Chapter 2

Linear Time Series

2.1 Moving Average Model

MA(q)-model: $X_t = \mu + \sum_{j=1}^{q} a_j \varepsilon_{t-j} + \varepsilon_t, \{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$.

★**Always** stationary

Example: **(k-period log-return)** $r_t[k] = r_t + r_{t-1} + \cdots + r_{t-k}$.

Its ACF is very simple to compute. e.g. MA(2) model:

$$X_t = \mu + \varepsilon_t + a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2}$$

$$X_{t-3} = \mu + \varepsilon_{t-3} + a_1\varepsilon_{t-4} + a_2\varepsilon_{t-5}$$

Hence, $\rho(3) = \rho(4) = \cdots = 0$. But the sample ACF will not be exactly zero. What is the **confidence limit**?

Theorem 1 : *Let X_t follow an MA(q) model. Then*

$$\sqrt{T}\widehat{\rho}(j) \xrightarrow{D} \mathcal{N}\Big(0, 1 + 2\sum_{i=1}^{q} \rho^2(i)\Big), \ j > q$$

Thus, for $j > q$, about 95% of sample correlations $\widehat{\rho}(j)$ fall in the interval

$$\pm\frac{1.96}{\sqrt{T}}\{1 + 2\sum_{i=1}^{q} \widehat{\rho}^2(i)\}^{1/2}.$$

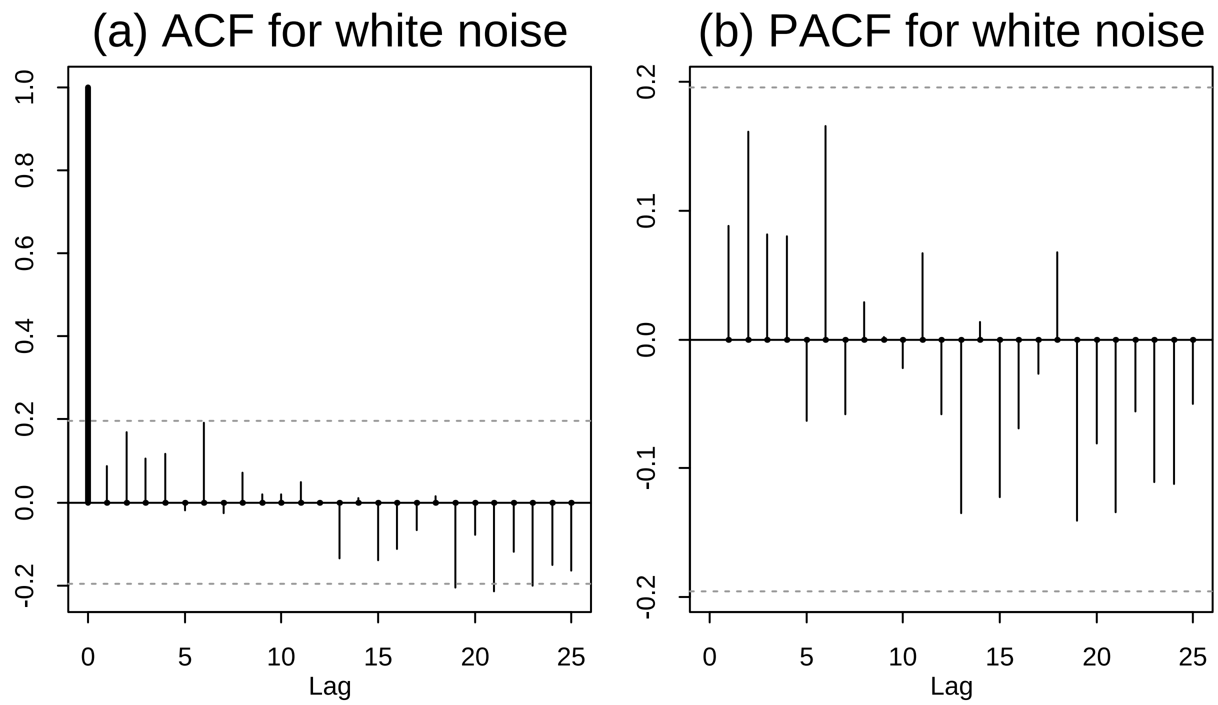

Figure 2.1: ACF and PACV for a simulated white noise series with T=100.

2.2 Autoregressive model

★a simple and useful class of models for forecasting returns.

AR(p)-model: $X_t = b_0 + b_1 X_{t-1} + \cdots + b_p X_{t-p} + \varepsilon_t$

Suppose that X_t is stationary with mean μ. Then

$$\mu = b_0 + (b_1 + \cdots + b_p)\mu \quad \Longrightarrow \quad \mu = \frac{b_0}{1 - b_1 - \cdots - b_p}.$$

Stationarity: Let $b(z) = 1 - b_1 z - \cdots - b_p z^p$ be the characteristic function. Then $b(B)X_t = b_0 + \varepsilon_t$, where B is a

Backshift operator: $B^k X_t = X_{t-k}, k = \pm 1, \pm 2, \cdots$.

The invertibility $X_t = b(B)^{-1}(b_0 + \varepsilon_t)$ requires technical conditions:

$b(z)$ **has roots outside the unit circle**.

In this case, $b(z)^{-1} = \sum_{j=0}^{\infty} c_j z^j$ and

$$X_t = \sum_{j=0}^{\infty} c_j B^j (b_0 + \varepsilon_t) = \sum_{j=0}^{\infty} c_j (b_0 + \varepsilon_{t-j})$$

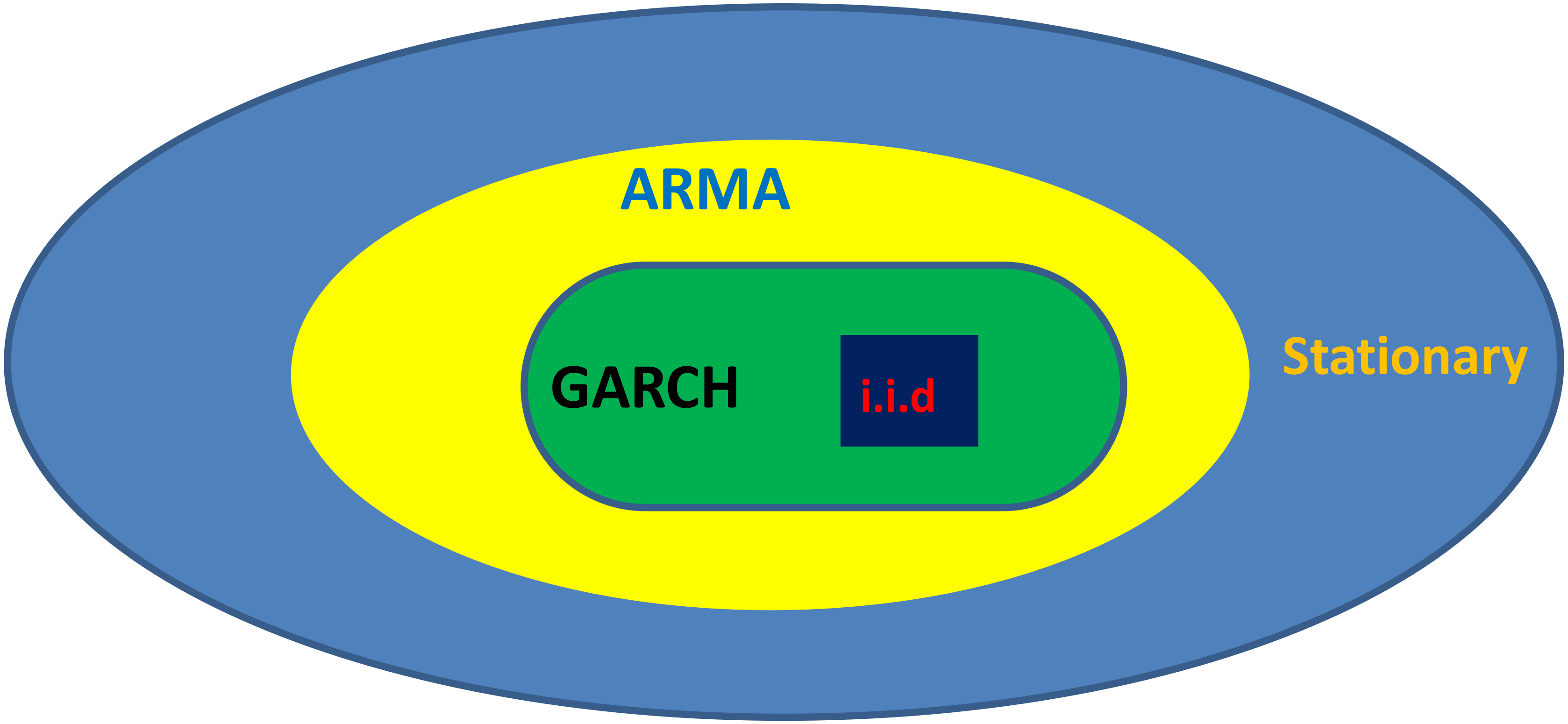

Figure 2.2: Relationship among different processes: Stationary processes are the largest set, followed by ARMA, GARCH, and white noises processes.

is stationary and **causal**.

Theorem 2 : *An AR(p) process is stationary and causal if*

$$\inf_{|z|\le 1} |b(z)| > 0.$$

Example 2. For the AR(1)-model: $X_t = bX_{t-1} + \varepsilon_t$, the characteristic function is $b(z) = 1 - bz$ and has a root $1/b$. Thus, when $|b| < 1$, the series is stationary. Now consider the following AR(3) model:

$$X_t = 0.8X_{t-1} - 0.5X_{t-2} + 0.4X_{t-3} + \varepsilon_t.$$

The characteristic function

$$\begin{aligned} b(z) &= 1 - 0.8z + 0.5z^2 - 0.4z^3 \\ &= (1 - 0.8z)(1 + 0.5z^2), \end{aligned}$$

which has all roots outside the unit circle. Thus, it is stationary.

