Simple AR models: (Regression with lagged variables.)

AR(1) model:

$$r_t = \phi_1 r_{t-1} + a_t,$$

where ϕ_1 is real number,which is referred to as parameters (to be estimated from the data in an application).

For example,

$$r_t = 0.2 r_{t-1} + a_t.$$

($r_{t-1} = 0.2 r_{t-2} + a_{t-1}$ and $r_{t+1} = 0.2 r_t + a_{t+1}$)

1. $r_t = a_t + \phi_1 a_{t-1} + \cdots + \phi_1^m a_{t-m} + \phi_1^{m+1} r_{t-m-1}$.

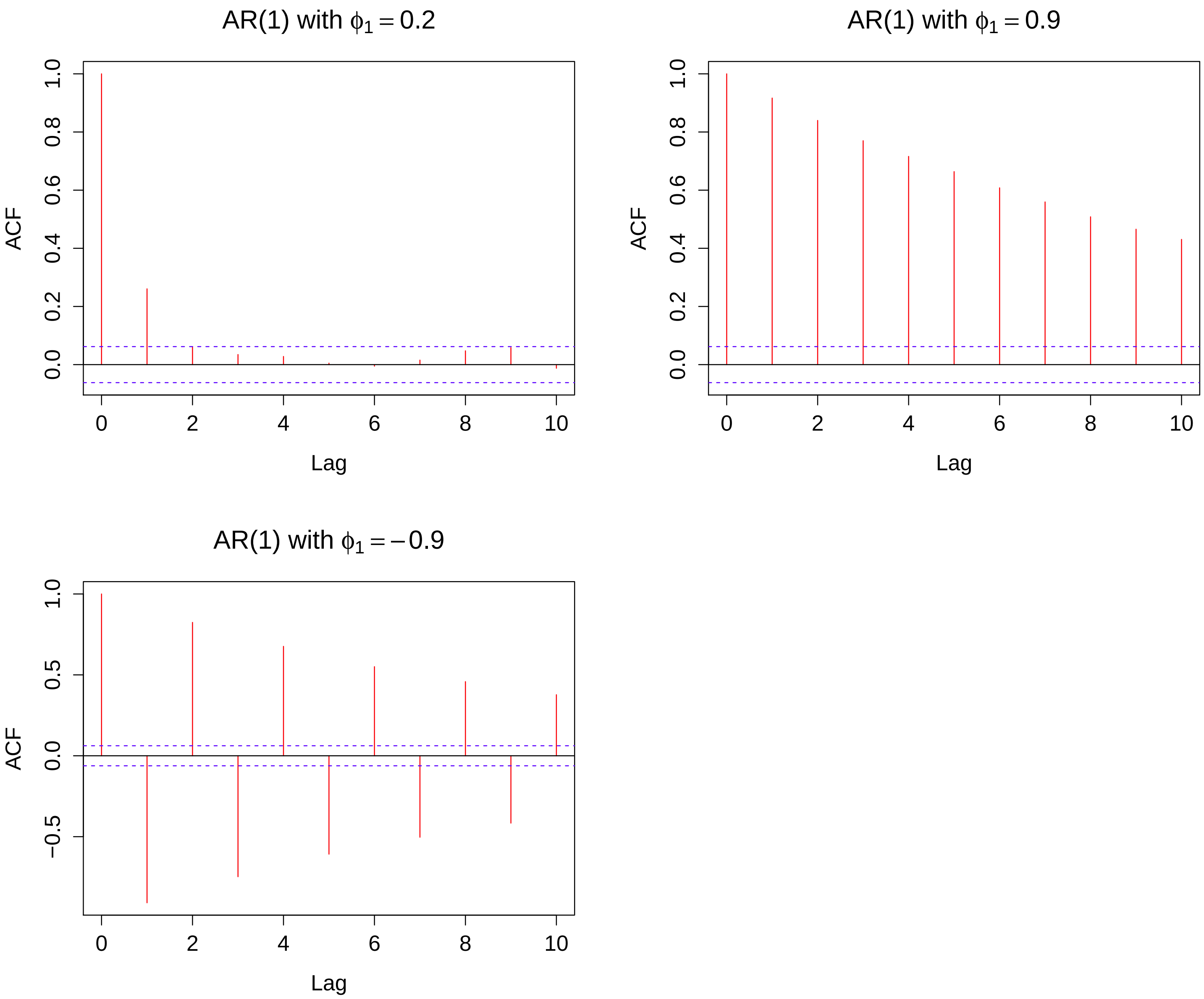
2. Stationarity: necessary and sufficient condition $|\phi_1| < 1$. Why?

