Times Series and Forecasting (VI)

Chapter 6. Model Specification

Jianhua Hu

hu.jianhua@mail.shufe.edu.cn

Spring 2013, SHUFE

6.1. Introduction

- In general, an ARIMA(p, d, q) process can be written as

$$\phi(B)(1-B)^d Y_t = \theta(B) e_t,$$

where the AR and MA characteristic operators are

$$\phi(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$$
$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)$$

and

$$(1-B)^d Y_t = \nabla^d Y_t$$

is the series of dth differences.

Tasks for this chapter

- In this chapter, we discuss how to choose sensible values of p, d, and q for an observed (or suitably transformed) time series.
- We want our choices to be consistent with the underlying structure of the observed data.
- Bad choices of p, d, and q lead to bad models, which, in turn, lead to bad predictions (forecasts) of future values.

6.2. Sample autocorrelation function (ACF)

- For $Y_1, Y_2, \cdots, Y_n$, the sample ACF at lag k, is given by

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2},$$

  where $\overline{Y}$ is the sample mean of $Y_1, Y_2, \cdots, Y_n$.

- The sample ACF r_k is an estimator (estimate) of the true (population) ACF ρ_k.

- The statistic r_k has a sampling distribution which we would like to investigate so that we can quantify the uncertainty in values of r_k we see in practice.

Large-sample property of sample ACF

Theorem

Fix $m > 1$. For $j = 1, 2, \cdots, m$ and any stationary ARMA(p, q) process, $\sqrt{n}(r_j - \rho_j)$ converges to the normal distribution $N(0, c_{ij})$, that is,

$$\sqrt{n}(r_j - \rho_j) \xrightarrow{d} N(0, c_{jj}) \text{ as } n \longrightarrow \infty,$$

where

$$c_{jj} = \sum_{s=-\infty}^{\infty} \left(\rho_{s+j}^2 + \rho_{s-j}\rho_{s+j} - 4\rho_j\rho_s\rho_{s+j} + 2\rho_j^2\rho_s^2\right).$$

More details on large-sample results for sample ACF

- In practical terms, when n is large, the sample ACF r_j is approximately normally distributed by $N(\rho_j, c_{jj}/n)$; i.e.,

$$r_j \sim AN\left(\rho_j, \frac{c_{jj}}{n}\right).$$

- It can also be shown that when n is large,

$$\text{corr}(r_j, r_k) \approx \frac{c_{jk}}{\sqrt{c_{jj}c_{kk}}},$$

where c_{jk} is given by

$$\sum_{s=-\infty}^{\infty} \left(\rho_{s+j}\rho_{s+k} + \rho_{s-j}\rho_{s+k} - 2\rho_j\rho_s\rho_{s+k} - 2\rho_k\rho_s\rho_{s+j} + 2\rho_j\rho_k\rho_s^2\right).$$

Large-sample result of a white noise process

- For a white noise process, we have stationary ARMA(0,0) process. For any $k > 0$, $c_{kk} = 1$. For large n, we have

$$r_k \sim AN\left(0, \frac{1}{n}\right),$$

  a result discussed in Chapter 3.

- This explains why $\pm 2/\sqrt{n}$ serve as approximate error bounds for the sample ACF r_k.

- Values of r_k outside these bounds would be "unusual" under the white noise model.

- In addition, for $j \neq k$, corr$(r_j, r_k) \approx 0$.

Figure 6.1. Monte Carlo simulation from N(0,1) WN process

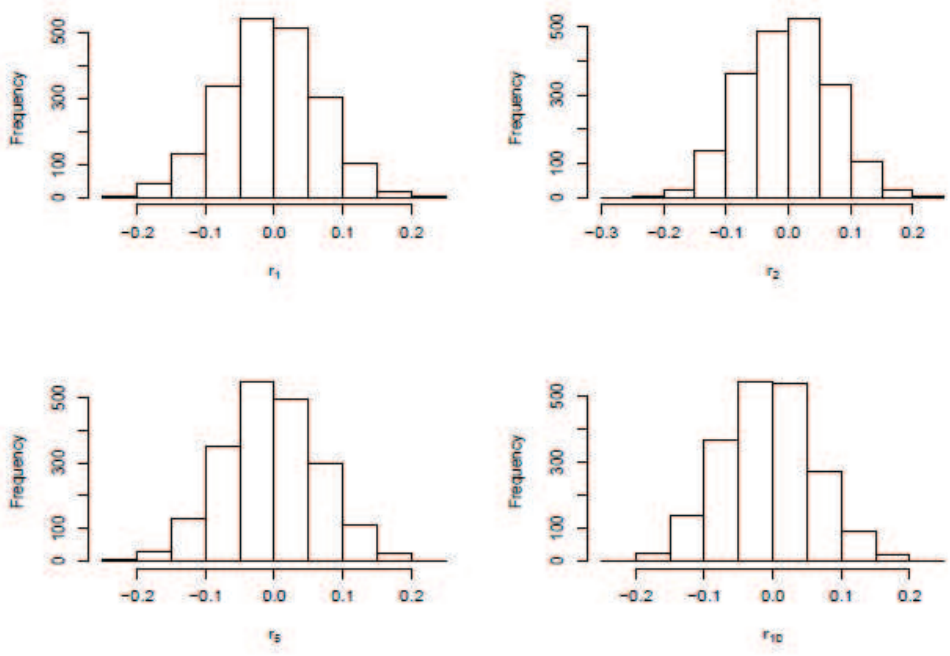

Figure 6.1: *Monte Carlo simulation. Histograms of sample autocorrelations based on* $B = 2000$ *Monte Carlo samples of size* $n = 200$ *taken from a* $\mathcal{N}(0, 1)$ *white noise process. Upper left:* r_1. *Upper right:* r_2. *Lower left:* r_5. *Lower right:* r_{10}. *The histograms are empirical representations of the true sampling distributions when* $n = 200$.

Large-sample result of AP(1)

- For a stationary AR(1) process,

$$Y_t = \phi Y_{t-1} + e_t,$$

the asymptotic distribution formulas again reduce considerably.

- We have, for large n,

$$r_k \sim AN\left(\rho_k, \sigma^2_{r_k}\right) = AN\left(\phi^k, \sigma^2_{r_k}\right),$$

and

$$\sigma^2_{r_k} = \frac{1}{n}\left[\frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k}\right].$$

Additional observations for the AR(1) process

- When ϕ is close to ± 1 (i.e., the "stationarity borderline"), the variance $\sigma^2_{r_k} = \text{var}(r_k)$ will be large since $1 - \phi^2$ is close to 0. It is hard to precisely estimate r_k when ϕ is close to ± 1.
- $\sigma^2_{r_k} = \text{var}(r_k)$ does tend to 0 as n increases to ∞.
- In general, r_j and r_k are asymptotically correlated; see pp 111 (CC). For example,

$$\text{corr}(r_1, r_2) \approx 2\phi \sqrt{\frac{1-\phi^2}{1+2\phi^2-3\phi^4}}.$$

Example 6.1. ACF of an AR(1) process

Example

Consider an AR(1) series with $n = 200$ and $\phi = 0.7$.

(a) Would you be surprised if $r_1 = 0.65$?

(b) Would $r_{10} = -0.15$ be unusual?

Solution of Example 6.1 (a)

- Due to $\rho_1 = \phi = 0.7$, theoretical result tells us that

$$r_1 \sim AN(0.7, \sigma^2_{r_1}),$$

where

$$\sigma^2_{r_1} = \frac{1}{200}\left[\frac{\left[1+(0.7)^2\right]\left[1-(0.7)^2\right]}{\left[1-(0.7)^2\right]} - 2(0.7)^2\right] \approx 0.0026.$$

- Approximately 95 percent CI of the estimates r_1 is

$$\left(0.7 - 2\sqrt{0.0026}, 0.7 + 2\sqrt{0.0026}\right) \Longrightarrow (0.598, 0.802).$$

Therefore, $r_1 = 0.65$ is not all that unusual. See Figure 6.2.

Solution of Example 6.1 (b)

- Upon first glance, $r_{10} = -0.15$ may look unusual, because $\rho_{10} = \phi^{10} = (0.7)^{10} \approx 0.0282 > 0$. However, perhaps r_{10} is negative due to natural sampling variability.
- Theorem tells us that $r_{10} \sim AN(0.0282, \sigma^2_{r_{10}})$, where

$$\sigma^2_{r_{10}} = \frac{1}{200}\left[\frac{\left[1+(0.7)^2\right]\left[1-(0.7)^{20}\right]}{\left[1-(0.7)^2\right]} - 2(0.7)^{20}\right] \approx 0.0145.$$

- 95 percent CI of r_{10} is

$$\left(0.282 - 2\sqrt{0.0145}, 0.282 + 2\sqrt{0.0145}\right) \Longrightarrow (-0.213, 0.269).$$

- So, $r_{10} = -0.15$ is not unusual either. Here, $\sigma^2_{r_{10}} \approx 6\sigma^2_{r_1}$. It is therefore harder to get a precise estimate of ρ_{10} than it is for ρ_1. See also Figure 6.2.

Figure 6.2. Monte Carlo simulation for an AR(1) process

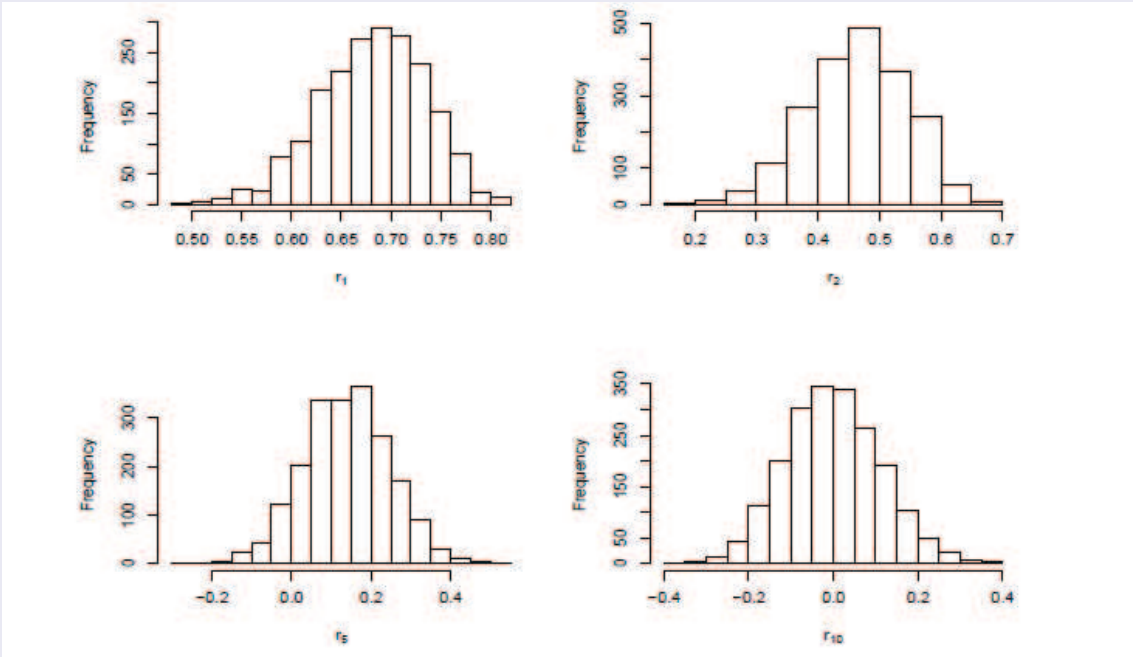

Figure 6.2: *Monte Carlo simulation. Histograms of sample autocorrelations based on* $B = 2000$ *Monte Carlo samples of size* $n = 200$ *taken from an* $AR(1)$ *process with* $\phi = 0.7$. *Upper left:* r_1. *Upper right:* r_2. *Lower left:* r_5. *Lower right:* r_{10}. *The histograms are empirical representations of the true sampling distributions when* $n = 200$.

