Chapter Three

Classical Univariate Linear Time Series Analysis

Tingguo Zheng (zhengtg@gmail.com)
Department of Statistics and Data Science, School of Economics, Xiamen University

(For Undergraduate Program, 2026-2027, Spring Semester)

- *Reference: “Analysis of Financial Time Series (3ed)” by Ruey S. Tsay, 2010. (Chapter 2)*

- **Outline:**
  - AR models
  - ARMA models
  - Model buliding
  - Nonstationarity and unit root testing
  - ARFIMA models
  - Seasonal ARMA models
  - Others

1. Linear time series models

1.1 Linear process

Definition 1 (Wold's decomposition theorem (Fuller, 1996)). Any covariance stationary time series $\{y_t\}$ has a linear process or infinite order *moving average* representation of the form

$$y_t = \mu + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}, \quad \psi_0 = 1, \quad \sum_{k=0}^{\infty} \psi_k^2 < \infty,$$

where ε_t is followed by the white noise process, i.e., $\varepsilon_t \sim WN(0, \sigma^2)$.

- The moving average weights are also called *impulse responses* since

$$\frac{\partial y_t}{\partial \varepsilon_{t-s}} = \frac{\partial y_{t+s}}{\partial \varepsilon_t} = \psi_s, \quad s = 1, 2, \ldots .$$

- For a stationary and ergodic time series $\lim_{s\to\infty}\psi_s = 0$ and the *long-run cumulative impulse response* $\sum_{s=0}^{\infty}\psi_s < \infty$.

- A plot of ψ_s against s is called the *impulse response function (IRF)*.

In the Wold form, it can be shown that

- (1) $E[y_t] = \mu$;

- (2) $\gamma_0 = \mathrm{Var}(y_t) = \sigma^2 \sum_{k=0}^{\infty}\psi_k^2$;

- (3) $\gamma_j = \mathrm{Cov}(y_t, y_{t-j}) = \sigma^2 \sum_{k=0}^{\infty}\psi_k\psi_{k+j}$;

- (4) $\rho_j = \frac{\sum_{k=0}^{\infty}\psi_k\psi_{k+j}}{\sum_{k=0}^{\infty}\psi_k^2}$.

1.2 ARMA model

The general Wold form is handy for theoretical analysis but is not practically useful for estimation purposes.

A very rich and practically useful class of *stationary* and *ergodic* processes is the *autoregressive-moving average (ARMA)* class of models made popular by Box and Jenkins (1976), which has the form

$$y_t - \mu = \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \quad \varepsilon_t \sim WN(0, \sigma^2).$$

- Define $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, the ARMA($p$, q) model may be compactly expressed using *lag polynomials*

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t.$$

- Similarly, the *Wold representation* in lag operator notation is

$$y_t = \mu + \psi(L)\varepsilon_t,$$

$$\psi(L) = \sum_{k=0}^{\infty} \psi_k L^k, \quad \psi_0 = 1$$

$$\psi(1) = \sum_{k=0}^{\infty} \psi_k,$$

where the lag polynomial $\psi(L) = \frac{\theta(L)}{\phi(L)}$ for the ARMA models.

- The ARMA(p, q) model is said to be *causal*, if the time series $\{y_t\}$ can be written as a one-sided linear process:

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \mu + \psi(L)\varepsilon_t,$$

where $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$; we set $\psi_0 = 1$.

2 Autoregressive Models

Definition 2 (Autoregressive (AR) Processes). If y_t satisfies

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

where c is an intercept term, and the ϕ_i are coefficients. $\{\varepsilon_t\}$ is assumed to be a white noise series with mean zero and variance σ^2, i.e., $\varepsilon_t \sim WN(0, \sigma^2)$. The process of y_t is called *an autoregressive process of order p*, denoted by *AR(p)*.

- In *lag operator* notation. *Let* $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, then

$$\phi(L) y_t = c + \varepsilon_t.$$

2.1 AR(1) Model

The stationary AR(1) process is given as

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim WN(0, \sigma^2)$ and $|\phi_1| < 1$, an intercept c is included to describe a nonzero mean of y_t.

- Stationarity: necessary and sufficient condition $|\phi_1| < 1$. Why?

- Mean: taking expectations,

$$E[y_t] = c + \phi_1 E[y_{t-1}] + E[\varepsilon_t] \quad \Rightarrow \quad \mu = \frac{c}{1 - \phi_1}. \tag{2}$$

- Mean corrected AR(1) model:

$$y_t - \mu = \phi_1 (y_{t-1} - \mu) + \varepsilon_t.$$

- MA representation or Wold representation: by *repeated substitution*

$$y_t - \mu = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \cdots = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}. \tag{3}$$

  - This implies that *y_{t-1} and ε_t are uncorrelated*, i.e., $\mathrm{Cov}((y_{t-1} - \mu), \varepsilon_t) = 0$.
  - The IRF decay at a *geometric rate*.

- Variance: the variance of this series is

$$\mathrm{Var}(y_t) = \phi_1^2 \mathrm{Var}(y_{t-1}) + \sigma^2 \quad \Rightarrow \quad \mathrm{Var}(y_t) = \gamma_0 = \sigma_y^2 = \frac{\sigma^2}{1 - \phi_1^2}.$$

- Autocorrelations:

$$\gamma_0 = \frac{\sigma^2}{1 - \phi_1^2}, \quad \gamma_j = \phi_1 \gamma_{j-1}, \quad \text{for } j = 1, 2, \ldots,,$$

$$\rho_j = \phi_1 \rho_{j-1} = \phi_1^j, \quad \text{for } j = 1, 2, \ldots, \text{(See Figure 1)}$$

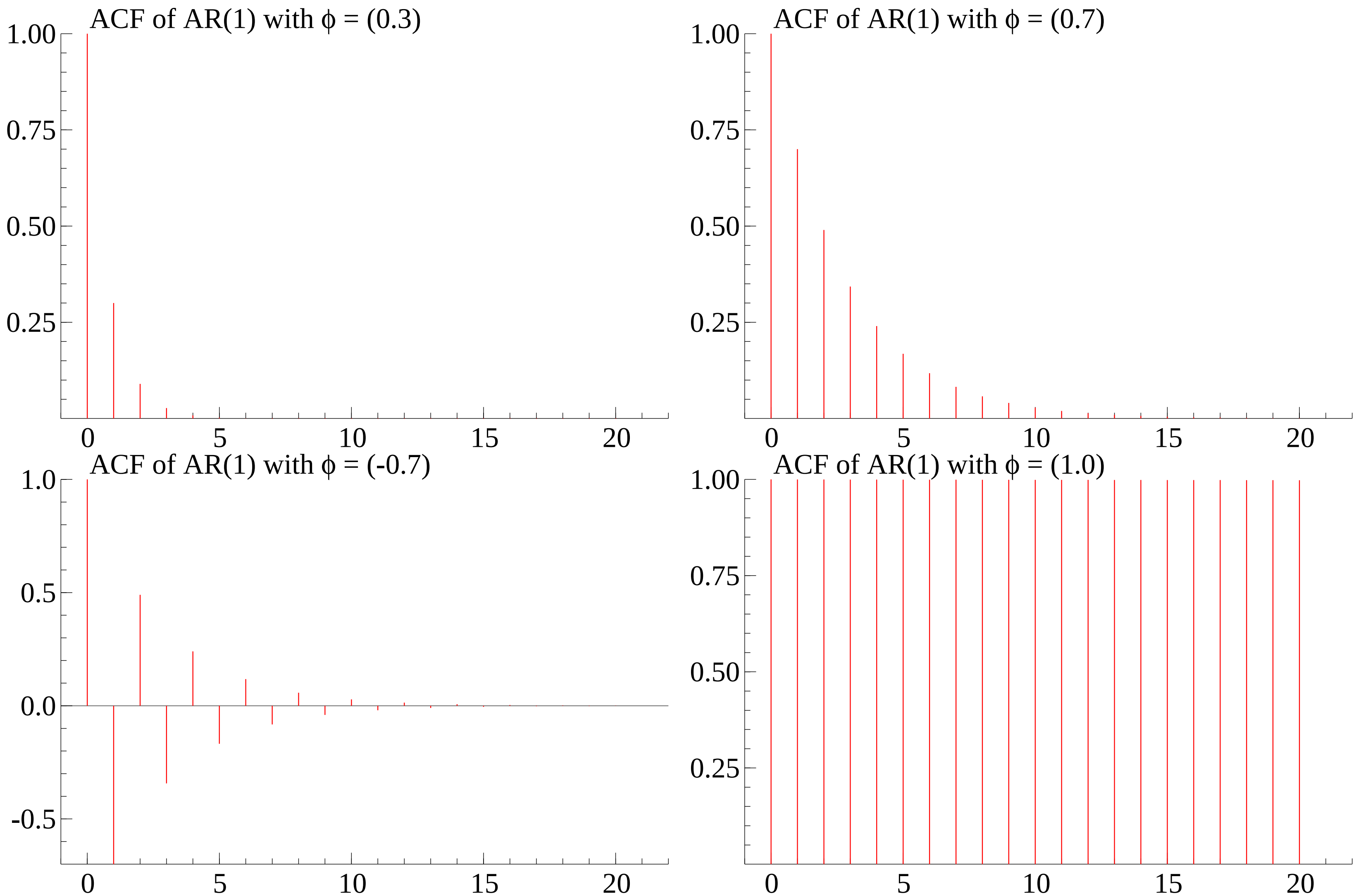


Figure 1 The ACF of AR(1) model.

- This stationary series $\{y_t\}$ exhibits *mean-reverting behavior*. Let $z_t = y_t - \mu$,

$$\Delta y_t = c + (\phi_1 - 1) y_{t-1} + \varepsilon_t \quad \Rightarrow \quad \Delta z_t = (\phi_1 - 1) z_{t-1} + \varepsilon_t,$$

$$\begin{cases} E[\Delta z_t | I_{t-1}] < 0, & \text{if } z_{t-1} > 0; \\ E[\Delta z_t | I_{t-1}] > 0, & \text{if } z_{t-1} < 0. \end{cases}$$

AR(1) Model with $\phi_1 = 1$

Consider again the AR(1) process. Rewrite it by *recursive substitution* as

$$y_t = \phi_1^t y_0 + \sum_{i=0}^{t-1} \phi_1^i c + \sum_{i=0}^{t-1} \phi_1^i \varepsilon_{t-i}, \tag{4}$$

from which it follows that $E[y_t] = \phi_1^t y_0 + \sum_{i=0}^{t-1} \phi_1^i c$.

- If $\phi_1 > 1$, y_t is *explosive*, in the sense that y_t *diverges to* $\pm\infty$.
- When $\phi_1 = 1$ and assuming $c = 0$, (4) can be rewritten as the RW process

$$y_t = y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i} = y_0 + \sum_{i=1}^{t} \varepsilon_i. \tag{5}$$

  From this expression it follows that $\gamma_{0,t} \equiv E[(y_t - y_0)^2] = t\sigma^2$ and in general $\gamma_{j,t} \equiv E[(y_t - y_0)(y_{t-j} - y_0)] = (t-j)\sigma^2$ and $\rho_{j,t} = (t-j)/t$ for all $k \geq 0$. See the lower right panel of Figure 1.

- Similarly, when $\phi_1 = -1$, it is not covariance stationarity.

2.2 AR(2) Model

The AR(2) process is given by

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \tag{6}$$

where $\varepsilon_t \sim WN(0, \sigma^2)$.

- Mean (given stationarity): $\mu = \dfrac{c}{1 - \phi_1 - \phi_2}$, provided that $\phi_1 + \phi_2 \neq 1$.
- Mean corrected AR(2) model: Using $c = (1 - \phi_1 - \phi_2)\mu$, we can rewrite the AR(2) model as
$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \varepsilon_t$$
- Moment equation: $\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}$ for $j > 0$. What is about $j = 0$?
- ACF: $\rho_0 = 1$, $\rho_1 = \phi_1/(1 - \phi_2)$, and $\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}$ for $j > 1$.

- Yule-Walker equation: $\rho_1 = \phi_1 + \phi_2\rho_1$ and $\rho_2 = \phi_1\rho_1 + \phi_2$, or equivalently,

$$\begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}.$$

- MA representation

$$y_t = \frac{c}{1-\phi_1-\phi_2} + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \psi_3\varepsilon_{t-3} + \cdots$$

where $1 + \psi_1 L + \psi_2 L^2 + \cdots = \frac{1}{1-\phi_1 L-\phi_2 L^2}$. How to calculate ψ_i?

- Second-order difference equation: $\rho_j - \phi_1\rho_{j-1} - \phi_2\rho_{j-2} = 0$, denoted by $(1 - \phi_1 B - \phi_2 B^2)\rho_j = 0$, where B is the *back-shift operator* such that $B\rho_j = \rho_{j-1}$.

- Second-order polynomial equation: $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0$.
  - The solutions of $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0$ are $\omega_{1,2} = \dfrac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$.

- In the time series literature, inverses of the two solutions are referred to as the characteristic roots of the AR(2) model. If denote λ_1 and λ_2 the inverse of ω_1 and ω_2 respectively, then

$$(1-\phi_1 B-\phi_2 B^2)\rho_j=(1-\lambda_1 B)(1-\lambda_2 B)\rho_j=(1-\omega_1^{-1}B)(1-\omega_2^{-1}B)\rho_j.$$

- The ACF is then a mixture of two exponential decays as the general solution of $(1-\phi_1 B-\phi_2 B^2)\rho_j=0$ has the following form:

$$\rho_j=c_1\lambda_1^j+c_2\lambda_2^j.$$

  where c_1 and c_2 are unknown parameters depending on the initial conditions.
- Consider the following two cases:
  1. If both roots ($\lambda_{1,2}$ or $\omega_{1,2}$) are real, then the ACF of AR(2) is a mixture of two *exponential decay*.
  2. If the roots are a complex conjugate pair, i.e. $\phi_1^2+4\phi_2<0$, then ACF would show a picture of *damping sine* and *cosine waves*; such roots usually give rise to the behavior of business cycles.

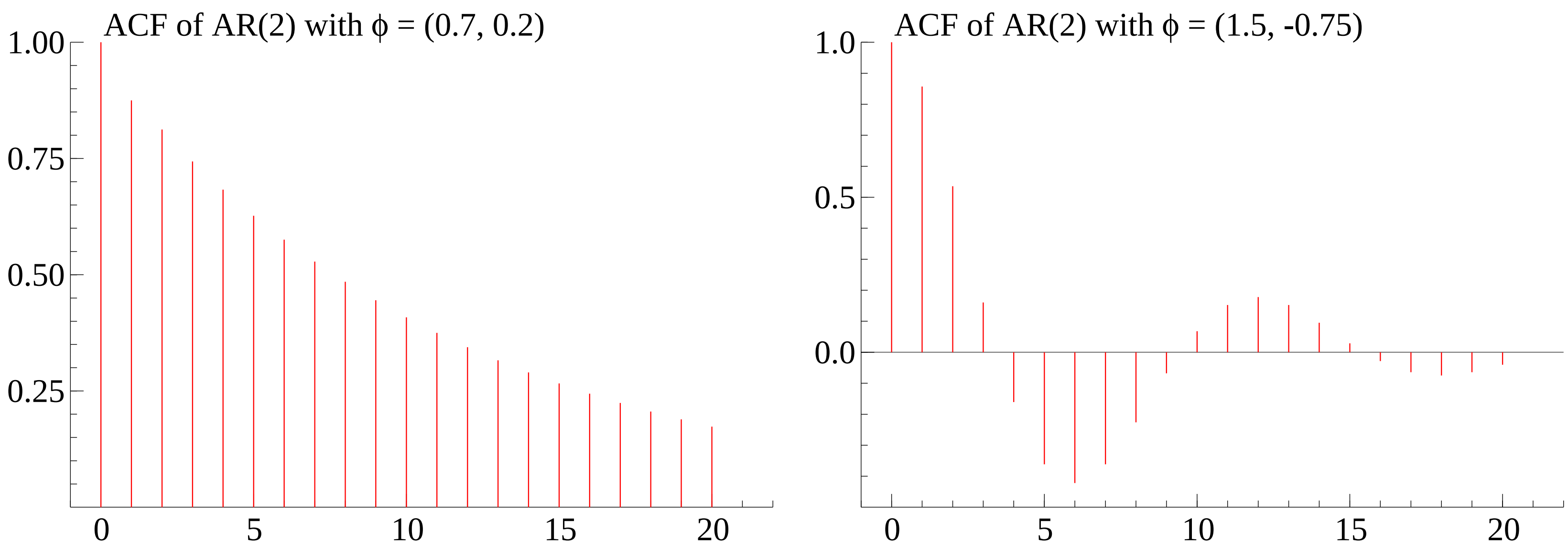


Figure 2 The ACF of AR(2) model.

– For an AR(2) model with a pair of complex characteristic roots, the *average length* of the *stochastic cycle* is

$$T_0 = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]}$$

where the cosine inverse is stated in *radians (弧度)*.

- **Stationarity condition:** all roots of $\phi(z) = 0$ lie outside the unit circle.