$$r_t = a_t + \sum_{i=1}^{\infty} \phi_1^i a_{t-i}.$$

3. Mean: $E(r_t) = 0$.

4. Variance: $Var(r_t) = \frac{\sigma_a^2}{1-\phi^2}$.

5. Autocorrelations: $\rho_1 = \phi_1$, $\rho_2 = \phi_1^2$, etc. In general, $\rho_k = \phi_1^k$ and ACF ρ_k decays exponentially as k increases,



ACF of AR(1) model.

- When $|\phi_1| < 1$, the series $\{r_t\}$ exhibits mean-reverting behavior:

  Let $\Delta r_t = r_t - r_{t-1} = (\phi_1 - 1)r_{t-1} + a_t$,

  $$E(\Delta r_t|F_{t-1}) = (\phi_1 - 1)r_{t-1} < 0, \text{ if } r_{t-1} > 0;$$
  $$E(\Delta r_t|F_{t-1}) = (\phi_1 - 1)r_{t-1} > 0, \text{ if } r_{t-1} < 0.$$

- When $\phi_1 = 1$, AR(1) model can be rewritten as the RW process:

  $$r_t = a_t + a_{t-1} + \cdots + a_1 + r_0.$$

- When $|\phi_1| > 1$, e.g. $\phi = 3$,

  $r_t = a_t + 3a_{t-1} + 3^2 a_{t-2} + \cdots + 3^{t-1}a_1 + 3^t r_0.$

  r_t is called the explosive process, in the sense that r_t diverges to ∞.

AR(1) model with a drift:

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t,$$

where ϕ_0 and ϕ_1 are real numbers.

$$Er_t = \frac{\phi_0}{1-\phi_1},$$

$Var(r_t)$ and ρ_k are not changed.

6. Forecast (minimum squared error):

(a) 1-step ahead forecast at time n, the forecast origin:

$$\widehat{r}_n(1) = \phi_0 + \phi_1 \widehat{r}_n.$$

(b) 1-step ahead forecast error:

$$e_n(1) = r_{n+1} - \widehat{r}_n(1) = a_{n+1}.$$

Thus, a_{n+1} is the un-predictable part of r_{n+1}. It is the shock at time $n+1$!

(c) Variance of 1-step ahead forecast error:

$$Var[e_n(1)] = Var(a_{n+1}) = \sigma_a^2.$$

(d) 2-step ahead forecast:

$$\widehat{r}_n(2) = \phi_0 + \phi_1 \widehat{r}_n(1).$$

(e) 2-step ahead forecast error:

$$e_n(2) = r_{n+2} - r_n(2) = a_{n+2} + \phi_1 a_{n+1}.$$

(f) Variance of 2-step ahead forecast error:

$$Var[e_n(2)] = (1 + \phi_1^2)\sigma_a^2.$$

which is greater than or equal to $Var[e_n(1)]$, implying that uncertainty in forecasts increases as the number of steps increases.

(g) Behavior of l-step ahead forecasts:

Forecasting error:

$$e_n(l) = r_{n+l} - \widehat{r}_n(l) = \sum_{j=0}^{l-1} \phi_1^j a_{n+l-j}.$$

Forecasting variance:

$$\text{Var}[e_n(l)] = \sigma_a^2 \sum_{j=0}^{l-1} \phi_1^{2j}.$$

Forecast interval (limit) (FI):

$$\left[\widehat{r}_n(l) - N_{\frac{\alpha}{2}}\sigma_a\sqrt{\sum_{j=0}^{l-1}\phi_1^{2j}},\ \widehat{r}_n(l) + N_{\frac{\alpha}{2}}\sigma_a\sqrt{\sum_{j=0}^{l-1}\phi_1^{2j}}\right]$$

where $N_{\frac{\alpha}{2}}$ is the $\alpha/2$-quantile of the standard normal distribution, i.e., $P(N > N_{\frac{\alpha}{2}}) = \alpha/2$.