Autocorrelation of AR(p): Since X_{t-k} is **causal**, depending only on ε_{t-k} and its past,

$$\mathrm{Cov}(b(B)X_t, X_{t-k}) = \mathrm{Cov}(b_0 + \varepsilon_t, X_{t-k}) = 0, \quad \text{for } k > 0$$

$$\Longrightarrow \gamma(k) - b_1\gamma(k-1) - \cdots - b_p\gamma(k-p) = 0 \quad \forall k > 0.$$

For $k = 0$, RHS $= \sigma^2$.

Solution: admits the form (homeowrk)

$$\gamma(k) = \alpha_1 z_1^{-k} + \cdots + \alpha_p z_p^{-k} = O(\{\min_j |z_j|\}^{-k}),$$

where $z_1, \cdots, z_p$ are the roots of $b(z)$ and $\{\alpha_j\}_{j=1}^k$ are constants. Hence, $\rho(k) \to 0$ exponentially fast **(short memories)**. For example, for AR(1): $X_t = bX_{t-1} + \varepsilon_t, \quad \rho(k) = b^{|k|}$.

ACF helps us identify a model. Fig 2.3 shows ACF for AR(1) with

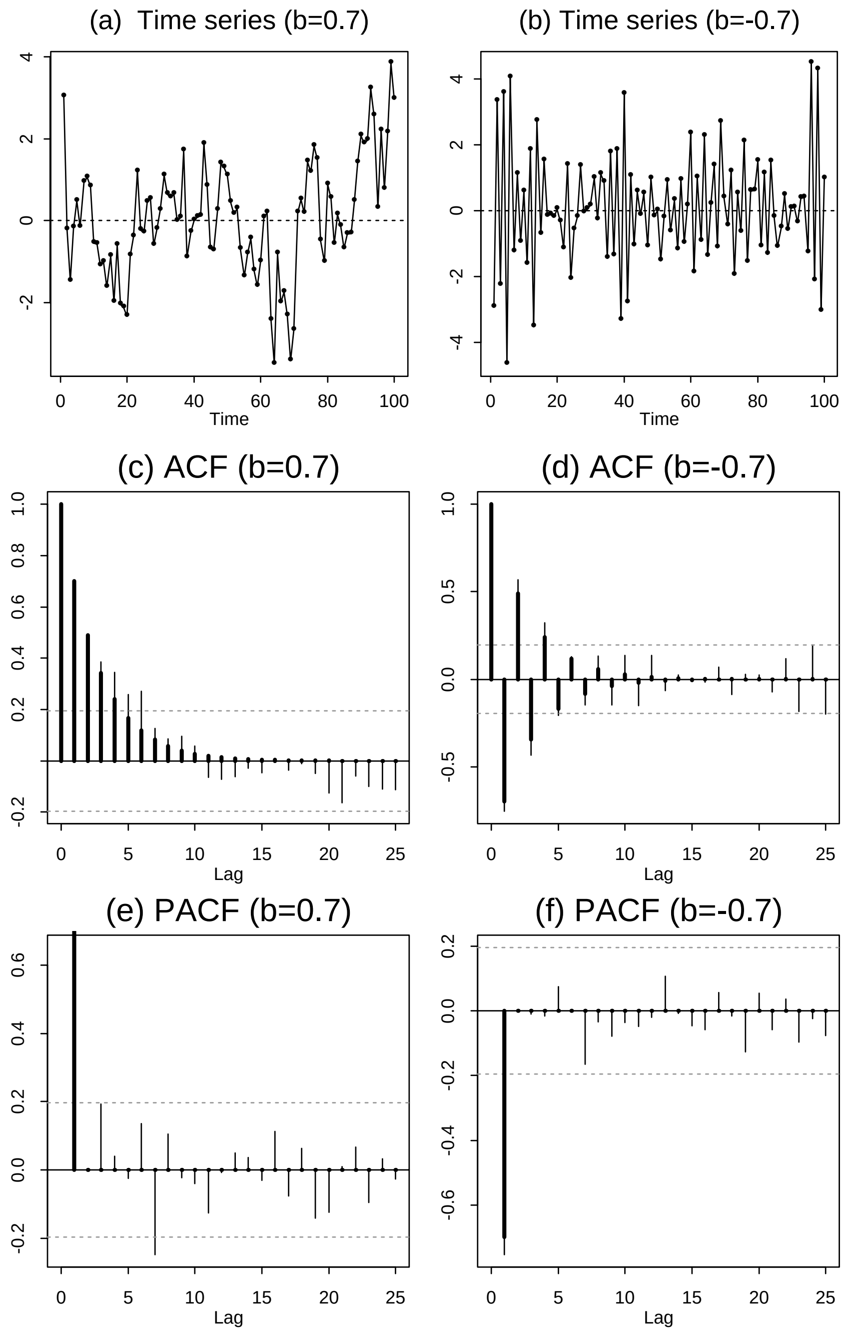

Figure 2.3: Simulated time series and their ACFs and PACFs, T= 100.

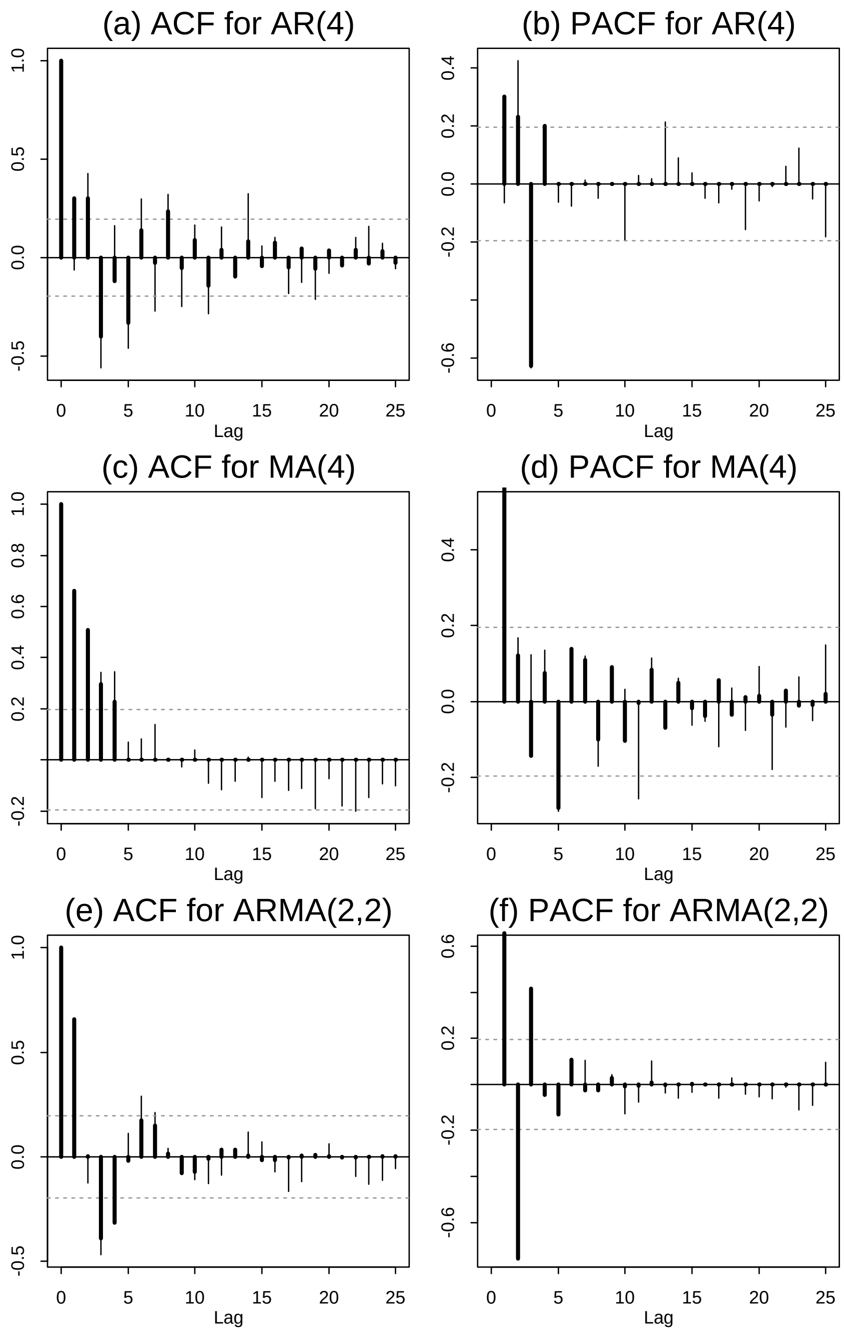

Figure 2.4: Simulated time series and their ACFs and PACFs, T= 100.

$b = 0.7$ or -0.7. Fig 2.4 shows ACF for

- AR(4): $X_t = 0.5X_{t-1} + 0.3X_{t-2} - 0.7X_{t-3} + 0.2X_{t-4} + \varepsilon_t$,
- MA(4): $X_t = \varepsilon_t + 0.6\varepsilon_{t-1} + 0.6\varepsilon_{t-2} + 0.3\varepsilon_{t-3} + 0.7\varepsilon_{t-4}$,
- ARMA(2,2): $X_t = 0.8X_{t-1} - 0.6X_{t-2} + \varepsilon_t + 0.7\varepsilon_{t-1} + 0.4\varepsilon_{t-2}$.

Partial autocorrelation: MA(q) model has a distinct rubric: $\rho(k) = 0, \ \forall k > q$. Do we have a similar measure? Yes, the partial autocovariance function $\pi(\cdot)$.

PACF: $\pi(1) = \text{Corr}(X_1, X_2) = \rho(1)$; and for $k \geq 2$,

$$\pi(k) = \text{Corr}(X_1, X_{k+1} | X_2, ..., X_k).$$

More precisely, it is $\text{Corr}(R_1, R_{k+1})$ where R_j = residual of X_j on $(X_2, \ldots, X_k)$.

