# 151163A - Financial Econometrics

## IIc. Stationary Processes: Estimation, Order Selection and Forecasting

CHEUNG Ying Lun

Capital University of Economics and Business

# Estimation of an AR($p$) process

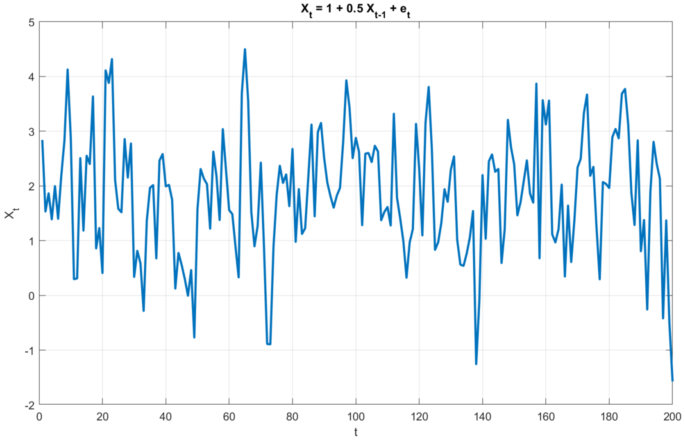
Conditional on the first $p$ observations, we have

$$X_t = a_0 + a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t, \qquad t = p+1, \ldots, T$$

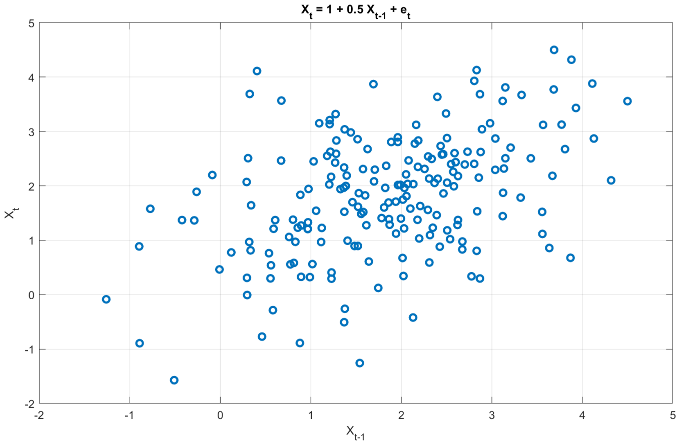
which is in form of a multiple linear regression and can be estimated by the least-squares method.

# Estimation of an AR($p$) process





# Estimation of an AR($p$) process



Suppose

$$y_i = \boldsymbol{\beta}'\mathbf{X}_i + e_i = \sum_{k=1}^{p} \beta_k X_{ki} + e_i, \qquad i = 1, \ldots, N.$$

Then, we can estimate $\boldsymbol{\beta}$ by solving the optimization problem

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} (y_i - \boldsymbol{\beta}'\mathbf{X}_i)^2.$$

The solution is given by

$$\widehat{\boldsymbol{\beta}} = \left( \sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \left( \sum_{i=1}^{N} \mathbf{X}_i y_i \right)$$

# Estimation of an AR($p$) process

We solve the optimization problem

$$(\widehat{a}_0, \ldots, \widehat{a}_p) = \underset{\{a_0,\ldots,a_p\}}{\arg\min} \sum_{t=p+1}^{T} (X_t - a_0 - a_1 X_{t-1} - \cdots - a_p X_{t-p})^2$$

which yields

$$\widehat{\mathbf{a}} = \left[ \sum_{t=p+1}^{T} \mathbf{Z}_t \mathbf{Z}_t' \right]^{-1} \left[ \sum_{t=p+1}^{T} \mathbf{Z}_t X_t \right]$$

where $\mathbf{Z}_t = (1, X_{t-1}, \ldots, X_{t-p})'$.

The fitted model is

$$\widehat{X}_t = \widehat{a}_0 + \widehat{a}_1 X_{t-1} + \cdots + \widehat{a}_p X_{t-p}$$

and the associated residual is

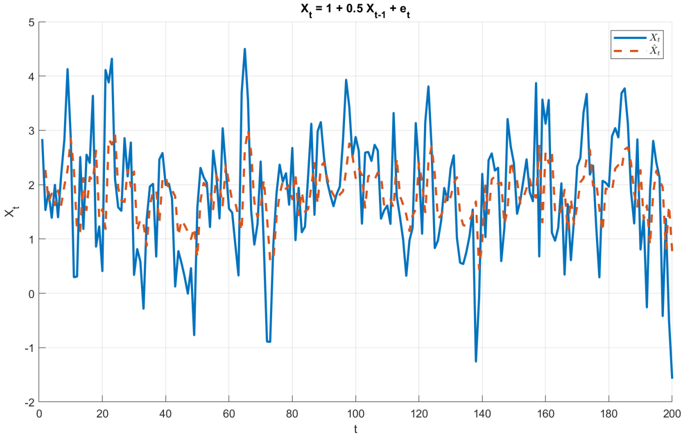
$$\widehat{\varepsilon}_t = X_t - \widehat{X}_t.$$