When $\alpha = 0.05$, $N_{\frac{\alpha}{2}} = 1.96$.

7. A compact form: $(1 - \phi_1 B)r_t = \phi_0 + a_t$.

AR(2) model:

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t,$$

$$or\ (1 - \phi_1 B - \phi_2 B^2) r_t = a_t.$$

1. Stationarity condition:

all the roots of $(1 - \phi_1 x - \phi_2 x^2) = 0$ lie outside the unit circle.

Decompose $1 - \phi_1 z - \phi_2 z^2 = (1 - \alpha_1 z)(1 - \alpha_2 z)$.

Then $|\alpha_1| < 1$ and $|\alpha_2| < 1$.

$$(1 - \alpha_1 B)(1 - \alpha_2 B) r_t = a_t.$$

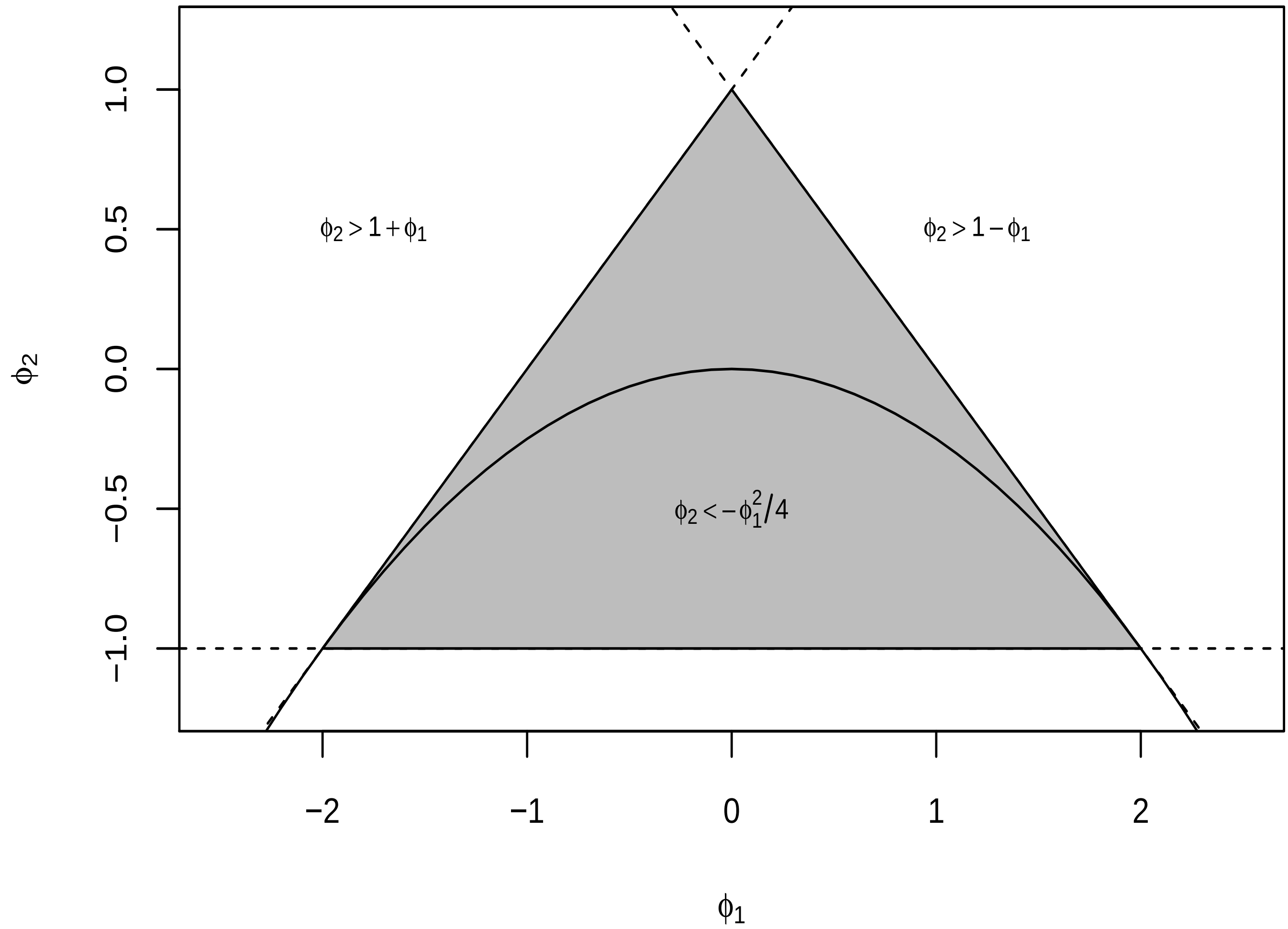
Let $u_t = (1 - \alpha_2 B) r_t$. Then

$$u_t = \alpha_1 u_{t-1} + a_t = a_t + \sum_{i=1}^{\infty} \alpha_1^i a_{t-i}.$$