- Consider the AR(2) process

$$(1 - \phi_1 L - \phi_2 L^2)(y_t - \mu) = (1 - \lambda_1 L)(1 - \lambda_2 L)(y_t - \mu) = \varepsilon_t.$$

This needs an expansion of $(1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}$. Assuming that $\lambda_1 \neq \lambda_2$, we have

$$\frac{1}{(1-\lambda_1 z)(1-\lambda_2 z)} = \frac{1}{\lambda_2 - \lambda_1}\left(\frac{\lambda_2}{1-\lambda_2 z} - \frac{\lambda_1}{1-\lambda_1 z}\right).$$

1. If $|\lambda_1| < 1$ and $|\lambda_2| < 1$ (stationary condition), the expansion is

$$\frac{1}{(1-\lambda_1 z)(1-\lambda_2 z)} = \frac{1}{\lambda_2 - \lambda_1}\left(\lambda_2 \sum_{j=0}^{\infty} \lambda_2^j z^j - \lambda_1 \sum_{j=0}^{\infty} \lambda_1^j z^j\right)$$

which leads to the *causal solution*

$$y_t = \mu + \frac{1}{\lambda_2 - \lambda_1}\left[\sum_{j=0}^{\infty}(\lambda_2^{j+1} - \lambda_1^{j+1})\varepsilon_{t-j}\right].$$

2. If $|\lambda_1| > 1$ and $|\lambda_2| < 1$, the expansion is

$$\frac{1}{(1-\lambda_1 z)(1-\lambda_2 z)} = \frac{1}{\lambda_2 - \lambda_1}\left\{\frac{\lambda_2}{1-\lambda_2 z} - \frac{\lambda_1}{(\lambda_1 z)[(\lambda_1 z)^{-1} - 1]}\right\}$$
$$= \frac{1}{\lambda_2 - \lambda_1}\left[\lambda_2 \sum_{j=0}^{\infty} \lambda_2^j z^j + z^{-1} \sum_{j=0}^{\infty} \lambda_1^{-j} z^{-j}\right]$$

which leads to the *non-causal solution*

$$y_t = \mu + \frac{1}{\lambda_2 - \lambda_1}\left(\sum_{j=0}^{\infty} \lambda_2^{j+1} \varepsilon_{t-j} + \sum_{j=0}^{\infty} \lambda_1^{-j} \varepsilon_{t+1+j}\right).$$

Example 1. Assume the AR(2) process is

$$y_t = 0.6y_{t-1} + 0.2y_{t-2} + \varepsilon_t \quad \text{or} \quad (1 - 0.6L - 0.2L^2)y_t = \varepsilon_t$$

We first solve the polynomial

$$z^2 + 3z - 5 = 0,$$

and get two roots $z_1 = 1.2926$ and $z_2 = -4.1925$. Recall that $\lambda_1 = \frac{1}{z_1} = 0.84$ and $\lambda_2 = \frac{1}{z_2} = -0.24$. So we can factorize the lag polynomial to be:

$$\begin{aligned}
(1 - 0.6L - 0.2L^2)y_t &= (1 - 0.84L)(1 + 0.24L)y_t \\
\Rightarrow \quad y_t &= (1 - 0.84L)^{-1}(1 + 0.24L)^{-1}\varepsilon_t \\
&= \frac{1}{-0.24 - 0.84}\left[\sum_{j=0}^{\infty}[(-0.24)^{j+1} - 0.84^{j+1}]\varepsilon_{t-j}\right].
\end{aligned}$$

Example 2. For a fitted AR(2) model to one monthly data series denoted by

$$y_t - 1.3512y_{t-1} + 0.4612y_{t-2} = \varepsilon_t.$$

Clearly, $\phi_1^2 + 4\phi_2 = 1.3512^2 - 4 \times 0.4612 = -0.0191 < 0$, which implies the existence of stochastic cycles.

The *average length* of the stochastic cycles for this AR(2) model is approximately

$$T_0 = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]} = 61.71 \quad \text{months},$$

which is about 5 years.

2.3 AR(p) Model

The AR(p) process is

$$(1 - \phi_1 L - \cdots - \phi_p L^p) y_t = c + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2).$$

- **Stationary condition:** It can be shown that the AR(p) is *stationary* and *ergodic* provided the roots of the p-order *polynomial equation*

$$1 - \phi_1 z - \cdots - \phi_p z^p = 0, \quad \text{or} \quad (1 - \alpha_1 z) \cdots (1 - \alpha_p z) = 0 \tag{7}$$

  lie *outside* the complex unit circle (have modulus greater than one).

  – A *necessary condition* for stationarity that is useful in practice is that $|\phi_1 + \cdots + \phi_p| < 1$.

  – When the largest of the α_is is equal to 1, i.e., $z = 1$, we say that the AR(p) process has a unit root.

- **Unconditional mean** (given stationarity): $\mu = c/(1 - \phi_1 - \cdots - \phi_p)$ and thus $\phi(L)(y_t - \mu) = \varepsilon_t$.

- **Autocovariance function**: For simplicity, assume that $c = \mu = 0$. Then, multiplying both sides of the AR model by y_{t-j}, and taking expectation, we have

$$E[(y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p})y_{t-j}] = E(y_{t-j}\varepsilon_t).$$

  Therefore we have

$$\gamma_j - \phi_1\gamma_{j-1} - \cdots - \phi_p\gamma_{j-p} = \begin{cases} \sigma^2, & \text{for } j = 0; \\ 0, & \text{for } j > 0. \end{cases} \tag{8}$$

- **Autocorrelation function**: From autocovariance function, we have

$$\rho_j - \phi_1\rho_{j-1} - \cdots - \phi_p\rho_{j-p} = \begin{cases} \frac{\sigma^2}{\gamma_0}, & \text{for } j = 0; \\ 0, & \text{for } j > 0. \end{cases}$$

The ACF of a stationary AR(p) model would show a mixture of damping sine and cosine patterns and exponential decays depending on the nature of its characteristic roots. See Figure 2.

- **Yule-Walker equation**: Consider the above equations jointly for $j = 1, 2, \ldots, p$, we have the p-order Yule-Walker equation

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{2-p} & \rho_{1-p} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{3-p} & \rho_{2-p} \\ \cdot & \cdot & \cdot & \ddots & \cdot & \cdot \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}.$$

- **MA representation**: The AR model can be written as the MA(∞) representation

$$y_t - \mu = \frac{1}{\phi(L)}\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i},$$

where the ψ_i's are referred to as the ψ-weight of the model and also called the impulse response function. The ψ-weights can be obtained as follows:

$$\frac{1}{\phi(L)} = 1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \cdots$$

$$1 = (1 - \phi_1 L - \cdots - \phi_p L^p)(1 + \psi_1 L + \psi_2 L^2 + \cdots)$$

$$\psi_1 = \phi_1$$

$$\psi_2 = \phi_1 \psi_1 + \phi_2$$

$$\psi_3 = \phi_1 \psi_2 + \phi_2 \psi_1 + \phi_3$$

$$\vdots$$

$$\psi_p = \phi_1 \psi_{p-1} + \phi_2 \psi_{p-2} + \cdots + \phi_{p-1} \psi_1 + \phi_p$$

$$\psi_j = \sum_{i=1}^{p} \phi_i \psi_{j-i}, \quad \text{for } j > p.$$

Again, the ψ-weights satisfying the difference equation $\phi(L)\psi_j = 0$ for $j > p$, so that they also decay exponentially to zero as j goes to infinity.

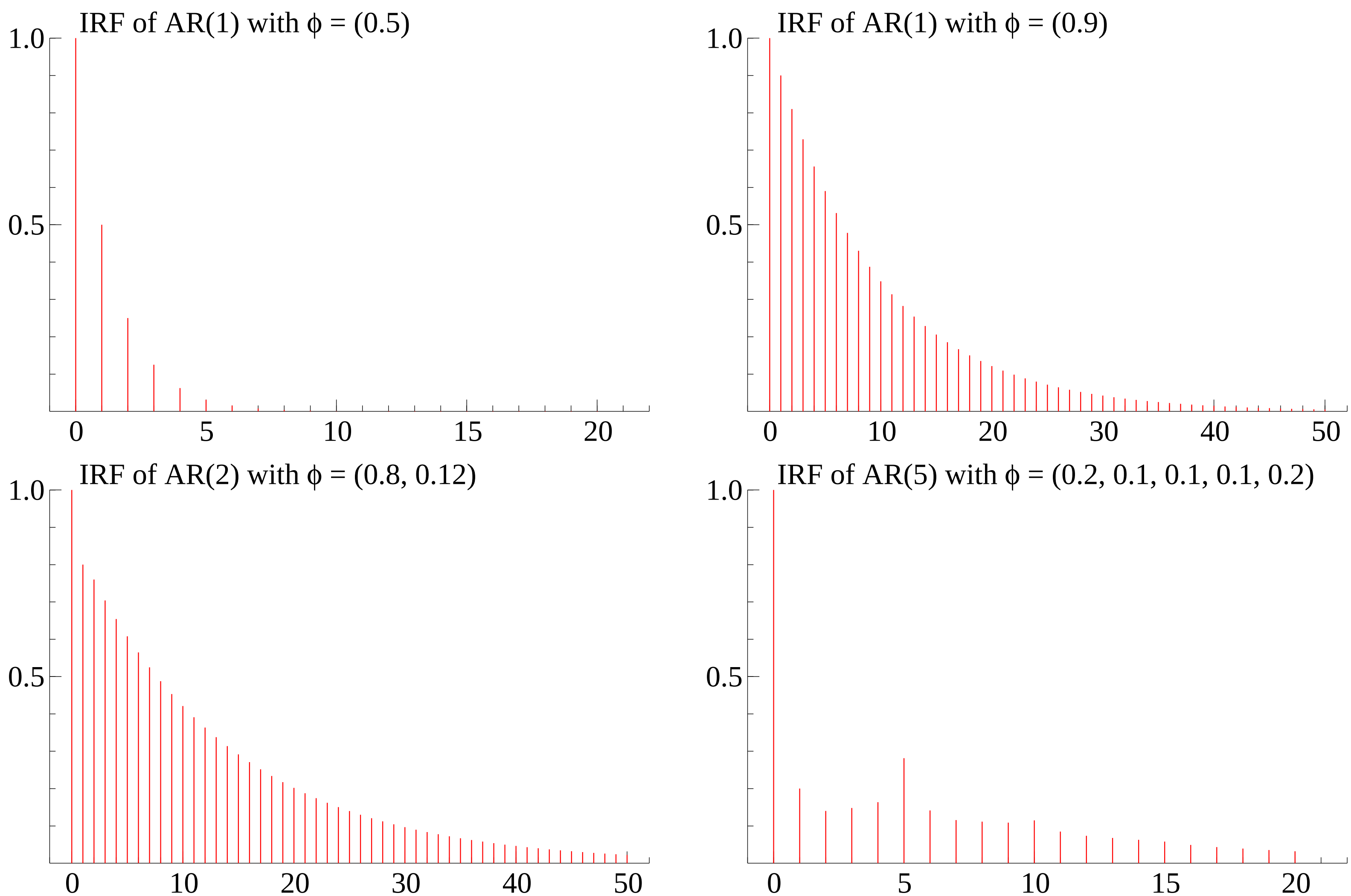


Figure 3 Impulse response functions for AR processes with different lags

2.4 Identifying AR Models in Practice – Order determination

(1) Partial Autocorrelation Function

Definition 3. The *partial autocorrelation function (PACF)* is defined by the following conditional correlation

$$P_k = \phi_{kk} = \mathrm{corr}(y_t, y_{t+k} | y_{t+1}, \ldots, y_{t+k-1}).$$

This definition is equivalent to say that

$$\begin{aligned} P_k &= \mathrm{corr}(y_t - E(y_t | y_{t+1}, \ldots, y_{t+k-1}), y_{t+k} - E(y_{t+k} | y_{t+1}, \ldots, y_{t+k-1})) \\ &= \mathrm{corr}(y_t - \hat{y}_t, y_{t+k} - \hat{y}_{t+k}). \end{aligned}$$

On linear regression theory, $\hat{y}_t$ and $\hat{y}_{t+k}$ are the best linear estimates of y_t and y_{t+k} (respectively) based on the values of $y_{t+1}, \ldots, y_{t+k-1}$.

Computing the PACF

According to the above definitions, the partial autocorrelation coefficient of order k is computed as the plim of the least squares estimator of the coefficient ϕ_{kk} in

$$y_t = \phi_{k1} y_{t-1} + \cdots + \underbrace{\phi_{kk}}_{P_k} y_{t-k} + \varepsilon_t \tag{9}$$

where y_t is assumed to be zero mean.

How to estimate ϕ_{kk}? The OLS estimator is not suitable.

Multiplying both sides of Eq. (9) by y_{t-j} ($j > 0$), and taking expectation, we can get

$$\gamma_j - \phi_{k1}\gamma_{j-1} - \phi_{k2}\gamma_{j-2} - \cdots - \phi_{kk}\gamma_{j-k} = 0.$$

Consider the above (moment) equations jointly for $j = 1, 2, \ldots, k$, we have

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}.$$

Or in reduced form

$$\boldsymbol{\gamma}_k = \Gamma \boldsymbol{\phi}_k.$$

Similarly, we can replace all γs with ρs, the corresponding result is

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_0 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}.$$

By the Cramer rule, we have

$$\phi_{kk} = \frac{\det \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & \rho_0 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{bmatrix}}{\det \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_0 \end{bmatrix}}$$

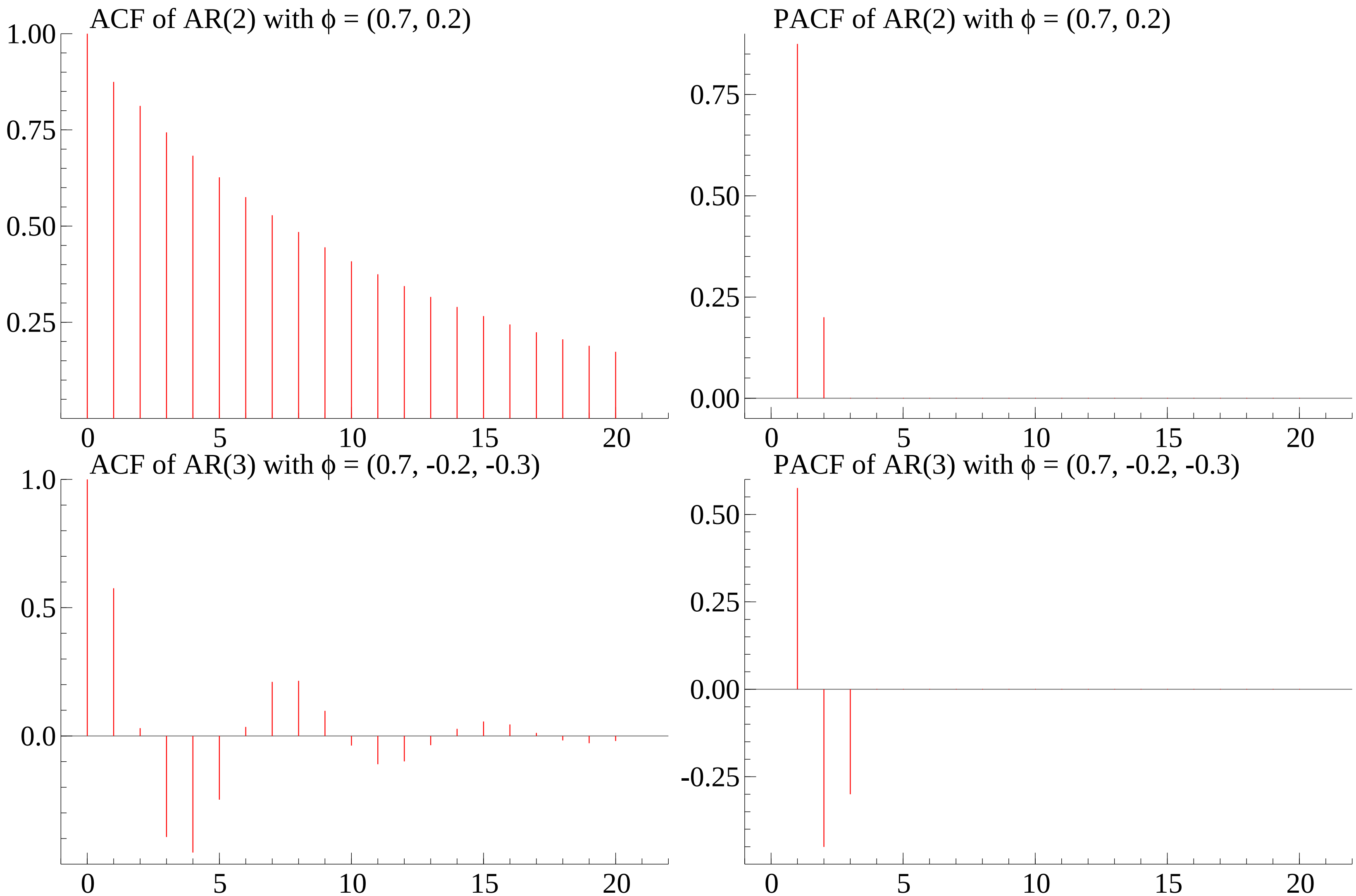


Figure 4 The PACF of AR(p) model.

Levinson-Durbin Algorithm*

A recursive way of computing the regression coefficients in equation (9) from the autocorrelation function was given by Levinson (1947) and Durbin (1960).

The Durbin-Levinson algorithm updates the coefficients the from $k-1$st order model to those of the kth order model as follows:

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j} \tag{10}$$

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j} \tag{11}$$

for $j = 1, \ldots, k-1$.

Sample PACF

- Suppose we have observed a time series $y_1, \ldots, y_n$;
- We can use series to estimate $\gamma_0, \ldots, \gamma_{n-1}$ using standard estimators $\hat{\gamma}_0, \ldots, \hat{\gamma}_{n-1}$.
- We can plug these into Γ and $\boldsymbol{\gamma}_k$ to obtain estimators

$$\hat{\Gamma} = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \gamma_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0 \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{\gamma}}_k = \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{bmatrix}.$$

- Since $\hat{\Gamma}$ is positive definite, we have $\hat{\boldsymbol{\phi}}_k = \hat{\Gamma}^{-1}\hat{\boldsymbol{\gamma}}_k$, where the last element of $\hat{\boldsymbol{\phi}}_k$ is $\hat{\phi}_{kk}$, which defines sample SPACF.

PACF for AR(p) Processes

- PACF is a useful tool to help identify the AR(p) model.
- Ramsey (1974): if $\phi_{p,p} \neq 0$ and $\phi_{j,j} = 0$ for $j > p$ for any Gaussian process, it is necessarily an AR(p) process.
- Its asympotic distribution is $\sqrt{n}\hat{\phi}_{kk} \overset{A}{\sim} N(0,1)$.
- Comparing $\hat{\phi}_{k,k}$ to $\pm 1.96/\sqrt{n}$ and decide if $\hat{\phi}_{k,k}$ is preferable.