The variance of the residuals is given by

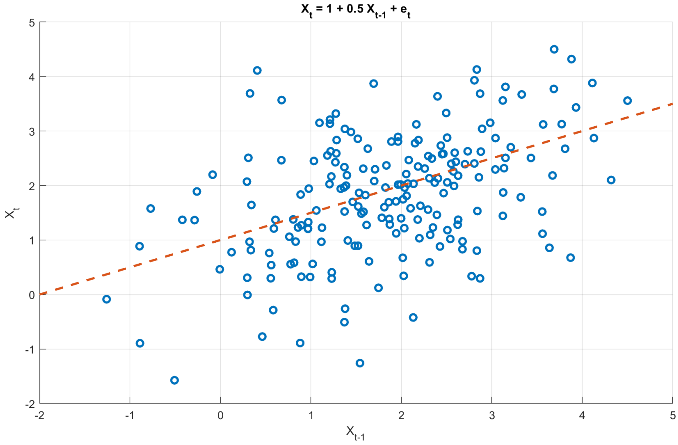
$$\widehat{\sigma}^2 = \frac{1}{T - 2p - 1} \sum_{t=p+1}^{T} \widehat{\varepsilon}_t^2.$$

# Fitted value





# Fitted value



# Order Selection

# Model checking



If the model is adequate, then the residual series should behave as a white noise. One can check if the autocorrelation function of the residual series is different from zero. One can also apply the Portmanteau test,

$$Q(m) = T(T+2)\sum_{j=1}^{m}\frac{\widehat{\rho}_{\widehat{\varepsilon}}(j)^2}{T-j}$$

If the residual series shows serial correlation (i.e., non-zero autocorrelation), then one may need to increase the AR order.

# Model checking

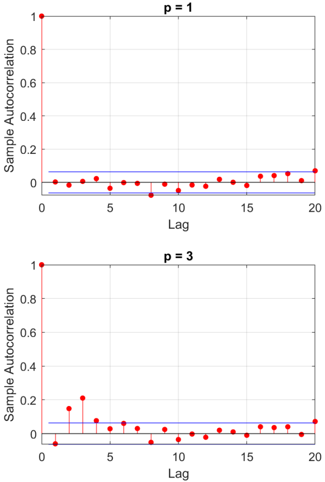
Simulate four series:

$$
\begin{aligned}
X_t &= 0.2X_{t-1} + \varepsilon_t \\
X_t &= 0.2X_{t-1} + 0.2X_{t-2} + \varepsilon_t \\
X_t &= 0.2X_{t-1} + 0.2X_{t-2} + 0.2X_{t-3} + \varepsilon_t \\
X_t &= 0.2X_{t-1} + 0.2X_{t-2} + 0.2X_{t-3} + 0.2X_{t-4} + \varepsilon_t
\end{aligned}
$$

for $t = 1, \ldots, 1000$. Then we estimate an AR(1) model and find the autocorrelations of the residual series.

# Autocorrelation of residual series





# Partial autocorrelation function



The partial autocorrelation function (PACF) can be obtained by consecutively estimating the following models

$$
\begin{aligned}
X_t &= a_{0,1} + a_{1,1} X_{t-1} + e_{1t} \\
X_t &= a_{0,2} + a_{1,2} X_{t-1} + a_{2,2} X_{t-2} + e_{2t} \\
X_t &= a_{0,3} + a_{1,3} X_{t-1} + a_{2,3} X_{t-2} + a_{3,3} X_{t-3} + e_{3t} \\
&\vdots
\end{aligned}
$$

The estimates $\widehat{a}_{1,1}$, $\widehat{a}_{2,2}$, $\widehat{a}_{3,3}$, and so on, are the sample PACF.