$$r_t = \alpha_2 r_{t-1} + u_t = u_t + \sum_{j=1}^{\infty} \alpha_2^j u_{t-j}.$$

Stationarity condition is equivalent to

$$\begin{cases} \phi_2 + \phi_1 < 1, \\ \phi_2 - \phi_1 < 1, \\ -1 < \phi_2 < 1. \end{cases}$$



Stationary region for AR(2) model.

2. Mean: $E(r_t) = 0$

3. ACF: $\rho_0 = 1$, $\rho_1 = \frac{\phi_1}{1-\phi_2}$,

$$\rho_l = \phi_1 \rho_{l-1} + \phi_2 \rho_{l-2}, \;\; l \geq 2.$$

The solutions of $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0$:

$$\omega_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}.$$

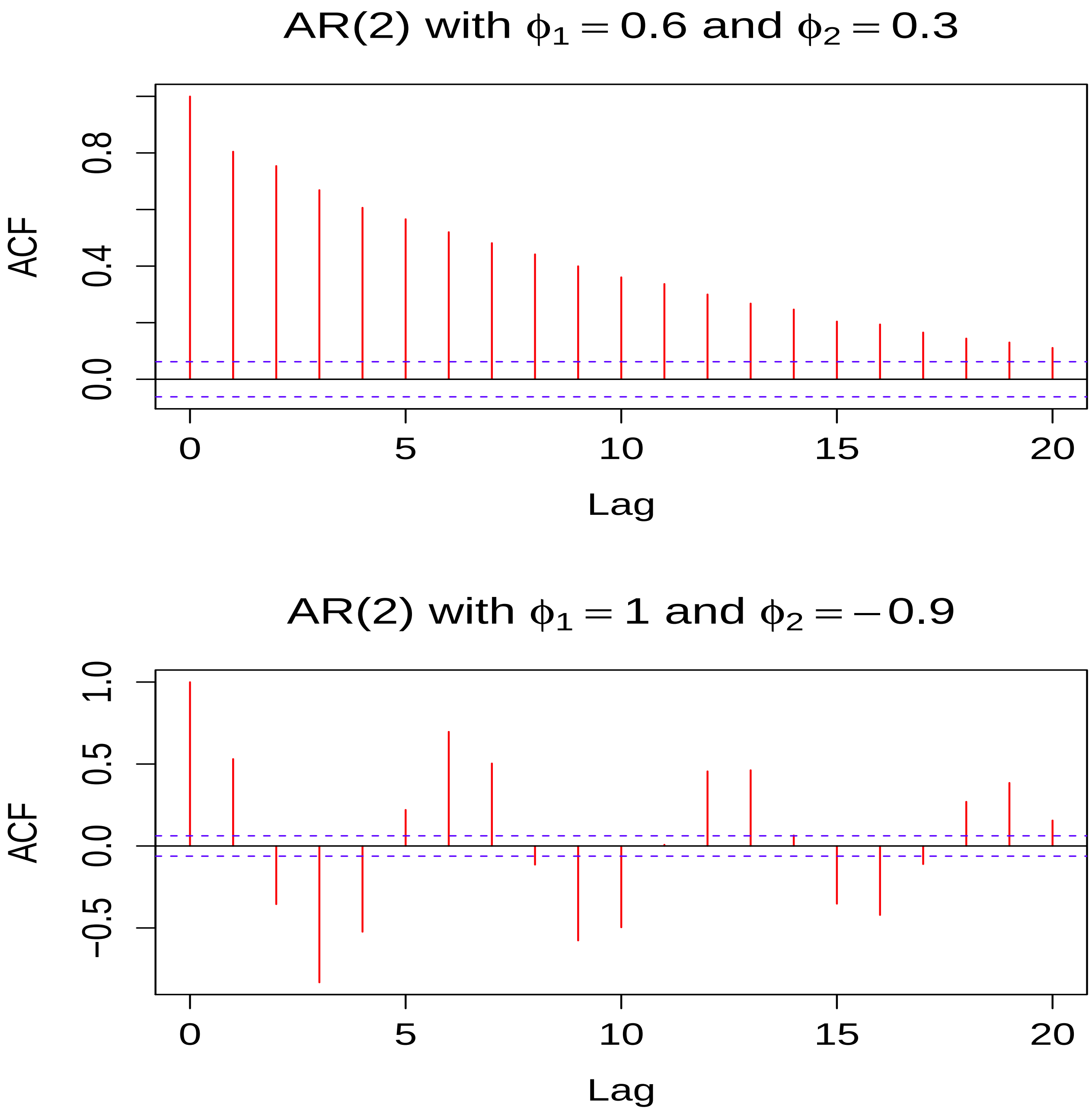
The ACF is a mixture of two exponential decays as the general solution of

