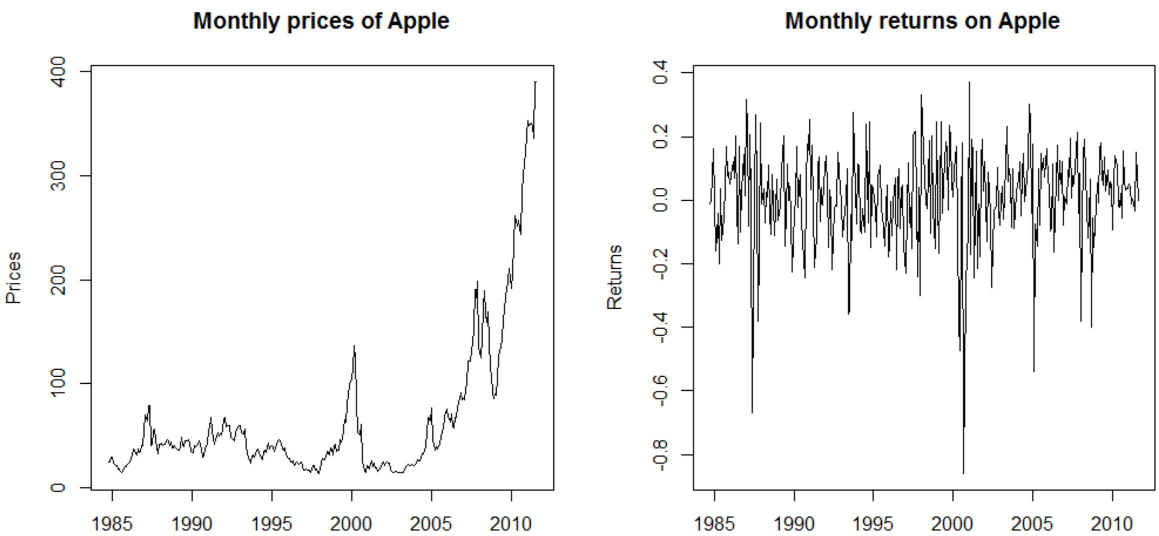
Lecture 2. Linear Time Series Models

1. Introduction

A time series is a set of "repeated" observations on the same random variables, ordered in time. Typical examples include stock prices and returns.

Example 1.1. Monthly prices of and returns on Apple stock

```
> mydat <- read.csv("data1.csv", header = T)
> mydat[1, ]
       Date Price      Return
1 10/1/1984 24.87  -0.9952229
> prc <- ts(mydat$Price, start = c(1984, 10), freq = 12)
> rtn <- ts(log(1 + mydat$Return/100), start = c(1984, 10), freq = 12)
> par(mfrow = c(1, 2))
> plot(prc, ylab = "Prices", xlab = "", main = "Monthly prices of Apple")
> plot(rtn, ylab = "Returns", xlab = "", main = "Monthly returns on Apple")
```



k. x_{t-k} x_t

$k=0.$ $\gamma_0 = cov(x_t, x_t) = Var(x_t).$

2. Stationarity

Definition 2.1. The kth-order *autocovariance* is defined by

$\gamma_{-k} = cov(x_t, x_{t+k}) = \gamma_k$

$$\gamma_k = Cov[x_t, x_{t-k}]$$

$t \in$ data

for $k = 0, 1, \ldots$. By definition, it satisfies that (a) $\gamma_0 = Var[x_t]$ and (b) $\gamma_{-k} = \gamma_k$.

Definition 2.2. A time series $\{x_t\}$ is *stationary* if (a) $E[x_t] = \mu$, which is constant and (b) $Cov[x_t, x_{t-k}] = \gamma_k$, which only depends on k. For a stationary time series $\{x_t\}$, hence, both the mean of x_t and the autocovariance between x_t and x_{t-k} are finite and time invariant.

If $cov(x_t, x_{t-k}) = kt$ then x_t is not time-invariant

Remark 2.3. The WLLN and CLT establish the asymptotic properties of estimators. The WLLN and CLT are valid only for "independent" observations, but time-series data are by nature dependent. So, one must rely on alternative versions of the WLLN and CLT that assume stationarity, thereby maintaining the WLLN- and CLT-related results.

Definition 2.4. The kth-order *autocorrelation* is defined by

$Var(x_t) = constant.$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov[x_t, x_{t-k}]}{Var[x_t]}$$

for $k = 0, 1, \ldots$. The collection of autocorrelations is called the *autocorrelation function* (ACF).

stationary 가정 있어야는
같다는 가정하여 estimate.

- The kth-order sample autocorrelation is computed as

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

estimator

for $0 \leq k < T - 1$.