Large-sample result of MA(1)

- Consider an invertible MA(1) process $Y_t = e_t - \theta e_{t-1}$.
- When $k = 1$, for large n,

$$r_1 \sim AN(\rho_1, \sigma^2_{r_1}),$$

where $\rho_1 = -\theta/(1+\theta^2)$ and

$$\sigma^2_{r_1} = \frac{1}{n}\left(1 - 3\rho_1^2 + 4\rho_1^4\right) = \frac{1}{n}\left[\frac{1+\theta^2+4\theta^4+\theta^6+\theta^8}{(1+\theta^2)^4}\right].$$

- For all $k > 1$, because $\rho_k = 0$, when n is large,

$$r_k \sim AN(0, \sigma^2_{r_k}),$$

where

$$\sigma^2_{r_k} = \frac{1}{n}(1+2\rho_1^2) = \frac{1}{n}\left[\frac{1+4\theta^2+\theta^4}{(1+\theta^2)^2}\right].$$

- Sample autocorrelations r_j and r_k are also correlated. For example, $\text{corr}(r_1, r_2) \approx \frac{2\rho_1(1-\rho_1^2)}{\sqrt{(1-3\rho_1^2+4\rho_1^4)(1+2\rho_1^2)}}$.

Figure 6.3. Monte Carlo simulation of a MA(1)

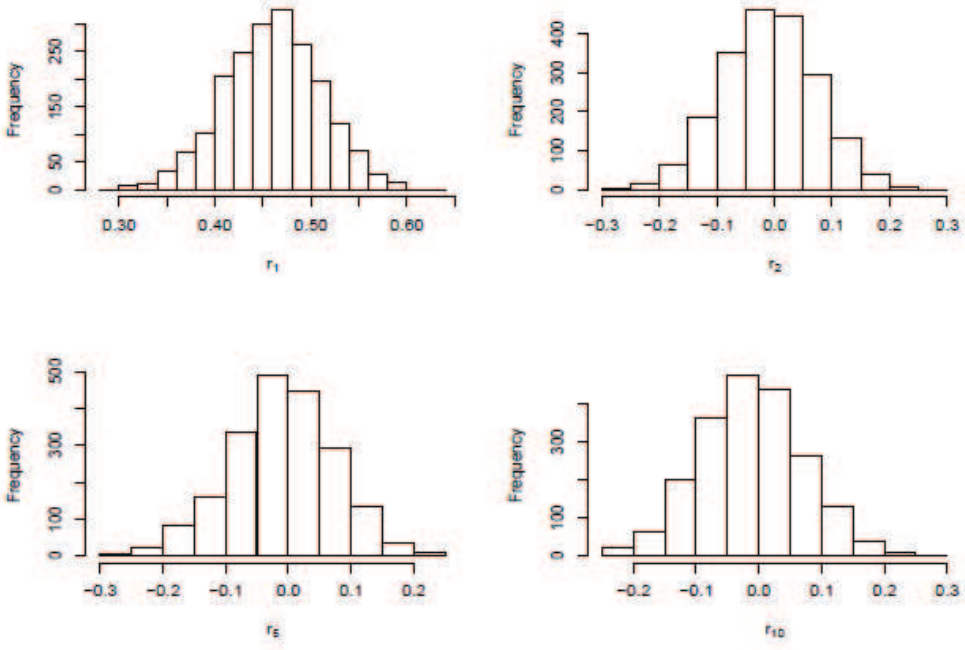

Figure 6.3: *Monte Carlo simulation. Histograms of sample autocorrelations based on* $B = 2000$ *Monte Carlo samples of size* $n = 200$ *taken from an MA*(1) *process with* $\theta = -0.7$. *Upper left:* r_1. *Upper right:* r_2. *Lower left:* r_5. *Lower right:* r_{10}. *The histograms are empirical representations of the true sampling distributions when* $n = 200$.

Large-sample result of MA(q)

Theorem

For an invertible MA(q) process,

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q},$$

the sample autocorrelation r_k, for all $k > q$, satisfies

$$r_k \sim AN\left[0, \frac{1}{n}\left(1 + 2\sum_{j=1}^{q} \rho_j^2\right)\right],$$

when n is large.

Large-sample level α test for MA(q)

- A natural (large-sample) level α test for

$$H_0: \text{ MA}(q) \text{ process is appropriate}$$

versus

$$H_1: \text{ MA}(q) \text{ process is not appropriate}$$

can be carried out by computing

$$Z = \frac{r_{q+1}}{\sqrt{\frac{1}{n}\left(1 + 2\sum_{j=1}^{q} r_j^2\right)}} \sim AN(0,1).$$

- Rejecting H_0 and being in favor of H_1 if $|Z| > z_{\alpha/2}$.
- Note that r_j has been used as an estimate for ρ_j in the standard error of the test statistic above.

Example 6.2. Which MA model is appropriate?

Example

- From a time series of $n = 200$ observations, we calculate $r_1 = -0.49$, $r_2 = 0.31$, $r_3 = -0.13$, $r_4 = 0.07$, and $|r_k| < 0.09$ for $k > 4$.
- Which moving average (MA) model is most consistent with these sample autocorrelations?

Solution of example 6.2

- Test

$$H_0: \text{ MA(1) process is appropriate}$$

versus

$$H_1: \text{ MA(1) process is not appropriate.}$$

- Compute

$$z = \frac{r_2}{\sqrt{\frac{1}{n}(1+2r_1^2)}} = \frac{0.31}{\sqrt{\frac{1}{200}[1+2(-0.49)^2]}} \approx 3.60.$$

- Rejecting H_0 because the p-value is $Pr(|Z| > 3.60) \approx 0.0003$. We conclude that the MA(1) model is not appropriate.

Solution of example 6.2

- Test

$$H_0: \text{ MA(2) process is appropriate}$$

versus

$$H_1: \text{ MA(2) process is not appropriate}$$

- Compute

$$z = \frac{r_3}{\sqrt{\frac{1}{n}(1 + 2r_1^2 + 2r_2^2)}} = \frac{-0.31}{\sqrt{\frac{1}{200}[1 + 2(0.49)^2 + 2(0.31)^2]}} \approx -1.42$$

- We can not reject H_0 because the p-value is $Pr(|Z| > 1.42) \approx 0.16$.
- Therefore, an MA(2) model is not inconsistent with these sample autocorrelations.

Example 6.3. Use SAS to simulate two MA(1) models

Example

Use SAS or R to generate data from two moving average processes:

- $Y_t = e_t - 0.5e_{t-1}$; *MA(1), with* $\theta = 0.5$
- $Y_t = e_t - 0.5e_{t-1} + 0.5e_{t-2}$; *MA(2), with* $\theta_1 = 0.5$ *and* $\theta_2 = -0.5$,

where $e_t \sim \text{iid} N(0, 1)$ and $n = 100$.

- The realized time series and the corresponding sample ACFs are displayed in In Figure 6.4.

Solution of Example 6.3

- WN margin of error bounds are $\pm 2/\sqrt{n} = \pm 2/\sqrt{100}$.
- The more precise error bounds are from

$$r_k \sim AN\left[0, \frac{1}{n}\left(1 + 2\sum_{j=1}^{q}\rho_j^2\right)\right].$$

- For each lag k, the (estimated) standard error bounds are

$$\pm 1.96\sqrt{\frac{1}{100}\left(1 + 2\sum_{j=1}^{k-1}r_j^2\right)}.$$

  That is, error bounds at lag k are computed assuming that the MA$(k-1)$ model is appropriate.
- Values of r_k exceeding these bounds are deemed to be statistically significant.

Figure 6.4. Simulated MA(1) and MA(2)

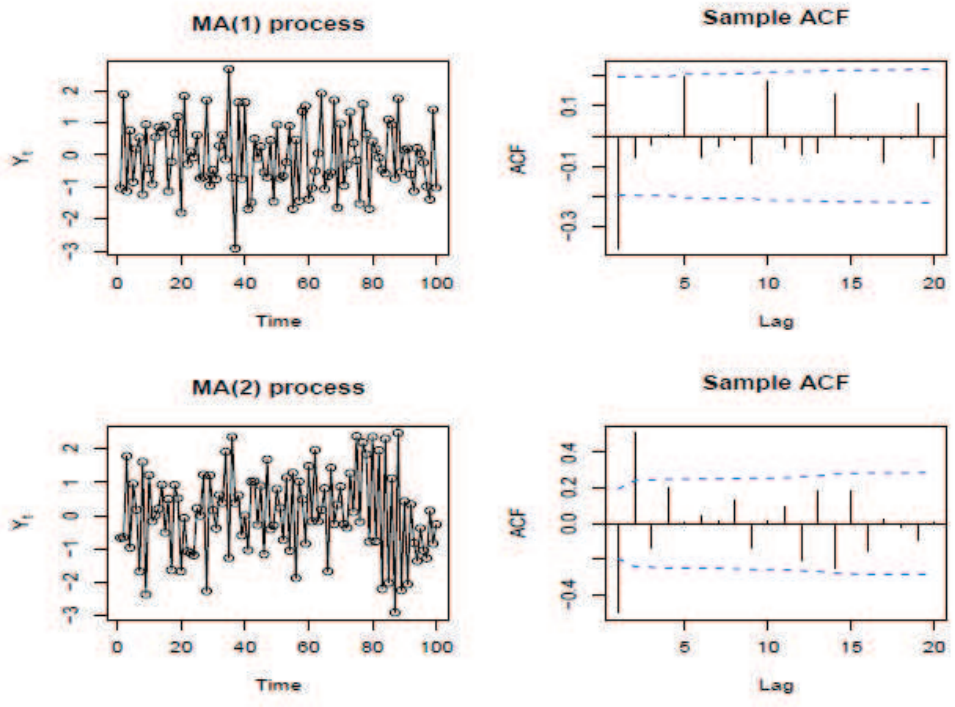

Figure 6.4: *Simulated MA(1) and MA(2) processes with* $n = 100$ *and* $\sigma_e^2 = 1$. *Moving average error bounds are used in the corresponding sample ACFs; not the white noise error bounds* $\pm 2/\sqrt{n}$.

