5 Forecasting

5.1 Introduction

Introduction

- **Forecasting**: The prediction problem from the observed values of a time series at past points $X_1, \cdots, X_t$ predict the value at some specific future time point $X_{t+h}$.

- Given $X_{t+h}$, $t$ is called the forecasting origin and the positive integer $h$ is the forecasting horizon.

- We refer to $\hat{X}_{t+h}$ as the $h$–step ahead forecast of $X_t$ at the forecast origin $t$.

- The accuracy of $\hat{X}_{t+h}$ is measured in terms of the smallness of the quantity $E\left(X_{t+h} - \hat{X}_{t+h}\right)^2$.

Mean squared risk minimization as a basis of forecasting

- $E\left(X_{t+h} - \hat{X}_{t+h}\right)^2$ is minimized by choosing $\hat{X}_{t+h}$ as the conditional mean of $X_{t+h}$, given the values of $X_1, \cdots, X_t$, i.e.

$$E\left(X_{t+h} \mid X_1 \cdots X_t\right).$$

- Thus, the forecast is the conditional expectation

$$\hat{X}_{t+h} = E\left(X_{t+h} \mid X_1, \cdots, X_t\right),$$

and the associated forecast error is

$$e_{t+h} = X_{t+h} - \hat{X}_{t+h}.$$ 

Types of Forecasts

1. **Subjective** - e.g. an expert opinion.

2. **Univariate** - the forecast $h$–steps ahead for the process $X_t$ depends only on the past values of $X_t$.

3. **Multivariate** - other predictors (explanatory variables) are used; those are not independent variables in the usual regression sense!

1. **Automatic forecast** - no human intervention.


3. Automatic forecasts are usually univariate.

4. Subjective and multivariate forecasts are always non-automatic.
Univariate Forecasting

- **Exponential smoothing**: for $0 < \alpha < 1$, we have

$$\hat{X}_N(1) = \sum_{j=1}^{\infty} \alpha(1-\alpha)^j X_{n-j}.$$

- Another representation ($e_N = X_N - \hat{X}_{N-1}(1)$):

$$\hat{X}_N(1) = \alpha X_N + (1-\alpha)\hat{X}_{N-1}(1) = \alpha e_N + \hat{X}_{N-1}(1).$$

- A famous exponential smoothing model is ARIMA(0,1,1):

$$X_t = X_{t-1} + Z_t - \alpha Z_{t-1}.$$

- The $\alpha$-values can be estimated using a simple minimization of squared errors procedure.

Selecting the parameter $\alpha$

- Use past data.

- For a specific $\alpha$, define one-step-ahead forecast errors recursively:

$$e_i = X_i - \hat{X}_{i-1}(1), \quad i = 1, 2, 3, \ldots.$$

- Compute $\sum_{i=2}^{N} e_i^2$ for different values of $\alpha$ and choose

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=2}^{N} e_i^2.$$

Example: (Exponential Smoothing)

```r
x <- arima.sim(list(order=c(0,1,1), ma=-0.8), n=100)
x.ima <- HoltWinters(x, beta=FALSE, gamma=FALSE)
plot(x)
plot(x.ima)
```

![Holt-Winters filtering](image.png)

Purdue University
General Holt-Winters Procedure

- Developed in 1960’s by C.C. Holt and P.R. Winters.
- For a series with no trend $T_t$, exponential smoothing represents updating of the local mean level of the series $L_t = \alpha X_t + (1 - \alpha)L_{t-1}$.
- In reality, a joint update is made possible by
  $$L_t = \alpha X_t + (1 - \alpha)(L_{t-1} + T_{t-1}).$$
  $$T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1}.$$  
- The $h$–step ahead forecast is
  $$\hat{X}_N(h) = (L_t + hT_t), \quad h = 1, 2, 3, \cdots.$$  
- Note that that the procedure now has two smoothing parameters.

1–Step Ahead Forecast of AR(1)
Consider the AR(1) model: $X_t = \alpha X_{t-1} + Z_t$. At time $t + 1$, we have
$$X_{t+1} = \alpha X_t + Z_{t+1}.$$  
So,
$$\hat{X}_{t+1} = E(X_{t+1} | X_t, \cdots, X_1)$$
$$= E(\alpha X_t | X_t, \cdots) + E(Z_{t+1} | X_t, \cdots).$$
But,
$$E(2X | X) = 2X,$$
and
$$E(Z_{t+1} | X_t, \cdots) = E(Z_{t+1}) = 0$$
because $Z_{t+1} \sim IDD(0, \sigma^2)$ has no relationship with $X_t$. Therefore,
$$\hat{X}_{t+1} = \alpha X_t.$$

