Lecture 6: Fitting Time Series Models In The Time Domain

Walid Sharabati

Purdue University

Latest Update
September 19, 2016
Estimating Autocovariance and Autocorrelation Functions

Recall,

\[ c_k = \sum_{d=1}^{N-k} \frac{(X_t - \bar{X}_t)(X_{t+k} - \bar{X}_t)}{N}. \]

Also, the estimated ac.f. is

\[ \gamma_k = \text{cov}(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]. \]

Thus,

\[ \lim_{N \to \infty} E(c_k) = \gamma_k. \]

- The sample autocorrelation function \( r_k = \frac{c_k}{c_0} \) estimates \( \rho_k \).

Suppose \( X_1, \ldots, X_N \) are iid observations, then

\[ E(r_k) \approx -\frac{1}{N}, \quad \text{and} \quad \text{Var}(r_k) \approx \frac{1}{N}. \]

\( r_k \) is asymptotically \( N \left( -\frac{1}{N}, \frac{1}{N} \right) \).
Estimating Autocovariance and Autocorrelation Functions

- To check for randomness by plotting, use 95% confidence limits at $\left[-\frac{1}{N} \pm \frac{2}{\sqrt{N}}\right]$, which are further approximated to $\pm \frac{2}{\sqrt{N}}$.

**Figure**: The correlogram of 100 independent normally distributed observations. The dotted lines are at $\pm 0.2$. 
Estimating The Mean of $X_t$

- First note that $\text{Var}(\bar{X}_t) = \frac{\text{Var}(X_t)}{N}$, i.e. $\sigma_{\bar{X}_t} = \frac{\sigma_X}{\sqrt{N}}$ for IID observations.

- For correlated observations we have
  \[
  \text{Var}(\bar{X}_t) = \frac{\sigma_X^2}{N} \left[ 1 + 2 \sum_{k=1}^{N-1} \left( 1 - \frac{k}{N} \right) \rho_k \right] \approx \frac{\sigma_X^2}{N} \left[ 1 + 2 \sum_{k=1}^{N} \rho_k \right].
  \]
  as $N \to \infty$, $\frac{k}{N} \to 0$, $1 - \frac{k}{N} \to 1$.

- The approximated $100(1 - \alpha)\%$ confidence interval for $\mu$ is
  \[
  \bar{X}_t \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma_X}{\sqrt{N}} \cdot \sqrt{1 + 2 \sum_{k=1}^{N} \rho_k}, \quad \text{where} \quad \bar{X}_t = \frac{1}{N} \sum_{i=1}^{N} X_i.
  \]

- For the AR(1) process $X_t = \alpha X_{t-1} + Z_t$, $Z_t \sim \text{IID}(0, \sigma_Z^2)$.
  \[
  \text{Var} \left( \bar{X}_t \right) = \frac{\sigma_X^2}{N} \cdot \left[ \frac{1 + \alpha}{1 - \alpha} \right].
  \]
Estimating The Mean of $X_t$

**Example**

Consider the following stationary AR(1) model with mean $\mu$

$$X_t - \mu = \alpha(X_{t-1} - \mu) + Z_t,$$

where $|\alpha| < 1$ and $\{Z_t\}_{-\infty}^{\infty} \sim N(0, \sigma_Z^2)$.

Find a 95% C.I. for $\mu$. 
Estimating The Mean of $X_t$

**Answer**

First note that for AR(1):

$$\gamma_k = \frac{\sigma_Z^2}{1 - \alpha^2} \alpha^k, \quad \sigma_X^2 = \gamma_0 = \frac{\sigma_Z^2}{1 - \alpha^2}, \quad \rho_k = \alpha^k.$$

$$\text{Var} (\bar{X}_t) = \frac{\sigma_X^2}{N} \left[ 1 + 2 \sum_{r=1}^{N} \rho_r \right] = \frac{\sigma_X^2}{N} \left[ 1 + 2 \sum_{r=1}^{N} \alpha^r \right]$$

$$= \frac{\sigma_X^2}{N} \left[ 2 \sum_{r=0}^{N} \alpha^r - 1 \right] = \frac{\sigma_X^2}{N} \left[ \frac{2}{1 - \alpha} - 1 \right] = \frac{\sigma_X^2}{N} \left[ \frac{1 + \alpha}{1 - \alpha} \right]$$

$$= \frac{1}{N} \cdot \frac{\sigma_Z^2}{1 - \alpha^2} \cdot \frac{1 + \alpha}{1 - \alpha} = \frac{\sigma_Z^2}{N (1 - \alpha)^2}.$$