6.3. Why consider the partial PACF?

- For MA(q) models, $ACF = \begin{cases} \text{nonzero}, & \text{if } k \leq q \\ 0, & \text{if } k > q. \end{cases}$.
  - Therefore, the ACF provides a considerable amount of information about the order of the dependence when the process is truly a moving average.
- For AR process, the ACF alone tells us little about the order of the dependence.
- It is therefore worthwhile to pursue a function for AR models similar to the ACF for MA models. This function is called the partial autocorrelation function (PACF).

Motivation for PACF

- To set our ideas, consider a stationary (zero mean) AR(1)

$$Y_t = \phi Y_{t-1} + e_t.$$

- The autocovariance between Y_t and Y_{t-2} is given by

$$\gamma_2 = \text{Cov}(Y_t, Y_{t-2}) = \text{Cov}(\phi^2 Y_{t-2} + \phi e_{t-1} + e_t, Y_{t-2}) = \phi^2 \gamma_0.$$

  - If Y_t is an MA(1) process, then $\gamma_2 = \text{Cov}(Y_t, Y_{t-2}) = 0$.
  - This is not true for an AR(1) process because Y_t depends on Y_{t-2} through Y_{t-1}.

- Considering the quantities $Y_t - \phi Y_{t-1}$ and $Y_{t-2} - \phi Y_{t-1}$ yields

$$\text{Cov}(Y_t - \phi Y_{t-1}, Y_{t-2} - \phi Y_{t-1}) = \text{Cov}(e_t, Y_{t-2} - \phi Y_{t-1}) = 0.$$

Motivation for PACF (continued)

- In the AR(1) case, if ϕ is known, then the quantity

$$Y_t - \phi Y_{t-1}$$

  can be thought of the prediction error from regressing Y_t on Y_{t-1} (with no intercept).
- Similarly, the quantity $Y_{t-2} - \phi Y_{t-1}$ can be thought of the prediction error from regressing Y_{t-2} on Y_{t-1}, again with no intercept.
- Both of these prediction errors are uncorrelated with the intervening variable Y_{t-1}. To see why, note that

$$\text{Cov}(Y_t - \phi Y_{t-1}, Y_{t-1}) = \gamma_1 - \phi\gamma_0 = 0,$$

  because $\gamma_1 = \phi\gamma_0$ in the AR(1) model. Similarly,

$$\text{Cov}(Y_{t-2} - \phi Y_{t-1}, Y_{t-1}) = \gamma_1 - \phi\gamma_0 = 0.$$

Motivation for PACF (continued)

- Consider a stationary, zero mean AR(2) process

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t.$$

- Similarly, suppose that we "break" the dependence between Y_t and Y_{t-3} in the AR(2) process by removing the effects of both Y_{t-1} and Y_{t-2}.
- Considering the quantities $Y_t - \phi_1 Y_{t_1} - \phi_2 Y_{t-2}$ and $Y_{t-3} - \phi_1 Y_{t_1} - \phi_2 Y_{t-2}$ yields

$$\begin{aligned} &\mathrm{Cov}(Y_t - \phi_1 Y_{t_1} - \phi_2 Y_{t-2}, Y_{t-3} - \phi_1 Y_{t_1} - \phi_2 Y_{t-2}) \\ &= \mathrm{Cov}(e_t, Y_{t-3} - \phi_1 Y_{t_1} - \phi_2 Y_{t-2}) = 0. \end{aligned}$$

Motivation for PACF (continued)

- In the AR(2) case, if ϕ_1 and ϕ_2 are known, then the quantity

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}$$

  can be thought of the prediction error from regressing Y_t on Y_{t-1} and Y_{t-2} (with no intercept).
- Similarly, the quantity

$$Y_{t-3} - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}$$

  can be thought of the prediction error from regressing Y_{t-3} on Y_{t-1} and Y_{t-2}, again with no intercept.
- Both of these prediction errors are uncorrelated with the intervening variables Y_{t-1} and Y_{t-2}.

The partial autocorrelation function

- For a zero mean time series, let $\hat{Y}_t^{(k-1)}$ denote the population regression of Y_t on the variables $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-(k-1)}$, that is,

$$\hat{Y}_t^{(k-1)} = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_{k-1} Y_{t-(k-1)}.$$

- Let $\hat{Y}_{t-k}^{(k-1)}$ denote the population regression of Y_{t-k} on the variables $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-(k-1)}$, that is,

$$\hat{Y}_{t-k}^{(k-1)} = \beta_1 Y_{t-(k-1)} + \beta_2 Y_{t-(k-2)} + \cdots + \beta_{k-1} Y_{t-1}.$$

- $\beta_1, \cdots, \beta_{k-1}$ are determined by minimizing the mean squares error of prediction.

The partial autocorrelation function

- The partial autocorrelation function (PACF) of a stationary process $\{Y_t\}$, denoted by ϕ_{kk}, for $k = 1, 2, \cdots$, satisfies $\phi_{11} = \rho_1$ (by convention) and

$$\phi_{kk} = \text{corr}\left(Y_t - \hat{Y}_t^{(k-1)}, Y_{t-k} - \hat{Y}_{t-k}^{(k-1)}\right),$$

for $k = 2, 3, \cdots$,.

Remarks for the partial autocorrelation function

- The quantities $\hat{Y}_t^{(k-1)}$ and $\hat{Y}_{t-k}^{(k-1)}$ both are linear functions of the intervening variables $Y_{t-1}, \cdots, Y_{t-(k-1)}$.
- The quantities $Y_t - \hat{Y}_t^{(k-1)}$ and $Y_{t-k} - \hat{Y}_{t-k}^{(k-1)}$ are called the prediction errors.
- The PACF at lag k is defined to be the correlation between these errors.
- If the underlying process Y_t is normal, then

$$\phi_{kk} = \text{corr}\left(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \cdots, Y_{t-(k-1)}\right),$$

  the correlation between Y_t and Y_{t-k}, conditional on the intervening variables $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-(k-1)}$.
- In essence, by conditioning, ϕ_{kk} measures the correlation between Y_t and Y_{t-k} after removing the linear effects of Y_{t-1}, Y_{t-2}, $\cdots$, $Y_{t-(k-1)}$.

PCAF of AR(1)

- We now revisit our AR(1) calculations. Consider the model

$$Y_t = \phi Y_{t-1} + e_t.$$

- In AR(1), $Y_t - \phi Y_{t-1}$ and $Y_{t-2} - \phi Y_{t-1}$ are the prediction errors from regressing Y_t on Y_{t-1} and Y_{t-2} on Y_{t-1}, resp.
- With $k = 2$, the general expressions

$$\hat{Y}_t^{(k-1)} = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_{k-1} Y_{t-(k-1)}$$
$$\hat{Y}_{t-k}^{(k-1)} = \beta_1 Y_{t-(k-1)} + \beta_2 Y_{t-(k-2)} + \cdots + \beta_{k-1} Y_{t-1}$$

become

$$\hat{Y}_t^{(2-1)} = \phi Y_{t-1}$$
$$\hat{Y}_{t-2}^{(2-1)} = \phi Y_{t-1}.$$

PCAF of AR(1) (continued)

- Therefore, we have shown that for the AR(1) model,

$$\phi_{22} = \text{corr}\left(Y_t - \hat{Y}_t^{(2-1)}, Y_{t-2} - \hat{Y}_{t-2}^{(2-1)}\right) = 0.$$

because

$$\begin{aligned}&\text{Cov}\left(Y_t - \hat{Y}_t^{(2-1)}, Y_{t-2} - \hat{Y}_{t-2}^{(2-1)}\right)\\ =&\text{Cov}\left(Y_t - \phi Y_{t-1}, Y_{t-2} - \phi Y_{t-1}\right) = 0.\end{aligned}$$

- For the AR(1) model, it can be shown mathematically that $\phi_{11} \neq 0$ (it is taken to be ρ_1 by convention) and that

$$\phi_{22} = \phi_{33} = \phi_{44} = \cdots = 0.$$

That is, $\phi_{kk} = 0$, for all $k > 1$.

PCAF of AR(2)

- Revisit our AR(2) calculations. Consider the model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t.$$

- $Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}$ and $Y_{t-3} - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}$ are the prediction errors from regressing Y_t on Y_{t-1} and Y_{t-2} and Y_{t-3} on Y_{t-1} and Y_{t-2}, respectively.
- With $k = 3$, the general expressions

$$\hat{Y}_t^{(k-1)} = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_{k-1} Y_{t-(k-1)}$$
$$\hat{Y}_{t-k}^{(k-1)} = \beta_1 Y_{t-(k-1)} + \beta_2 Y_{t-(k-2)} + \cdots + \beta_{k-1} Y_{t-1}$$

  become

$$\hat{Y}_t^{(3-1)} = \phi_1 Y_{t-1} + \phi_2 Y_{t-2}$$
$$\hat{Y}_{t-3}^{(3-1)} = \phi_1 Y_{t-1} + \phi_2 Y_{t-2}.$$

- For the AR(2) model,

$$\phi_{33} = \text{corr}\left(Y_t - \hat{Y}_t^{(3-1)}, Y_{t-3} - \hat{Y}_{t-3}^{(3-1)}\right) = 0$$

PCAF of AR(2) (continued)

- For the AR(2) model, it can be shown mathematically that $\phi_{11} \neq 0$, $\phi_{22} \neq 0$, and that

$$\phi_{33} = \phi_{44} = \phi_{55} = \cdots = 0.$$

That is, $\phi_{kk} = 0$, for all $k > 2$.

PCAF of AR(p)

- In general, for the AR(p) model, we have the following results:
  - $\phi_{11} \neq 0, \phi_{22} \neq 0, \cdots, \phi_{pp} \neq 0$
  - $\phi_{kk} = 0$, for all $k > p$.
- In other words, for the AR(p) model, the PACF "drops off" to zero after the pth lag.
- Therefore, the PACF can help to determine the order of an AR(p) process just like the ACF helps to determine the order of an MA(q) process!

Example 6.4. Use SAS to simulate two AR processes

Example

Use SAS or R to generate observations from two autoregressive processes:

(i) $Y_t = -0.7Y_{t-1} + e_t$; AR(1), with $\phi = -0.7$

(ii) $Y_t = 0.3Y_{t-1} + 0.6Y_{t-2} + e_t$; AR(2), with $\phi_1 = 0.3$ and $\phi_2 = 0.6$,

where $e_t \sim \text{iid} N(0, 1)$ and $n = 100$.

Figure 6.5 displays the simulated time series from each AR model, the sample ACF, and the sample PACF.