Prop:$\pi(k) =$ *the last regression coefficient of*

$$\min_{\beta} \mathrm{E}(X_{t+k+1} - \beta_0 - \beta_1 X_{t+k} - \cdots - \beta_k X_{t+1})^2.$$

Namely, $\pi(k) = b_{kk}$ where $(b_{k1}, \ldots, b_{kk})$ *is the regression coef.*

Model identification: For an AR(p) model, it can easily be shown that (homework)

$$\pi(k) = 0 \qquad \forall k > p.$$

Prop: If $\{X_t\}$ follows an AR(p) model with i.i.d. white noise, then

$$T^{1/2}\widehat{\pi}(k) \stackrel{D}{\rightarrow} N(0,1), \text{ for } k > p.$$

Example 3. For the monthly returns of the CRSP value-weighted index (Jan., 1926 — Dec. 1997, $T = 864$), it is computed that

p	1	2	3	4	5	6	7	8	9	10
PACF	0.11	-0.02	**-0.12**	0.04	**0.07**	-0.06	0.02	0.06	0.06	-0.01
AIC	-5.807	-5.805	-5.817	-5.816	-5.819	-5.821	-5.819	-5.820	-5.821	-5.818

$\text{SE} = \frac{1}{\sqrt{T}} = 0.034$, $\quad$ **2**SE = **0.068**, $\quad \widehat{p} = 3$ or 5.

Parameter estimation

The AR coefficient is estimated by the least-squares:

$$\min_{\boldsymbol{b}} \sum_{t=p+1}^{T} \Big(X_t - b_0 - b_1 X_{t-1} - \cdots - b_p X_{t-p} \Big)^2 .$$

This is an auto-regression problem:

$$\min_{\boldsymbol{b}} \sum_{t=p+1}^{T} \Big(\overbrace{\underbrace{X_t}_{Y_t} - b_0 - b_1 \underbrace{X_{t-1}}_{X_{t,1}} - \cdots - b_p \underbrace{X_{t-p}}_{X_{t,p}}}^{\text{part that can not be predicted}} \Big)^2 .$$

Residuals: $\widehat{\varepsilon}_t = X_t - \widehat{b}_0 - \widehat{b}_1 X_{t-1} - \cdots - \widehat{b}_p X_{t-p}$

♣provide raw materials for model checking;

♣ideal fit if the residual series behaves like white noise series.

Residual variance:

$$\widehat{\sigma}^2 = \frac{\text{RSS}}{(T-p)-(p+1)} = \frac{1}{T-2p-1}\sum_{t=p+1}^{T}\widehat{\varepsilon}_t^2$$

Example 3 (Cont). For the CRSP index, an AR(3) fit results in

$$\begin{array}{llllll} r_t & = 0.0103 & + \ 0.104 r_{t-1} & - \ 0.010 r_{t-2} & - \ 0.120 r_{t-3} & + \ \varepsilon_t \\ \text{SE} & \ \ \ 0.002 & \ \ \ 0.034 & \ \ \ 0.034 & \ \ \ 0.034 & \end{array}$$

For example, to test whether $H_0 : b_0 = 0$ (or the mean return is zero), we compute the t-statistic $0.0103/0.002 = 5$ and hence its associated P-value is $2\Phi(-5) = 0$. We have strong evidence against H_0, namely, the monthly returns are positive.

— The time series is weakly dependent (small b's).

— b_0 is significantly positive. The expect return (monthly) is

$$\widehat{\mu} = \frac{\widehat{b}_0}{1 - \widehat{b}_1 - \widehat{b}_2 - \widehat{b}_3} = 0.01 \quad \text{(statistically significant?)}$$

Annualized expected return $= (1 + 0.01)^{12} - 1 \approx 12.6\%$.

Actual annualized return (1/1926—12/1997): 10.53%

— For checking the white noise of the residual series,

$$Q(10) = 15.8, \qquad d.f. = 10, \qquad \text{p-value} = 10.55\%.$$

With $T = 864$, the null hypothesis that the white noise of residuals is very reasonable.

— What is the SE of $\widehat{\mu}$?

Delta method: When $\widehat{\boldsymbol{b}}$ is a consistent estimate of $\boldsymbol{b}$, then $\widehat{\mu} = f(\widehat{\boldsymbol{b}}) \approx f(\boldsymbol{b}) + f'(\boldsymbol{b})^T(\widehat{\boldsymbol{b}} - \boldsymbol{b})$. Thus,

$$\mathrm{var}(\widehat{\mu}) \approx \mathbf{f}'(\boldsymbol{b})^{\mathbf{T}} \mathrm{var}(\widehat{\boldsymbol{b}}) \mathbf{f}'(\boldsymbol{b}).$$

The matrix $\mathrm{var}(\widehat{\boldsymbol{b}})$ is usually given by software packages.

More rigorously, if $\sqrt{T}(\widehat{\boldsymbol{b}} - \boldsymbol{b}) \xrightarrow{L} N(0, \Sigma)$, then

$$\sqrt{\mathbf{T}}[\mathbf{f}(\widehat{\boldsymbol{b}}) - \mathbf{f}(\boldsymbol{b})] \xrightarrow{\mathbf{L}} \mathbf{N}(\mathbf{0}, \mathbf{f}'(\boldsymbol{b})^{\mathbf{T}} \boldsymbol{\Sigma} \mathbf{f}'(\boldsymbol{b})).$$

Namely, $\mathrm{AVar}(f(\widehat{\boldsymbol{b}})) = f'(\widehat{\boldsymbol{b}})^T \mathrm{AVar}(\widehat{\boldsymbol{b}}) f'(\widehat{\boldsymbol{b}})$.

Example 3 (Cont). $\mu = f(\boldsymbol{b}) = \frac{b_0}{1-b_1-b_2-b_3}$. Thus,

$$f'(\boldsymbol{b}) = (a, b_0, b_0, b_0)^T / a^2, \qquad a = 1 - b_1 - b_2 - b_3.$$

Evaluation of the gradient at the estimates gives

$$f'(\widehat{\boldsymbol{b}}) = (0.9746, 0.0098, 0.0098, 0.0098)^T.$$

Suppose that the estimated variance of $\widehat{\boldsymbol{b}}$ is (from software output)

$$\boldsymbol{S} = 1000^{-2} \begin{pmatrix} 2^2 & 34 & 0 & 0 \\ 34 & 34^2 & 0 & 0 \\ 0 & 0 & 34^2 & 0 \\ 0 & 0 & 0 & 34^2 \end{pmatrix}.$$

Then, the asymptotic variance is given by $f'(\widehat{\boldsymbol{b}})^T \boldsymbol{S} f'(\widehat{\boldsymbol{b}}) = 4.7804 * 10^{-6}$. Hence, $\mathrm{SE}(\widehat{\mu}) = \sqrt{4.7804 * 10^{-6}} = 0.2186\%$. Therefore, 95% confidence interval for the monthly return is

$$1\% \pm 1.96 * 0.2186\% = 1\% \pm .4285\%.$$

Order Selection

Aim: To select p to minimize an estimated PE. Let L be the maximum order to be fitted

Akaike information Criterion:

— AIC(p)= −2 (max log-likelihood)+ 2(No. parameters)

$= (T-L)\log(\widehat{\sigma}_p^2) + \mathbf{2}(p+1) + \text{constant}, \quad \text{for} \quad p = 0, 1, \ldots, L,$

where $\widehat{\sigma}_p^2$ is the estimated residual variance. The first part measures the lack of fit and the second part **penalizes** the complexity of the model.

— $\text{AICC}(p) = (T-L)\log(\widehat{\sigma}_p^2) + \frac{\mathbf{2(T-L)}}{\mathbf{T-L-p-2}}(p+1)$, a correction of AIC.

Bayesian information criterion:

$$\text{BIC}(p) = (T-L)\log(\widehat{\sigma}_p^2) + (p+1)\log(\mathbf{T}-\mathbf{L}).$$

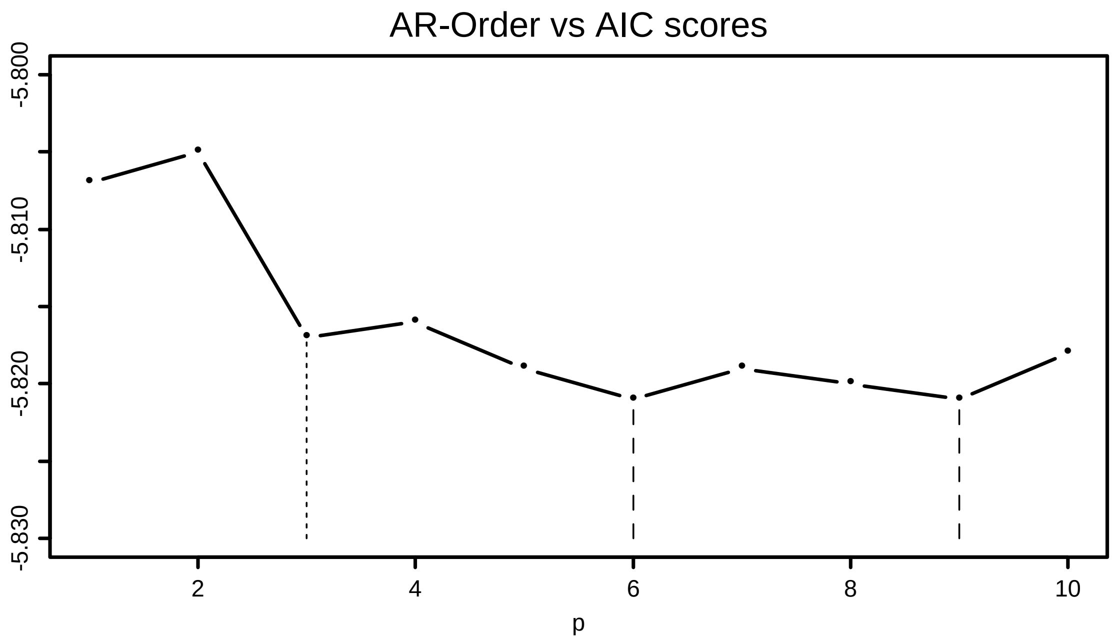

Figure 2.5: order versus AIC for the CRSP value weighted data

For CRSP value-weighted index,

$\widehat{p} = 3$ — the point where AIC stops deceasing dramatically;

$\widehat{p} = 6$ or 9 — minimum.