Thus, a 95% C.I. for $\mu$ is $\bar{X}_t \pm 1.96 \cdot \frac{1}{\sqrt{N}} \cdot \frac{\sigma_Z}{1 - \alpha}$. 
Estimating The Mean of $X_t$

Exercise
The sample mean $\bar{X}_{100} = 0.271$ was computed from a sample of size 100 generated from an AR(1) process with mean $\mu$, variance $\sigma^2_Z = 2$ and $\alpha = 0.6$.

1. Construct a 95% C.I. for $\mu$.
2. Are the data compatible with the hypothesis that $\mu = 0$?

Answer
1. $(-0.422, 0.964)$.
2. Yes, it is compatible since 0 falls in the 95% range.
It has been shown that under certain conditions

\[ \hat{\rho}_k = r_k \sim N \left( \rho_k, \frac{W}{N} \right), \]

where \( W \) is the covariance matrix (Bartlett’s formula).

In particular, for a pure random process if \( X_t = Z_t \sim IID(0, \sigma^2) \) then \( r_k \sim N \left(0, \frac{1}{N}\right) \) for \( k \neq 0 \). A 95% C.I. for \( \rho_k \) is

\[
\left( r_k - 1.96 \cdot \frac{1}{\sqrt{N}}, r_k + 1.96 \cdot \frac{1}{\sqrt{N}} \right) \approx \left( r_k - \frac{2}{\sqrt{N}}, r_k + \frac{2}{\sqrt{N}} \right)
\]

This provides a method for testing whether time series is white noise.
C.I. for the Autocorrelation Function of a Stationary Time Series

2 MA(q): If \( X_t = Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \), then

\[
r_k \sim N \left( \rho_k = 0, \frac{1}{N} \cdot \left[ 1 + 2 \sum_{i=1}^{q} \rho_i^2 \right] \right), \quad k > q.
\]

MA(1):

\[
\hat{\rho}_k = r_k \sim N \left( 0, \frac{1}{N} \left[ 1 + 2 \rho_1^2 \right] \right), \quad k > 1.
\]

\[
\hat{\rho}_1 = r_1 \sim N \left( \rho_1, \frac{1}{N} \cdot \left[ 1 - 3\rho_1^2 + 4\rho_4^1 \right] \right), \quad k = 1.
\]

A 95% confidence interval for \( \rho_1 \) is

\[
r_1 \pm 1.96 \cdot \sqrt{\frac{1 - 3\rho_1^2 + 4\rho_4^1}{N}}
\]

Exercise

Consider the MA(1) with \( r_1 = 0.6 \), and \( N = 100 \). Find a 95% C.I. for \( \rho_1 \).
3 AR(1): If \( X_t = \alpha X_{t-1} + Z_t \), then

\[
r_k \sim N \left( \rho_k, \frac{1}{N} \left[ \frac{(1 - \alpha^{2k}) (1 + \alpha^2)}{1 - \alpha^2} - 2k \cdot \alpha^{2k} \right] \right), \quad \text{for } k = 1, 2, \cdots.
\]

A 95% confidence interval for \( \rho_k \) is given by

\[
r_k \pm 1.96 \cdot \sqrt{\frac{1}{N} \left[ \frac{(1 - \alpha^{2k}) (1 + \alpha^2)}{1 - \alpha^2} - 2k \cdot \alpha^{2k} \right]}.
\]

For large \( k \), \( \alpha^{2k} \to 0 \).

\[
r_k \sim N \left( \rho_k, \frac{1}{N} \cdot \frac{1 + \alpha^2}{1 - \alpha^2} \right).
\]

A 95% confidence interval for \( \rho_k \) is given by

\[
r_k \pm 1.96 \cdot \sqrt{\frac{1}{N} \left( \frac{1 + \alpha^2}{1 - \alpha^2} \right)}.
\]
Example
Suppose a sample of size 100 from an AR(1) process gives $r_1 = 0.638$. Are the data consistent with the hypothesis that $\alpha = 0.7$?