1–Step Ahead Forecast of AR($p$)
Consider the AR($p$) model: $X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t$. At time $t + 1$, we have
$$X_{t+1} = \alpha_1 X_t + \cdots + \alpha_p X_{t-p+1} + Z_{t+1}.$$  
So,
$$\hat{X}_{t+1} = E(X_{t+1} | X_t, \cdots, X_1)$$
$$= E(\alpha_1 X_t + \cdots + \alpha_p X_{t-p+1} + Z_{t+1} | X_t, \cdots, X_1)$$
$$= \alpha_1 X_t + \cdots + \alpha_p X_{t-p+1},$$
and the associated forecast error is
$$e_{t+1} = X_{t+1} - \hat{X}_{t+1} = Z_{t+1}.$$  
Consequently, the variance of the 1–step ahead forecast error is
$$\text{Var}(e_{t+1}) = \text{Var}(Z_{t+1}) = \sigma_z^2.$$  
If $Z_t$ is normally distributed, then a 95% CI for the forecast (PI) is $\hat{X}_{t+1} \pm 1.96 \cdot \sigma$. 

Purdue University
2–Step Ahead Forecast of AR(1)
Consider the AR(1) model: \( X_t = \alpha X_{t-1} + Z_t \). At time \( t + 2 \), we have
\[
X_{t+2} = \alpha X_{t+1} + Z_{t+2}.
\]
So,
\[
\hat{X}_{t+2} = E(X_{t+2}|X_t, \ldots, X_1) = E(\alpha X_{t+1}|X_t \ldots) + E(Z_{t+2}).
\]
\[
= \alpha \cdot \hat{X}_{t+1} = \alpha \cdot \alpha X_t = \alpha^2 \cdot X_t.
\]

Exercise
Find \( E(e_{t+2}) \) and \( \text{Var}(e_{t+2}) \).

2–Step Ahead Forecast of AR(\( p \))
Consider the AR(\( p \)) model: \( X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t \). At time \( t + 2 \), we have
\[
X_{t+2} = \alpha_1 X_{t+1} + \cdots + \alpha_p X_{t-p+2} + Z_{t+2}.
\]
So,
\[
\hat{X}_{t+2} = E(X_{t+2}|X_t, \ldots, X_1) = E(\alpha_1 X_{t+1} + \cdots + \alpha_p X_{t-p+2}|X_t \cdots) + E(Z_{t+2}).
\]
\[
= \alpha_1 \hat{X}_{t+1} + \alpha_2 X_t + \cdots + \alpha_p X_{t-p+2}.
\]

Exercise
Find \( E(e_{t+2}) \) and \( \text{Var}(e_{t+2}) \).

Multiple \( h \)–Step Ahead Forecast of AR(\( p \))
In general, for the AR(\( p \)) model: \( X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t \), the \( h \)–step ahead forecast is explicitly calculated recursively as
\[
\hat{X}_{t+h} = \alpha_1 \hat{X}_{t+1} + \cdots + \alpha_p \hat{X}_{t+h-p},
\]
where \( \hat{X}_{t+h-i} \) are the observed values \( X_{t+h-i} \) for \( i \geq h \).
\[
h - i \leq 0 \quad \Rightarrow \quad \hat{X} \text{ is approximated by } X \text{ for } t, t-1, t-2, \cdots.
\]
Let \( h = 2 \), then
\[
\hat{X}_{t+2} = \alpha_1 \hat{X}_{t+1} + \alpha_2 \hat{X}_{t+h-1} + \cdots + \alpha_p \hat{X}_{t+h-p+2}.
\]
Summary

- First step ahead forecast:
  \[ AR(1) : \hat{X}_{t+1} = \alpha X_t. \]
  \[ AR(p) : \hat{X}_{t+1} = \alpha_1 X_t + \cdots + \alpha_p X_{t-p+1}. \]

- Second step ahead forecast:
  \[ AR(1) : \hat{X}_{t+2} = \alpha^2 X_t. \]
  \[ AR(p) : \hat{X}_{t+2} = \alpha_1 (\alpha_1 X_t + \cdots + \alpha_p X_{t-p+1}) + \alpha_2 X_t + \cdots + \alpha_p X_{t-p+2}. \]

- Multi step ahead forecast:
  \[ AR(p) : \hat{X}_{t+h} = \alpha_1 \hat{X}_{t+h-1} + \cdots + \alpha_p \hat{X}_{t+h-p}. \]
  \[ h = 2 : \hat{X}_{t+2} = \alpha_1 \hat{X}_{t+1} + \alpha_2 X_t + \cdots + \alpha_p X_{t-p+2}. \]