$$\rho_l = \frac{c_1}{\omega_1^l} + \frac{c_2}{\omega_2^l},$$

where c_1 and c_2 are unknown parameters depending on the initial conditions.

(a). If both roots ($\omega_{1,2}$) are real, then the ACF of AR(2) is a mixture of two exponential decay.

(b). 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.



The ACF of AR(2) model.

4. Stochastic business cycle: if $\phi_1^2 + 4\phi_2 < 0$, then r_t shows characteristics of business cycles with average length

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

where the cosine inverse is stated in radian. If we denote the solutions of the polynomial as $a \pm bi$, where $i = \sqrt{-1}$, then we have $\phi_1 = 2a$ and $\phi_2 = -(a^2 + b^2)$ so that

$$k = \frac{2\pi}{cos^{-1}(a/\sqrt{a^2 + b^2})}.$$

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

$$r_t - 1.3512 r_{t-1} + 0.4612 r_{t-1} = a_t.$$

Clearly, $\phi_1^2 - 4\phi_2 = 1.35122 - 4 \times 0.4612 = -0.0191 < 0$, which implies the existence of stochastic cycles. The average length of the stochastic cycles for this AR(2) model is approximately

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

which is about 5 years.

5. Forecasts: Similar to AR(1) models

AR(2) model with a drift:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t,$$

$$E(r_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2},$$

others are unchanged.

Building an AR model–a general form:

$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \phi_p r_{t-p} + a_t$

Stationarity condition:

all the roots of $\phi_p(z) \equiv 1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p z^p = 0$ lie outside the unit circle.

1. Order specification

Using Partial ACF (PACF): (naive, but effective)

Definition:
Let r_t be a stationary TS process. The conditional correlation

$$
\begin{aligned}
&Corr(r_t, r_{t+k} | r_{t+1}, \cdots, r_{t+k-1}) \\
&\qquad = \frac{Cov[(r_t - \widehat{r}_t)(r_{t+k} - \widehat{r}_{t+k})]}{\sqrt{Var(r_t - \widehat{r}_t) Var(r_{t+k} - \widehat{r}_{t+k})}}
\end{aligned}
$$

is called the PACF of r_t and r_{t+k}, denoted by ϕ_{kk}, where $\widehat{Z}_t$ and $\widehat{Z}_{t+k}$ is the best linear estimate of Z_{t+k} given $\{Z_{t+1}, \cdots, Z_{t+k-1}\}$ in mean square errors as $k \geq 2$. $\widehat{Z}_t = \widehat{Z}_{t+k} = EZ_t$ as $k = 1$.