Remarks:

— AIC usually selects too many parameters;

— BIC can select too few parameters;

— BIC and AICC penalize more heavily on **complexity** than AIC;

2.3 Prediction

Best prediction: find f to minimize the prediction error:

$$X_T(m) = \arg\inf_f \mathrm{E}(X_{T+m} - f(X_T, X_{T-1}, \cdots))^2 = \mathrm{E}_T X_{T+m}.$$

For the AR(p) model, the best one-step predictor

$$\begin{aligned} X_T(1) &= \mathrm{E}_T(b_0 + b_1 X_T + \cdots + b_p X_{T+1-p} + \varepsilon_{T+1}) \\ &= b_0 + b_1 X_T + \cdots + b_p X_{T+1-p}, \end{aligned}$$

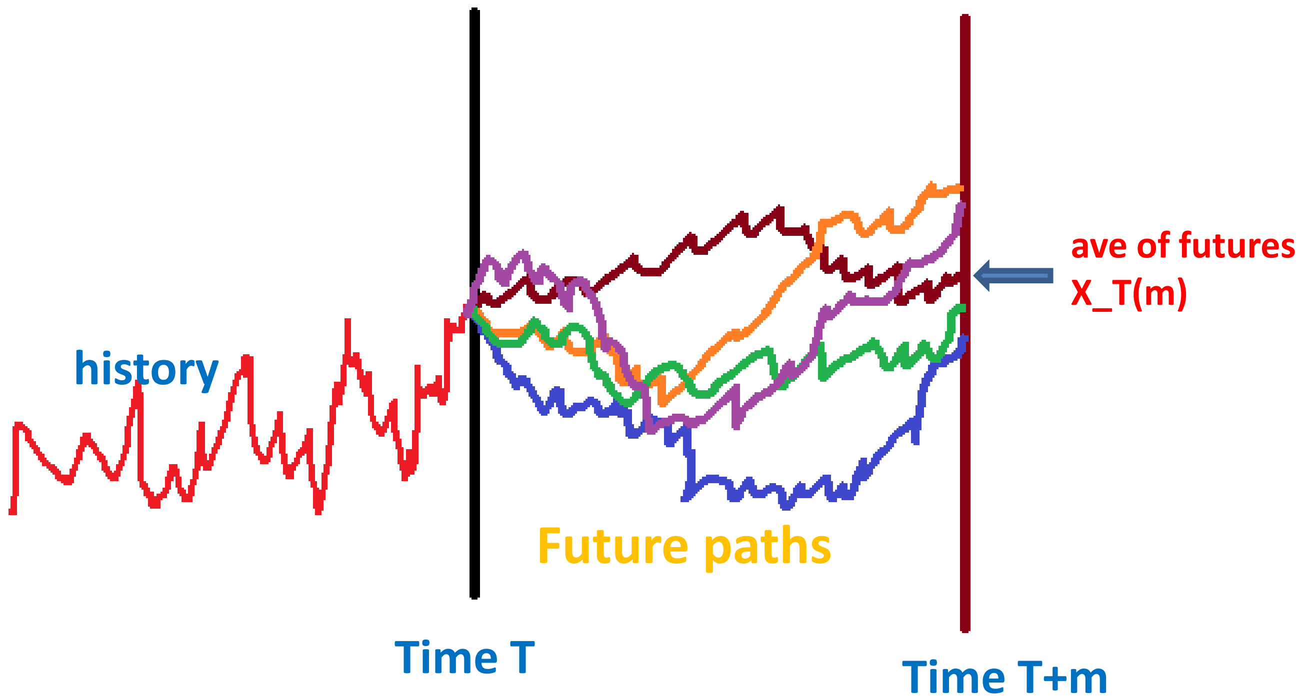

Figure 2.6: Illustration of the best prediction of X_{T+m} given $X_1, \cdots, X_T$, which is the average of possible values.

which coincides with the best linear predictor. In practice, one step-forecasting is

$$\widehat{X}_T(1) = \widehat{b}_0 + \widehat{b}_1 X_T + \cdots + \widehat{b}_p X_{T+1-p}.$$

Size of prediction error: Ignoring error in estimated coefficients, it is σ. There are statistical methods to account for that.

2-step ahead prediction: Note that

$$X_{T+2} = b_0 + b_1 X_{T+1} + \cdots + b_p X_{T+2-p} + \varepsilon_{T+2}.$$

Hence, the two-step prediction is

$$X_T(2) = b_0 + b_1 X_T(1) + \cdots + b_p X_{T+2-p},$$

which is linear predictor and hence coincides with the best linear prediction. The 2-step prediction error is the part involving **future** noises:

$$e_T(2) = X_{T+2} - X_T(2) = \varepsilon_{T+2} + b_1 \varepsilon_{T+1}.$$

The size of the two-step prediction error is

$$\sqrt{\mathrm{var}(e_T(2))} = \sqrt{(1 + b_1^2)\sigma^2},$$

which is larger than one-step.

Multi-step forecast: ★Iterated AR(p); ★direct LS (better)

(equivalent under AR(p) models, better for other stationary models)

Example 4: Consider the CRSP data (Jan. 26 – Dec. 97), the following AR(5) model is used to predict the monthly log-return

$$r_t = 0.0075 + 0.103 r_{t-1} + 0.002 r_{t-2} - 0.114 r_{t-3} + 0.032 r_{t-4} + 0.084 r_{t-5} + \varepsilon_t$$

with $\sigma = 0.054, T = 858$.

Results: The actual and forecasted one are summarized in the following table and Fig. 2.7.

Step	1	2	3	4	5	6
Forecast	0.0071	-0.0008	0.0086	0.0154	0.0141	0.0100
SE	0.0541	0.0545	0.0545	0.0549	0.0549	0.0550
Actual	0.0762	-0.0365	0.0580	-0.0341	0.0311	0.0183

2.4 Unit root

Fundamental question: Whether asset prices are predictable.

Consider the log-price of SP500 (Jan. 72 – Dec. 99): $X_t = \log(S_t)$, where S_t = index at t. The autocorrelation of $\{X_t\}$ features strong

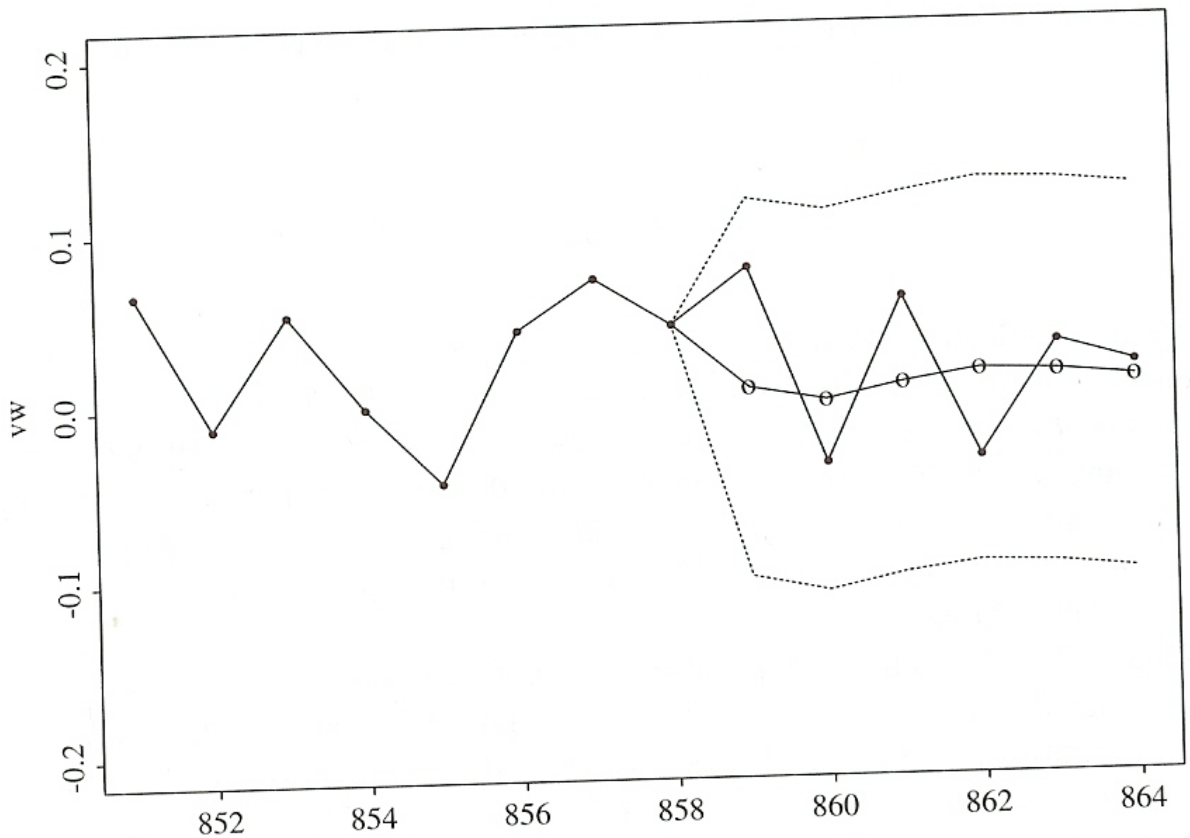

Figure 2.7: Out sample forecast

persistence. From PACF plot, it is reasonable to assume that

$$X_t = X_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \quad \text{white noise.}$$

It is an AR(1) model whose characteristic function has unit root.