Forecasting Recruitment Series

library(astsa)
regr <- ar.ols (rec, order=2, demean=FALSE, intercept=TRUE)
fore <- predict(regr, n.ahead=24)
plot(rec, fore$pred, col=1:2, xlim=c(1980,1990), ylab='Recruitment')
lines(fore$pred, type='p', col=2)
lines(fore$pred+fore$se, lty='dashed', col=4)
lines(fore$pred-fore$se, lty='dashed', col=4)

Forecasting AR(1)

\[ X_t = \alpha_1 X_{t-1} + Z_t. \]
\[ \hat{X}_{t+1} = \alpha_1 X_t, \]
where \( t + h - 1 = t + 1 - 1 = t \).
\[ \hat{X}_{t+2} = \alpha_1 \hat{X}_{t+1} = \alpha_1^2 X_t, \]
where \( t + h - 1 = t + 2 - 1 = t + 1 \).
\[ \hat{X}_{t+3} = \alpha_1 \hat{X}_{t+2} = \alpha_1^3 X_t. \]
\[ \vdots \]
\[ \hat{X}_{t+h} = \alpha_1^h X_t. \]
The Residual of the Forecast $\hat{X}_{t+h}$ of an AR(1)

\[ e_{t+h} = X_{t+h} - \hat{X}_{t+h} = \alpha_1 X_{t+h-1} + Z_{t+h} - \alpha_1^h X_t. \]

\[ = \alpha_1 (\alpha_1 X_{t+h-2} + Z_{t+h-1}) + Z_{t+h} - \alpha_1^h X_t. \]

\[ = \cdots \]

\[ = \alpha_1^h X_t + \alpha_1^{h-1} Z_{t+1} + \alpha_1^{h-2} Z_{t+2} + \cdots + \alpha_1 Z_{t+h-1} + Z_{t+h} - \alpha_1^h X_t \]

\[ = \alpha_1^{h-1} Z_{t+1} + \alpha_1^{h-2} Z_{t+2} + \cdots + \alpha_1 Z_{t+h-1} + Z_{t+h}. \]

The Variance of the Residual $e_{t+h}$

\[
\text{Var}(e_{t+h}) = \text{Var} \left( \alpha_1^{h-1} Z_{t+1} + \alpha_1^{h-2} Z_{t+2} + \cdots + \alpha_1 Z_{t+h-1} + Z_{t+h} \right) \\
= \sigma^2 \left[ 1 + \alpha_1^2 + \cdots + \alpha_1^{2(h-2)} + \alpha_1^{2(h-1)} \right] \\
= \sigma^2 \frac{1 - \alpha_1^{2h}}{1 - \alpha_1^2}.
\]

Note that

\[
\sum_{k=1}^{n} a^{k-1} = 1 + a + a^2 + \cdots + a^{n-1} = \frac{1 - a^n}{1 - a}.
\]

This forecast exhibits the classic property of \textbf{mean reversion} where prices and returns eventually move back toward the mean or average.

Forecasting MA(1)

\[
MA(1) : \quad X_t = Z_t + \beta_1 Z_{t-1}, \quad (q = 1).
\]

\[
\hat{X}_{t+h} = \sum_{i=1}^{q=1} \beta_i Z_{t+h-i}, \quad i = q = 1.
\]

\[
\hat{X}_{t+1} = \beta_1 Z_t. \quad \hat{X}_{t+2} = \beta_1 \cdot Z_{t+1} \quad \text{not happened yet}
\]

Forecasting more than 1-step ahead results in the mean as the forecast due to the nature of MA(1) process.
The Residual of the Forecast $\hat{X}_{t+1}$ of an MA(1)

\[ e_{t+1} = X_{t+1} - \hat{X}_{t+1} = Z_{t+1} + \beta_1 Z_t - \beta_1 Z_t = Z_{t+1}. \]
\[ \text{Var}(e_{t+1}) = \text{Var}(Z_{t+1}) = \sigma_z^2. \]

A 95% PI for the forecast $X_{t+1}$ is

\[ \hat{X}_{t+1} \pm 1.96 \cdot \sigma \]
\[ e_{t+2} = X_{t+2} - \hat{X}_{t+2} = Z_{t+2} + \beta_1 Z_{t+1} - 0. \]
\[ \text{Var}(e_{t+2}) = (1 + \beta_1^2) \cdot \sigma_z^2. \]

A 95% PI for the forecast $X_{t+2}$ is

\[ \hat{X}_{t+2} \pm 1.96 \cdot \sigma_z \cdot \sqrt{1 + \beta_1^2} = \pm 1.96 \cdot \sigma_z \cdot \sqrt{1 + \beta_1^2}. \]