Figure 6.5. Sample ACF and PACF for Example 6.4

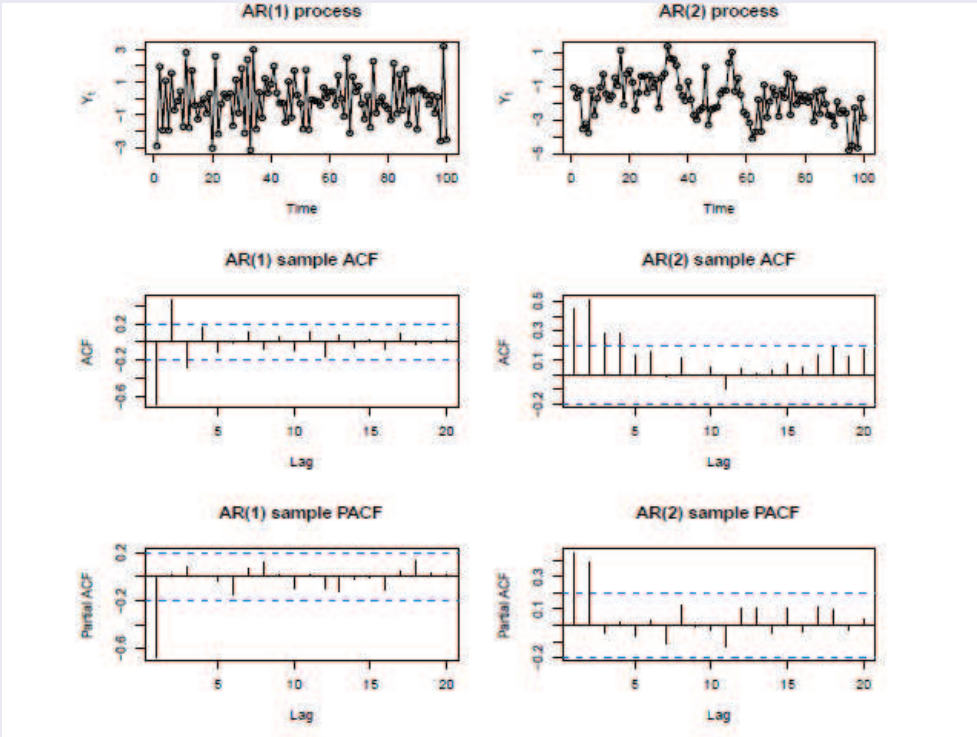

Figure 6.5: *Left: AR(1) simulation with $n = 100$ and $e_t \sim iid\,\mathcal{N}(0,1)$, sample ACF (middle), and sample PACF (bottom). Right: AR(2) simulation with $n = 100$ and $e_t \sim iid\,\mathcal{N}(0,1)$, sample ACF (middle), and sample PACF (bottom).*

Remarks for sample ACF and PACF of Example 6.4

- The sample PACF, which we will discuss shortly, is an estimate of the true (population) PACF.
- Note that the sample PACF for the AR(1) simulation declares ϕ_{kk} insignificant for $k > 1$. That is, the estimates of ϕ_{kk}, for $k > 1$, are all within the margin of error bounds. Similarly, the sample PACF for the AR(2) simulation declares ϕ_{kk} insignificant for $k > 2$.
- We will soon discuss the theory of why the PACF error bounds here are correct.

How does sample PACF behave MA(q) process?

- How does the PACF behave for moving average processes?
- To answer this, consider the MA(1) model,

$$Y_t = e_t - \theta e_{t-1}.$$

- It can be shown that, for $k \geq 1$,

$$\phi_{kk} = \frac{\theta^k(\theta^2 - 1)}{1 - \theta^{2(k+1)}}.$$

- Because $|\theta| < 1$ (our invertibility condition), note that the PACF for the MA(1) process decays to zero as the lag k increases, much like the ACF decays to zero for the AR(1). The same thing happens in higher order MA models.
- In general, the PACF for an MA process behaves like the ACF for an AR process of the same order.

Comparison between ACF and PACF for AR(p) and MA(q)

- The following table succinctly summarizes the behavior of the ACF and PACF for MA and AR processes.

	AR(p)	MA(q)
ACF	Tails off	Cuts off after lag q
PACF	Cuts off after lag p	Tails off

- Therefore, the ACF is the key tool to help determine the order of a MA process. The PACF is the key tool to help determine the order of an AR process.
- For ARMA processes, we need a different tool (coming up).

Example 6.5. Use SAS to simulate two MA processes

Example

Use SAS or R to generate observations from two moving average processes:

- $Y_t = e_t - 0.5e_{t-1}$; MA(1), with $\theta = 0.5$
- $Y_t = e_t - 0.5e_{t-1} + 0.5e_{t-2}$; MA(2), with $\theta_1 = 0.5$ and $\theta_2 = -0.5$,

where $e_t \sim \text{iid} N(0, 1)$ and $n = 100$.

- Figure 6.6 displays the simulated time series from each MA model, the sample ACF, and the sample PACF.

Figure 6.6. ACF and PACF for MA(1) and MA(2)

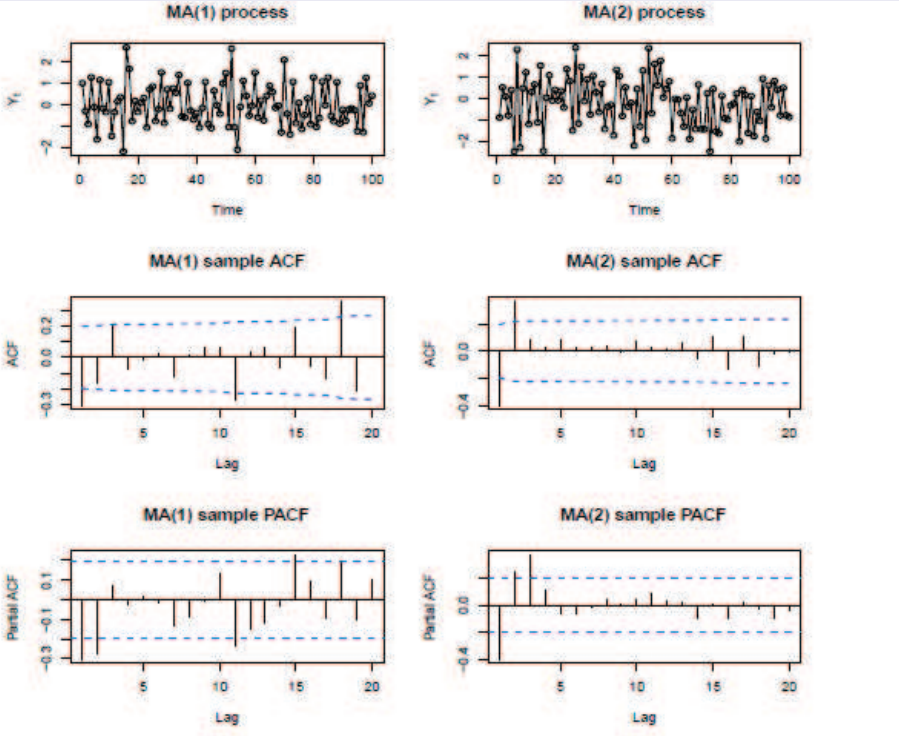

Figure 6.6: *Left: MA(1) simulation with $n = 100$ and $e_t \sim iid\,\mathcal{N}(0,1)$, sample ACF (middle), and sample PACF (bottom). Right: MA(2) simulation with $n = 100$ and $e_t \sim iid\,\mathcal{N}(0,1)$, sample ACF (middle), and sample PACF (bottom).*

Remarks

- MA margin of error bounds are used in the sample ACFs.
- The sample ACFs for the MA(1) and MA(2) simulations agree with what we would expect from the theory; that is, all $\rho_k = 0$, for $k > q$. The "significant" autocorrelations in the MA(1) sample ACF at high lags are probably false alarms.
- The sample PACFs appear to display patterns that we would expect to see in the ACFs for AR processes. The PACFs are not helpful in determining the order of an MA process.

Computation for ARMA

- For any stationary ARMA process, it is possible to compute the theoretical values ϕ_{kk}, for $k = 1, 2, \cdots$,.
- For a fixed k, we have the following Yule-Walker equations:

$$\begin{aligned}
\rho_1 &= \phi_{k,1} + \rho_1\phi_{k,2} + \rho_2\phi_{k,3} + \cdots + \rho_{k-1}\phi_{k,k} \\
\rho_2 &= \rho_1\phi_{k,1} + \phi_{k,2} + \rho_1\phi_{k,3} + \cdots + \rho_{k-2}\phi_{k,k} \\
&\vdots \\
\rho_k &= \rho_{k-1}\phi_{k,1} + \rho_{k-2}\phi_{k,2} + \rho_{k-3}\phi_{k,3} + \cdots + \phi_{k,k},
\end{aligned}$$

  where

$$\begin{aligned}
\rho_j &= \text{corr}(Y_t, Y_{t-j}) \\
\phi_{k,j} &= \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}, \qquad j = 1, 2, \cdots, k-1 \\
\phi_{kk} &= \text{corr}(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \cdots, Y_{t-(k-1)}).
\end{aligned}$$

- For known $\rho_1, \rho_2, \cdots, \rho_k$, we can solve this system for $\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,k-1}$, and keep the value of ϕ_{kk}.

Estimation for the partial autocorrelation ϕ_{kk}

- The partial autocorrelation ϕ_{kk} can be estimated by taking the Yule-Walker equations and substituting r_k in for the true autocorrelations ρ_k, that is,

$$
\begin{aligned}
r_1 &= \phi_{k,1} + r_1\phi_{k,2} + r_2\phi_{k,3} + \cdots + r_{k-1}\phi_{k,k} \\
r_2 &= r_1\phi_{k,1} + \phi_{k,2} + r_1\phi_{k,3} + \cdots + r_{k-2}\phi_{k,k} \\
&\vdots \\
r_k &= r_{k-1}\phi_{k,1} + r_{k-2}\phi_{k,2} + r_{k-3}\phi_{k,3} + \cdots + \phi_{k,k}.
\end{aligned}
$$

  This system can then be solved for $\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,k-1}, \phi_{k,k}$ as before.

- The solutions are estimates $\widehat{\phi}_{k,1}, \widehat{\phi}_{k,2}, \cdots, \widehat{\phi}_{k,k-1}, \widehat{\phi}_{k,k}$. This can be done for each $k = 1, 2, \cdots,$.

Test for the partial autocorrelation ϕ_{kk} of AR(p)

- If the AR(p) model is correct, then for large n and all $k > p$,

$$\widehat{\phi}_{k,k} \sim AN\left(0, \frac{1}{n}\right).$$

- We can use $\pm z_{\alpha/2}/\sqrt{n}$ as "critical points" to test, at level α,

$$\begin{aligned} H_0: &\quad \text{AR(p) model is appropriate} \\ &\quad \text{versus} \\ H_1: &\quad \text{AR(p) model is not appropriate.} \end{aligned}$$

- Similarly to how we tested whether or not a specific MA model was appropriate using the sample autocorrelations.

Summary: Identification of ARIMA Processes

- White noise process
  - $ARIMA(0,0,0)$: ACF: no significant spikes; PACF: no significant spikes.
- Integrated process
  - $ARIMA(0,1,0)$: ACF: slow attenuation; PACF: 1 spike at order of differencing.