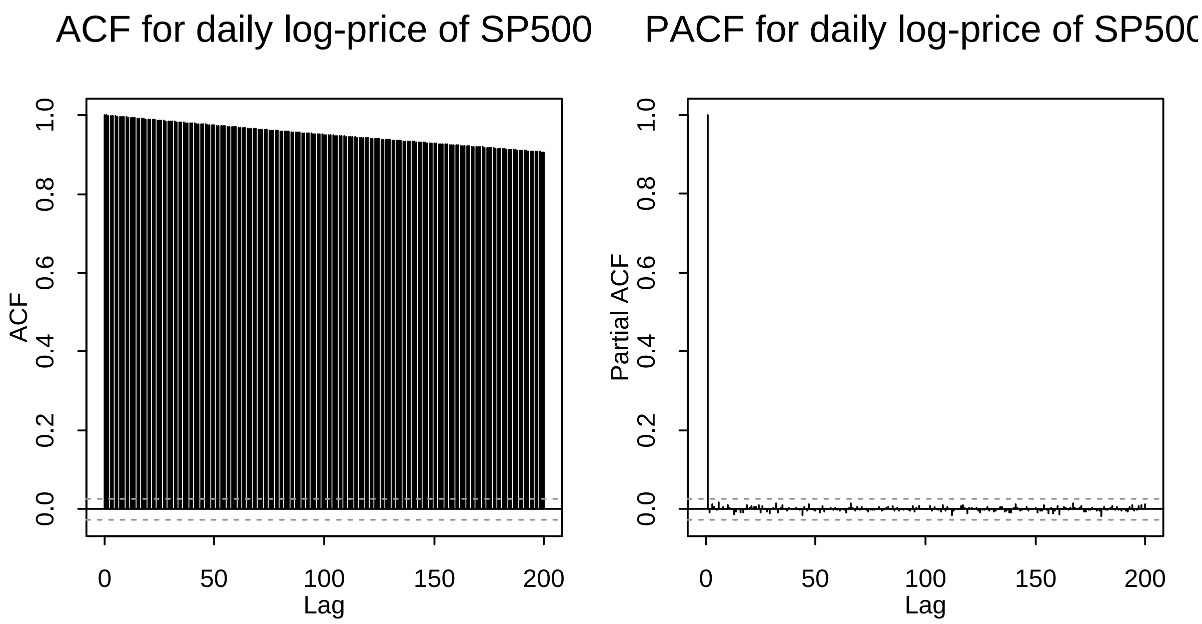

Figure 2.8: ACF and PACF for daily log-price of SP500 index.

Note that $X_t = X_0 + \sum_{i=1}^{t} \varepsilon_i$, which is

— nonstationary: $\text{var}(X_t) = \text{var}(X_0) + t\text{var}(\varepsilon)$

— a random walk without drift.

The random walk with a drift is defined as

$$X_t = \mu + X_{t-1} + \varepsilon_t = \mu t + X_0 + \sum_{i=1}^{t} \varepsilon_i.$$

Unit root test: Embed the random walk into the AR(1) model:

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad \text{without drift;}$$

$$X_t = \mu + \rho X_{t-1} + \varepsilon_t \text{ with drift.}$$

The random walk hypothesis: $H_0 : \rho = 1$.

Estimation of ρ by the least-squares method:

- **Model without drift**: $\min_\rho \sum_{t=2}^T (X_t - \rho X_{t-1})^2$ gives

$$\widehat{\rho} = \frac{\sum_{t=2}^T X_t X_{t-1}}{\sum_{t=2}^T X_{t-1}^2}.$$

- **Model with drift**: $\min_{\mu,\rho} \sum_{t=2}^T (X_t - \mu - \rho X_{t-1})^2$ gives

$$\begin{aligned} \widetilde{\mu} &= \bar{X}_2 - \widetilde{\rho}\bar{X}_1 \\ \widetilde{\rho} &= \frac{\sum_{t=2}^T (X_t - \bar{X}_2)(X_{t-1} - \bar{X}_1)}{\sum_{t=2}^T (X_{t-1} - \bar{X}_1)^2}, \end{aligned}$$

where $\bar{X}_2 = (T-1)^{-1}\sum_{t=2}^T X_t$ and $\bar{X}_1 = (T-1)^{-1}\sum_{t=2}^T X_{t-1}$.

Dickey-Fuller coefficient test: Reject H_0 at $\alpha = 5\%$, when

$$T(\widehat{\rho} - 1) < -8.347 \text{ without drift;}$$

$$T(\widetilde{\rho} - 1) < -13.96 \text{ with drift.}$$

Dickey-Fuller t-test: the t-statistic for ρ from the linear fit to $X_t = \mu + \rho X_{t-1} + \varepsilon_t$ (more frequently used).

Critical values fo Dickey-Fuller tests

Level α	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
With drift, **coef-test**	-20.21*	-13.96*	-11.33*
With drift, **t-test**	-3.459	-2.874	-2.573
Without drift, coef-test	-14.07*	-8.347*	-5.862*
Without drift, t-test	-2.574	-1.941	-1.616

* based on 10,000 simulation with $T = 1000$

Null distributions: Let $\{W_t\}$ be a Brownian motion and $\widetilde{W}_t = W_t - \int_0^1 W_t dt$. Then

$$T(\widehat{\rho} - 1) \xrightarrow{L} \frac{\int_0^1 W_t dW_t}{\int_0^1 W_t^2 dt}, \qquad T(\widetilde{\rho} - 1) \xrightarrow{L} \frac{\int_0^1 \widetilde{W}_t dW_t}{\int_0^1 \widetilde{W}_t^2 dt}.$$

Example 5. For the S&P 500 daily log-prices, $T = 5348$. For testing against random-walk **without a drift**, it was computed that

$$\widehat{\rho} = 1.0006, \qquad T(\widehat{\rho} - 1) = 3.2088.$$

The random walk hypothesis can not be rejected.

Similarly, for random walk hypothesis **with a drift** (reasonable),

$$\widetilde{\rho} = 0.9997106, \qquad T = 5348, \qquad T(\widetilde{\rho} - 1) = -1.5479,$$

which is bigger than the critical value -13.96. We can not reject the random walk hypothesis.

For **monthly** log-prices, $T = 214$. It can be computed that

$$\widehat{\rho} = 1.0050, \qquad T(\widehat{\rho} - 1) = 0.321,$$

and

$$\widetilde{\rho} = 0.99339, \qquad T(\widetilde{\rho} - 1) = -1.4135.$$

Both do not provide strong evidence against H_0.

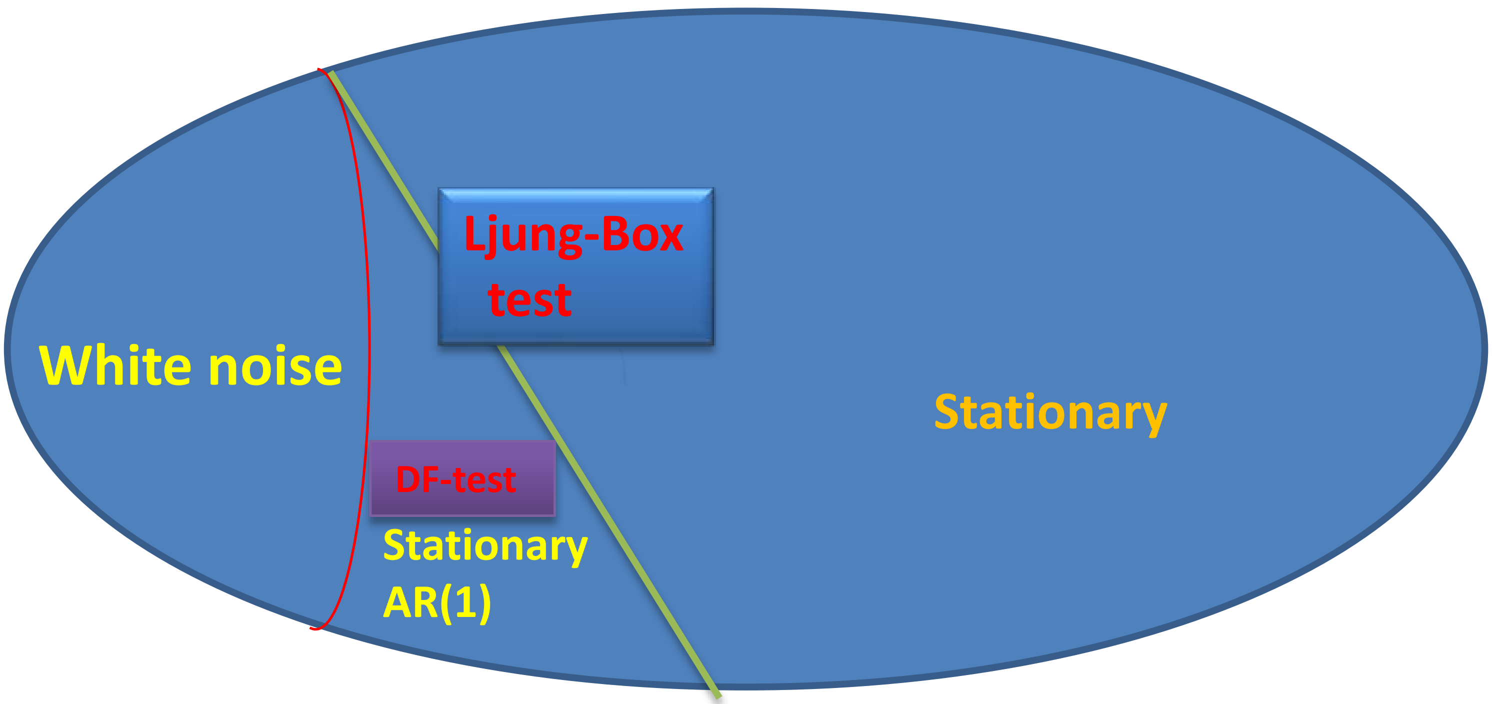

Figure 2.9: In terms of returns, the null hypotheses of both Ljung-Box and Dickey-Fuller tests are the same. However, the alternative of Ljung-Box is larger.

Remarks: Dickey-Fuller and Ljung-Box tests both validate some aspects of the efficient market hypothesis. They are based on very different model assumptions.