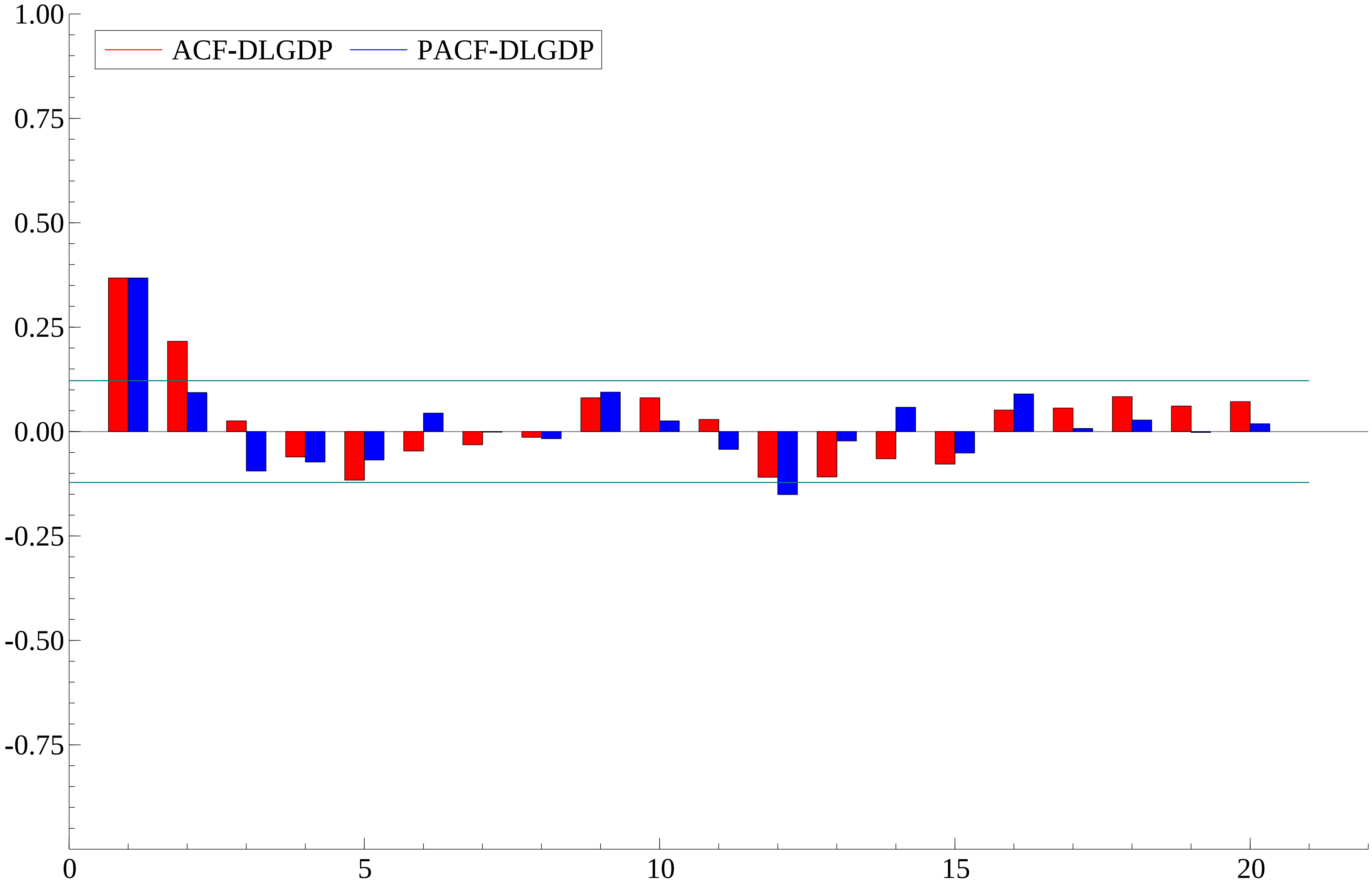


Figure 5 The PACF of US real GDP growth rate (1947Q1-2014Q1).

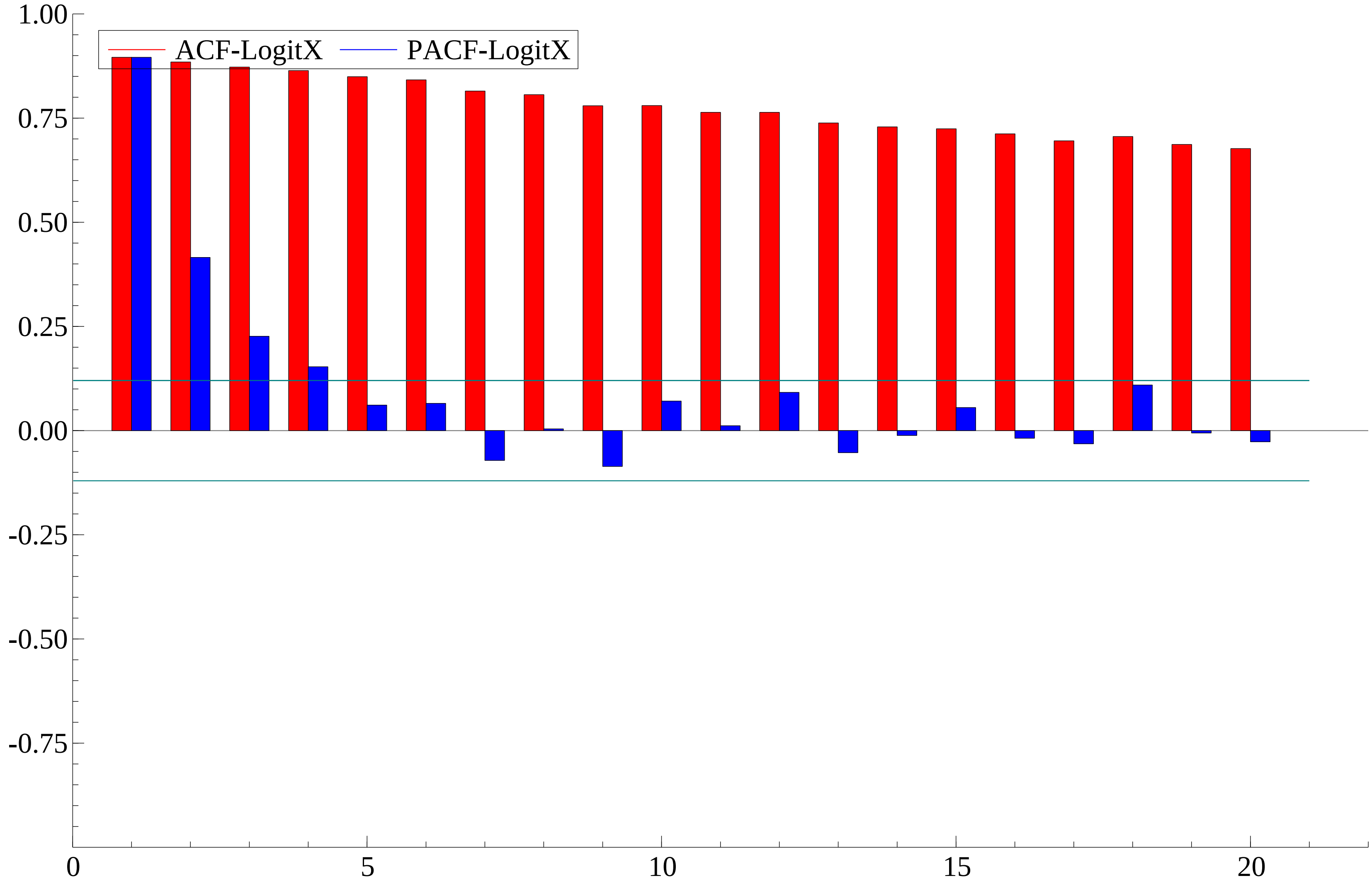


Figure 6 The PACF of weekly market share data of Crest toothpaste (January 1958 - April 1963) (with logit transformation).

(2) Information Criteria or Model Selection Criteria

The idea is to fit all AR(p) models with orders $p \leq p_{\max}$ and choose the values of p which minimizes some *model selection criteria*.

Model selection criteria for AR(p) models under Gaussian assumption have the form

$$\text{MSC}(p) = \ln(\tilde{\sigma}^2(p)) + c_T \cdot \psi(p),$$

- $\tilde{\sigma}^2(p)$ is the least square (LS) estimation of $\mathrm{Var}(\varepsilon_t) = \sigma^2$ without a degrees of freedom correction from the AR(p) model

- c_T is a sequence indexed by the sample size T

- $\psi(p)$ is a *penalty function* which penalizes large AR(p) models.

The two most common *information criteria* are the *Akaike (AIC)* and *Schwarz-Bayesian (BIC)*: (Akaike, 1973; Schwarz, 1978)

$$\begin{aligned} \text{AIC}(p) &= \ln(\tilde{\sigma}^2(p)) + \frac{2p}{T}, \\ \text{BIC}(p) &= \ln(\tilde{\sigma}^2(p)) + \frac{p \ln T}{T}. \end{aligned}$$

- The BIC estimates the order *consistently* under fairly general conditions if the true orders p are less than or equal to $p_{\max}$.

- The AIC criterion asymptotically *overestimates* the order, since the penalty term is smaller for AIC than BIC ($2 < \ln n$).

Example 3. Consider the monthly simple returns of CRSP value-weighted index from January 1926 to December 2008, with $T = 996$ observations. The following table gives the first 12 lags of sample PACF and information criteria of the series.

p	1	2	3	4	5	6
PACF	0.115	-0.030	$-\mathbf{0.102}$	0.033	0.062	-0.050
AIC	-5.838	-5.837	-5.846	-5.845	-5.847	-5.847
BIC	$-\mathbf{5.833}$	-5.827	-5.831	-5.825	-5.822	-5.818
p	7	8	9	10	11	12
PACF	0.031	0.052	**0.063**	0.005	-0.005	0.011
AIC	-5.846	-5.847	$-\mathbf{5.849}$	-5.847	-5.845	-5.843
BIC	-5.812	-5.807	-5.805	-5.798	-5.791	-5.784

- PACF order determination:
  - With $T = 996$, the asymptotic standard error of the sample PACF is 0.032.
  - $p = 3$ or 9 using the 5% significant level; $p = 3$ using the 1% significant level.

- The AIC values are close to each other with minimum -5.849 occurring at $p = 9$, suggesting that an AR(9) model is preferred.

- The BIC attains its minimum value -5.833 at $p = 1$ with -5.831 as a close second at $p = 3$. Thus, the BIC selects an AR(1) model for the value-weighted return series.

- Different methods yield different choice of p. Generally speaking, we will employ residual checking for further model modification.

2.5 Parameter Estimation: $c, \phi_1, \ldots, \phi_p, \sigma^2$

(1) Yule-Walker estimator

For the AR(p), the first $p+1$ equations of (8) lead to the following

$$\gamma_j = \phi_1 \gamma_{j-1} + \cdots + \phi_j \gamma_{j-p}, \quad j = 1, \ldots, p \tag{12}$$

$$\sigma^2 = \gamma_0 - \phi_1 \gamma_1 - \cdots - \phi_p \gamma_p. \tag{13}$$

In matrix notation, the Yule-Walker equations are

$$\Gamma_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p, \quad \sigma^2 = \gamma_0 - \boldsymbol{\phi}' \boldsymbol{\gamma}_p, \tag{14}$$

where $\Gamma_p = \{\gamma_{k-j}\}_{j,k=1}^{p}$ is a $p \times p$ matrix, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ is a $p \times 1$ vector, and $\boldsymbol{\gamma}_p = (\gamma_1, \ldots, \gamma_p)'$ is a $p \times 1$ vector.

Using the method of moments, we solve

$$\hat{\boldsymbol{\phi}} = \hat{\Gamma}_p^{-1}\hat{\boldsymbol{\gamma}}_p, \quad \hat{\sigma}^2 = \hat{\gamma}_0 - \hat{\boldsymbol{\gamma}}_p'\hat{\Gamma}_p^{-1}\hat{\boldsymbol{\gamma}}_p. \tag{15}$$

These estimators are typically called the *Yule-Walker estimators*.

For calculation purposes, the YW estimates with sample ACF are

$$\hat{\boldsymbol{\phi}} = \hat{\mathbf{R}}_p^{-1}\hat{\boldsymbol{\rho}}_p, \quad \hat{\sigma}^2 = \hat{\gamma}_0\left[1 - \hat{\boldsymbol{\rho}}_p'\hat{\mathbf{R}}_p^{-1}\hat{\boldsymbol{\rho}}_p\right]. \tag{16}$$

where $\hat{\mathbf{R}}_p = \{\hat{\rho}_{k-j}\}_{j,k=1}^p$ is a $p \times p$ matrix, and $\hat{\boldsymbol{\rho}}_p = (\hat{\rho}_1, \ldots, \hat{\rho}_p)'$ is a $p \times 1$ vector.

If the sample size is large

$$\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} N(\mathbf{0}, \sigma^2\Gamma_p^{-1}), \quad \hat{\sigma}^2 \xrightarrow{p} \sigma^2. \tag{17}$$

(2) Maximum likelihood estimator

AR(1) model

Let

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t,$$

where $|\phi| < 1$ and $\varepsilon_t \sim iidN(0, \sigma^2)$.

Given the data $y_1, \ldots, y_n$, we seek the likelihood

$$L(\mu, \phi, \sigma^2) = f(y_1, \ldots, y_n | \mu, \phi, \sigma^2).$$

In the case of an AR(1), we may write the likelihood as

$$L(\mu, \phi, \sigma^2) = f(y_1) f(y_2|y_1) \cdots f(y_n|y_{n-1}) = f(y_1) \prod_{t=2}^{n} f(y_t|y_{t-1}).$$

Since $y_t | y_{t-1} \sim N(\mu + \phi(y_{t-1} - \mu), \sigma^2)$, we have

$$f(y_t | y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{[y_t - \mu - \phi(y_{t-1} - \mu)]^2}{2\sigma^2} \right\}.$$

To find $f(y_1)$, we can use the causal representation

$$y_1 = \mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{1-j} \sim N(\mu, \sigma^2/(1 - \phi^2))$$

Finally, for an AR(1), the likelihood is

$$L(\mu, \phi, \sigma^2) = (2\pi\sigma^2)^{-n/2}(1 - \phi^2)^{1/2} \exp\left\{ -\frac{S(\mu, \phi)}{2\sigma^2} \right\}, \qquad (18)$$

with the *unconditional sum of squares* $S(\mu, \phi) = (1 - \phi^2)(y_1 - \mu)^2 + \sum_{t=2}^{n}[y_t - \mu - \phi(y_{t-1} - \mu)]^2$.

Taking the first-order condition (FOC), the MLE of σ^2 is

$$\hat{\sigma}^2 = n^{-1} S(\hat{\mu}, \hat{\phi}), \tag{19}$$

where $\hat{\mu}$ and $\hat{\phi}$ are the MLEs of μ and ϕ, respectively.

- If we replace n by $n-2$, we would obtain the unconditional least estimate of σ^2. (The LS estimation of μ and ϕ is by minimizing $S(\mu, \phi)$)

- If we take logs, replace σ^2 by $\hat{\sigma}^2$ in (19), and ignore constants, $\hat{\mu}$ and $\hat{\phi}$ are the values that minimize the criterion function

$$l(\mu, \phi) = \log[n^{-1} S(\mu, \phi)] - n^{-1} \log(1 - \phi^2)$$

that is, $l(\mu, \phi) \propto -2 \log L(\mu, \phi, \hat{\sigma}^2)$. This criterion function is sometimes called the *profile or concentrated likelihood*.

- If dropping the term in the likelihood that causes the nonlinearity, (here the first term of y_1 in $S(\mu, \phi)$), the *conditional likelihood* becomes

$$L(\mu, \phi, \sigma^2 | y_1) = (2\pi\sigma^2)^{-(n-1)/2} \exp\left\{ -\frac{S_c(\mu, \phi)}{2\sigma^2} \right\},$$

where the conditional sum of square is

$$S_c(\mu, \phi) = \sum_{t=2}^{n} [(y_t - \mu) - \phi(y_{t-1} - \mu)]^2.$$

The *conditional MLE* of σ^2 is

$$\hat{\sigma}^2 = S_c(\hat{\mu}, \hat{\phi})/(n-1),$$

and $\hat{\mu}$ and $\hat{\phi}$ are the values that minimize $S_c(\mu, \phi)$.

- Conditional least square.
  - Let $\alpha = \mu(1-\phi)$, the conditional sum of squares can be written as

$$S_c(\mu, \phi) = \sum_{t=2}^{n} [y_t - (\alpha + \phi y_{t-1})]^2.$$

    The problem is now the linear regression problem.
  - Following the results from least square estimation (LS), we have $\hat{\alpha} = \bar{y}_{(2)} - \hat{\phi}\bar{y}_{(1)}$, where $\bar{y}_{(1)} = (n-1)^{-1}\sum_{t=1}^{n-1} y_t$, and $\bar{y}_{(2)} = (n-1)^{-1}\sum_{t=2}^{n} y_t$, and the conditional estimates are then

$$\hat{\mu} = \frac{\bar{y}_{(2)} - \hat{\phi}\bar{y}_{(1)}}{1-\hat{\phi}} \qquad \hat{\phi} = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_{(2)})(y_{t-1} - \bar{y}_{(1)})}{\sum_{t=2}^{n}(y_{t-1} - \bar{y}_{(1)})^2}. \tag{20}$$

AR(p) model

For general AR(p) models, maximum likelihood estimation, unconditional least squares, and conditional least squares follow analogously to the AR(1) example.

A more detailed introduction of the maximum likelihood estimator for the AR(p) model can be found in the textbook “Time Series Analysis” by Hamilton (1994).

2.6 Model Checking

(1) Residual autocorrelation

Three common methods are used to test for residual autocorrelation.

1. Sample ACF of the residuals, given by

$$r_k(\hat{\varepsilon}) = \frac{\sum_{t=k+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=k+1}^{T} \hat{\varepsilon}_t^2} \sim N(0, 1/n), \tag{21}$$

2. Joint significance of the first m residual autocorrelations. The test-statistic developed by Ljung and Box (1978), given by

$$\mathrm{LB}(m) = T(T+2) \sum_{k=1}^{m} (T-k)^{-1} r_k^2(\hat{\varepsilon}) \sim \chi^2(m-p). \tag{22}$$

(2) Homoscedasticity

Neglecting heteroscedasticity of the residuals leads to

- ordinary t-statistics cannot be used.
- many diagnostic tests, such as tests for nonlinearity, are affected.
- confidence intervals for forecasts can no longer be computed in the usual manner.

Heteroscedasticity-consistent t-statistic: Davidson and MacKinnon (1985) and Wooldridge (1990, 1991). In standard statistical or econometric software, it provides *heteroscedasticity-consistent t-statistic* for parameter estimates.

Portmanteau test statistic on (standardized) squared residuals developed by McLeod and Li (1983):

$$\mathrm{McL}(m) = T(T+2)\sum_{k=1}^{m}(T-k)^{-1}r_k^2(\hat{\varepsilon}^2). \tag{23}$$

When applied to the residuals from an ARMA(p, q) model, the McL test has an asymptotic $\chi^2(m-p-q)$ distribution, again provided that m/T is small and m is moderately large.