Formula: $\phi_{11} = \rho_1$,

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ & & \cdots & & & \\ & & \cdots & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ & & \cdots & & & \\ & & \cdots & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

Sample PACF: given observations $r_1, \ldots, r_n$,

$\rho_0, \ldots, \rho_{n-1}$ are estimated by $\widehat{\rho}_0, \ldots, \widehat{\rho}_{n-1}$.

ϕ_{kk} is estimated by

$$\widehat{\phi}_{kk} = \frac{\begin{vmatrix} 1 & \widehat{\rho}_1 & \widehat{\rho}_2 & \cdots & \widehat{\rho}_{k-2} & \widehat{\rho}_1 \\ \widehat{\rho}_1 & 1 & \widehat{\rho}_1 & \cdots & \widehat{\rho}_{k-3} & \widehat{\rho}_2 \\ & & \cdots & & & \\ & & \cdots & & & \\ \widehat{\rho}_{k-1} & \widehat{\rho}_{k-2} & \widehat{\rho}_{k-3} & \cdots & \widehat{\rho}_1 & \widehat{\rho}_k \end{vmatrix}}{\begin{vmatrix} 1 & \widehat{\rho}_1 & \widehat{\rho}_2 & \cdots & \widehat{\rho}_{k-2} & \widehat{\rho}_{k-1} \\ \widehat{\rho}_1 & 1 & \widehat{\rho}_1 & \cdots & \widehat{\rho}_{k-3} & \widehat{\rho}_{k-2} \\ & & \cdots & & & \\ & & \cdots & & & \\ \widehat{\rho}_{k-1} & \widehat{\rho}_{k-2} & \widehat{\rho}_{k-3} & \cdots & \widehat{\rho}_1 & 1 \end{vmatrix}}$$

- Key feature: PACF cuts off at lag p for an AR(p) model, i.e. the lag-p sample PACF should not be zero.

- $\widehat{\phi}_{k,k}$ should be close to zero for all $j > p$.