— Dickey-Fuller test is a **parametric** test on the log-price process $\{X_t\}$. The basic assumption is that $\{X_t\}$ follows an AR(1) model.

— Ljung-Box test is a **nonparametric** test for uncorrelatedness based on the returns $\{r_t\}$.

The null hypotheses are the same.

2.5 ARMA Processes

Assume that the series $\{X_t\}$ has zero mean (being removed).

ARMA model: $\{X_t\} \sim \text{ARMA}(p, q)$, if

$$X_t = \underbrace{b_1 X_{t-1} + \cdots b_p X_{t-p}}_{\text{obseved returns}} + \varepsilon_t + \underbrace{a_1 \varepsilon_{t-1} + \cdots + a_q \varepsilon_{t-q}}_{\text{relized noises}},$$

where $\{\varepsilon_t\} \sim \mathcal{WN}(0, \sigma^2)$. It can be written as

$$b(B) X_t = a(B) \varepsilon_t,$$

where $b(z) = 1 - b_1 z - \cdots - b_p z^p$, $a(z) = 1 + a_1 z + \cdots + a_q z^q$.

Why ARMA? More parsimonious representation! For example, a simple model $X_t = 0.6 X_{t-1} + \varepsilon_t + 0.3 \varepsilon_{t-1}$ would require a high order of AR or MA process to approximate it.

Identification: The order as defined above is not identifiable:

$$(1 - 0.7B) b(B) X_t = (1 - 0.7B) a(B) \varepsilon_t$$

is ARMA$(p+1, q+1)$. Thus, we assume $a(z)$ and $b(z)$ do not share the same root.

Stationarity: When

$$\inf_{|z|\le 1} |b(z)| > 0,$$

the ARMA model can be written as

$$X_t = b(B)^{-1} a(B)\varepsilon_t = d_0\varepsilon_t + d_1\varepsilon_{t-1} + \cdots ,$$

for some coefficients $\{d_i\}$, an infinite order of MA and hence is stationary.

Yule-Walker equation: Note that

$$\text{Cov}\{b(B)X_t, X_{t-k}\} = \text{Cov}\{a(B)\varepsilon_t, X_{t-k}\} = 0, \quad \text{if} \quad k > q.$$

Thus,

$$\gamma(k) - b_1\gamma(k-1) - \cdots - b_p\gamma(k-p) = 0, \text{for } k > q$$

Short memory: Stationary ARMA model has short memories. In fact, $\rho(k) \to 0$ exponentially fast.

Example 6. Consider the ARMA(1,1) model. For $k > 1$, we have

$$\gamma(k) = b_1\gamma(k-1) = \cdots = b_1^{k-1}\gamma(1).$$

Using $X_t = b_1 X_{t-1} + \varepsilon_t + a_1\varepsilon_{t-1}$, we have

$$\gamma(1) = \text{Cov}(X_t, X_{t-1}) = \text{Cov}(b_1 X_{t-1} + a_1\varepsilon_{t-1}, X_{t-1}) = b_1\gamma(0) + a_1\sigma^2.$$

Note that

$$\gamma(0) = \text{var}(X_t) = b_1^2\gamma(0) + (1+a_1^2)\sigma^2 + 2a_1b_1\sigma^2.$$

Hence,

$$\gamma(0) = \frac{1 + a_1^2 + 2a_1 b_1}{1 - b_1^2}\sigma^2$$

and

$$\gamma(1) = \frac{a_1 + b_1(1 + a_1^2 + a_1 b_1)}{1 - b_1^2}\sigma^2, \qquad \gamma(k) = \gamma(1) b_1^{k-1}, \text{for } k \geq 1$$

Estimation: For given orders p and q, the parameters can be estimated by the **Quasi Maximum Likelihood** (QML) method assuming that $\{\varepsilon_t\} \sim i.i.d \;\; \mathcal{N}(0, \sigma^2)$. Then, $\mathbf{X}_T \equiv (X_1, \cdots, X_T)' \sim \mathcal{N}(0, \mathbf{\Sigma})$. The likelihood function (the density function of $\mathbf{X}_T$) is

$$L(\mathbf{a}, \mathbf{b}, \sigma^2) \propto |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{X}_T' \mathbf{\Sigma}^{-1} \mathbf{X}_T\right\}.$$

Order Selection: Let $\widehat{\ell}(p,q) = \max_{\mathbf{a},\mathbf{b},\sigma} \log L(\mathbf{a},\mathbf{b},\sigma)$. Then

$$\begin{aligned} \mathrm{AIC}(p,q) &= -2\widehat{\ell}(p,q) + \mathbf{2}(p+q+1); \\ \mathrm{AICC}(p,q) &= -2\widehat{\ell}(p,q) + \frac{\mathbf{2T}}{\mathbf{T-p-q-2}}(p+q+1); \\ \mathrm{BIC}(p,q) &= -2\widehat{\ell}(p,q) + \log(\mathbf{T-L})(p+q+1). \end{aligned}$$

Example 7: Consider the daily return of SP500 index. Fitting the ARMA(1,1) model results in

$$(r_t - 0.0399) = -0.3379(r_{t-1} - 0.0399) + \varepsilon_t - 0.359\varepsilon_{t-1}.$$

	ARMA(1,1)	AR(1)	AR(2)	AR(3)	MA(1)	MA(2)	MA(3)
AIC	15962	15961	15952	15945	15964	15957	15952
σ	1.159	1.159	1.157	1.156	1.159	1.157	1.156

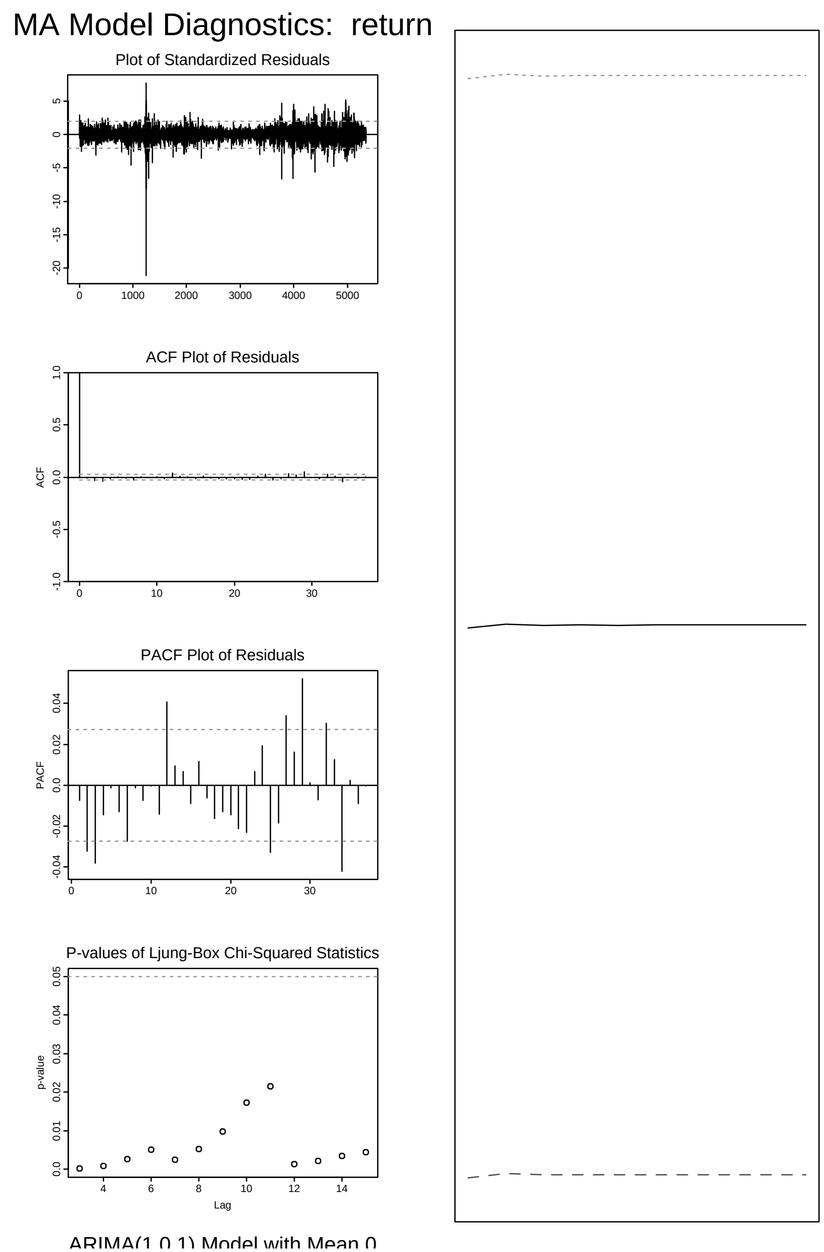

Figure 2.10: ARMA fit to S&P500 returns.

Prediction: Let $X_T(m) = \mathrm{E}_T(X_{T+m})$. Then, from

$$X_{T+m} = \sum_{j=1}^{p} b_j X_{T+m-j} + \varepsilon_{T+m} + \sum_{j=1}^{q} a_j \varepsilon_{T+m-j},$$

we have

$$X_T(m) = \sum_{j=1}^{p} b_j X_T(m-j) + \sum_{j=1}^{q} a_j \varepsilon_T(m-j),$$

where

$$\varepsilon_T(i) = \mathrm{E}_T\{\varepsilon_{T+i}\} = \begin{cases} 0 & \text{if} \quad i > 0 \\ \varepsilon_{T+i} & \text{if} \quad i \le 0 \end{cases}.$$

Thus, starting from the one-step ahead prediction, we compute the two-step ahead prediction and so on.

Example 8. Consider an ARMA(1,1) model

$$X_t - bX_{t-1} = \varepsilon_t - a\varepsilon_{t-1}.$$

For $m \geq 2$, it is easy to see

$$X_T(m) - bX_T(m-1) = 0 \Longrightarrow \widehat{X}_{T+m} \equiv X_T(m) = b^{m-1} X_T(1).$$

The long-term forecasting is nearly impossible.