(3) Normality

Defining the j-th moment of the estimated (standardized) residuals as

$$\hat{m}_j = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^j,$$

the skewness and kurtosis of $\hat{\varepsilon}_t$ can be calculated as

$$\widehat{SK}_{\hat{\varepsilon}} = \frac{\hat{m}_3}{\sqrt{\hat{m}_2^3}}, \quad \text{and} \quad \widehat{K}_{\hat{\varepsilon}} = \frac{\hat{m}_4}{\hat{m}_2^2}.$$

Under the null hypothesis of normality, $\sqrt{T/6}\cdot\widehat{SK}_{\hat{\varepsilon}} \sim N(0,1)$ and $\sqrt{T/24}\cdot(\widehat{K}_{\hat{\varepsilon}}-3) \sim N(0,1)$.

A joint test for normality (Jarque and Bera, 1987) is then given by

$$\text{JB} = \frac{T}{6}\widehat{SK}_{\hat{\varepsilon}}^2 + \frac{T}{24}(\widehat{K}_{\hat{\varepsilon}}-3)^2 \sim \chi^2(2). \tag{24}$$

2.7 Forecasting

Our goal is to predict future values of a time series, y_{t+h}, $h = 1, 2, \ldots$, based on the data collected to the present, $I_t = \{y_t, y_{t-1}, \ldots\}$.

Let $y_{t+h|t}$ denote a forecast of y_{t+h} made at time t, which has an associated *forecast error* or *prediction error* $\varepsilon_{t+h|t}$,

$$\varepsilon_{t+h|t} = y_{t+h} - y_{t+h|t}.$$

Obviously, many different forecasts $y_{t+h|t}$ exists. The *optimal forecast* $y_{t+h|t}$ which minimizes the *mean squared error of the forecast*,

$$\text{MSE}(\varepsilon_{t+h|t}) \equiv E[\varepsilon^2_{t+h|t}] = E[(y_{t+h} - y_{t+h|t})^2].$$

Theorem 1. *The minimum MSE forecast (best forecast) of y_{t+h} based on I_t is*

$$y_{t+h|t} = E[y_{t+h}|I_t]. \tag{25}$$

Proof.

$$E[(y_{t+h} - y_{t+h|t})^2] = E\left\{[y_{t+h} - E(y_{t+h}|I_t) + E(y_{t+h}|I_t) - y_{t+h|t}]^2\right\}$$

$$= E\left\{[y_{t+h} - E(y_{t+h}|I_t)]^2\right\} + 2E\left\{[y_{t+h} - E(y_{t+h}|I_t)][E(y_{t+h}|I_t) - y_{t+h|t}]\right\}$$

$$+ E\left\{[E(y_{t+h}|I_t) - y_{t+h|t}]^2\right\}$$

$$= E\left\{[y_{t+h} - E(y_{t+h}|I_t)]^2\right\} + E\left\{[E(y_{t+h}|I_t) - y_{t+h|t}]^2\right\}$$

Taking the minimization with different values of $y_{t+h|t}$, we have

$$y_{t+h|t} = E(y_{t+h}|I_t).$$

Thus, the minimum MSE is $E[(y_{t+h} - y_{t+h|t})^2] = E[(y_{t+h} - E(y_{t+h}|I_t))^2]$. □

Forecasting AR(p) model

- **1-Step-Ahead Forecast:** From the AR(p) model, we have

$$y_{t+1} = c + \phi_1 y_t + \cdots + \phi_p y_{t+1-p} + \varepsilon_{t+1}, \quad \varepsilon_t \sim IID(0, \sigma^2).$$

  – Under the *MSE loss function*, the *point forecast* of y_{t+1} given I_t is

$$y_{t+1|t} = E[y_{t+1}|I_t] = c + \phi_1 y_t + \cdots + \phi_p y_{t+1-p},$$

  – The *1-step-ahead forecast error* is

$$\varepsilon_{t+1|t} = y_{t+1} - y_{t+1|t} = \varepsilon_{t+1}.$$

  – The variance of *1-step-ahead forecast error* is $\mathrm{Var}[\varepsilon_{t+1|t}] = \mathrm{Var}(\varepsilon_{t+1}) = \sigma^2$.

  – If at is *normally distributed*, then a *95% 1-step-ahead interval forecast* of y_{t+1} is $y_{t+1|t} \pm 1.96\sigma$.

- **2-Step-Ahead Forecast:** From the AR(p) model, we have

$$y_{t+2} = c + \phi_1 y_{t+1} + \cdots + \phi_p y_{t+2-p} + \varepsilon_{t+2}. \tag{26}$$

  – Taking conditional expectation, we have

$$y_{t+2|t} = E[y_{t+2}|I_t] = c + \phi_1 y_{t+1|t} + \phi_2 y_t + \cdots + \phi_p y_{t+2-p}, \tag{27}$$

  – The associated forecast error

$$\varepsilon_{t+2|t} = y_{t+2} - y_{t+2|t} = \phi_1(y_{t+1} - y_{t+1|t}) + \varepsilon_{t+2} = \varepsilon_{t+2} + \phi_1 \varepsilon_{t+1}.$$

  – The variance of the forecast error is $\mathrm{Var}[\varepsilon_{t+2|t}] = (1 + \phi_1^2)\sigma^2$.
  – If ε_t is normally distributed, then a 95% 1-step-ahead interval forecast of y_{t+1} is $y_{t+2|t} \pm 1.96\sigma\sqrt{1 + \phi_1^2}$.
  – It is interesting to see that $\mathrm{Var}[\varepsilon_{t+2|t}] \geq \mathrm{Var}[\varepsilon_{t+1|t}]$, <u>meaning that as the forecast horizon increases the uncertainty in forecast also increases.</u>

- **Multistep-Ahead Forecast:** In general, we have

$$y_{t+h} = c + \phi_1 y_{t+h-1} + \cdots + \phi_p y_{t+h-p} + \varepsilon_{t+h}. \tag{28}$$

  - Taking conditional expectation, we have

$$y_{t+h|t} = E[y_{t+h}|I_t] = c + \sum_{i=1}^{p} \phi_i y_{t+h-i|t}, \tag{29}$$

    where $y_{t+\ell|t} = y_{t+\ell}$ if $\ell \leq 0$. This forecast can be computed recursively using forecasts $y_{t+i|t}$ for $i = 1, \ldots, h-1$.
  - The h-step-ahead forecast error is

$$\varepsilon_{t+h|t} = y_{t+h} - y_{t+h|t}.$$

  - The forecasting interval is not straightforward for the AR(p) model. We will use the method similar to that of the ARMA model.

Example 4 (*Forecasting CRSP returns, see pages 56-57 in Tsay's textbook)*. The fitted AR(3) model is

$$r_t = 0.0098 + 0.1024 r_{t-1} - 0.0201 r_{t-2} - 0.1090 r_{t-3} + a_t,$$

where $\hat{\sigma}_a^2 = 0.054$.

- Forecast origin is $t = 984$, which is December 2007.
- Forecasts are denoted by "$\circ$" and actual observations by "$\bullet$".
- Two dashed lines denote two standard error limits of the forecasts.

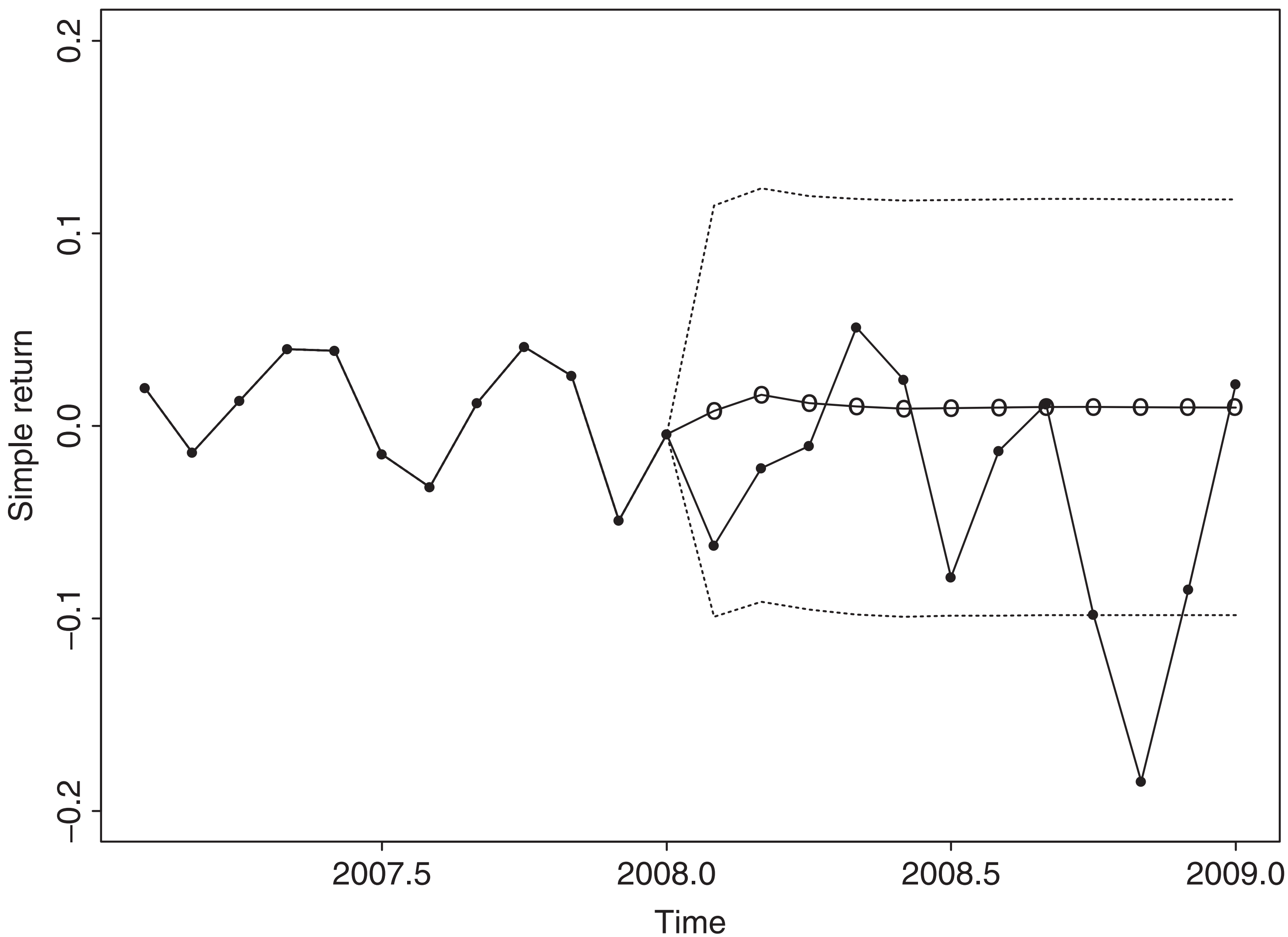


Figure 7 Plot of 1- to 12-step-ahead out-of-sample forecasts for monthly simple returns of CRSP value-weighted index.

Example 5 (*Forecasting the U.S. growth rate, see Figure 8*).

Figure 8 Forecasting US GDP growth with 8 periods.

3. Autoregressive Moving-Average (MA) Models

3.1 Moving average (MA) models

MA(1) model

The MA(1) model has the form

$$y_t = c + \varepsilon_t + \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim WN(0, \sigma^2).$$

- Invertibility condition: $|\theta| < 1$.

- Mean: $E[y_t] = \mu = c$.

- Autocovariance function: $\gamma_0 = \sigma^2(1 + \theta^2)$, $\gamma_1 = \sigma^2\theta$, $\gamma_j = 0$ for $j > 1$.

- Autocorrelation function: $\rho_j = \begin{cases} 1, & \text{for } j = 0; \\ \theta/(1+\theta^2), & \text{for } j = 1; \\ 0, & \text{for } j > 1. \end{cases}$

- The ACF of an MA(1) process *cuts off* at lag one, and the maximum value of this correlation is ± 0.5.

- There is an identification problem with the MA(1) model since $\theta = 1/\theta$ produce the same value of ρ_1.

- In the invertible MA(1), the error term ε_t has an infinite order AR representation (let $\theta^* = -\theta$)

$$\varepsilon_t = (y_t - \mu) + \theta^*(y_{t-1} - \mu) + \theta^{*2}(y_{t-2} - \mu) + \cdots = \sum_{j=0}^{\infty} \theta^{*j}(y_{t-j} - \mu).$$

- **PACF of MA(1):** Generally, it is possible to show

$$P_k = \frac{-(-\theta)^k(1-\theta^2)}{1-\theta^{2(k+1)}}, \qquad \text{for } k \geq 1.$$

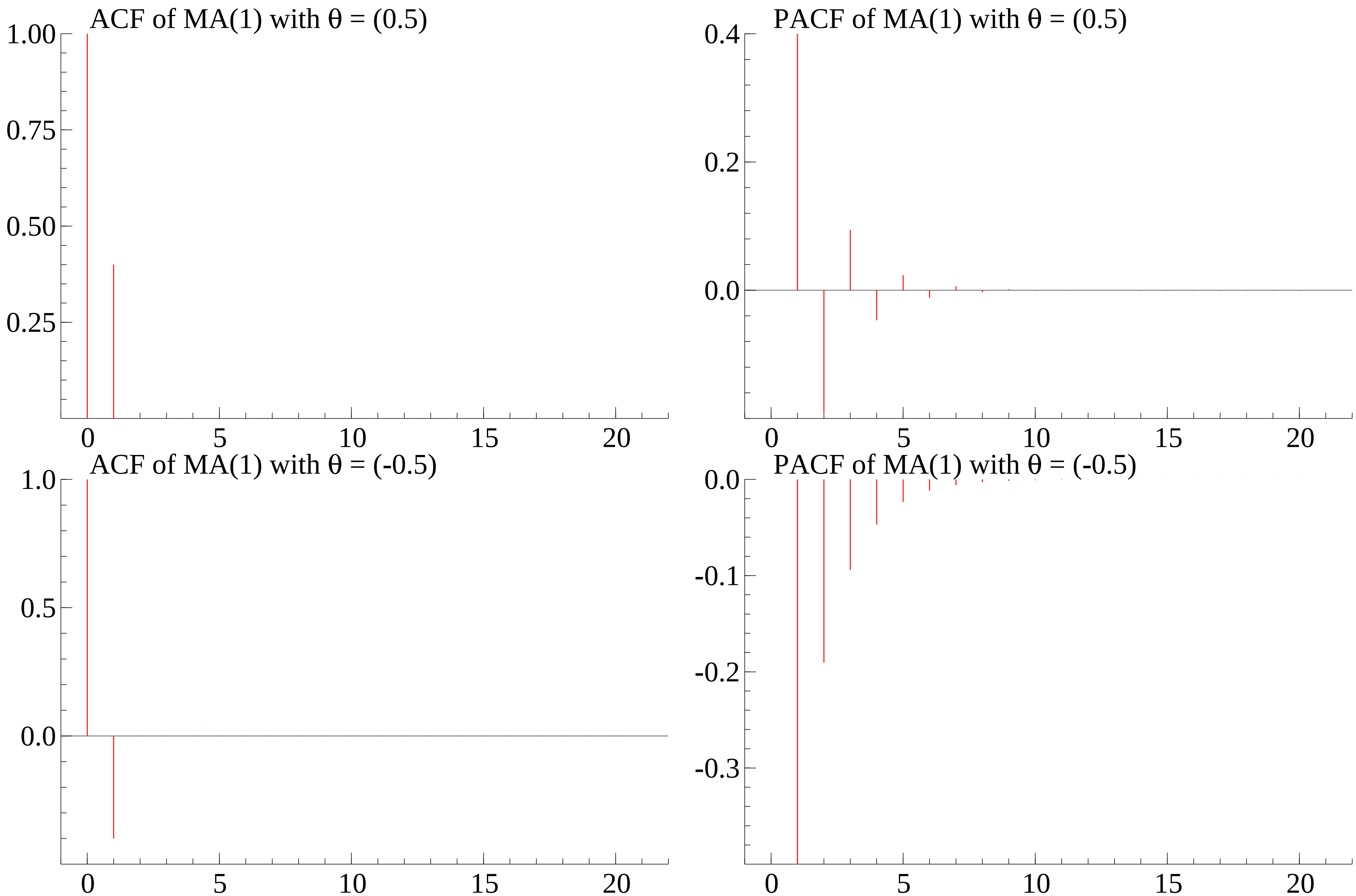


Figure 9 The ACF and PACF of MA(1) model.

MA(q) model

The MA(q) model has the form

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim WN(0, \sigma^2).$$

- **Stationarity:** Finite order MA models provided $\theta_1, \ldots, \theta_q$ are finite are stationary.

- **Mean:** The constant term c is the mean of the process, i.e., $\mu = c$.

- **Invertibility**(necessary and sufficient condition): It is *invertible* if all of the roots of the MA *characteristic polynomial* $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = 0$ lie *outside* the complex unit circle.

- **Autocovariance function:** Again for simplicity, assume $\mu = 0$.

$$\gamma_j = \begin{cases} \sigma^2(1+\theta_1^2+\cdots+\theta_q^2) & \text{for } j=0, \\ \sigma^2 \sum_{i=0}^{q-j} \theta_i\theta_{i+j} & \text{for } j=1,\ldots,q, \\ 0 & \text{for } j>q, \end{cases}$$

where $\theta_0 = 1$.

- **Autocorrelation function:** $\rho_j = \frac{\gamma_j}{\gamma_0}$. Again, from the autocovariance function we have

$$\rho_j = \begin{cases} \neq 0, & \text{for } j=0,\ldots,q; \\ 0, & \text{for } j>q. \end{cases}$$

In other words, the ACF of an MA(q) is non-zero up to lag q and is zero afterwards. This is a special feature of MA processes and it provides a convenient way to identify an MA model in practice.