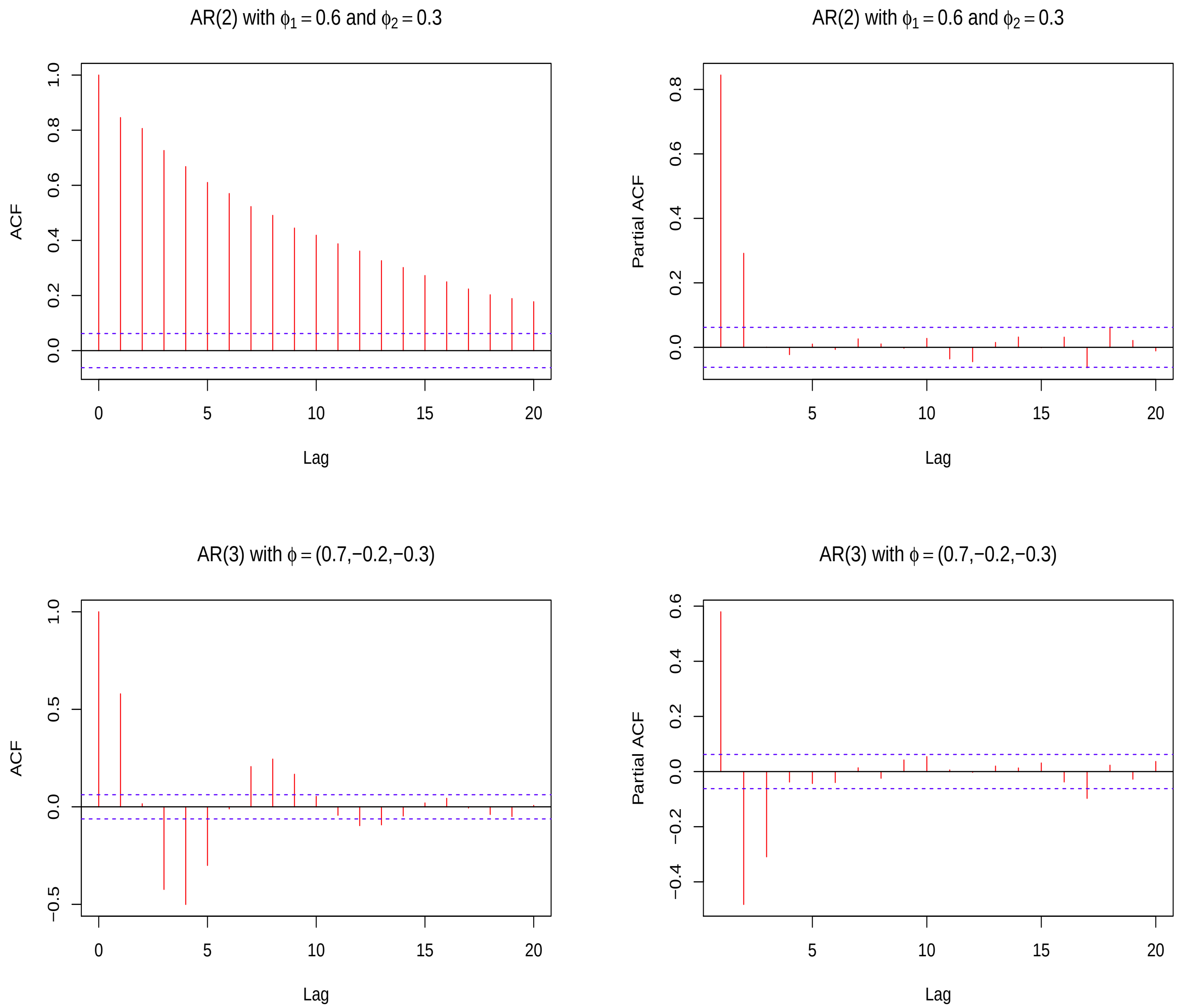
- Its distribution is

$$\widehat{\phi}_{kk} \sim N(0, \frac{1}{n}).$$

  Comparing $\widehat{\phi}_{k,k}$ to $\pm 1.96/\sqrt{n}$ and decide if $\phi_{k,k} = 0$ is preferable.

(See Text (p. 40) for details)

– Illustration



PACF of AR(2) and AR(3) models.

2. Estimation: least squares estimator (LSE) or maximum likelihood estimator (MLE)

Assume $\{r_1, \cdots, r_n\}$ is from AR(p) model with parameter $\tilde{\phi}_0 = (\phi_{00}, \phi_{01}, \cdots, \phi_{0p})'$, i.e.

$$r_t = \phi_{00} + \phi_{01} r_{t-1} + \phi_{02} r_{t-2} + \cdots + \phi_{0p} r_{t-p} + a_t$$

$\tilde{\phi}_0$ is called the true parameter.

Note that

$$r_t \neq \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \phi_p r_{t-p} + a_t$$

if $\phi_i \neq \phi_{0i}$.

LSE:

$$\begin{aligned} a_t(\phi) &= r_t - \phi_0 - \phi_1 r_{t-1} - \phi_2 r_{t-2} - \cdots - \phi_p r_{t-p} \\ &= r_t - \widetilde{Z}'_{t-1}\phi, \end{aligned}$$

where $\widetilde{Z}_{t-1} = (1, r_{t-1}, r_{t-2}, \cdots, r_{t-p})'$, and $\phi = (\phi_0, \phi_1, \cdots, \phi_p)'$ is called unknown parameter.

$$a_t(\phi) = a_t \ if\ and\ only\ if\ \phi = \tilde{\phi}_0$$

$a_t(\phi)$ is called residuals.

$$S_*(\phi) = \sum_{t=1}^{n} a_t^2(\phi)$$

$$\frac{\partial S_*(\phi)}{\partial \phi} = 2 \sum_{t=1}^{n} \frac{\partial a_t(\phi)}{\partial \phi} a_t(\phi)$$

$$= -2 \sum_{t=1}^{n} \widetilde{Z}_{t-1} r_t + 2 \sum_{t=1}^{n} \widetilde{Z}_{t-1} \widetilde{Z}'_{t-1} \phi$$