Let us now consider the one-step forecasting. First of all, $X_T(1) = bX_T - a\varepsilon_T$. Note that

$$\begin{aligned} \varepsilon_t &= (1-aB)^{-1}(X_t - bX_{t-1}) \\ &= \sum_{j=0}^{\infty} a^j B^j (X_t - bX_{t-1}) \\ &= X_t + a\sum_{j=1}^{\infty} a^{j-1} X_{t-j} - b\sum_{j=1}^{\infty} a^{j-1} X_{t-j} \\ &= X_t - (b-a)\sum_{j=1}^{\infty} a^{j-1} X_{t-j}. \end{aligned}$$

Hence,

$$X_T(1) = (b-a)\{X_T + aX_{T-1} + a^2 X_{T-2} + \cdots\}.$$

When $b = 1$ [nonstationary, ARIMA(0,1,1)], we have

$$X_T(1) = (1-a)\{X_T + aX_{T-1} + a^2 X_{T-2} + \cdots\} \quad \text{(exponential smoothing)}$$

2.6 ARIMA model

When a financial time series appears to have a slow-varying **time trend**, a common practice is to remove the trend by differencing. e.g. Let X_t be the log-price of SP500 index, and

$$r_t = X_t - X_{t-1} = (1 - B)X_t$$

be the log-return. We hope to model r_t by an ARMA(p, q) model. Since X_t is the integration of r_t, the model is called an autoregressive integrated moving average (ARIMA) process. Note that the ARIMA model for $(d > 0)$ is non-stationary.

ARIMA: Let $Y_t = (1 - B)^d X_t$. If

$$b(B)Y_t = a(B)\varepsilon_t \Longleftrightarrow b(B)(1 - B)^d X_t = a(B)\varepsilon_t,$$

$\{X_t\}$ is called ARIMA model with order p, d and q, denoted by $\{X_t\} \sim \text{ARIMA}(p, d, q)$.

The techniques for ARMA model can readily be extended.

2.7 Persistence and Long Memory Processes

For ARMA models, $|\rho(k)| \le Cr^k, \quad r < 1, \quad k = 0, 1, 2, \cdots.$ There also exists a class models with

$$\rho(k) \sim Ck^{2d-1}, \quad \text{as} \quad k \to \infty, \quad d < 0.5.$$

For $d \in (0, 0.5), \sum |\rho(k)| = \infty$. Such a process is called a **long memory process**.

e.g. ACF of log-price of S&P500 appears to decay slowly;

e.g. ACF of absolute log-return of S&P500 appears persistent.

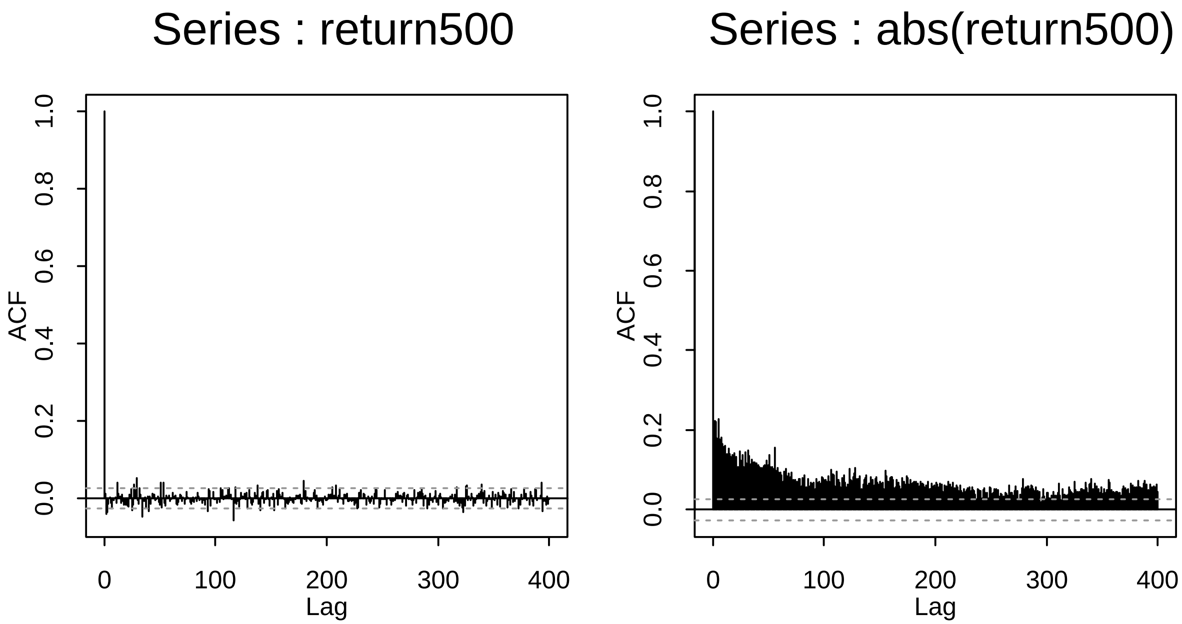

Figure 2.11: ACFs of the log-returns and absolute log-returns of the S&P 500 index.

A family of models with $\rho(k) \sim Ck^{2d-1}$ is given by the **fractional difference**:

$$(1-B)^d X_t = \varepsilon_t, \quad -0.5 < d < 0.5.$$

Here,

$$(1-B)^d = 1 - dB + \frac{d(1-d)}{2!}B^2 - \cdots - (-1)^k \frac{d(1-d)\cdots(k-1-d)}{k!}B^k - \cdots.$$

Thus, X_t admits an AR(∞) representation

$$X_t - dX_{t-1} + \cdots + (-1)^k \frac{d(1-d)\cdots(k-1-d)}{k!}X_{t-k} - \cdots = \varepsilon_t.$$

Properties. For the process defined above,

(i) $\rho(k) = \dfrac{d(1+d)\cdots(k-1+d)}{(1-d)(2-d)\cdots(k-d)} \sim Ck^{2d-1}, \quad k \to \infty.$

(ii) PACF: $\quad \pi(k) = d/(k-d)$.

(iii) The Fourier transform of the ACF admits

$$f(\omega) \sim w^{-2d}, \quad \text{as} \quad \omega \to 0.$$

— the spectral density function

— d can be estimated from log-periodogram with ω small.

FARIMA: In general, if $\{X_t\}$ satisfies

$$b(B)(1-B)^d X_t = a(B)\varepsilon_t,$$

then it is called an FARIMA(p, d, q) $\Longleftrightarrow$ Fractional ARIMA model.

Note that

$$b(B)X_t = a(B)(1-B)^{-d}\varepsilon_t \Longleftrightarrow b(B)X_t = a(B)\eta_t,$$

where $\eta_t = (1-B)^{-d}\varepsilon_t$ is a long-memory process. Thus, FARIMA can be viewed as an ARMA model driven by a long memory noise.

2.8 Seasonal Models

Some financial time series exhibits periodic behavior. e.g. quarterly earning per share of IBM and Johnson and Johnson 1984-2013. The features of earning data include serial correlation and seasonality.

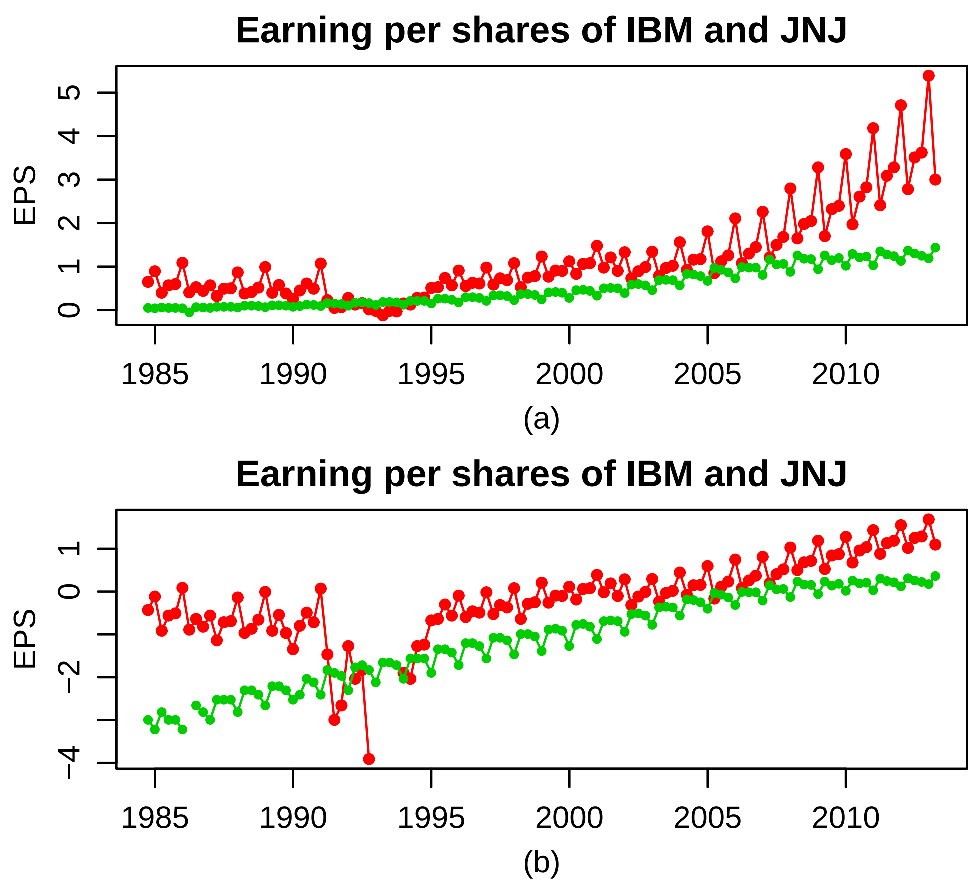

Figure 2.12: Quarterly earnings per share of IBM (red) and Johnson and Johnson (green) from 1984– 2013. Top panel: earning per share; bottom panel: logarithm of earings per share.