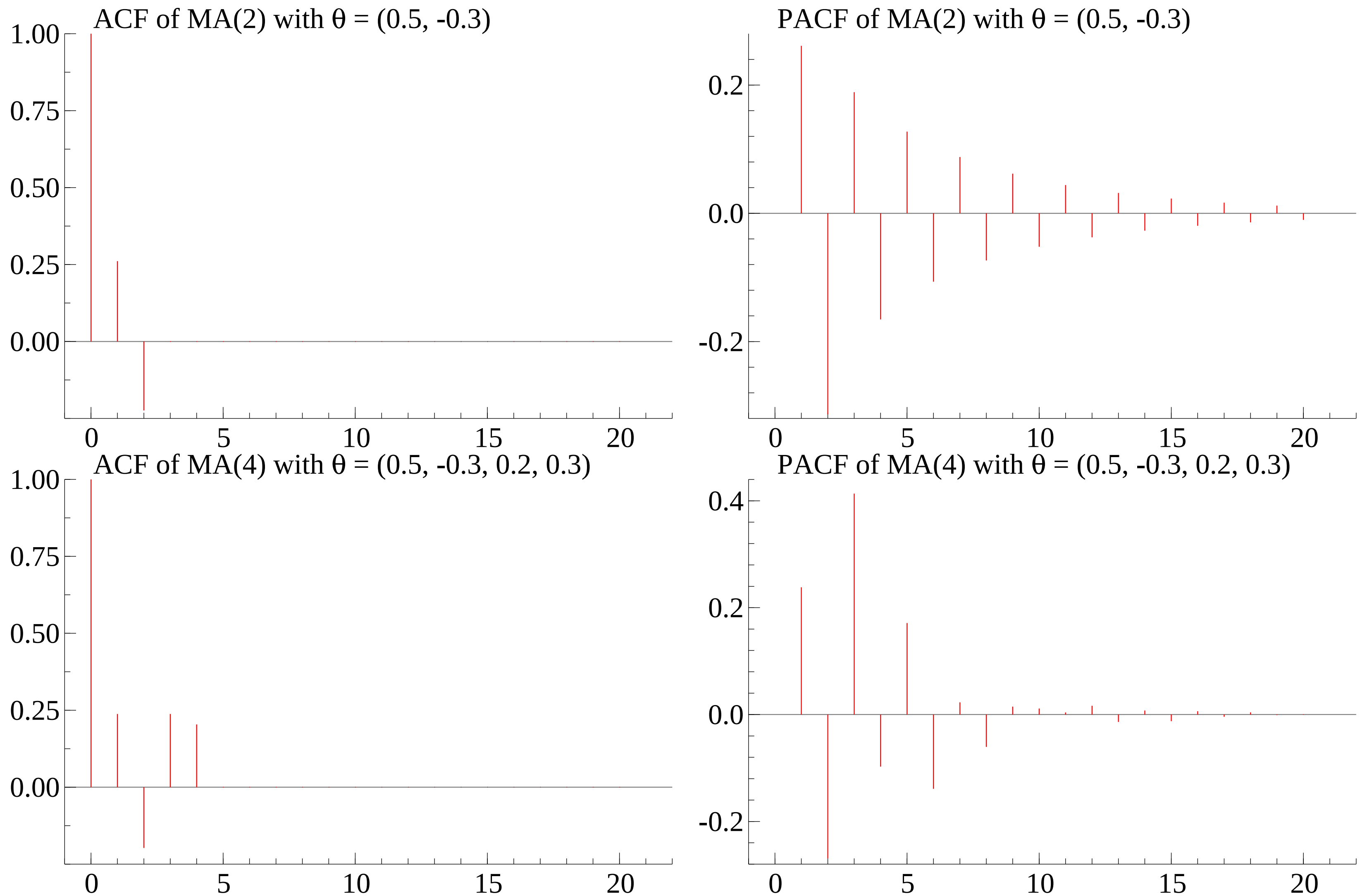


Figure 10 The ACF and PACF of MA(q) model.

3.2 Mixed ARMA Processes

ARMA(1,1) model

The ARMA(1,1) model is defined by

$$y_t = c + \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}.$$

- Stationarity condition: $|\phi| < 1$.

- Invertibility condition: $|\theta| < 1$.

- Mean: $\mu = \frac{c}{1-\phi}$.

- ACF: Multiplying the model by y_{t-j} and taking expectation yields:

$$\gamma_j - \phi\gamma_{j-1} = \begin{cases} [1+\theta(\phi+\theta)]\sigma^2, & \text{if } j=0; \\ \theta\sigma^2, & \text{if } j=1; \\ 0, & \text{if } j>1, \end{cases}$$

by employing the fact that $E(\varepsilon_t y_t) = E[\varepsilon_t(\phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1})] = \sigma^2$ and $E(\varepsilon_{t-1} y_t) = E[\varepsilon_{t-1}(\phi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1})] = \phi\sigma^2 + \theta\sigma^2 = (\phi+\theta)\sigma^2$.

  – Solving the first two equations produces

$$\gamma_0 = \frac{1+2\theta\phi+\theta^2}{1-\phi^2}\sigma^2$$

  and using the last recursively shows

$$\gamma_j = \frac{(1+\theta\phi)(\phi+\theta)}{1-\phi^2}\phi^{j-1}\sigma^2 \qquad \text{for } j \geq 1.$$

– The autocorrelation function (ACF) can then be computed as

$$\rho_j = \frac{(1+\theta\phi)(\phi+\theta)}{1+2\theta\phi+\theta^2}\phi^{j-1} \qquad \text{for } j \geq 1.$$

– The pattern here is similar to that for AR(1) model, except for the first term.

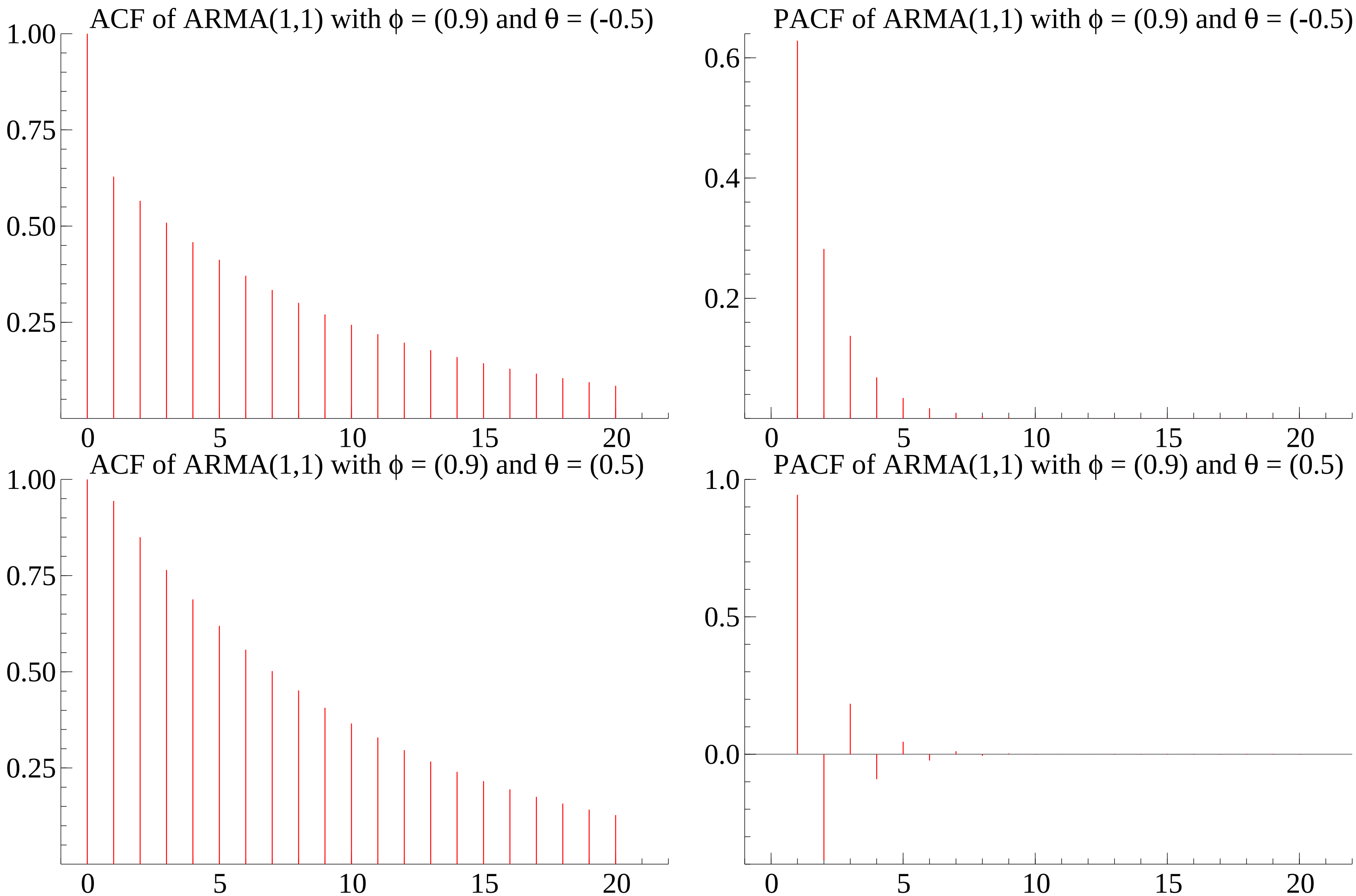


Figure 11 The ACF and PACF of ARMA$(1, 1)$ model.

ARMA(p, q) model

The regression formulation of the ARMA(p, q) model is

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

and its lag operator notation is

$$\phi(L) y_t = c + \theta(L) \varepsilon_t,$$

where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$. *It is assumed that the polynomials $\phi(L)$ and $\theta(L)$ do not have common factors.*

Again, for a stationary process, we can rewrite the model as

$$\phi(L)(y_t - \mu) = \theta(L) \varepsilon_t.$$

The Properties of ARMA model can be summarized as follows:

- **Stationarity:** The ARMA process is stationary if and only if all roots of the AR characteristic polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0$ lie outside the complex unit circle.

- **Invertibility:** All roots of the MA characteristic polynomial $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = 0$ lie outside the unit circle.

- **AR representation:** $\pi(L) y_t = \frac{c}{\theta(1)} + \varepsilon_t$, where $\pi(L) = \frac{\phi(L)}{\theta(L)}$. The π-weight π_i can be obtained by equating the coefficients of L^i in $\pi(L)\theta(L) = \phi(L)$.

- **MA representation:** $y_t = \frac{c}{\phi(1)} + \psi(L)\varepsilon_t$, where $\psi(L) = \frac{\theta(L)}{\phi(L)}$. Again, the ψ-weight can be obtained by equating the coefficients.

- **Moments:** (Assume stationarity)
  - **Mean:** $\mu = \dfrac{c}{1 - \phi_1 - \cdots - \phi_p}$.
  - **Autocovariance function:** (Assume $\mu = 0$) Using the result

$$E(y_t \varepsilon_{t-j}) = \begin{cases} \sigma^2, & \text{for } j = 0; \\ \psi_j \sigma^2, & \text{for } j > 0; \\ 0, & \text{for } j < 0. \end{cases}$$

and the same technique as before, we have

$$\gamma_j - \phi_1 \gamma_{j-1} - \cdots - \phi_p \gamma_{j-p} = \begin{cases} (1 + \theta_1 \psi_1 + \cdots + \theta_q \psi_q) \sigma^2, & \text{for } j = 0; \\ (\theta_j + \theta_{j+1} \psi_1 + \cdots + \theta_q \psi_{q-j}) \sigma^2, & \text{for } j = 1, \ldots, q; \\ 0, & \text{for } j > q, \end{cases}$$

where $\psi_0 = 1$ and $\theta_j = 0$ for $j > q$.

– **Autocorrelation function:** the correlation coefficient ρ_j satisfies that

$$\rho_j - \phi_1 \rho_{j-1} - \cdots - \phi_p \rho_{j-p} = 0, \quad \text{for } j > q,$$

then the ACF satisfy the difference equation $\phi(L)\rho_j = 0$ for $j > q$ with $\rho_1, \ldots, \rho_q$ as initial conditions.

– **Generalized Yule-Walker equation:** Consider the above equations of ACF for $j = q+1, q+2, \ldots, q+p$, we have

$$\begin{bmatrix} \rho_{q+1} \\ \rho_{q+2} \\ \vdots \\ \rho_{q+p} \end{bmatrix} = \begin{bmatrix} \rho_q & \rho_{q-1} & \cdots & \rho_{q+2-p} & \rho_{q+1-p} \\ \rho_{q+1} & \rho_q & \cdots & \rho_{q+3-p} & \rho_{q+2-p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{q+p-1} & \rho_{q+p-2} & \cdots & \rho_{q+1} & \rho_q \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix},$$

which is referred to as a p-order generalized Yule-Walker equation for the ARMA(p, q) process. It can be used to solve for ϕ_i's give ρ_i's.

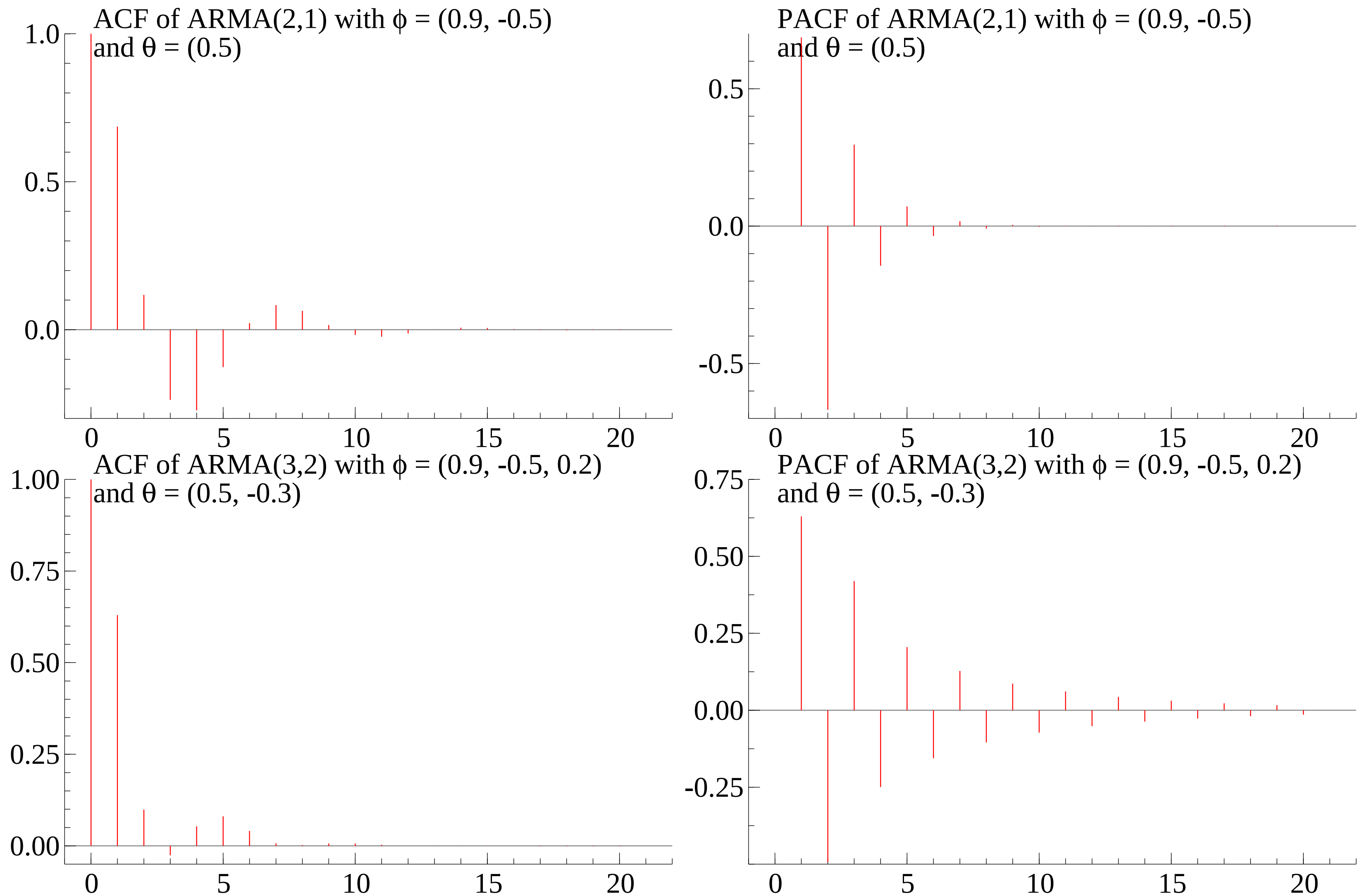


Figure 12 The ACF and PACF of ARMA(p, q) model.

3.3 Identifying ARMA Models

(1) Extended ACF (EACF) – Tsay and Tiao (1984, JASA).

Consider the ARMA(p, q) model given by

$$(1 - \phi_1 L - \cdots - \phi_p L^p)(y_t - \mu) := w_t = (1 + \theta_1 L + \cdots + \theta_q L^q)\varepsilon_t$$

Its basic idea is based on the *"generalized" Yule-Walker equation*. Conceptually, it involves two steps:

- In the first step, we attempt to obtain *consistent estimates* of AR coefficients. Given such estimates, we can transform the ARMA series into a *pure MA process*.

- The second step then uses the *sample ACF* of the transformed MA process to identify the MA order q.

Again,

$$(1 - \phi_1 L - \cdots - \phi_p L^p)(y_t - \mu) := w_t = (1 + \theta_1 L + \cdots + \theta_q L^q)\varepsilon_t$$

The generalized Yule-Walker equation is that,

$$\begin{bmatrix} \rho_{q+1} \\ \rho_{q+2} \\ \vdots \\ \rho_{q+p} \end{bmatrix} = \begin{bmatrix} \rho_q & \rho_{q-1} & \cdots & \rho_{q+2-p} & \rho_{q+1-p} \\ \rho_{q+1} & \rho_q & \cdots & \rho_{q+3-p} & \rho_{q+2-p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{q+p-1} & \rho_{q+p-2} & \cdots & \rho_{q+1} & \rho_q \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix},$$

$$\begin{bmatrix} \rho_{q+2} \\ \rho_{q+3} \\ \vdots \\ \rho_{q+p+1} \end{bmatrix} = \begin{bmatrix} \rho_{q+1} & \rho_q & \cdots & \rho_{q+3-p} & \rho_{q+2-p} \\ \rho_{q+2} & \rho_{q+1} & \cdots & \rho_{q+4-p} & \rho_{q+3-p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{q+p} & \rho_{q+p-1} & \cdots & \rho_{q+2} & \rho_{q+1} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix},$$

To make use of the EACF for model specification, we consider the two-way table:

AR	MA (or q)					
p	0	1	2	3	4	$\cdots$
0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	$\cdots$
1	$\rho_{1,1}$	$\rho_{1,2}$	$\rho_{1,3}$	$\rho_{1,4}$	$\rho_{1,5}$	$\cdots$
2	$\rho_{2,1}$	$\rho_{2,2}$	$\rho_{2,3}$	$\rho_{2,4}$	$\rho_{2,5}$	$\cdots$
3	$\rho_{3,1}$	$\rho_{3,2}$	$\rho_{3,3}$	$\rho_{3,4}$	$\rho_{3,5}$	$\cdots$
$\vdots$	$\vdots$			$\vdots$		

Suppose that z_t is an ARMA(1,1) model, then the EACF table is

AR	MA (or q)						
p	0	1	2	3	4	5	$\cdots$
0	X	X	X	X	X	X	$\cdots$
1	X	O	O	O	O	O	$\cdots$
2	*	X	O	O	O	O	$\cdots$
3	*	*	X	O	O	O	$\cdots$
4	*	*	*	X	O	O	

where "X" and "O" denote non-zero and zero quantities, respectively, "*" represents a quantity which can assume any value between -1 and 1.