Answer
For AR(1), $r_1 = \alpha$, a 95% C.I. for $\rho_1$ is $r_1 \pm 1.96 \sqrt{\text{Var}(r_1)}$. But,

$$\text{Var}(r_1) = \frac{1}{N} \left[ \frac{(1 - \alpha^2) (1 + \alpha^2)}{1 - \alpha^2} - 2\alpha^2 \right].$$

Hence,

$$r_1 \pm 1.96 \cdot \sqrt{\frac{1}{N} (1 - \alpha^2)}.$$

Since the 95% C.I. covers 0.7, it is consistent.
A 95% confidence interval for $\mu = E(X_t)$ is given by

$$\bar{X}_t \pm 1.96 \cdot \sqrt{Var(\bar{X}_t)},$$

where $Var(\bar{X}_t) = \frac{\sigma^2 X}{N} \left[ 1 + 2 \sum_{k=1}^{N} \rho_k \right]$ for any time series.

A 95% confidence interval for $\rho_k$ is given by

$$r_k \pm 1.96 \cdot \sqrt{Var(r_k)},$$

where $r_k \sim N(\rho_k, Var(r_k))$.

- **PR:** $Var(r_k) = \frac{1}{N}$.
- **MA(q):** $Var(r_k) = \frac{1}{N} \cdot \left[ 1 + 2 \sum_{k=1}^{q} \rho_k^2 \right]$.
- **AR(1):** $Var(r_k) = \frac{1}{N} \left[ \frac{(1-\alpha^2k)(1+\alpha^2)}{1-\alpha^2} - 2k \cdot \alpha^2 \right]$.

Random Series, MA(q), MA(1), AR(1).
Estimating Parameters of an AR Process

1. Ordinary Regression Model. (Approximately, $\hat{\mu} = \bar{X}$)

$$X_t - \bar{X}_t = \alpha_1(X_{t-1} - \bar{X}_t) + \cdots + \alpha_p(X_{t-p} - \bar{X}_t) + Z_t.$$

1. AR($p$) is viewed as a multiple regression model.

   1. If $p = 1$ then $\hat{\alpha}_1 = \frac{\sum_{t=1}^{N-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{N-1}(X_t - \bar{X})^2} = \frac{\text{Cov}(X_t, X_{t+1})}{\text{Var}(X_t)} = r_1,$

      where $r_k = \text{sample autocorrelation function of lag } k,$ and

      $$r_k \xrightarrow{\text{estimates}} \rho_k.$$

   2. If $p = 2$ then $\hat{\alpha}_1 = \frac{r_1(1-r_2)}{1-r_1^2},$ and $\hat{\alpha}_2 = \frac{r_2-r_1^2}{1-r_1^2}.$

   3. If the true model is AR(1), then $r_2 \approx r_1^2$ since $\rho_2 = \rho_1^2 = \alpha_1^2.$ Thus, $\hat{\alpha}_1 \approx r_1$ and $\hat{\alpha}_2 \approx 0.$

Example

Fitting an AR model to the quarterly growth of the US-GNP (continued).

```r
model1 <- ar(gnp, method = 'mle')
model1
```
2 Yule-Walker Equations.