Summary: Identification of some ARIMA Processes

- Autoregressive processes
  - $ARIMA(1,0,0)$ for $\phi > 0$: ACF: exponential decay, positive spikes; PACF: 1 positive spike at lag 1.
  - $ARIMA(1,0,0)$ for $\phi < 0$: ACF: oscillating decay, begins with negative spike; PACF: 1 negative spike at lag 1.
  - $ARIMA(2,0,0)$ for $\phi_1, \phi_2 > 0$: ACF: exponential decay, positive spikes; PACF: 2 positive spikes at lags 1 and 2.
  - $ARIMA(2,0,0)$ for $\phi_1 < 0, \phi_2 > 0$: ACF: oscillating exponential decay; PACF: 1 negative spike at lag 1, 1 positive spike at lag 2.

Summary: Identification of ARIMA Processes

- Moving average processes
  - $ARIMA(0,0,1)$ for $\theta > 0$: ACF: 1 negative spike at lag 1; PACF: exponential decay of negative spikes.
  - $ARIMA(0,0,1)$ for $\theta < 0$: ACF: 1 positive spike at lag 1; PACF: oscillating decay of positive and negative spikes.
  - $ARIMA(0,0,2)$ for $\theta_1, \theta_2 > 0$: ACF: 2 negative spikes at lags 1 and 2; PACF: exponential decay of negative spikes.
  - $ARIMA(0,0,2)$ for $\theta_1, \theta_2 < 0$: ACF: 2 positive spikes at lags 1 and 2; PACF: oscillating decay of positive and negative spikes.

Summary: Identification of ARIMA Processes

- ARMA processes
  - $ARIMA(1,0,1)$ for $\phi > 0$, $\theta > 0$: ACF: exponential decay of positive spikes; PACF: exponential decay of positive spikes.
  - $ARIMA(1,0,1)$ for $\phi > 0$, $\theta < 0$: ACF: exponential decay of positive spikes; PACF: oscillating decay of positive and negative spikes.
  - $ARIMA(1,0,1)$ for $\phi < 0$, $\theta > 0$: ACF: oscillating decay; PACF: exponential decay of negative spikes

    $ARIMA(1,0,1)$ for $\phi < 0$, $\theta < 0$: ACF: oscillating decay of negative and positive spikes; PACF: oscillating decay of negative and positive spikes.

6.4. The extended sample autocorrelation function (ESACF)

- The ACF $\longrightarrow$ determine the order of MA(q) because $\rho_k = 0$, for all lags $k > q$.
- The PACF $\longrightarrow$ determine the order of AR(p) because $\phi_{kk} = 0$, for all lags $k > p$.
- In sample ACF and sample PACF, check if values of r_k and $\widehat{\phi}_{kk}$, respectively, are consistent with this theory.
- Two formal testing procedures $\longrightarrow$ determine if a given MA(q) or AR(p) model is appropriate.

ESACF for ARMA model

- The problem, however, is that neither the sample ACF nor sample PACF $\longrightarrow$ determine an ARMA.
- A new function $\longrightarrow$ is needed to identify the orders of an ARMA(p, q) process.
- The new function is said to be the extended autocorrelation function (ESACF).

Motivation for the ESACF

- A stationary ARMA(p, q) process can be expressed as

$$\phi(B)Y_t = \theta(B)e_t,$$

where the AR and MA characteristic operators are

$$\phi(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$$
$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right).$$

- Our discussion starts with letting

$$W_t \equiv \phi(B)Y_t = Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p}$$

follows an MA(q) model, that is,

$$W_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) e_t.$$

- Of course, the $\{W_t\}$ process is not observed because W_t depends on $\phi_1, \phi_2, \cdots, \phi_p$, which are unknown parameters.

Motivation for the ESACF (continued)

Strategy:

- Regress Y_t on $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}$ and use ordinary least squares to fit the no-intercept model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t.$$

  where ϵ_t denotes a error term.
- This produces estimates $\widehat{\phi}_1, \widehat{\phi}_2, \cdots, \widehat{\phi}_p$ and we compute

$$\widehat{W}_t = Y_t - \widehat{\phi}_1 Y_{t-1} - \widehat{\phi}_2 Y_{t-2} - \cdots - \widehat{\phi}_p Y_{t-p}.$$

- These values (merely the residuals from the regression) serve as proxies for the true $\{W_t\}$ process, and we could now treat these residuals as our "data".

Motivation for the ESACF (continued)

- Construct the sample ACF for the $\widehat{W}_t$ data so that we can learn about the order q of the MA part of the process.
- For example, if we fit an AR(2) model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

and the residuals $\widehat{W}_t$ look to follow an MA(2) process, then this would suggest that a mixed ARMA(2,2) model is worthy of consideration.

Motivation for the ESACF (continued)

Problem:

- We have just laid out a sensible strategy on how to select candidate ARMA models; i.e., choosing values for p and q.
- The problem is that ordinary least squares regression estimates $\widehat{\phi}_1, \widehat{\phi}_2, \cdots, \widehat{\phi}_p$ are inconsistent estimates of $\phi_1, \phi_2, \cdots, \phi_p$ when the underlying process is ARMA(p, q).
- Inconsistency means that these estimates $\widehat{\phi}_1, \widehat{\phi}_2, \cdots, \widehat{\phi}_p$ estimate the wrong things (in a large-sample sense).
- Therefore, the strategy that we have just described could lead to incorrect identification of p and q.

Motivation for the ESACF (continued)

Adjustment:

- Describe an "algorithm" to repair the approach just outlined.
  0. Consider using OLS to fit the same no-intercept AR(p) model

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t.$$

  - If the true process is an ARMA(p, q), then the LS estimate from the regression, say,

$$\widehat{\phi}_1^{(0)}, \widehat{\phi}_2^{(0)}, \cdots, \widehat{\phi}_p^{(0)}$$

    will be inconsistent and the ls residuals

$$\widehat{\epsilon}_t^{(0)} = Y_t - \widehat{\phi}_1^{(0)} Y_{t-1} - \widehat{\phi}_2^{(0)} Y_{t-2} - \cdots - \widehat{\phi}_p^{(0)} Y_{t-p}$$

    will not be white noise.
  - In fact, if $q \geq 1$ (so that the true process is ARMA), then the residuals $\widehat{\epsilon}_t^{(0)}$ and lagged versions of them will contain information about the process $\{Y_t\}$.

Motivation for the ESACF (continued)

1. Because the residuals $\widehat{\epsilon}_t^{(0)}$ contain information about the value of q, we first fit the model

$$Y_t = \phi_1^{(1)} Y_{t-1} + \phi_2^{(1)} Y_{t-2} + \cdots + \phi_p^{(1)} Y_{t-p} + \beta_1^{(1)} \widehat{\epsilon}_{t-1}^{(0)} + \epsilon_t^{(1)},$$

Note that we have added the lag 1 residuals $\widehat{\epsilon}_{t-1}^{(0)}$ from the initial model fit as a predictor in the regression.

- If the order of the MA part of the ARMA process is truly $q = 1$, then the least squares estimates

$$\widehat{\phi}_1^{(1)}, \widehat{\phi}_2^{(1)}, \cdots, \widehat{\phi}_p^{(1)}$$

will be consistent, that is, they will estimate the true AR parameters in large samples.
- If $q > 1$, then the estimates will be inconsistent and the residual process $\{\widehat{\epsilon}_t^{(1)}\}$ will not be white noise.

Motivation for the ESACF (continued)

2. If $q > 1$, then the residuals from the most recent regression $\{\widehat{\phi}_t^{(1)}\}$ still contain information about the value of q, so we next fit the model

$$Y_t = \phi_1^{(2)} Y_{t-1} + \cdots + \phi_p^{(2)} Y_{t-p} + \beta_1^{(2)} \widehat{\epsilon}_{t-1}^{(1)} + \beta_2^{(2)} \widehat{\epsilon}_{t-2}^{(0)} + \epsilon_t^{(2)}.$$

Note that in this model, we have added the lag 2 residuals $\widehat{\epsilon}_{t-2}^{(0)}$ from the initial model fit as well as the lag 1 residuals $\widehat{\epsilon}_{t-1}^{(1)}$ from the most recent fit.

- If the order of the MA part of the ARMA process is truly $q = 2$, then the least squares estimates

$$\widehat{\phi}_1^{(2)}, \widehat{\phi}_2^{(2)}, \cdots, \widehat{\phi}_p^{(2)}$$

will be consistent, that is, they will estimate the true AR parameters in large samples.

- If $q > 2$, then the estimates will be inconsistent and the residual process $\{\widehat{\epsilon}_t^{(2)}\}$ will not be white noise.

Motivation for the ESACF (continued)

3. We continue this iterative process, at each step, adding the residuals from the most recent fit in the same fashion. For example, at the next step, we would fit

$$Y_t = \phi_1^{(3)} Y_{t-1} + \cdots + \phi_p^{(3)} Y_{t-p} + \beta_1^{(3)} \widehat{\epsilon}_{t-1}^{(2)} + \cdots + \beta_3^{(3)} \widehat{\epsilon}_{t-3}^{(0)} + \epsilon_t^{(3)}.$$

The next model fit would be

$$Y_t = \phi_1^{(4)} Y_{t-1} + \cdots + \phi_p^{(4)} Y_{t-p} + \beta_1^{(4)} \widehat{\epsilon}_{t-1}^{(3)} + \cdots + \beta_4^{(4)} \widehat{\epsilon}_{t-4}^{(0)} + \epsilon_t^{(4)},$$

and so on. We continue fitting higher order models until residuals (from the most recent fit) resemble a white noise process.

The extended ACF

- In practice, the true orders p and q of the ARMA(p, q) model are unknown and have to be estimated.
- Based on the strategy outlined, however, we can estimate p and q using a new type of function.
- For an AR(m) model fit, define the mth extended sample autocorrelation function (ESACF) $\widehat{\rho}_j^{(m)}$ as the sample ACF for the residual process

$$\widehat{W}_t^{(j)} = Y_t - \widehat{\phi}_1^{(j)} Y_{t-1} - \widehat{\phi}_2^{(j)} Y_{t-2} - \cdots - \widehat{\phi}_m^{(j)} Y_{t-m},$$

  for $m = 0, 1, 2, \cdots$, and $j = 0, 1, 2, \cdots$. Here, the subscript j refers to the iteration number in the aforementioned sequential fitting process (hence, j refers to the order the MA part).
- The value m refers to the AR part of the process. Usually the maximum values of m and j are taken to be 10 or so.

Tabular representation of the ESACF

- It is useful to arrange the estimates $\widehat{\rho}_j^{(m)}$ in a two-way table where one direction corresponds to the AR part and the other direction corresponds to the MA part.