In practice, the EACF is replaced by its sample EACF.

This sample EACF can be run with R, see the link: `http://www.inside-r.org/packages/cran/TSA/docs/eacf`

Table 1 Sample EACF for logarithmic S&P 500 realized volatility

EACF table								
$p \setminus q$	0	1	2	3	4	5	6	7
0	0.785	0.759	0.730	0.718	0.703	0.684	0.667	0.663
1	−0.437	0.009	−0.037	0.006	0.010	−0.003	−0.026	0.006
2	−0.423	−0.124	−0.041	0.013	0.015	−0.002	−0.032	0.010
3	−0.493	−0.150	−0.189	0.003	−0.007	0.002	−0.025	−0.017
4	−0.473	−0.179	−0.174	−0.007	−0.011	−0.007	−0.010	−0.012
5	−0.431	−0.285	−0.101	−0.049	0.220	0.008	−0.012	−0.004
6	−0.465	0.066	−0.161	−0.162	0.163	0.117	−0.013	−0.005

Simplified EACF table：X denotes significance，O denotes insignificance								
$q \setminus p$	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	O	X	O	O	O	O	O
2	X	X	X	O	O	O	O	O
3	X	X	X	O	O	O	O	O
4	X	X	X	O	O	O	O	O
5	X	X	X	X	X	O	O	O
6	X	X	X	X	X	X	O	O

(2) Information criteria

The information criteria discussed earlier can also be used to select ARMA(p, q) models.

Typically, for some prespecified positive integers P and Q, one computes AIC (or BIC) for ARMA(p, q) models, where $0 \leq p \leq P$ and $0 \leq q \leq Q$, and selects the model that gives the minimum AIC (or BIC).

3.4 Estimation methods

(1) Maximum Likelihood Estimation (MLE)

Let $\boldsymbol{\beta} = (\mu, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$ and $\mathcal{F}_t = \{y_t, y_{t-1}, \ldots, y_1, y_0, \ldots, y_{1-p}\}$.

We write the model in terms of the errors

$$\varepsilon_t(\boldsymbol{\beta}) = y_t - \sum_{j=1}^{p} \phi_j y_{t-j} - \sum_{k=1}^{q} \theta_k \varepsilon_{t-k}(\boldsymbol{\beta}). \tag{30}$$

In this case, the likelihood conditional on $\mathcal{F}_0 = \{y_0, \ldots, y_{1-p}\}$ is given by

$$L(\boldsymbol{\beta}, \sigma^2; \mathcal{F}_0) = \prod_{t=1}^{n} f(y_t | \mathcal{F}_{t-1}) = \prod_{t=1}^{n} f(\varepsilon_t(\boldsymbol{\beta}) | \mathcal{F}_{t-1}).$$

Then the maximization on $L(\boldsymbol{\beta}, \sigma^2; \mathcal{F}_0)$ yields the MLE.

(2) Conditional LS

As before, write $\boldsymbol{\beta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$, and for the ease of discussion, we will put $\mu = 0$.

For conditional least squares, we approximate the residual sum of squares by conditioning on $y_0, \ldots, y_{1-p}$ (if $p > 0$) and $\varepsilon_0 = \cdots = \varepsilon_{1-q} = 0$ (if $q > 0$).

Using this conditioning argument, the conditional error sum of squares is

$$S_c(\boldsymbol{\beta}) = \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\beta}).$$

Minimizing $S_c(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ yields the conditional least squares estimates.

3.5 Forecasting

(1) Forecasting MA(q) models

Consider the MA(q) model $y_t = \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}$, with $\theta_0 \equiv 1$.

Using (25) and the *IID* properties of ε_t, the *optimal forecast* is

$$y_{t+h|t} = \begin{cases} \sum_{i=h}^{q} \theta_i \varepsilon_{t+h-i}, & \text{for } h = 1, \ldots, q, \\ 0, & \text{for } h > q, \end{cases}$$

whereas the corresponding forecast error follows as

$$\varepsilon_{t+h|t} = \begin{cases} \sum_{i=0}^{h-1} \theta_i \varepsilon_{t+h-i}, & \text{for } h = 1, \ldots, q, \\ \sum_{i=0}^{q} \theta_i \varepsilon_{t+h-i}, & \text{for } h > q, \end{cases}$$

which can be simplified to $\varepsilon_{t+h|t} = \sum_{i=0}^{h-1} \theta_i \varepsilon_{t+h-i}$ by defining $\theta_i \equiv 0$ for $h > q$.

Given the assumptions on ε_t, it follows that

$$E[\varepsilon_{t+h|t}] = 0,$$

and for *the mean squared error (MSE) of the forecast*

$$\text{MSE}(\varepsilon_{t+h|t}) = E[\varepsilon^2_{t+h|t}] = \sigma^2 \sum_{i=0}^{h-1} \theta_i^2.$$

Assuming normality, a 95% *forecasting interval* for y_{t+h} is bounded by

$$\left(y_{t+h|t} - 1.96 \cdot \text{RMSE}(\varepsilon_{t+h|t}), \quad y_{t+h|t} + 1.96 \cdot \text{RMSE}(\varepsilon_{t+h|t})\right),$$

where RMSE($\varepsilon_{t+h|t}$) denotes the *square root* of MSE($\varepsilon_{t+h|t}$).

(2) Forecasting ARMA(p, q) model:

The ARMA(p, q) model for y_t is

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

or, in lag operator form $\phi_p(L) y_t = \theta_q(L) \varepsilon_t$.

- The true forecast value is

$$y_{t+h} = \phi_1 y_{t+h-1} + \cdots + \phi_p y_{t+h-p} + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \cdots + \theta_q \varepsilon_{t+h-q}.$$

- The minimum MSE forecast for y_{t+h} is

$$y_{t+h|t} = \phi_1 y_{t+h-1|t} + \cdots + \phi_p y_{t+h-p|t} + \theta_1 \varepsilon_{t+h-1|t} + \cdots + \theta_q \varepsilon_{t+h-q|t},$$

  where $y_{t+\ell|t} = y_{t+\ell}$ if $\ell \leq 0$; and $\varepsilon_{t+\ell|t} = 0$ if $\ell > 0$, otherwise $\varepsilon_{t+\ell|t} = \varepsilon_{t+\ell}$.

- It is convenient to rewrite the model as an MA(∞) model, that is, $y_t = \phi_p(L)^{-1}\theta_q(L)\varepsilon_t$ or

$$y_t = \varepsilon_t + \eta_1\varepsilon_{t-1} + \eta_2\varepsilon_{t-2} + \eta_3\varepsilon_{t-3} + \cdots .$$

Theorem 2 (See Hamilton (1994) page 74)**.** *The minimum MSE linear forecast (best linear predictor) of y_{t+h} based on I_t is*

$$y_{t+h|t} = \eta_h\varepsilon_t + \eta_{h+1}\varepsilon_{t-1} + \cdots , \quad \textit{with } \eta_0 \equiv 1.$$

- From which it follows that the *h-step-ahead prediction error* is given by

$$\varepsilon_{t+h|t} = \varepsilon_{t+h} + \eta_1\varepsilon_{t+h-1} + \cdots + \eta_{h-1}\varepsilon_{t+1} = \sum_{i=0}^{h-1} \eta_i\varepsilon_{t+h-i},$$

and the MSE of the forecast error is

$$\text{MSE}(\varepsilon_{t+h|t}) = \sigma^2 \sum_{i=0}^{h-1} \eta_i^2,$$

3.6. ARIMA(p, d, q) Models

- The specification of the ARMA(p, q) model assumes that y_t is stationary and ergodic.

- If y_t is a trending variable like an asset price or a macroeconomic aggregate like real GDP, then y_t *must be transformed to stationary form by eliminating the trend.*

- Box and Jenkins (1976) advocate removal of trends by *differencing*.

Let $\Delta = 1 - L$ denote the *difference operator*.

(1) If there is a linear trend in y_t then the first difference $\Delta y_t = y_t - y_{t-1}$ will not have a trend.

(2) If there is a quadratic trend in y_t, then Δy_t will contain a linear trend but the second difference $\Delta^2 y_t = (1 - 2L + L^2) y_t = y_t - 2y_{t-1} + y_{t-2}$ will not have a trend.

The class of ARMA(p, q) models where the trends have been transformed by *differencing d times* is denoted ARIMA(p, d, q).

3.7 Model Building

An effective procedure for building empirical time series models (ARIMA) is the *Box-Jenkins approach* (Box and Jenkins, 1976), which consists of three stages:

- Model specification,

- Estimation,

- Diagnostics checking.

Forecasts follow directly from the form of fitted model.

For ARIMA models, the modelling sequence can be usually described as follows:

(1) Input and plot the time series data and choose proper transformation (e.g., logarithm, logit, and *variance stablization transformation*).

(2) Compute and examine the sample ACF and the sample PACF of the data to further confirm a necessary *degree of differencing*.

- Check stationarity of the transformed time series data. If the transformed data has unit roots, then difference it until the differenced transformed data do not have unit root anymore. Denote d as the differencing times.

(3) Determine the order values of p and q preliminarily for the differenced series. For ARIMA models, there are two main approaches to model specification.

- The *"correlation" approach* such as the ACF, PACF and *extended autocorrelation function (EACF)*.

Function	Model	Feature
ACF	MA(q)	Cutting-off at lag q
PACF	AR(p)	Cutting-off at lag p
EACF	ARMA(p, q)	A triangle with vertex (p, q)

- The *information criterion approach* such as AIC, BIC, HQ.

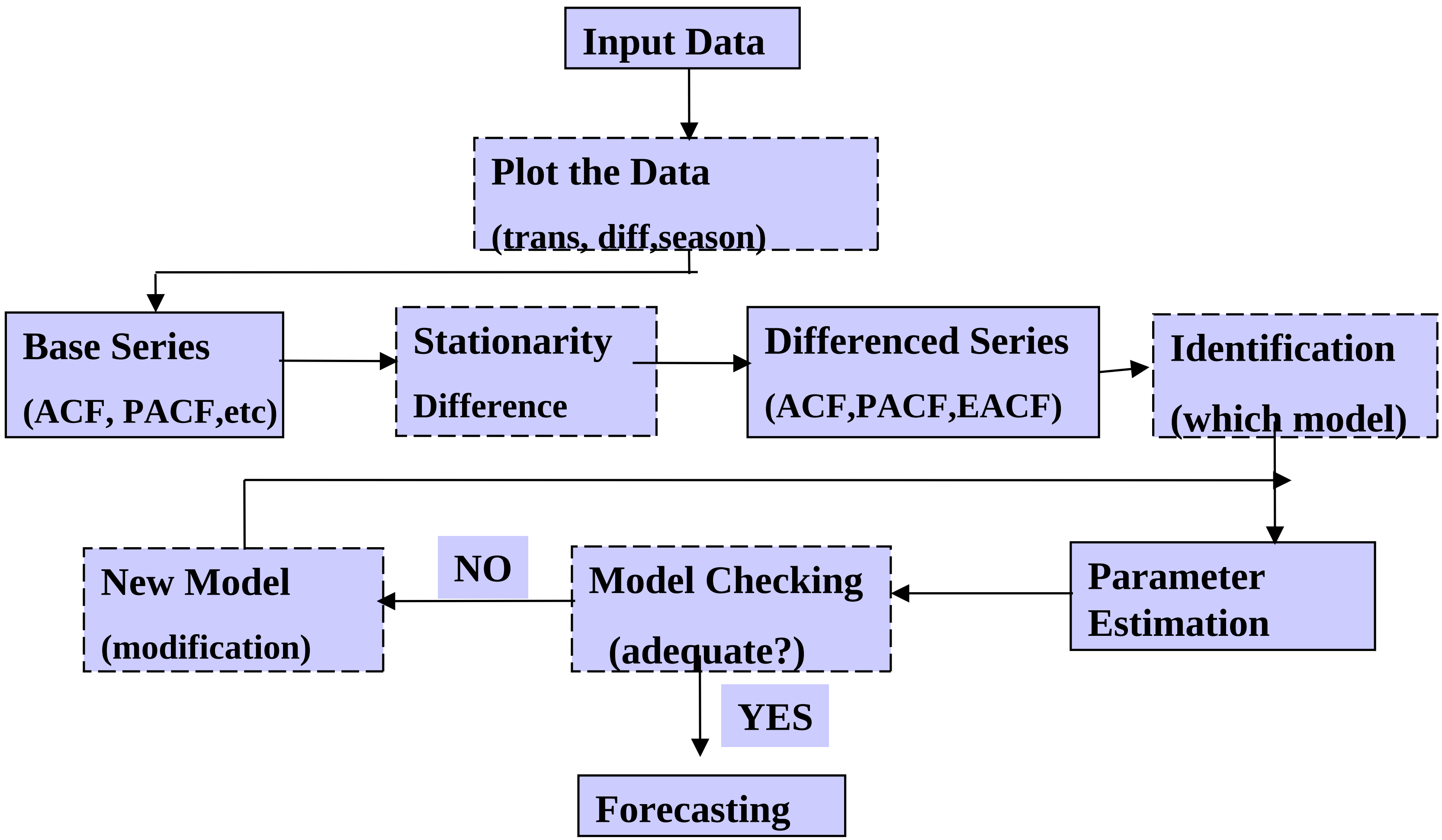
(4) Estimate the ARIMA(p, d, q) model for the transformed data.

(5) Perform *diagnostic analysis* to confirm whether the transformed ARIMA(p, d, q) model *adequately* describes the data (e.g. examine residuals from the fitted model).

(6) Respecify the ARIMA(p, d, q) model if necessary.

(7) Use the model for descriptive or forecasting purposes.

See the Figure.


Input Data
Plot the Data
(trans, diff,season)
Base Series
(ACF, PACF,etc)
Stationarity
Difference
Differenced Series
(ACF,PACF,EACF)
Identification
(which model)
New Model
(modification)
NO
Model Checking
(adequate?)
Parameter Estimation
YES
Forecasting

4 Unit Root Nonstationarity

So far we have focused on return and growth rate series that are stationary.

However, in some studies some series are nonstationary, such as

- Interest rates
- Foreign exchange rates
- The price series

In time series literature, such as nonstationary series is called unit-root nonstationary.

4.1 Random Walk

A time series $\{y_t\}$ is a random walk (RW) if it satisfies

$$y_t = y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2),$$

where y_0 is a real number denoting the starting value of the process.

- If y_t is the log price at date t, then y_0 could be the log price of the stock at its initial public offering (IPO) (i.e., the logged IPO price).

- If ε_t has a symmetric distribution around zero, then conditional on y_{t-1}, y_t has a $50 - 50$ chance to go up or down.

- A random-walk series is not weakly stationary, and it is a unit-root nonstationary time series.

- Under such a model, the stock price is not predictable or mean reverting.
  - The 1-step-ahead forecast of y_{t+1} given I_t is

$$y_{t+1|t} = E(y_{t+1}|I_t) = y_t$$

  - The 2-step-ahead forecast is

$$y_{t+2|t} = E(y_{t+2}|I_t) = E(y_{t+1} + \varepsilon_t|I_t) = y_{t+1|t} = y_t$$

  - In fact, for any forecast horizon $h > 0$, we have

$$y_{t+h|t} = y_t$$

- The MA representation of the random-walk model is

$$y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots$$

  - this representation is also called the *stochastic trend*
  - the h-step-ahead forecast error is

$$e_t(h) = \varepsilon_{t+h} + \cdots + \varepsilon_{t+1}$$

  - the variance of $e_t(h)$ is

$$\mathrm{Var}[e_t(h)] = h\sigma^2 \to \infty \qquad \text{as } h \to \infty$$

    which implies that an interval forecast of y_{t+h} will approach infinity.
  - The weight $\psi_i = 1$ for all i, indicating that the impact of ε_{t-i} on y_t does not decay over time. The sample ACFs are all approaching 1.

4.2 Random walk with drift

The log return series of a market index tends to have a small and positive mean. This implies that the model for the log price is

$$y_t = \mu + y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2),$$

where $\mu = E(y_t - y_{t-1})$ and the initial log price is y_0.

- The log price consists of a time trend and a stochastic trend

$$\begin{aligned}
y_1 &= \mu + y_0 + \varepsilon_1 \\
y_2 &= \mu + y_1 + \varepsilon_2 = 2\mu + y_0 + \varepsilon_2 + \varepsilon_1 \\
&\vdots \\
y_t &= \mu t + y_0 + \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1 = \mu t + y_0 + \sum_{i=1}^{t} \varepsilon_i
\end{aligned}$$

4.3 Trend-stationary time series

A closely related model that exhibits linear trend is the trend-stationary (TS) time series model,

$$y_t = \beta_0 + \beta_1 t + x_t,$$

where x_t is a stationary time series, e.g., a stationary AR(p) series.

- Here y_t grows linearly in time with rate β_1 and hence can exhibit behavior similar to that of a RW model with drift.