\[
\begin{align*}
\begin{pmatrix} R \\ \hat{\alpha} \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{\alpha} \end{pmatrix} = \begin{pmatrix} r \\ \hat{\alpha} \end{pmatrix} \Rightarrow \hat{\alpha} = R^{-1} r, \text{ where}
\end{align*}
\]

\[
R = \begin{pmatrix}
1 & r_1 & r_2 & \cdots & r_{p-1} \\
r_1 & 1 & r_1 & \cdots & r_{p-2} \\
r_2 & r_1 & 1 & \cdots & r_{p-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1
\end{pmatrix}, \quad \hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix}, \quad r = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix}
\]

\[
\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \cdots + \alpha_p \rho_{k-p}.
\]

\[
\begin{pmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\
\rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1
\end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix}
\]
Partial Autocorrelation Function

- When fitting an AR\( (p) \) model, the last coefficient \( \alpha_p \), denoted by \( \pi_p \), measures the excess correlation at lag \( p \), which is not accounted for by any AR\( (p-1) \) model.

- It is called the \( p \)-th partial autocorrelation coefficient, and when plotted against \( p \) gives the partial ac.f., \( r_p \) vs \( p \).

\[
\begin{align*}
AR(1) : \quad & X_t = \alpha X_{t-1} + Z_t, \quad \pi_1 = \alpha. \\
AR(2) : \quad & X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t, \quad \pi_2 = \alpha_2. \\
AR(3) : \quad & X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} + Z_t, \quad \pi_3 = \alpha_3.
\end{align*}
\]

- For AR(2): Given \( \rho_1, \rho_2 \) find \( \pi_1, \pi_2 \).
Partial Autocorrelation Function

Example

AR(1), $\rho_k = \alpha^k$, $\rho_1 = \alpha$.

$$\pi_1 = \rho_1 = \alpha.$$ 

What is $\pi_2 = \alpha_2$?

Yule-Walker: $\rho_k \Leftrightarrow \alpha$’s.

$$\alpha_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \pi_2.$$ 

If the real model is AR(1), i.e. $\rho_2 = \alpha^2 = \rho_1^2$, then $\alpha_2 = \pi_2 = 0$.

$$\pi_k = 0, \quad \text{for all } k > 1.$$
Partial Autocorrelation Function

Partial Autocorrelation Function for an AR(1), n=100

Partial Autocorrelation Function for an AR(2), n=100
Partial Autocorrelation Function

- $\pi_k$ is usually estimated by fitting AR process of successively higher order.
  - Taking $\hat{\pi}_1 = \hat{\alpha}_1$ when an AR(1) process is fitted.
  - Taking $\hat{\pi}_2 = \hat{\alpha}_2$ when an AR(2) process is fitted, and so on.
- Values of $\hat{\pi}_i$ which are outside the range $\pm \frac{2}{\sqrt{N}}$ are significantly different from zero at the 5% level.
- AR($p$): The partial ac.f. of an AR($p$) process “cuts off” at lag $p$, i.e. $\pi_{p+1}, \pi_{p+2}, \cdots$ are not significantly different from zero.
- Consider an AR(2) process that exhibits a pseudocyclic behavior

```r
pacf <- ARMAacf(ar=c(1.5,-0.75), ma=0.24, pacf=TRUE, lag.max=10)
plot(0:9, pacf, type='h', xlab='lag', ylim = c(-0.8,1))
lines(0:9, pacf, type='p')
abline(h=0)
```
Partial Autocorrelation Function

Example

Given, $\alpha_1$ and $\alpha_2$, find $\pi_1$ and $\pi_2$.

\[
AR(2) : \quad X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t.
\]

$\pi_2 = \alpha_2, \quad \pi_1 = \alpha$ in AR(1) and $\alpha = \rho_1$.

\[
\alpha_1, \alpha_2 \xrightarrow{\text{find}} \rho_1.
\]

Yule-Walker:

\[
\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2}.
\]

\[
\rho_1 = \alpha_1 \rho_0 + \alpha_2 \rho_1.
\]

\[
\rho_1 = \frac{\alpha_1}{1 - \alpha_2} = \pi_1.
\]
ACF: $\rho_k$ for MA($q$) process cuts off at lag $k = q$, i.e. $\rho_k = 0 \forall k > q$.