Theorem 2.5. *If $\{x_t\}$ is a sequence of iid random variables with the finite second moment, it shows*

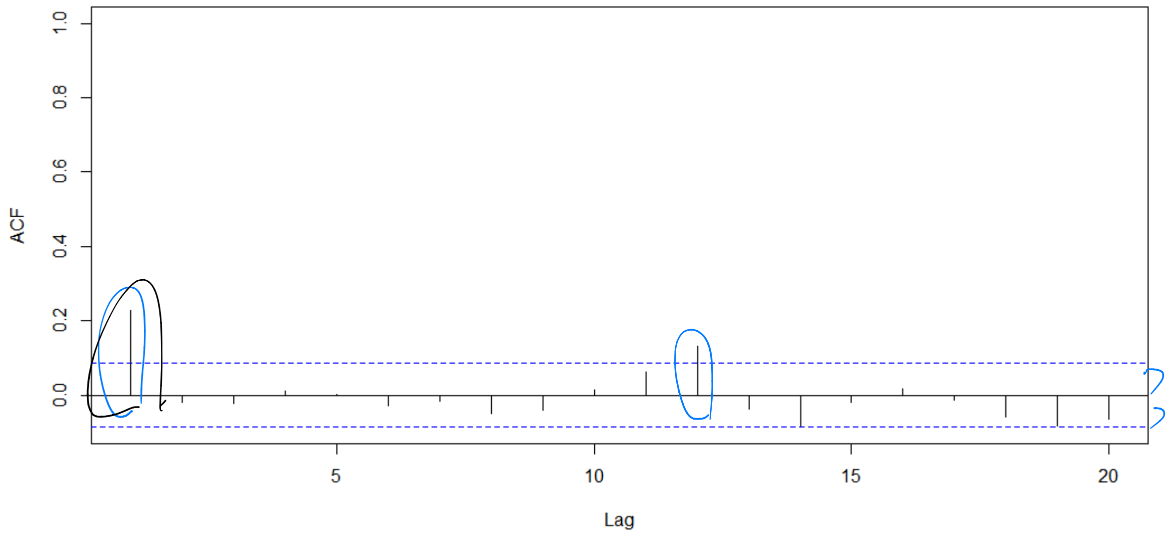
i.e. $E[X_t^2] \leq M$

$$\hat{\rho}_k \approx N\left(0, \frac{1}{T}\right),$$

where

Example 2.6. Consider the monthly simple returns of the Decile 10 portfolios of CRSP from January 1967 to December 2009.

```
> mydat <- read.table("data2.txt", header = T)
> mydat[1, ]
      date      dec1      dec2      dec9     dec10
1 19670131  0.068568  0.080373  0.180843 0.211806
> dec10 <- mydat$dec10
> acf(dec10, lag = 24, main = "", xlim = c(1, 20))
```



2.6

- The two horizontal lines denote the two standard error limits. The first-order autocorrelation is significantly different from zero at the 5% level.

time invariant mean and cov. ∴ ⇒ stationarity.

3. White Noise and Linear Time Series

Definition 3.1. A time series $\{x_t\}$ is a *white noise process*, denoted by $x_t \sim WN(0, \sigma^2)$, if (a) $E[x_t] = 0$ for all t, (b) $Cov[x_t, x_{t-k}] = 0$ for all t and $k \neq 0$, and (c) $Var[x_t] = \sigma^2$ for all t.

- If $x_t \sim WN(0, \sigma^2)$, the kth-order autocorrelations ρ_k for all $k > 0$ are zero. So, a test procedure that determines a given time series is a white noise process asks whether the time series has zero autocorrelations. If the null hypothesis $H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0$ is rejected in favor of the alternative hypothesis $H_1 : \rho_i \neq 0$ for some $i \in \{1, \ldots, m\}$, it concludes that the time series is not a white noise process.

Theorem 3.2. *In testing $H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0$ versus $H_1 : \rho_i \neq 0$ for some $i \in \{1, \ldots, m\}$, Ljung and Box (1978) propose that the test statistic is*

$$Q(m) = T(T+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{T-k} \approx \chi^2_{(m)}$$

under the null hypothesis.

Example 3.3. Consider the monthly returns of IBM stock from January 1967 to December 2009.

```
> mydat <- read.table("data3.txt", header = T)
> mydat[1, ]
      date       ibm       sp
1 19670131  0.075370  0.078178
> ibm <- mydat$ibm
```

```
> Box.test(ibm, lag = 12, type = "Ljung")
  Box-Ljung test
  data: ibm
  X-squared = 7.5666, df = 12, p-value = 0.818
```

- The Ljung-Box statistics with $m = 12$ cannot reject the null hypothesis of no serial correlations in the IBM stock returns. $\therefore$ ibm $\sim$ WN

Definition 3.4. A time series x_t is said to be *linear* if it can be written as

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \quad = \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots$$

where $\mu = E[x_t]$, $\psi_0 = 1$, and $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. $\quad \frac{\partial x_t}{\partial \varepsilon_{t-k}} = \psi_k$ shock, (news).

- The white noise term ε_t denotes the new information at time t and is often referred to as the *innovation* or *shock* at time t. //

$$x_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \psi_3 \varepsilon_{t-3} + \cdots$$

MA model

$\Rightarrow$ x_t is linear.

$\mu = E[x_t]$,

$\psi_0 = 1$,

$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$

$E[X_t] = a$

$E[X_{t+k}] = b.$

$Var[X_t] = c$

$Var[X_{t+k}] = d.$

$Cov[X_t, X_{t+k}]$

$= E[(X_t - E[X_t])(X_{t+k} - E[X_{t+k}])]$

$= E[X_t \cdot X_{t+k}] - a \cdot b.$

$c + d + 2Cov[X_t, X_{t+k}]$

$\shortparallel$

$Var[X_t + X_{t+k}] = Var[X_t] + Var[X_{t+k}] + 2Cov[X_t, X_{t+k}]$