- Major difference between the two models.
  - The RW model with drift: $E(y_t) = y_0 + \mu t$ and $\mathrm{Var}(y_t) = \sigma^2 t$
  - The TS model: $E(y_t) = \beta_0 + \beta_1 t$ and $\mathrm{Var}(y_t) = \mathrm{Var}(x_t) < \infty$

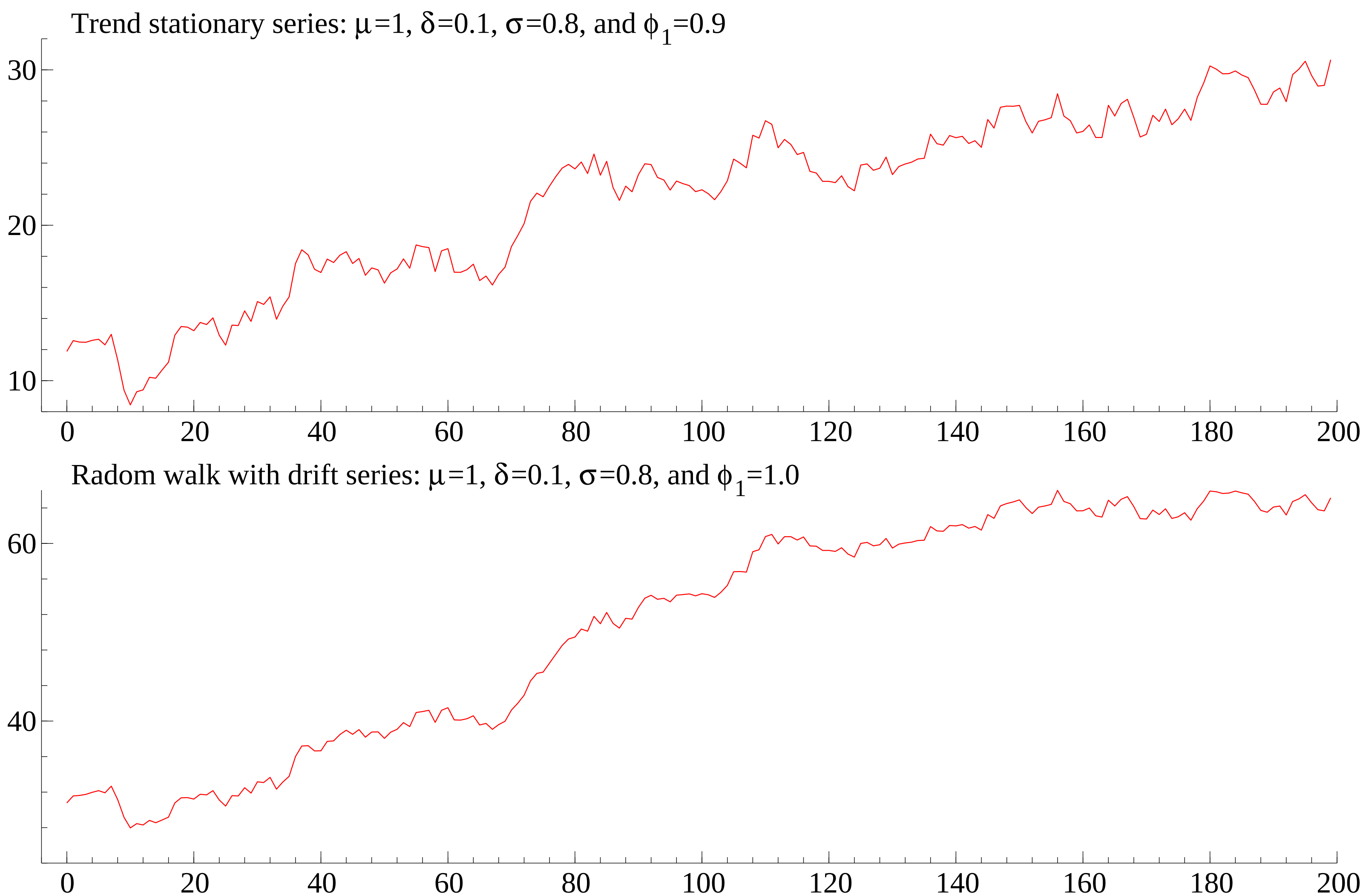


Figure 13 Simulate the process: $y_t = \mu + \delta t + \phi_1(y_{t-1} - \mu - \delta(t-1)) + \sigma W_t$.

4.4 General Unit-Root Nonstationary Models

Consider the following process with stochastic trend

$$y_t = \delta + y_{t-1} + \eta_t, \quad \eta_t = [\phi^*_{p-1}(L)]^{-1}\varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2) \qquad \textbf{(??)}$$

where η_t is weakly stationary, and $\phi^*_{p-1}(L)$ is a lag-polynomial of order $p-1$ which has all roots outside the unit circle.

- The model of y_t can be represented as *an AR(p) process*

$$\phi_p(L)y_t = c + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

where $\phi_p(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $c = \delta\phi^*_{p-1}(1)$. We have $\phi_p(L) = \phi^*_{p-1}(L)(1-L)$, which implies that this process has an unit root.

- A time series with a stochastic trend can be made stationary by applying the differencing filter $\Delta_1 = \Delta$. Therefore, in this case the time series y_t is called *difference-stationary (DS)*.

- A time series y_t that requires the first-differencing filter Δ to remove the stochastic trend is called a time series that is *integrated of order 1*, denoted by $I(1)$.

$I(2)$ series

An $I(2)$ time series needs the Δ_1 filter *twice* to become stationary. Intuitively, a time series with a growth rate that fluctuates as a random walk is an $I(2)$ series.

The impact of past shocks can be demonstrated by recursive substitution of lagged y_t in $\Delta^2 y_t = \delta + \varepsilon_t$ as

$$y_t = y_0 + z_0 t + \delta t(t+1)/2 + \sum_{i=1}^{t} \sum_{j=1}^{i} \varepsilon_j, \tag{31}$$

where y_0 and z_0 are values that depend on pre-sample observations.

This result shows that when δ is positive, an $I(2)$ time series displays explosive growth because of the $t(t+1)/2$ component. Haldrup (1998) gives a comprehensive survey of the analysis of $I(2)$ time series.

Quadratic trend

Again,

$$y_t = \mu^* + \delta^* t + \phi_1 y_{t-1} + \varepsilon_t, \tag{32}$$

When $\phi_1 = 1$ and $\delta^* \neq 0$ in Eq.(32), we say that $\{y_t\}$ is a random walk with *quadratic trend*.

4.5 Testing for unit roots

If $z = 1$ is a solution to the characteristic equation of the AR(p) model, it holds that

$$\phi_p(1) = 1 - \phi_1 - \phi_2 - \cdots - \phi_p = 0. \tag{33}$$

This shows that the sum of the AR parameters equals 1.

To test the empirical validity of the relevant parameter restriction, it is useful to decompose the AR(p) polynomial $\phi_p(L)$ as

$$\phi_p(L) = (1 - \phi_1 - \phi_2 - \cdots - \phi_p)L^i + \phi^*_{p-1}(L)(1 - L), \tag{34}$$

which holds for any $i \in \{1, 2, \ldots, p\}$. We have $\phi^*_{p-1}(L) = 1 - \alpha^*_1 L - \cdots - \alpha^*_{p-1} L^{p-1}$, where $\alpha^*_i = -\sum_{j=i+1}^{p} \phi_j = -\phi_{i+1} - \phi_{i+2} - \cdots - \phi_p$.

We prove the case when $i = 1$.

Proof. We have $\Delta y_t = y_t - y_{t-1} =$

$$= (\phi_1 - 1)y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

$$= (\phi_1 + \phi_2 - 1)y_{t-1} - \phi_2 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

$$= (\phi_1 + \phi_2 + \phi_3 - 1)y_{t-1} - \phi_2 \Delta y_{t-1} - \phi_3 y_{t-1} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

$$= (\phi_1 + \phi_2 + \phi_3 - 1)y_{t-1} - (\phi_2 + \phi_3)\Delta y_{t-1} - \phi_3 \Delta y_{t-2} + \phi_4 y_{t-4} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

$$= (\phi_1 + \cdots + \phi_p - 1)y_{t-1} - (\phi_2 + \phi_3 + \cdots + \phi_p)\Delta y_{t-1}$$

$$- (\phi_3 + \cdots + \phi_p)\Delta y_{t-2} - \cdots - \phi_p \Delta y_{t-p} + \varepsilon_t$$

$$= \rho y_{t-1} + \alpha_1^* \Delta y_{t-1} + \cdots + \alpha_{p-1}^* \Delta y_{t-p+1} + \varepsilon_t.$$

So let $\rho = \sum_{i=1}^{p} \phi_i - 1$ and $\alpha_i^* = -\sum_{j=i+1}^{p} \phi_j = -\phi_{i+1} - \phi_{i+2} - \cdots - \phi_p$, then

$$(1 - \phi_1 L - \cdots - \phi_p L^p)y_t = \varepsilon_t = -\rho L y_t + (1 - \alpha_1^* L - \cdots - \alpha_{p-1}^* L^{p-1})\Delta y_t$$

$$\phi_p(L)y_t = \left[(1 - \phi_1 - \cdots - \phi_p)L + \phi_{p-1}^*(L)(1-L)\right] y_t.$$

Setting $i=1$, an AR(2) polynomial, for example, can be written as

$$1-\phi_1 L-\phi_2 L^2=(1-\phi_1-\phi_2)L+(1+\phi_2 L)(1-L). \tag{35}$$

Replacing $\phi_2(L)$ by the right-hand side of (35), the AR(2) model can thus be written as

$$\phi^*_{p-1}(L)\Delta y_t=(\phi_1+\phi_2-1)y_{t-1}+\varepsilon_t, \tag{36}$$

with $\phi^*_{p-1}(L)=(1-\alpha^*_1 L)=(1+\phi_2 L)$.

When $\phi_1+\phi_2-1$ equals zero, (36) becomes an AR(1) model for Δy_t.

By the same token, we can represent the AR(p) model as

$$\phi_{p-1}^*(L)\Delta y_t = (\phi_1 + \cdots + \phi_p - 1)y_{t-1} + \varepsilon_t$$

or

$$\Delta y_t = (\phi_1 + \cdots + \phi_p - 1)y_{t-1} + \sum_{j=1}^{p-1} \alpha_j^* \Delta y_{t-j} + \varepsilon_t.$$

Augmented Dickey-Fuller (ADF) test

Based on (34), Dickey and Fuller (1979) propose to test for a unit root by testing the *statistical relevance* of y_{t-1} in the *auxiliary regression*

$$\Delta y_t = \rho y_{t-1} + \alpha_1^* \Delta y_{t-1} + \cdots + \alpha_{p-1}^* \Delta y_{t-(p-1)} + \varepsilon_t. \tag{37}$$

- The null hypothesis is $\rho = 0$ and the relevant alternative is $\rho < 0$, resulting in a *one-sided* t test-statistic, i.e., $t(\hat{\rho}) = \dfrac{\hat{\rho}}{\mathrm{std}(\hat{\rho})}$.

- Phillips (1987) derives the *asymptotic distribution* of $t(\hat{\rho})$, which is *nonstandard* because under the null hypothesis of a unit root, y_t is nonstationary and standard limit theory does not apply.

- The *critical values* for $t(\hat{\rho})$ have to be obtained through *Monte Carlo simulation*.

In general, it appears best to *include an intercept and linear trend* in the ADF regression. The ADF regression then becomes

$$\Delta y_t = \mu^{**} + \delta^{**} t + \rho y_{t-1} + \alpha_1^* \Delta y_{t-1} + \cdots + \alpha_{p-1}^* \Delta y_{t-(p-1)} + \varepsilon_t, \quad (38)$$

- Under the unit root hypothesis, ρ and δ^{**} are both equal to zero in the case of linear trend, and $\rho = \mu^{**} = 0$ in the case of *no trends*.

- To make the test independent of *nuisance parameters*, Dickey and Fuller (1981) develop a *joint F-test* for the hypothesis $\rho = \delta^{**} = 0$, and in the case of no trends, for $\rho = \mu^{**} = 0$.

- A common practical procedure, however, is to test $\rho = 0$ in (38) and to consider critical values depending on the type of deterministic regressors included.

ADF Test Procedure for Unit Root Testing

Our possible models are:

$$\Delta y_t = \rho y_{t-1} + \alpha_1^* \Delta y_{t-1} + \cdots + \alpha_{p-1}^* \Delta y_{t-(p-1)} + \varepsilon_t, \tag{39}$$

$$\Delta y_t = \rho y_{t-1} + \alpha_1^* \Delta y_{t-1} + \cdots + \alpha_{p-1}^* \Delta y_{t-(p-1)} + \mu^{**} + \varepsilon_t, \tag{40}$$

$$\Delta y_t = \rho y_{t-1} + \alpha_1^* \Delta y_{t-1} + \cdots + \alpha_{p-1}^* \Delta y_{t-(p-1)} + \mu^{**} + \delta^{**} t + \varepsilon_t, \tag{41}$$

and the procedure is as follows:

1. Graphically inspect y_t in levels.

2. If y_t appear to be trending, then run model (41) and test: $H_0 : \rho = 0$.

- Non-rejection: Random walk with quadratic trend (*highly unlikely, we should check whether the trend coefficient δ^{**} is equal to zero)*.
  - If $\delta^{**} = 0$, the series y_t is a random walk with *drift*.
- Rejection implies the process is trend stationary *(more likely)*.

3. If y_t appears to have no trend, then run model (40) and test: $H_0 : \rho = 0$.

- Non-rejection implies random walk with drift (*unlikely, also we should check whether the drift coefficient μ^{**} is equal to zero))*.
  - If $\mu^{**} = 0$, the series y_t is a random walk with *nonzero* initial value.
- Rejection implies stationary about non-zero level (more likely).

4. If y_t appears mean zero without trend, then run model (39) and test $H_0 : \rho = 0$.

- Non-rejection implies random walk (with zero initial value).
- Rejection implies stationary process with zero level.

5. Regression Models with Time Series Errors

Motivated examples:

- The market model in finance or capital asset pricing model that relates the excess return of an individual stock to that of a market index.

- The relationship between interest rates with different maturities. (The term structure of interest rates)

These examples lead naturally to the consideration of a linear regression in the form

$$y_t = \alpha + \beta x_t + e_t,$$

where y_t and x_t are two time series and e_t denotes the error term.

- If $\{e_t\}$ is a white noise series, then the least square (LS) method produces consistent estimates.

- In practice, the error term e_t is serially correlated and then the LS estimates of α and β may not be consistent.

- A regression model with time series errors is widely applicable in economics and finance, but it is one of the most commonly misused econometric models because the series dependence in e_t is often overlooked.

Example 6. We introduce the model by considering the relationship between two U.S. weekly interest rate series:

1. r_{1t}: the 1-year Treasury constant maturity rate;
2. r_{3t}: the 3-year Treasury constant maturity rate.

Both series have 2467 observations from January 5, 1962, to April 10, 2009, and are measured in percentages.

- A naive way to describe the relationship between the two interest rates is to use the simple model
$$r_{3t} = \alpha + \beta r_{1t} + e_t.$$
- The fitted model is
$$r_{3t} = \underset{(0.024)}{0.832} + \underset{(0.004)}{0.930}\, r_{1t} + e_t, \quad \hat{\sigma}_e = 0.523$$
with $R^2 = 96.5\%$.

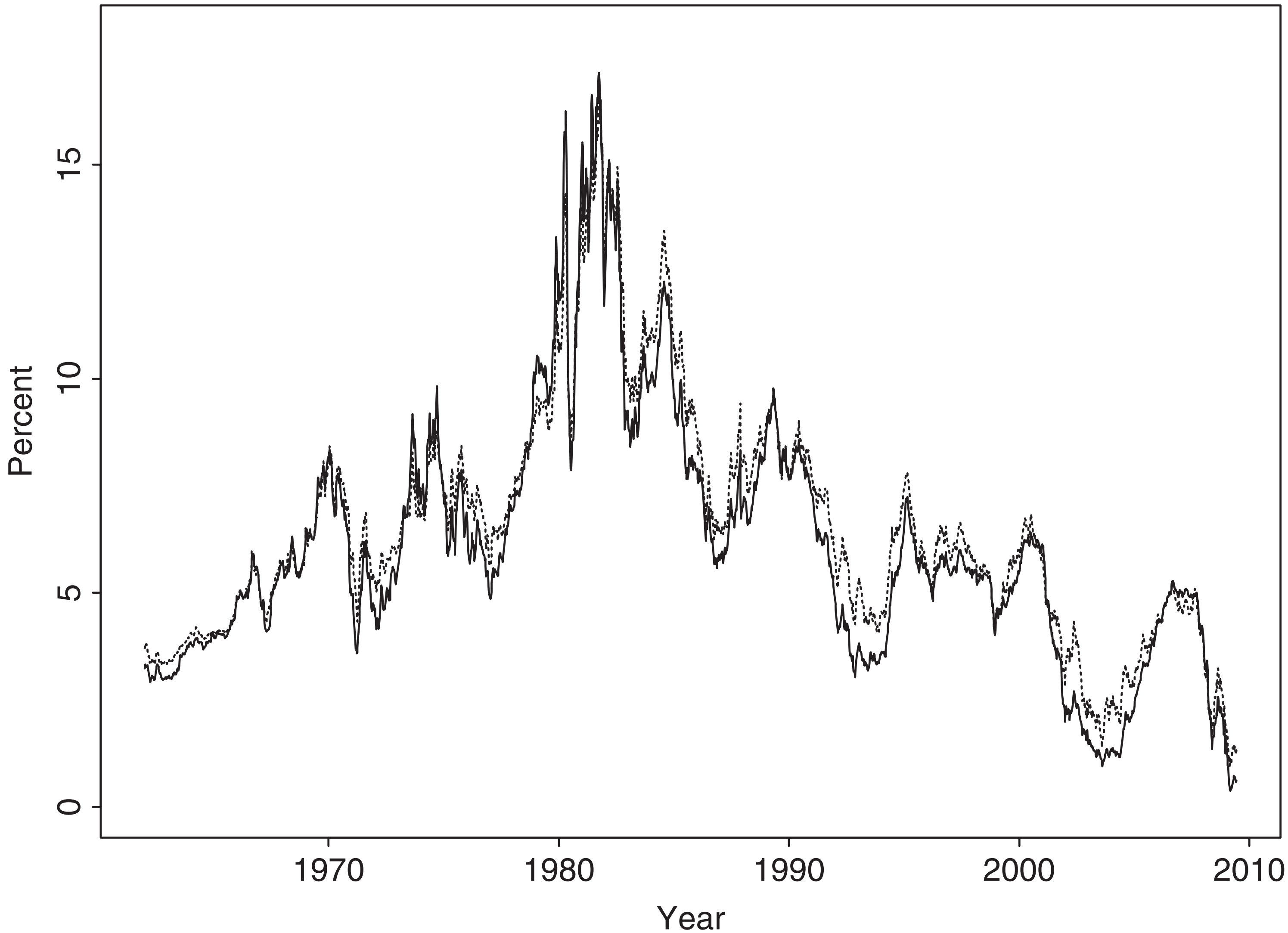


Figure 14 Time plots of U.S. weekly interest rates (in percentages) from January 5, 1962, to April 10, 2009. Solid line is Treasury 1-year constant maturity rate and dashed line Treasury 3-year constant maturity rate.