	MA					
AR	0	1	2	3	4	$\cdots$
0	$\widehat{\rho}_1^{(0)}$	$\widehat{\rho}_2^{(0)}$	$\widehat{\rho}_3^{(0)}$	$\widehat{\rho}_4^{(0)}$	$\widehat{\rho}_5^{(0)}$	$\cdots$
1	$\widehat{\rho}_1^{(1)}$	$\widehat{\rho}_2^{(1)}$	$\widehat{\rho}_3^{(1)}$	$\widehat{\rho}_4^{(1)}$	$\widehat{\rho}_5^{(1)}$	$\cdots$
2	$\widehat{\rho}_1^{(2)}$	$\widehat{\rho}_2^{(2)}$	$\widehat{\rho}_3^{(2)}$	$\widehat{\rho}_4^{(2)}$	$\widehat{\rho}_5^{(2)}$	$\cdots$
3	$\widehat{\rho}_1^{(3)}$	$\widehat{\rho}_2^{(3)}$	$\widehat{\rho}_3^{(3)}$	$\widehat{\rho}_4^{(3)}$	$\widehat{\rho}_5^{(3)}$	$\cdots$
4	$\widehat{\rho}_1^{(4)}$	$\widehat{\rho}_2^{(4)}$	$\widehat{\rho}_3^{(4)}$	$\widehat{\rho}_4^{(4)}$	$\widehat{\rho}_5^{(4)}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$

An important result of the extended ACF

Theorem

As $n \longrightarrow \infty$,

$$\widehat{\rho}_j^{(m)} \longrightarrow 0, \quad \textit{for} \;\; 0 \leq m - p < j - q$$
$$\widehat{\rho}_j^{(m)} \longrightarrow c \neq 0, \quad \textit{otherwise.}$$

Tsay and Tiao (1984) proposed the technique and Choi(1992) provides useful descriptions of the algorithm.

ESACF table for ARMA(1,1)

- The true large-sample ESACF table for an ARMA(1,1) process, for example, looks like

	MA						
AR	0	1	2	3	4	5	$\cdots$
0	x	x	x	x	x	x	$\cdots$
1	x	0	0	0	0	0	$\cdots$
2	x	x	0	0	0	0	$\cdots$
3	x	x	x	0	0	0	$\cdots$
4	x	x	x	x	0	0	$\cdots$
5	x	x	x	x	x	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Interpretation of the ESACF table of ARMA(1,1)

- In this table, the "0" entries correspond to the zero limits of $\widehat{\rho}_j^{(m)}$ when $0 \leq m-1 < j-1$.
- The "x" entries correspond to limits of $\widehat{\rho}_j^{(m)}$ which are nonzero.
- Therefore, the geometric pattern formed by the zeros is a "wedge" with a tip at (1,1).
- This tip corresponds to the values of $p = 1$ and $q = 1$ in the ARMA model.

Interpretation of the ESACF table of ARMA(2,2)

- The true large-sample ESACF table for ARMA(2,2) looks like

	MA						
AR	0	1	2	3	4	5	$\cdots$
0	x	x	x	x	x	x	$\cdots$
1	x	x	x	x	x	x	$\cdots$
2	x	x	0	0	0	0	$\cdots$
3	x	x	x	0	0	0	$\cdots$
4	x	x	x	x	0	0	$\cdots$
5	x	x	x	x	x	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

- In this table, the tip of the wedge is at the point (2,2).
- This tip corresponds to the values of $p = 2$ and $q = 2$ in the ARMA model.

Interpretation of the ESACF table (continued)

- The true large-sample ESACF table for an ARMA(2,1) process looks like

	MA						
AR	0	1	2	3	4	5	$\cdots$
0	x	x	x	x	x	x	$\cdots$
1	x	x	x	x	x	x	$\cdots$
2	x	0	0	0	0	0	$\cdots$
3	x	x	0	0	0	0	$\cdots$
4	x	x	x	0	0	0	$\cdots$
5	x	x	x	x	0	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

- In this table, the tip of the wedge is at the point (2,1).
- This tip corresponds to the values of $p = 2$ and $q = 1$ in the ARMA model.

Disclaimer for the ESACF table

- The tables shown above represent theoretical results for infinitely large sample sizes.
- With real data, we would not expect the tables to follow such a clear cut pattern.
- Remember, ESACF values $\widehat{\rho}_j^{(m)}$ are estimates, so they have inherent sampling variation.
- For some data sets, ESACF table may reveal 2 or 3 models which are consistent with the estimates.
- In other situations, ESACF may be completely ambiguous and give little or no information, especially if the sample size n is small.

Sample distribution

Theorem

If the residual process

$$\widehat{W}_t^{(j)} = \left(1 - \widehat{\phi}_1^{(j)} B - \widehat{\phi}_2^{(j)} B^2 - \cdots - \widehat{\phi}_m^{(j)} B^m\right) Y_t$$

truly is white noise, then ESACF estimate

$$\widehat{\rho}_j^{(m)} \sim AN\left(0, \frac{1}{n-m-j}\right), \text{ as } n \to \infty.$$

- *Therefore, we would expect 95 percent of the estimates* $\widehat{\rho}_j^{(m)}$ *to fall within* $\pm 1.96/\sqrt{n-m-j}$. *Values outside these cutoffs are classified with an "x" in ESACF. Values within these bounds are classified with a "0".*

Example 6.6. (a) an ARMA(1, 1) process and its ESACF

- The first simulation is an ARMA(1,1), with $n = 200$, $\phi = 0.6$, $\theta = -0.8$, and $e_t \sim \text{iid} N(0, 1)$.
- The ESACF produced from the simulation was

AR/MA	0	1	2	3	4	5	6	7	8	9	10
0	x	x	x	x	x	0	0	0	0	0	0
1	x	0	0	0	0	0	0	0	0	0	0
2	x	0	0	0	x	0	0	0	0	0	0
3	x	x	x	0	0	0	0	0	0	0	0
4	x	0	x	0	x	0	0	0	0	0	0
5	x	x	x	x	0	0	0	0	0	0	0
6	x	x	0	x	x	0	0	0	0	0	0
7	x	x	0	x	0	x	0	0	0	0	0

Example 6.6. Interpretation for output in (a)

Interpretation:

- This sample ESACF agrees largely with the theory, which says that there should be a wedge of zeros with tip at (1,1); the "x"s at (2,4) and (4,4) may be false positives.

- If one is willing to additionally assume that the "x" at (3,2) is a false positive, then an ARMA(2,1) model would also be deemed consistent with these estimates.

Example 6.6. (b) An ARMA$(2, 2)$ process and its ESACF

- The second simulation is an
  - ARMA(2,2), with $n = 200$, $\phi_1 = 0.5$, $\phi_2 = -0.5$, $\theta_1 = -0.8$, $\theta_2 = 0.2$, and $e_t \sim$ iid$N(0, 1)$.
- The sample ESACF produced from the simulation was

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12
0	x	x	x	0	x	0	0	0	0	0	x	0	0
1	x	x	x	0	x	0	0	0	0	0	x	0	x
2	x	0	0	0	0	0	0	0	0	0	x	0	0
3	x	x	0	0	0	0	0	0	0	0	0	0	0
4	x	x	0	x	x	0	0	0	0	0	0	0	0
5	x	x	x	x	0	0	0	0	0	0	0	0	0
6	x	x	x	x	0	0	0	0	x	0	0	0	0
7	x	0	x	x	x	0	0	0	0	0	0	0	0

Example 6.6. Interpretation for output in (b)

Interpretation:

- This ESACF also agrees largely with the theory, which says that there should be a wedge of zeros with tip at (2,2).
- If one is willing to additionally assume that the "x" at (4,3) is a false positive, then an ARMA(2,1) model would also be deemed consistent with these estimates.

6.5. Nonstationarity

- In general, an ARIMA(p, d, q) process can be written as

$$\phi(B)(1-B)^d Y_t = \theta(B)e_t,$$

where the AR and MA characteristic operators are

$$\phi(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$$
$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)$$

and

$$(1-B)^d Y_t = \nabla^d Y_t.$$

- Three functions can help us identify possible values for p and q in stationary ARMA processes.
  - The sample ACF $\rightarrow$ the order q in an MA process.
  - The sample PACF $\rightarrow$ the order p in an AR process.
  - The ESACF $\rightarrow$ the orders p and q in an ARMA process.

Non-stationary processes

- For a series of data, a clear indicator of nonstationarity is that the sample ACF exhibits a very slow decay across lags.
- This occurs because in a nonstationary process, the series tends to "hang together" and display "trends".

Differencing

- When the sample ACF for a series decays very slowly, take first differences.
- If the sample ACF for the first differences resembles that a stationary ARMA process (the ACF decays quickly), then take $d = 1$ in the ARIMA(p, d, q) family and use the ACF, PACF, and ESACF (on the first differences) to identify plausible values of p and q.
- If the sample ACF for the first differences still exhibits a slow decay across lags, take second differences and use $d = 2$.
- There should rarely be a need to consider values of $d > 2$.
- If a transformation is needed, implement the transformation up front before taking any differences.

Overdifferencing occurs when we take d to be too large

- Consider an invertible IMA(1,1), $Y_t = Y_{t-1} + e_t - \theta e_{t-1}$.
- The first differences $\nabla Y_t = Y_t - Y_{t-1} = e_t - \theta e_{t-1}$, a stationary and invertible MA(1) process.
- The 2nd differences $\nabla^2 Y_t = \left[1 - (1+\theta)B + \theta B^2\right] e_t$ causes a not-invertible process, because $\theta(x) = 1 - (1+\theta)x + \theta x^2$ has the unit root $x = 1$.
- we difference an invertible MA(1) process into a non-invertible one.
- If a process is not invertible, the parameters in the model can not be estimated uniquely.

Dickey and Fuller unit root test

- The sample ACF may be subjective in "borderline cases".
- The time series could be nonstationary because of random walk, drift, or trend.
- Dickey and Fuller (1979) proposed a test methodology for testing whether or not an observed time series is stationary.

Dickey and Fuller unit root test

- One way to test this is to evaluate a regression that nests a mean, a lagged term (to test for difference stationarity), and a deterministic trend term (to test for trend stationarity) in one model:

$$y_t = \alpha + \rho y_{t-1} + \beta t + e_t$$

and taking the first difference of the y_t yields

$$\nabla y_t = \alpha + (\rho - 1) y_{t-1} + \beta t + e_t.$$

- This model forms the basis of the Dicker-Fuller test. The test parameter distributions depend on the sample size and which terms are in the model.
- The three model contexts are those of (1) a pure random walk, (2) random walk plus drift, or (3) the combination of deterministic trend, random walk, and drift.

Dickey and Fuller unit root test

- The application of the Dickey-Fuller test depends on the regression context in which the lagged dependent variable is tested.
- To detect nonstationarity, the DF test entails a regression of a series on a first lag of itself to determine whether the regression coefficient of the lagged term is essentially equal to unity and is significant,
  - under conditions (cases) of no constant, that is, zero constant,
  - some nonzero constant, that is, zero trend, or
  - some nonzero constant plus a deterministic trend coefficient.