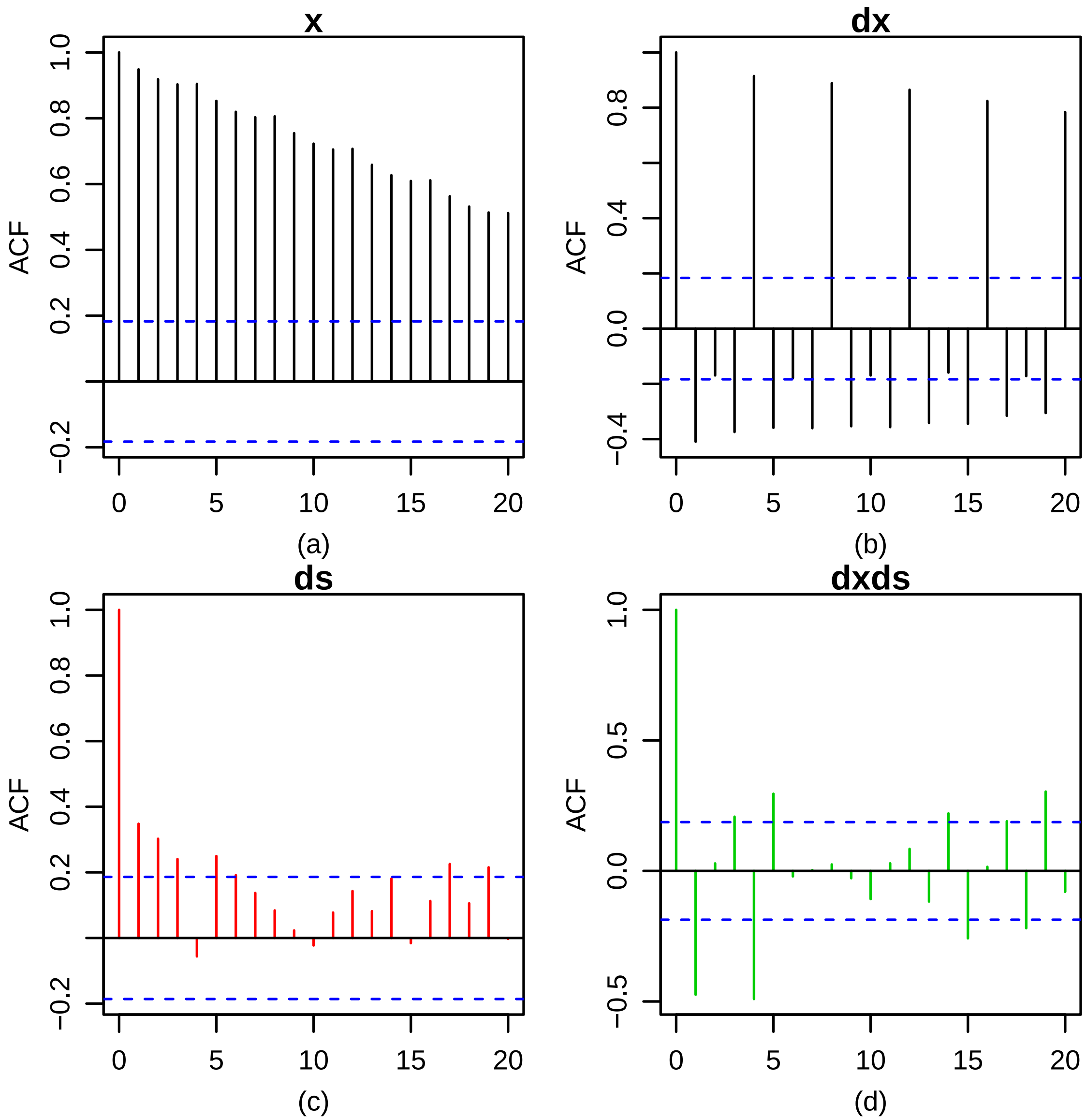

Figure 2.13: Autocorrelation functions for the logarithm of earning per share of Johnson and Johnson from 1984-2013. (a) ACF of the original time series $\{X_t\}$. (b) ACF of de-trend series $\{\nabla X_t\}$; (c) ACF of de-seasonalized series $\{\nabla_4 X_t\}$; (d) ACF of de-trend and de-seasonalized data $\{(1-B^4)(1-B)X_t.\}$

Differencing:

— Serial correlation is weakened by a differencing: $\Delta X_t = X_t - X_{t-1}$. The ACF is given in Fig 2.13, labeled (dx).

— Seasonality is handled by a seasonal difference: $\Delta_4 = (1 - B^4)$. The ACF is given in Fig 2.13, labeled (ds).

— Serial correlation and seasonality are handled by

$$(1 - B^4)(1 - B)X_t = X_t - X_{t-1} - X_{t-4} + X_{t-5}.$$

The resulting series has ACF given in Fig 2.13(d).

After the above preprocessing, we can construct an ARMA model.

2.9 Summary

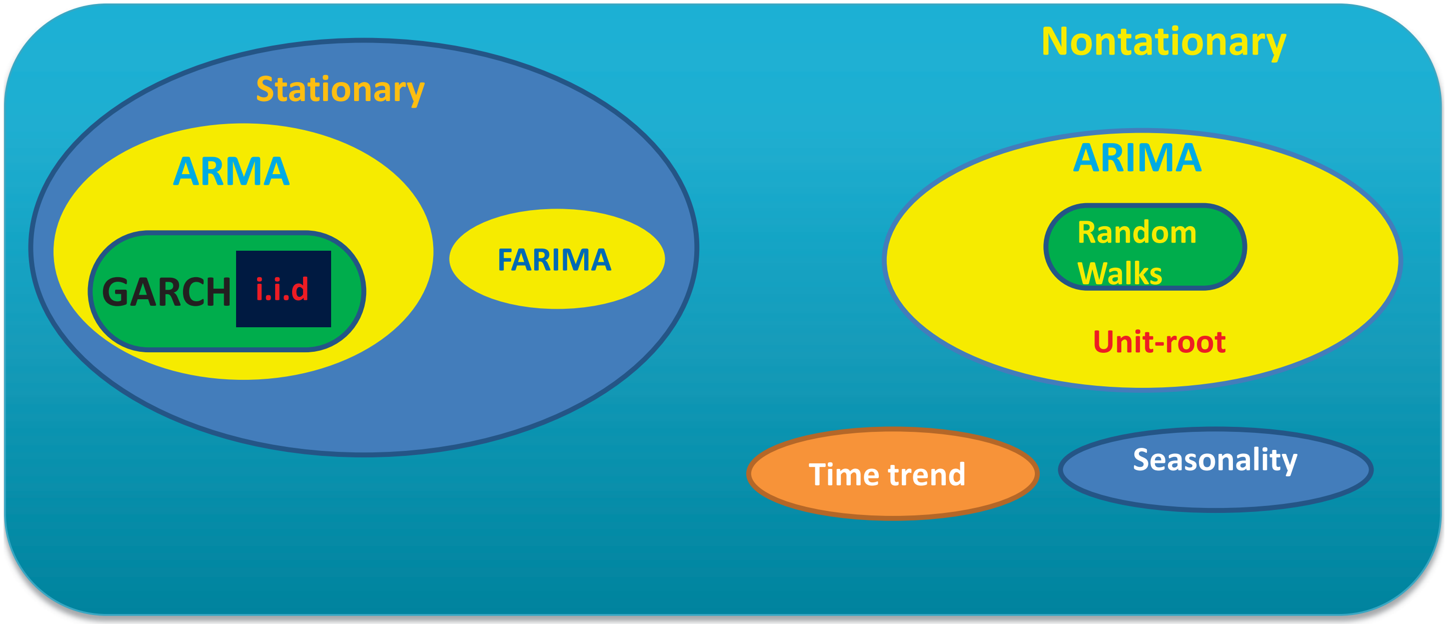

Figure 2.14: Summary of stochastic processes introduced in the chapter.

Stochastic processes: Stationary + nonstationary. Both classes are too wide to be useful.

Specification of Stationary: ★Stationary ARMA (short memory)★FARIMA (long memory).

White Noise: is a special stationary process.

In ARMA models, the white noise process is not specified but will be further specified as the GARCH for volatility modeling. See Figure 2.15 for further classification of white noise process.

Specification of Nonstationarity. ARIMA model (unit roots) with the random walk as a specific example. Time series with seasonality and trend can not be stationary.

ARIMA models: the difference of time series is ARMA

★ MA models are always stationary and are strong stationary if noises are independent;

★ ARMA models are stationary if roots outside unit circle

★ Yule-Walker equation for acf functions

★ ACF and PACF for model identification; AIC/BIC for model selection

★ prediction and prediction errors; •iterative AR(p) or ARMA versus direct least-squares

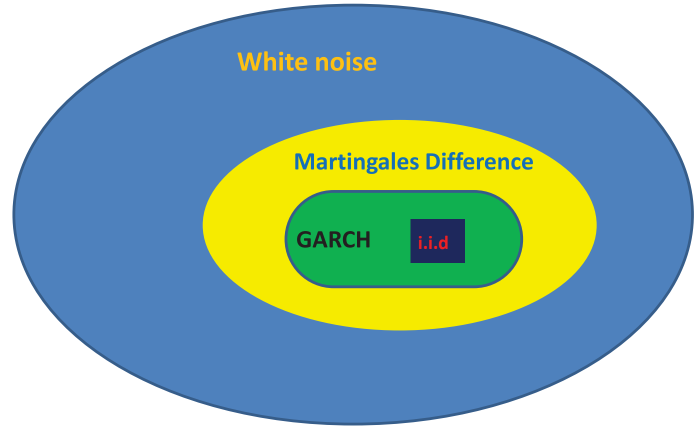

Figure 2.15: Relationship among different white noise processes: Margingales, GARCH, and i.i.d. white noises processes.

Statistics:

★ MLE gives the most efficient estimator, but not robust to model misspecifications

★ Hessian matrix gives precision matrix of parameter estimation

★ Delta method gives standard errors of nonlinear functionals of parameters