- However, the model is seriously inadequate.
  - Figure 15 shows that the sample ACF of the residuals is highly significant and decays slowly, showing the pattern of a unit-root nonstationary time series.
  - If the two interest rate series are unit-root nonstationary, the behavior of the residuals indicates that the two interest rates are not cointegrated.
  - In other words, the data fail to support the hypothesis that there exists a long-term equilibrium between the two interest rates.
- This leads to the consideration of the change series of interest rates and consider the linear regression

$$c_{3t} = \beta c_{1t} + e_t.$$

where $c_{1t} = r_{1t} - r_{1,t-1} = \Delta r_{1t}$ and $c_{3t} = r_{3t} - r_{3,t-1} = \Delta r_{3t}$.

- The fitted model for differenced series is

$$c_{3t} = \underset{(0.0073)}{0.792}\, c_{1t} + e_t, \quad \hat{\sigma}_e = 0.0690.$$

with $R^2 = 82.5\%$.

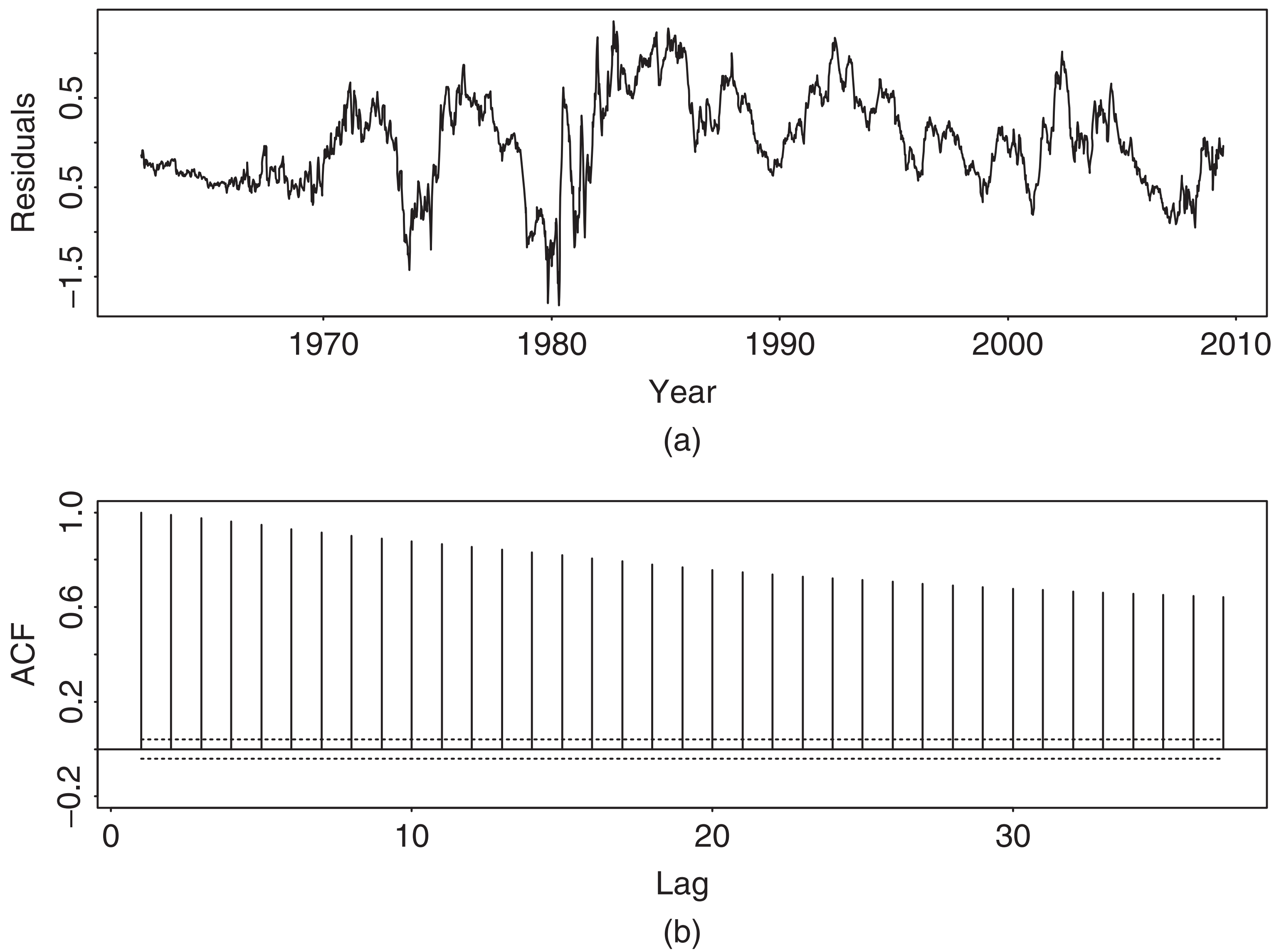


Figure 15 Residual series of linear regression for two U.S. weekly interest rates: (a) time plot and (b) sample ACF.

- The ACF of the residuals shown in Figure 16 shows some significant serial correlations in the residuals, but magnitudes of the correlations are much smaller.
- Because residuals of the model are serially correlated, we shall identify a simple ARMA model for the residuals.
- From the sample ACF of the residuals shown in Figure 16, we specify an MA(1) model for the residuals and modify the linear regression model to

$$c_{3t} = \beta c_{1t} + e_t, \quad e_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

  - This is a simple example of linear regression with time series errors.
  - If the time series model used is stationary and invertible, then one can estimate the model jointly via the maximum-likelihood method.
- The fitted model is then

$$c_{3t} = \underset{(0.0075)}{0.794}\, c_{1t} + e_t, \quad e_t = \varepsilon_t + \underset{(0.0196)}{0.1823}\, \varepsilon_{t-1} \quad \hat{\sigma}_e = 0.0678.$$

with $R^2 = 83.1\%$. The model no longer has a significant lag-1 residual ACF.

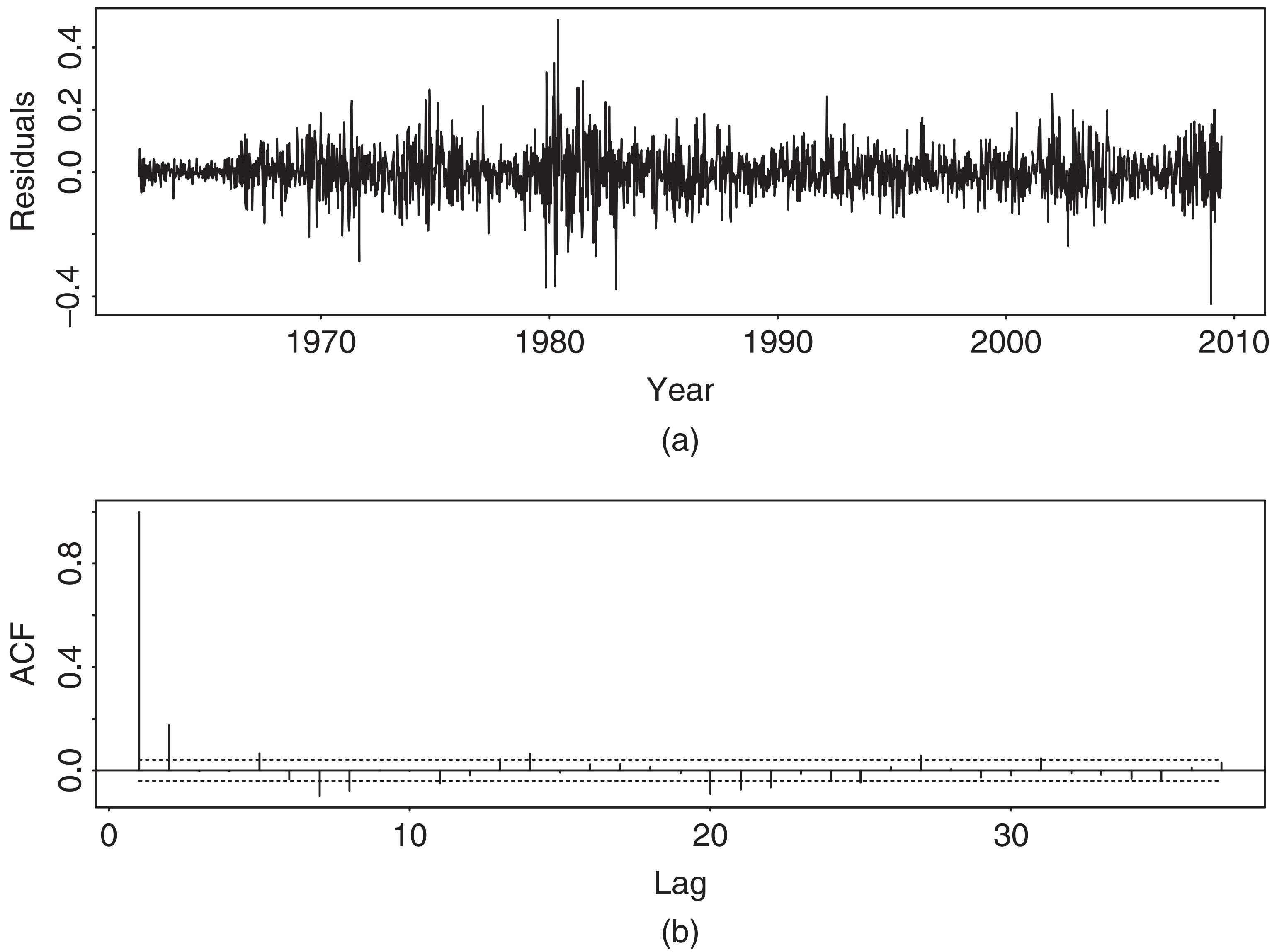


Figure 16 Residual series of linear regression for two differenced series of U.S. weekly interest rates: (a) time plot and (b) sample ACF.

Summary

A general procedure for analyzing linear regression models with time series errors:

1. Fit the linear regression model and check serial correlations of the residuals.

2. If the residual series is unit-root nonstationary, take the first difference of both the dependent and explanatory variables. Go to step 1. If the residual series appears to be stationary, identify an ARMA model for the residuals and modify the linear regression model accordingly.

3. Perform a joint estimation via the maximum-likelihood method and check the fitted model for further improvement.

6. Long Memory Time Series

If a time series y_t is $I(0)$ then its ACF declines at a *geometric rate*. As a result, $I(0)$ process have *short memory* since observations far apart in time are essentially independent.

Conversely, if y_t is $I(1)$ then its ACF declines at a *linear rate* and observations far apart in time are not independent.

In between $I(0)$ and $I(1)$ processes are so-called *fractionally integrated* $I(d)$ process where $0 < d < 1$. The ACF for a fractionally integrated processes declines at a *polynomial (hyperbolic) rate*, which implies that observations far apart in time may exhibit *weak but non-zero correlation*. This weak correlation between observations far apart is often referred to as *long memory*.

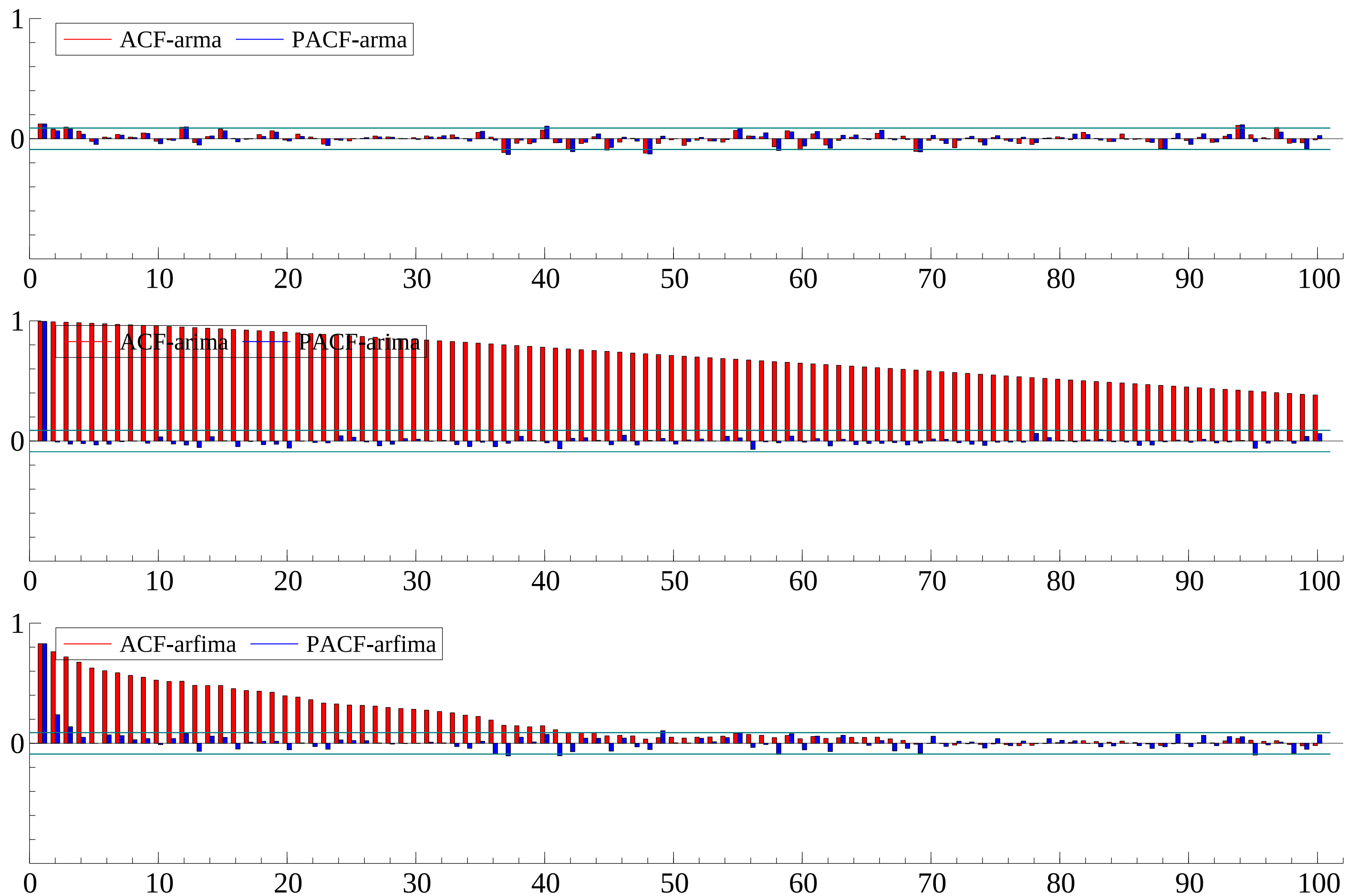


Figure 17 ACF and PACF for ARMA(1,1), ARIMA(1,$d = 1$,1), and ARFIMA(1,$d = 0.48$,1) with $\phi_1 = 0.6$ and $\theta_1 = -0.5$.

Theorem 3 (*Newton's Binomial Formula*). *For any positive integer n and real x, y we have*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k. \tag{42}$$

Newton's binomial formula extends to the following theorem, which can be found in most calculus texts:

Theorem 4. *For any a we have*

$$(1+x)^a = \sum_{k=0}^{\infty} \binom{a}{k} x^k. \tag{43}$$

Proof. It suffices to observe that the right hand side of (43) is the Taylor expansion of the left-hand side, $\binom{a}{k}$ equals $\frac{1}{k!}\frac{d^{(k)}}{dx^k}(1+x)^a$ for $x = 0$. $\square$

Fractionally integrated process

A fractionally integrated white noise process y_t has the form

$$(1-L)^d y_t = \varepsilon_t, \quad \varepsilon_t \sim WN(0,\sigma^2), \tag{44}$$

where the *differencing operator* $(1-L)^d$ has the binomial series expansion representation (valid for any $d > -1$)

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k, \tag{45}$$

where the binomial coefficients $\binom{d}{k}$ are defined by

$$\binom{d}{k} = \frac{d(d-1)(d-2)\cdots(d-k+1)}{k!},$$

and we have

$$(-1)^k \binom{d}{k} = \frac{-d(1-d)(2-d)\cdots(k-1-d)}{k!} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

since $\Gamma(t+1) = t\Gamma(t)$ and $k!$ (read as k factorial) is

$$k! = \Gamma(k+1) = 1 \cdot 2 \cdot 3 \cdots (k-1) \cdot k, \quad \text{with } 0! = 1.$$

Using these definitions, we can rewrite (45) as

$$\begin{aligned}
(1-L)^d &= 1 - dL + \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 + \cdots \\
&= \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k.
\end{aligned}$$

- If $d = 1$ then y_t is a random walk and if $d = 0$ then y_t is white noise.

- If $d < 0.5$, then y_t is a weakly stationary process and has the infinite MA representation

$$y_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}$$

with

$$\psi_k = \frac{d(1+d)\cdots(k-1+d)}{k!} = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}.$$

- If $d > -0.5$, then y_t is invertible and has the infinite AR representation

$$y_t = \sum_{i=1}^{\infty} \pi_i y_{t-i} + \varepsilon_t$$

with

$$\pi_k = \frac{-d(1-d)\cdots(k-1-d)}{k!} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}.$$

- For $-0.5 < d < 0.5$ it can be shown that

$$\rho_k = \frac{d(1+d)\cdots(k-1+d)}{(1-d)(2-d)\cdots(k-d)} = \frac{\Gamma(k+d)}{\Gamma(1+k-d)}\frac{\Gamma(1-d)}{\Gamma(d)},$$

for $k = 1, 2, \ldots$. In particular, $\rho_1 = d/(1-d)$ and

$$\rho_k \simeq \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \propto k^{2d-1}$$

as $k \to \infty$ so that the ACF for y_t declines *hyperbolically* to zero at a speed that depends on d.

- For $-0.5 < d < 0.5$, the PACF of y_t is $\phi_{k,k} = d/(k-d)$ for $k = 1, 2, \ldots$.

- Further, it can be shown y_t is stationary and ergodic for $0 < d < 0.5$ and that the variance of y_t is infinite for $0.5 \le d < 1$.

ARFIMA(p, d, q) model

A fractionally integrated process with stationary and ergodic ARMA(p, q) errors

$$(1 - L)^d y_t = u_t, \quad u_t \sim ARMA(p, q)$$

is called an *autoregressive fractionally integrated moving average* (ARFIMA) process, i.e.

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t.$$

Estimation routines for general ARFIMA models, which include additional AR and MA parts in (44) are proposed in Sowell (1992) and Beran (1995).

Reading

The readers are recommend to refer to ”A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima package 1.04 for Ox” by JURGEN A. DOORNIK AND MARIUS OOMS.