Dickey and Fuller unit root test for no constant case

- Consider the case called the autoregressive no constant model,

$$y_t = \rho y_{t-1} + e_t \text{ or } \nabla y_t = (\rho - 1) y_{t-1} + e_t.$$

- $H_0 : \quad \rho = 1$ versus $H_1 : \rho < 1$ for a pure random walk.
- Use LS method to the model and get estimator of ρ. The

$$\tau_{DF} = \frac{\hat{\rho} - 1}{se(\hat{\rho})} = \frac{\sum_{t=1}^{n} y_{t-1} e_t}{se(\hat{\rho}) \sum_{t=1}^{n} y_{t-1}^2}.$$

- If e_t is WN with $\sigma_e^2 > 2$, $\tau_{DF} \longrightarrow$ standard Brownian motion as $T \longrightarrow \infty$, see Chan and Wei(1988), and Phillips (1987).
- The τ_{DF} resembles a $t-$test. SAS has performed its own Monte Carlo study with 10^8 replications. $|\tau_{DF}| \geq \tau_1$ is a rule for rejection of H_0 where τ_1 is the critical value of this first case.

Dickey and Fuller unit root test

- The second Dickey-Fuller case is a random walk plus a drift.
- Consider the regression model, also called the AR(1) with constant model (Greene,1997),

$$y_t = \alpha + \rho y_{t-1} + e_t.$$

- The first hypothesis: $H_0^{1st}: \ \rho = 1$ vs $H_1^{1st}: \rho < 1$
- The second hypothesis: $H_0^{2nd}: \alpha = 0 \ \& \ \rho = 1$, a joint F test.
- The statistic is

$$\tau_{DF} = \frac{\hat{\rho} - 1}{se(\hat{\rho})}.$$

- τ_{DF} for testing $(\rho - 1)/se$ resembles a $t-$test.
- An F test for $\alpha = 0$ and $\rho = 1$ is available in SAS.

Dickey and Fuller unit root test

- The third Dickey-Fuller case is one with a context of random walk plus drift in addition to a deterministic linear trend.
- Consider the regression model

$$y_t = \alpha + \rho y_{t-1} + \beta t + e_t.$$

- The first hypothesis: $H_0^{1st}: \ \rho = 1$ vs $H_1^{1st}: \rho < 1$
- The third hypothesis: $H_0^{3rd}: \rho = 1 \ \& \ \beta = 0$, a joint F test.
- The statistic is

$$\tau_{DF} = \frac{\hat{\rho} - 1}{se(\hat{\rho})}.$$

- τ_{DF} for testing $(\rho - 1)/se$ resembles a $t-$test. The statistic become even more nonstandard (more biased).
- An F test is available for $\rho = 1$ and $\beta = 0$ in SAS.

Augmented Dickey and Fuller unit root test

- Consider the model

$$Y_t = \rho Y_{t-1} + X_t,$$

where $\{X_t\}$ is a stationary AR(k) process, that is,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_k X_{t-k} + e_t.$$

- Therefore,

$$\begin{aligned} Y_t =& \alpha + \rho Y_{t-1} + \beta t \\ &+ \phi_1 (Y_{t-1} - \rho Y_{t-2}) + \cdots + \phi_k (Y_{t-k} - \rho Y_{t-k-1}) + e_t. \end{aligned}$$

Augmented Dickey and Fuller unit root test

- The augmented Dickey-Fuller (ADF) unit root test therefore tests

$$H_0: \ \rho = 1 \ \ (\text{nonstationarity})$$

versus

$$H_1: \ \rho < 1 \ \ (\text{stationarity}).$$

Augmented Dickey and Fuller unit root test (continued)

- Dickey and Fuller advocated that this test could be carried out using least squares regression.
- When $\rho = 1$, the model under $H_0 : \ \rho = 1$ is

$$\nabla Y_t = (\rho - 1)Y_{t-1} + \phi_1 \nabla Y_{t-1} + \cdots + \phi_k \nabla Y_{t-k} + e_t.$$

- We regress ∇Y_t on Y_{t-1}, ∇Y_{t-1}, $\cdots$, ∇Y_{t-k}.
- Let $a = \rho - 1$. Then, $a = 0$ is equivalent to $\rho = 1$.

Augmented Dickey and Fuller unit root test (continued)

- Decide between H_0 and H_1 by examining $\hat{a}$.
- If the least squares regression estimate of a is significantly different from 0, we reject H_0 and conclude that the process is stationary.
- If the least squares regression estimate of a is not significantly different from 0, we do not reject H_0.

Remark: Augmented Dickey and Fuller unit root test

- The test statistic and its large-sample distr^n are complicated.
- The test statistic is similar to the t test statistic from OLS, however, the large-sample distribution is not t.
- The only thing we need to do is choose a value of k in the model

$$\nabla Y_t = (\rho - 1)Y_{t-1} + \phi_1 \nabla Y_{t-1} + \cdots + \phi_k \nabla Y_{t-k} + e_t,$$

  that is, the value k is the order of the AR process for ∇Y_t.
- Under $H_0 : \ \rho = 1$, the value of $a = 0$.

How to determine the order k for ADF unit root test

- Determine the "best value" of k using model selection criteria that we will discuss in the next subsection.
- Cromwell *et al.* (1994), assuming normality of the residuals, give the formula for the likelihood ratio test:

$$LR = n \ln\left(\frac{\sigma_{k-1}^2}{\sigma_k^2}\right) \sim \chi^2(1),$$

where σ_i^2 is the residual variance of model i, for test of

$$H_0 : AR(k-1) \text{ model} \quad \text{vs } H_1 : AR(k) \text{ model.}$$

6.6.1. Akaike's Information Criterion (AIC)

- Akaike's Information Criterion is used more generally for model selection in statistics (not just in the analysis of time series data).
- It was developed by Hirotsugu Akaike in 1974, under the name of "an information criterion" AIC.
- Herein, we restrict attention to its use in selecting candidate stationary ARMA(p, q) models.

AIC

- The Akaike's Information Criterion (AIC) says to select the ARMA(p, q) model which minimizes

$$\text{AIC} = -2\ln(L) + 2k,$$

where $\ln(L)$ is the natural logarithm of the estimated likelihood function (computed under some probability distribution for $Y_1, Y_2, \cdots, Y_n$) and k is the number of parameters in the model (excluding the white noise variance).

AIC for a stationary no-intercept ARMA(p, q)

- In a stationary no-intercept ARMA(p, q) model, there are $k = p + q$ parameters.
- The likelihood function gives (loosely speaking) the "probability of the data", so we would like for it to be as large as possible. This is equivalent to wanting $-2\ln(L)$ to be as small as possible.
- The $2k$ term serves as a penalty, namely, we do not want models with too many parameters (adhering to the Principle of Parsimony).

AIC for ARMA(p, q)

- The AIC is an estimator of the expected Kullback-Leibler divergence, which measures the closeness of a candidate model to the truth. The smaller this divergence, the better the model. See pp 130 (CC).
- A problem arises in that AIC is a biased estimator of the expected KL divergence in ARMA(p, q) models. An alternative AIC statistic which corrects for this bias is

$$\text{AIC}_C = \text{AIC} + \frac{2(k+1)(k+2)}{n-k-2}.$$

6.6.2. Bayesian Information Criterion (BIC)

- The Bayesian Information Criterion (BIC) says to select the ARMA(p, q) model which minimizes

$$\text{BIC} = -2\ln(L) + k\ln(n),$$

  where $\ln(L)$ is the natural logarithm of the estimated likelihood function and k is the number of parameters in the model (excluding the white noise variance).

- Bayesian Information Criterion also is called Schwarz's Bayesian Information Criterion (SBC).

BIC for a stationary no-intercept ARMA(p, q) model

- In a stationary no-intercept ARMA(p, q) model, there are $k = p + q$ parameters.
  - Both AIC and BIC require the maximization of the log likelihood function (which requires a distributional assumption for the data). We will discuss maximum likelihood estimation in Chapter 7.
  - When compared to AIC, BIC offers a stiffer penalty for overparameterized models.

6.7. Summary for the first stage of ARMA modeling

1. The first step of the first stage for ARIMA model selection: Plot the data in a time series plot and identify an appropriate transformation if needed.
   - By examining the time series plot, we can get a good idea about whether the series contains a trend, seasonality, outliers, nonconstant variance, etc.
   - This understanding often provides a basis for postulating a possible data transformation (e.g. Box-Cox family *et al*).
   - Transformations should always be performed before taking any data differences.

The second step of model selection

2. Compute the sample ACF and the sample PACF of the original series (or transformed series) and further confirm the need for differencing.
    - For very slowly decaying sample ACF, take first differences.
    - Implement ADF unit root tests (or PP unit root test or RW test) for stationarity on the original or transformed series.
    - In a borderline case, differencing is generally recommended.
    - Higher order differencing may be needed after an ADF test for stationarity of the first differences is implemented.
    - Some authors argue that the consequences of overdifferencing are much less serious than those of underdifferencing. However, overdifferencing can create model identifiability problems.

The third step of model selection

3. Compute the sample ACF, the sample PACF, and ESACF of the original, properly transformed, properly differenced, or properly transformed/differenced series to identify the orders of p and q.
    - Usually, p and q are not larger than 4 (excluding seasonal models, which we have not yet to discuss).
    - Use your knowledge of the patterns for theoretical versions of these functions, that is,
        - the ACF for an MA(q) drops off after lag q
        - the PACF for an AR(p) drops off after lag p
        - the "tip" in the ESACF identifies the proper ARMA(p, q) model.

The third step of model selection (continued)

- To build a reasonable ARIMA model, ideally, we need a minimum of about $n = 50$ observations, and the number of sample ACF and PACF to be calculated should be about $n/4$ (a rough guideline). SAS gives ACF, PACF of 24 lags. It might be hard to identify an adequate model with smaller data sets.
- "The art of model selection is very much like the method of an FBI's agent criminal search. Most criminals disguise themselves to avoid being recognized." This is also true of the ACF, PACF, and ESACF. Sampling variation can disguise the theoretical ACF,PACF and ESACF patterns.
- BIC and AIC can also be used to identify models which are consistent with the data.

Next two stages

- It is unlikely, after going through all of this, that the statistician is going to be able to identify one model that is a "clear-cut" choice. It is more likely that a small number of candidate models have been identified from the steps above.

- Next two stags: With (hopefully small) set of candidate models, then we move forward to
  - parameter estimation (fit) (Chapter 7),
  - model diagnostics (model checking) (Chapter 8),
  - finally forecasting (Chapter 9).

Appendix: the three stages of ARIMA modeling

The analysis performed by PROC ARIMA is divided into three stages, corresponding to the stages described by Box and Jenkins (1976). They are

- the identification stage,
- the estimation and diagnostic checking stage
- the forecasting stage.

In the identification stage

- use the IDENTIFY statement to specify the response series and identify candidate ARIMA models for it.
- The IDENTIFY statement reads time series that are to be used in later statements, possibly differencing them, and computes ACF, IACF, PACF, and cross correlations.
- Stationarity tests, e.g., ADF, can be performed to determine if differencing is necessary.
- The analysis of the IDENTIFY statement output usually suggests one or more ARIMA models, e.g., from esacf or scan and BIC, that could be fit. Options enable you to test for stationarity and tentative ARMA order identification.