Note that $\frac{\partial S_*(\phi)}{\partial \phi} |_{\phi=\widehat{\phi}} = 0$, we have

$$\widehat{\phi} = \left(\sum_{t=1}^{n} \widetilde{Z}_{t-1} \widetilde{Z}'_{t-1} \right)^{-1} \left(\sum_{t=1}^{n} \widetilde{Z}_{t-1} r_t \right).$$

$$\frac{\partial^2 S_*(\phi)}{\partial \phi \partial \phi'} = 2 \sum_{t=1}^{n} \widetilde{Z}_{t-1} \widetilde{Z}'_{t-1} > 0.$$

$$a_t(\widehat{\phi}) = r_t - \widehat{\phi}_0 - \widehat{\phi}_1 r_{t-1} - \widehat{\phi}_2 r_{t-2} - \cdots - \widehat{\phi}_p r_{t-p}.$$

is called the residual.

$$\widehat{\sigma}_a^2 = \frac{1}{n} \sum_{t=1}^{n} a_t(\widehat{\phi})^2.$$

Theory:

$$\sqrt{n}(\widehat{\phi} - \phi_0) \sim N(0, \widehat{\Omega}).$$

$$\sqrt{n}(\widehat{\phi}_i - \phi_{0i}) \sim N(0, \widehat{\sigma}_{ii}).$$

where

$$\widehat{\Omega} = \left(\sum_{t=1}^{n} \widetilde{Z}_{t-1} \widetilde{Z}'_{t-1} \right)^{-1} \widehat{\sigma}_a^2$$

and $\widehat{\sigma}_{ii}$ is the (i,i)-element of $\widehat{\Omega}$.

In R, s.e. is $\sqrt{\widehat{\sigma}_{ii}/n}$.

$$H_0 : \phi_{0i} = 0 \ v.s. \ H_a : \phi_{0i} \neq 0$$

If $|\widehat{\phi}_i| > 1.96 s.e. \approx 2 s.e.$, then reject H_0. ϕ_{0i} is considered to non-zero at significance level $\alpha = 0.05$.

3. Model checking:

Residual should be close to white noise if the model is adequate. Sample ACF of Residuals is

$$\widehat{\rho}_k = \frac{\sum_{t=1}^{T-k}[a_t(\widehat{\phi}) - \bar{a}][a_{t+k}(\widehat{\phi}) - \bar{a}]}{\sum_{t=1}^{T}[a_t(\widehat{\phi}) - \bar{a}]^2},$$

where $\bar{a} = \sum_{t=1}^{n} a_t(\widehat{\phi})/n$. If model is correct, then the ACF of $\{\widehat{\rho}_k\}$ is closed to that of $\{a_t\}$, i.e.,

$$\widehat{\rho}_k \approx 0.$$

Joint test (Ljung-Box statistics):

$$Q(m) = T(T+2)\sum_{k=1}^{m} \frac{\widehat{\rho}_k^2}{T-k} \sim \chi^2_{m-g},$$

where g is the number of AR coefficients used in the model.

$$H_0: \text{ model correct } vs\ H_a: \text{ model incorrect.}$$

Reject H_0 if

$$Q(m) > \chi^2_{m-g}(\alpha) \text{ or } p = P(\chi^2_{m-g} > Q(m)) < \alpha.$$

4. Model selection:

(i). Akaike information criterion

$$AIC(l) = \ln(\widehat{\sigma}_l^2) + \frac{2l}{T},$$

for an AR(l) model, where $\widehat{\sigma}_l^2$ is the MLE of residual variance.

Find the AR order p with

$$AIC(p) = \min_{l \in \{0, \cdots, P\}} AIC(l),$$

where P is a given integer large enough.

(ii). BIC criterion:

$$BIC(l) = \ln(\widehat{\sigma}_l^2) + \frac{l \ln(T)}{T}.$$

Find the AR order p with

$$BIC(p) = \min_{l \in \{0, \cdots, P\}} BIC(l).$$

Needs a constant term? Check the sample mean.

Many software packages available, e.g. R, SCA, Splus, Eviews, SAS, SPSS, etc.