PACF: $\pi_k$ for AR($p$) process cuts off at lag $k = p$, i.e. $\pi_k = 0 \forall k > p$.

PACFs for AR(1) and AR(2) models up to lag 4.

1. AR(1):
   \[
   X_t = \alpha_1 X_{t-1} + Z_t,
   \]
   \[
   \pi_1 = \alpha_1, \pi_2 = 0, \quad \rightarrow \quad \{\alpha_1, 0, 0, 0\}.
   \]

2. AR(2):
   \[
   X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t,
   \]
   \[
   \pi_1 = \frac{\alpha_1}{1 - \alpha_2}, \pi_2 = \alpha_2, \pi_3 = 0, \quad \rightarrow \quad \left\{\frac{\alpha_1}{1 - \alpha_2}, \alpha_2, 0, 0\right\}.
   \]
The AR(1) and Economic and Financial Time Series (Review)

- The AR(1) model is a good description for the following time series
  - Interest rates.
  - Growth rate of macroeconomic variables.
    - Real GDP, industrial production.
    - Money, and velocity.
    - Real wages, and unemployment.

- Autoregressive processes (random walk) are useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock.

- Yule used an AR process to describe the phenomena of sunspot numbers and the behavior of a simple pendulum.

- The AR(1) process is sometimes called a Markov process because the value of $X_t$ is determined from $X_{t-1}$ and the present “shock,” but no earlier $X_t$’s are needed.
Many time series are nonstationary, but, due to equilibrium forces, different parts of the series behave very much alike except for the difference in the local mean level.

Box and Jenkins refer to this kind of nonstationary behavior as homogenous nonstationary.

The mean of such series changes in a random fashion, but the behavior of the series otherwise is independent of its level.
Analysis of El Nino and Fish Population

- For a period of 453 months (1950-1987) the recruitment index (number of new fish) in the Pacific ocean has been recorded.
- The associated SOI (Southern Oscillation Index) measures changes in atmospheric pressure, related to sea surface temperatures, in the central Pacific

```r
# Least squares fit
plot(rec, ylab='', xlab='', main='Recruitment)
acf2(rec, 48)
(regr = ar.ols(rec, order=2, demean=FALSE, intercept=TRUE))
regr@asy.se.coef

# Yule-Walker fit
rec.yw <- ar.yw(rec, order=2)
rec.yw$x.mean
rec.yw$ar
sqrt( diag(rec.yw$asy.var.coef) )
```
Information criteria: an alternative way of estimating the order of an AR($p$) process

- AIC (Akaike Information Criterion):

$$AIC = -\frac{2l}{T} + \frac{2p}{T},$$

where $p$ is the number of parameters, $T$ is the sample size and $l$ is the log-likelihood.

- For a Gaussian AR($p$) process,

$$AIC(l) = log\hat{\sigma}^2 + \frac{2p}{T},$$

where $\hat{\sigma}^2$ is the estimated error variance.

- For a Gaussian AR($p$) model, BIC (Bayesian Information Criterion)

$$BIC = log\hat{\sigma}^2 + p \frac{log(T)}{T}.$$  

Of the two, AIC has a tendency to overfit for small and moderate sample sizes.
Example

gnp <- scan('I:/pu.data/Desktop/Stat420/Datasets/dgnp82.txt', header=F, sep='')
ord <- ar(gnp, method='mle')
ord$aic
ord$order
Estimating MA($q$) processes

- Seemingly, a method of moments can be used ... .
- If $X_t = Z_t + \beta Z_{t-1}$, solve

$$\frac{\beta}{1 + \beta^2} = r_1$$

for $\beta$.
- Even for an invertible MA(1), it is possible to have $|r_1| > 0.5$.

```r
set.seed(2)
ma1 <- arima.sim(list(order=c(0,0,1), ma=0.9), n=50)
ma1
acf(ma1)
acf(ma1, plot=FALSE)[1]
```
- Can show that this estimator is grossly inefficient.