In the estimation and diagnostic checking stage

- use the ESTIMATE statement to specify the ARIMA model to fit to the variable specified in the IDENTIFY statement, and to estimate the parameters of that model.
- The ESTIMATE statement also produces diagnostic statistics to help you judge the adequacy of the model. Significance tests for parameter estimates indicate whether some terms in the model may be unnecessary. Goodness-of-fit statistics aid in comparing this model to others.
- Tests for white noise residuals indicate whether the residual series contains additional information that might be utilized by a more complex model.
- The OUTLIER statement provides another useful tool to check whether the currently estimated model accounts for all the variation in the series. If the diagnostic tests indicate problems with the model, you try another model, then repeat the estimation and diagnostic checking stage.

In the forecasting stage

- use the FORECAST statement to forecast future values of the time series and to generate confidence intervals for these forecasts from the ARIMA model produced by the preceding ESTIMATE statement.

Example 6.7. A simulated IMA(1,1) model

- Generates a pseudo-random sample of 100 periods from the ARIMA(0,1,1) process

$$Y_t = Y_{t-1} + e_t - 0.8e_{t-1}, e_t \sim iidN(0,1).$$

- the sas program for data step is follows:
  title1 'Simulated IMA(1,1) time series';
  data aa; y = 0.9; e1 = 0; do t =0 to 100;
  e =rannor(0); y = y + e - 0.8*e1; output;
  e1 =e; end; run;
- The following ARIMA procedure statements identify:
  proc arima data=aa;
  identify var=y esacf scan minic
  stationarity=(adf=(0,1,2,3,4,5)); run;

Example 6.7. A simulated IMA(1,1) model

- Nonstationarity is reflected in a pattern of significant autocorrelations that do not decline quickly with increasing lag, not in the size of the autocorrelations.
- ESACF and SCAN suggest ARMA(0 1, 1) model.
- Autocorrelation check for white noise displays that the sample autocorrelation for the first 24 lags are significant different from zero.
- The output shows the nonstationarity of the process from the behavior of the sample ACF.

Example 6.7. A simulated IMA(1,1) model

The following identify statement is required.

```
identify var=y(1) esacf scan minic stationarity=(adf=(0,1,2,3,4,5)); run;
```

- The second IDENTIFY statement differences the series. This output shows autocorrelation, inverse autocorrelation, and partial autocorrelation functions typical of MA(1) processes.
- ESACF, SCAN and Autocorrelation check for white noise suggest MA(1) model for differenced simulated data.
- ADF unit root test displays the stationarity of differenced data.
- Therefore, ARMA(0,1,1) is a suitable model for the simulated series data.

Example 6.7. A simulated IMA(1,1) model

The following estimate statement for the differening data is required.

estimate q=1 plot;

- The estimate statement fits an ARIMA(0,1,1) model to be simulated data. In the case, parameter estimates are

$$\hat{\mu} = 0, \quad \hat{\theta} = 0.79706 \text{ and } \hat{\sigma}^2 = 0.808849.$$

- The estimated ARIMA(0,1,1) model is as follows:

$$Y_t = Y_{t-1} + e_t - 0.79706 e_{t-1} \text{ with } e_t \sim N(0, 0.808849)$$

- The true values of parameter are

$$\mu = 0, \quad \theta = 0.8 \text{ and } \sigma^2 = 1.$$

- Independence and normality can also be tested.

Example 6.8. A simulated ARIMA(1,0,0) model with a trend

- Generates a pseudo-random sample of 100 periods from the ARIMA(1,0,0) process with atrend

$$\begin{cases} Y_t = -2 + 0.6t + X_t, \\ X_t = 0.6X_{t-1} + e_t, e_t \sim iidN(0,1). \end{cases}$$

- the sas program for data step is follows:
  title1 'Simulated AIMA(1,0,0) time series';
  data bb; x1=0; do time=-5 to 100; x=.6*x1+normal(0);
  y=-2+.6*time+x; if time>0 then output; x1=x; end; run;
- The following ARIMA procedure statements identify:
  proc arima data=bb;
  identify var=y esacf scan minic
  stationarity=(adf=(0,1,2,3,4,5)); run;

Example 6.8. A simulated ARIMA(1,0,0) model with a trend

- Nonstationarity is reflected in a pattern of significant autocorrelations that do not decline quickly with increasing lag, not in the size of the autocorrelations.
- PACF suggests ARMA(1, 0, 0) model.
- ESACF and BIC suggest ARMA(1, 0, 0) model or ARIMA(0,1,0).
- Autocorrelation check for white noise displays that the sample autocorrelation for the first 24 lags are significant different from zero.
- The ADF root test displays thate $\alpha \neq 0$, $\beta \neq 0$ and $\rho \neq 1$ for the simulated ARMA(1,0,0) data.
- There is a trend for the data.

Example 6.8. A simulated ARIMA(1,0,0) model with a trend

- It is not suitable for us to use ARIMA procedure. It is necessary for us to use proc autoreg procedure to fit the simulated data

- SAS codes are as follows
  proc autoreg data=bb; model y = time/nlag=6 method=ml dwprob dw=6 backstep; output out=resdat2 r=resid2 ucl=ucl lcl=lcl p=forecast pm=ytrend; title2 'AR model of the data'; run;

Example 6.8. A simulated ARIMA(1,0,0) model with a trend

- One good way to determine the order of autoregressive error is to employ the backward elimination procedure. This output reveals the autocorrelation, the standard error, and the T ratio for each of the parameters tested. Significant autocorrelation is found in lag 1.
- The model suggested by the OLS estimation is

$$\begin{cases} Y_t = -2.0158 + 0.6023t + X_t \\ X_t = 0.607X_{t-1} + e_t. \end{cases}$$

with $e_t \sim N(0, 2.06208)$ with AIC=358.1388.

Example 6.8. A simulated ARIMA(1,0,0) model with a trend

- The MLE output and the estimated equation is

$$\begin{cases} Y_t = -2.0585 + 0.602t + X_t \\ X_t = 0.6194 X_{t-1} + e_t. \end{cases} \qquad (*)$$

  with $e_t \sim N(0, 1.29188)$ and AIC=312.8388.

- For the estimated model, the Durbin-Warson test and normality test show that e_t is iid normal series.
- The fitted model (*) is best in AIC or BIC.

Example 6.9. Global temperature deviation data

- We have seen from the time plot that global temperature deviation data shows a trend with the change of time.
- the sas program for procedure step is follows:

```
title1 'Global temperature deviation data';
proc arima data=TS6.figure68;
identify var=temperatureV esacf scan
stationarity=(adf=(0,1,2,3,4,5,6); run;
```

Example 6.9. Global temperature deviation data

- Nonstationarity is reflected in a pattern of significant autocorrelations that do not decline quickly with increasing lag, not in the size of the autocorrelations.
- Autocorrelation check for white noise displays that the sample autocorrelation for the first 24 lags are significant different from zero.
- ADF unit root test shows that there is not strong evidence against hull hypothesis H_0: a unit root.
- ESACF and SCAN suggest ARMA(0 1, 2) model, ARMA(1, 1, 1) model and ARMA(0, 2, 1) model.

Example 6.9. Global temperature deviation data

- The following identify statement is required.

```
identify var=temperatureV(1)
esacf
scan
minic
stationarity=(adf=(0,1,2,3,4,5,6));
run;
```

Example 6.9. Global temperature deviation data

- ADF unit root test in the second IDENTIFY statement displays the stationarity of differenced data. No further differencing is necessary. ARMA(0,2,1) model can be omitted.
- ESACF and SCAN both suggest MA(2) for differenced data with a minimum BIC value: $BIC(0,2) = -4.47801$.
- Therefore, ARMA(0,1,2) may be the best suitable model for the global temperature deviation data.
- SCAN also suggests ARMA(1,1) for differenced data with a BIC value: $BIC(1,1) = -4.4581$.
- Further check for ARMA(0,1,2) or ARMA(1,1,1) model can be implemented by ESTIMATE stage.

Example 6.9. Global temperature deviation data

The following estimate statement is required.

```
estimate q=2 method=ml; run;
estimate p=1 q=1 method=ml; run;
```

- For ARMA(0,1,2) model, the estimated model shown in sas output is as follows:

$$(1-B)Y_t = (1 - 0.4268B - 0.24721B^2)e_t$$

with $BIC = -216.683$ and $\hat{\sigma}^2 = 0.011865$.

- Correlation of $\hat{\theta}_1$ and $\hat{\theta}_2$ is -0.554.
- The test in correlation check of residuals for the first 6 lag sample ACF shows that the statistic value is 3.49, occurence with 0.4788 probability.

Example 6.9. Global temperature deviation data

- For ARMA(1,1,1) model, the estimated model shown in sas output is as follows:

$$(1 - 0.33245)(1 - B)Y_t = (1 - 0.78858B)e_t$$

with $BIC = -214.86$ and $\hat{\sigma}^2 = 0.012022$.

- Correlation of $\hat{\phi}_1$ and $\hat{\theta}_1$ is 0.788.
- The test in correlation check of residuals for the first 6 lag sample ACF shows that the statistic value is 5.57, occurence with 0.234 probability.

Example 6.9. Global temperature deviation data

- Compared two estimated model, ARMA(0,2,1) model is better to fit the data
- The estimated model has the following equation

$$Y_t = Y_{t-1} + e_t - 0.4268e_{t-1} - 0.24721e_{t-2}.$$

- There may be a trend in the global temperature deviation data. Using proc autoreg procedure to fit the data yields the following equation

$$\begin{cases} Y_t = 0.000012t + X_t \\ X_t = 0.5302X_{t-1} + 0.2213X_{t-4} + e_t, \end{cases} \quad (**)$$

with $e_t \sim N(0, 0.01117)$ and $BIC = -222.65294$.

Conclusion for global temperature deviation data

- The global temperature follows the following pattern

$$\begin{cases} Y_t = 0.000012t + X_t \\ X_t = 0.5302X_{t-1} + 0.2213X_{t-4} + e_t. \end{cases}$$

- Non-stationarity is a bad new for opponents of global warming.
- Statistical conclusion supports the claim that the overall trend is caused by some human-induced interface, for example, human being greedy activities.

Example 6.10. Los Angeles annual rainfall data

- The sas program for data step is follows:

  Data TS6.figure69; infile 'C:\Data _CC \larain.dat';
  input rainfall @@; t=_N_-1;
  if _N_ < 2 then delete;
  year=intnx('year','1jan1878'd, _N_-1);
  format year year4.; run;

- The procedure step is written by using an identify statement.

  proc arima data=TS6.figure69;
  identify var=rainfall scan esacf minic stationarity=(adf=6);
  run;

Example 6.10. Los Angeles annual rainfall data

- ACF and PACF shows that the data seems to be from a white noise.
- Autocorrelation check for white noise displays that the sample autocorrelation for the first 24 lags are not significant different from zero.
- ADF unit root test shows that there is strong evidence against hull hypothesis H_0: a unit root.
- ESACF and SCAN suggest ARMA(0,0,0) model.

Example 6.10. Los Angeles annual rainfall data

- Differencing data makes thing worse.
- I suggest that you use Box-Cox transformation or log transformation for the original data and use ARIMA procedure again.

chapter 6 is over

Thank you for your attention!