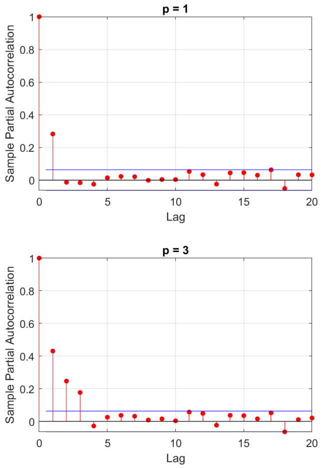
For a stationary normal AR($p$) process, it can be shown that the sample PACF has the following properties:

- $\widehat{a}_{p,p}$ converges to $a_p$ as the sample size $T$ goes to infinity.
- $\widehat{a}_{j,j}$ converges to zero for all $j > p$.
- The asymptotic variance of $\widehat{a}_{j,j}$ is $1/T$ for all $j > p$.

Therefore, for an AR($p$) process, the sample PACF cuts off at lag $p$.

# Partial autocorrelation function





# Information criteria



Two well-known information criteria are

$$AIC(j) = \ln(\widetilde{\sigma}_j^2) + \frac{2j}{T}$$
$$BIC(j) = \ln(\widetilde{\sigma}_j^2) + \frac{j \ln T}{T}$$

where

$$\widetilde{\sigma}_j^2 = \frac{1}{T-j} \sum_{t=j+1}^{T} \varepsilon_t(j)^2$$

and $\varepsilon_t(j)$ is the residual of a fitted AR$(j)$ model.

The first term $\ln(\widetilde{\sigma}_j^2)$ measures how well an AR($j$) model fits the data, while the second term is the penalty term, which penalizes model complexity.

- If $j < p$, increasing $j$ can significantly improve model fit, thus reducing AIC and BIC.
- if $j > p$, increasing $j$ does not improve model fit, therefore AIC and BIC become larger due to the penalty term.

Therefore, we choose $j$ such that AIC or BIC is minimized.

# AIC



| $j$ | $p = 1$ | $p = 2$ | $p = 3$ | $p = 4$ |
|---|---|---|---|---|
| 0 | 0.011 | 0.066 | 0.187 | 0.450 |
| 1 | **-0.041** | -0.012 | 0.052 | 0.150 |
| 2 | -0.039 | **-0.039** | -0.002 | 0.045 |
| 3 | -0.037 | -0.037 | **-0.036** | -0.010 |
| 4 | -0.036 | -0.034 | -0.033 | **-0.034** |
| 5 | -0.033 | -0.033 | -0.032 | -0.032 |
| 6 | -0.033 | -0.034 | -0.033 | -0.032 |
| 7 | -0.032 | -0.032 | -0.032 | -0.030 |
| 8 | -0.030 | -0.030 | -0.030 | -0.030 |
| 9 | -0.029 | -0.028 | -0.028 | -0.028 |
| 10 | -0.030 | -0.028 | -0.027 | -0.026 |

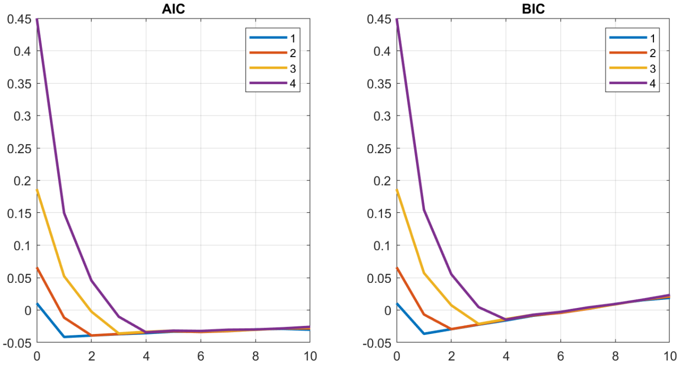
# BIC



| $j$ | $p = 1$ | $p = 2$ | $p = 3$ | $p = 4$ |
|---|---|---|---|---|
| 0 | 0.011 | 0.066 | 0.187 | 0.450 |
| 1 | **-0.037** | -0.007 | 0.057 | 0.155 |
| 2 | -0.029 | **-0.029** | 0.007 | 0.055 |
| 3 | -0.022 | -0.022 | **-0.021** | 0.005 |
| 4 | -0.016 | -0.015 | -0.014 | **-0.015** |
| 5 | -0.009 | -0.008 | -0.007 | -0.007 |
| 6 | -0.004 | -0.004 | -0.003 | -0.003 |
| 7 | 0.002 | 0.002 | 0.003 | 0.004 |
| 8 | 0.009 | 0.009 | 0.009 | 0.009 |
| 9 | 0.015 | 0.016 | 0.016 | 0.016 |
| 10 | 0.019 | 0.021 | 0.022 | 0.023 |

# AIC and BIC





# Forecasting

# One-step-ahead forecast



From the AR($p$) model, we have

$$X_{t+1} = a_0 + a_1 X_t + \cdots + a_p X_{t+1-p} + \varepsilon_{t+1}.$$

The optimal point forecast is given by the conditional expectation

$$\widehat{X}_t(1) = \mathbb{E}\left[X_{t+1} | \mathcal{I}_t\right] = a_0 + a_1 X_t + \cdots + a_p X_{t+1-p}.$$

where $\mathcal{I}_t = \{X_t, X_{t-1}, \ldots\}$. The estimation error is

$$\widehat{\varepsilon}_t(1) = X_{t+1} - \widehat{X}_t(1) = \varepsilon_{t+1}.$$

# One-step-ahead forecast

The variance of the forecast error is $\text{var}(\widehat{\varepsilon}_t(1)) = \text{var}(\varepsilon_{t+1}) = \sigma^2$. If $\varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$, then the 95% one-step-ahead interval forecast is

$$\widehat{X}_t(1) \pm 1.96\sigma$$

From the AR($p$) model, we have

$$X_{t+2} = a_0 + a_1 X_{t+1} + a_2 X_t + \cdots + a_p X_{t+2-p} + \varepsilon_{t+2}.$$

The optimal point forecast is given by

$$\begin{aligned}
\widehat{X}_t(2) &= \mathbb{E}\left[X_{t+2} | X_t, \ldots\right] \\
&= a_0 + a_1 \mathbb{E}\left[X_{t+1} | X_t, \ldots\right] + a_2 X_t + \cdots + a_p X_{t+2-p} \\
&= a_0 + a_1 \widehat{X}_t(1) + a_2 X_t + \cdots + a_p X_{t+2-p}
\end{aligned}$$

# Two-step-ahead forecast



The estimation error is

$$\begin{aligned}\widehat{\varepsilon}_t(2) &= X_{t+2} - \widehat{X}_t(2) \\ &= a_1(X_{t+1} - \widehat{X}_t(1)) + \varepsilon_{t+2} \\ &= a_1\varepsilon_{t+1} + \varepsilon_{t+2}\end{aligned}$$

Therefore, the variance of the forecast error is

$$\mathrm{var}(\widehat{\varepsilon}_t(2)) = (1 + a_1^2)\sigma^2$$

Notice that $\mathrm{var}(\widehat{\varepsilon}_t(1)) \le \mathrm{var}(\widehat{\varepsilon}_t(2)) \le \mathrm{var}(X_t)$.

# $h$-step-ahead forecast

In general, we have

$$X_{t+h} = a_0 + a_1 X_{t+h-1} + \cdots + a_p X_{t+h-p} + \varepsilon_{t+h}.$$

The $h$-step-ahead forecast is given by

$$\widehat{X}_t(h) = a_0 + a_1 \widehat{X}_t(h-1) + \cdots + a_p \widehat{X}_t(h-p),$$

where $\widehat{X}_t(i) = X_{t+i}$ if $i \leq 0$.

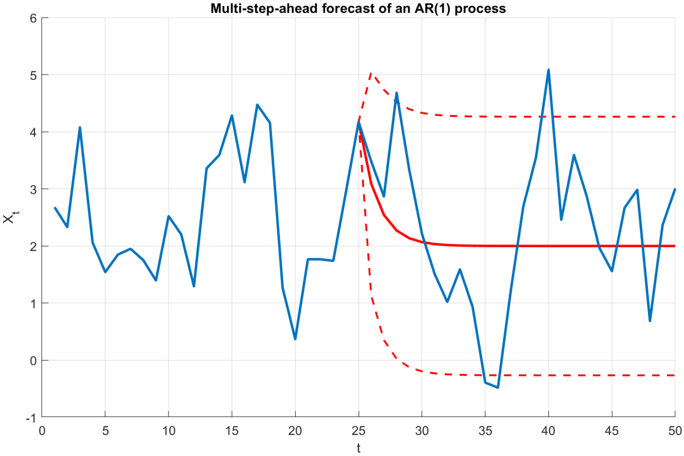
Consider the AR(1) model

$$X_t = 1 + 0.5X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \left(0, \sigma^2\right).$$

Find the $h$-step-ahead forecast and the respective variances of the forecast errors for $h = 1, 2$. What if $h \to \infty$?

# Example




Multi-step-ahead forecast of an AR(1) process
6
5
4
3
2
1
0
-1
X_t
0
5
10
15
20
25
30
35
40
45
50
t