Forecasting MA(2)

\[ MA(2): \quad X_t = Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2}, \quad (q = 2). \]
\[ \hat{X}_{t+h} = \sum_{i=1}^{2} \beta_i \cdot Z_{t+h-i}. \]
\[ \hat{X}_{t+1} = \sum_{i=1}^{2} \beta_i \cdot Z_{t+1-i} = \beta_1 Z_t + \beta_2 Z_{t-1}. \]
\[ \hat{X}_{t+2} = \sum_{i=1}^{2} \beta_i Z_{t+2-i} = \beta_1 Z_{t+1} + \beta_2 Z_t = \beta_2 Z_t. \]

The Residual of the Forecasts $\hat{X}_{t+1}$ and $\hat{X}_{t+2}$ of an MA(2)

\[ e_{t+1} = X_{t+1} - \hat{X}_{t+1} = Z_{t+1} + \beta_1 Z_t + \beta_2 Z_{t-1} - \beta_1 Z_t - \beta_2 Z_{t-1} = Z_{t+1}. \]
\[ \text{Var}(e_{t+1}) = \text{Var}(Z_{t+1}) = \sigma_z^2. \]

\[ e_{t+2} = X_{t+2} - \hat{X}_{t+2} = Z_{t+2} + \beta_1 Z_{t+1} + \beta_2 Z_t - \beta_2 Z_t = Z_{t+2} + \beta_1 Z_{t+1}. \]
\[ \text{Var}(e_{t+2}) = (1 + \beta_1^2) \cdot \sigma_z^2. \]

A 95% PI for the forecast $X_{t+2}$ is

\[ \hat{X}_{t+2} \pm 1.96 \cdot \sigma_z \cdot \sqrt{1 + \beta_1^2}. \]
MA forecasting for the daily log-return of the value-weighted index

da = read.table('d-ibmvwewsp6203.txt')
dim(da)
vw <- log(1+da[,3])*100
acf(vw, lag.max=10)
m1 <- arima(vw, order=c(0,0,1))
m1
tsdia(m1)
predict(m1,5)

Forecasting AR(2)

\[ \hat{X}_{t+2} = \alpha_1 \hat{X}_{t+1} + \alpha_2 X_t = \alpha_1 (\alpha_1 X_t + \alpha_2 X_{t-1}) + \alpha_2 X_t. \]

\[ e_{t+1} = X_{t+1} - \hat{X}_{t+1} \]
\[ = \alpha_1 X_t + \alpha_2 X_{t-1} + Z_{t+1} - (\alpha_1 X_t + \alpha_2 X_{t-1}) \hat{X}_{t+1} = Z_{t+1}. \]
\[ Var(e_{t+1}) = \sigma_z^2. \]

\[ e_{t+2} = X_{t+2} - \hat{X}_{t+2} \]
\[ = \alpha_1 X_{t+1} + \alpha_2 X_t + Z_{t+2} - \hat{X}_{t+2}. \]
\[ = \alpha_1 (\alpha_1 X_t + \alpha_2 X_{t-1} + Z_{t+1}) + \alpha_2 X_t + Z_{t+2} - \alpha_1 (\alpha_1 X_t + \alpha_2 X_{t-1}) - \alpha_2 X_t, \]
\[ = \alpha_1 Z_{t+1} + Z_{t+2}. \]
\[ Var(e_{t+2}) = (1 + \alpha_1^2) \cdot \sigma_z^2. \]

Forecasting ARMA(p, q)

Consider the following ARMA(p, q) process:

\[ X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + Z_t + \sum_{i=1}^{q} \beta_i Z_{t-i}. \]

Forecasts are explicitly calculated recursively as

\[ \hat{X}_{t+h} = \sum_{i=1}^{p} \alpha_i \hat{X}_{t+h-i} + \sum_{i=1}^{q} \beta_i Z_{t+h-i}, \]

where \( \hat{X}_{t+h-i} \) are the observed values \( X_{t+h-i} \) for \( i \geq h \) and the \( Z \)'s that have not yet happened are replaced by zeroes.

\[ t^\text{th} - i > t \rightarrow 0, \quad i \geq h. \]
Forecasting ARIMA($p, d, q$)

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t.$$  

Define, $\phi^*(B) = \phi(B)(1 - B)^d$. Then,

$$X_t = \sum_{i=1}^{p+d} \alpha_i^* X_{t-i} + Z_t + \sum_{i=1}^{q} \beta_i Z_{t-i}.$$  

Example 1.

$$X_t = \frac{1.8}{\alpha_1^*} X_{t-1} + \frac{0.8}{\alpha_2^*} X_{t-2} + Z_t + \frac{0.3}{\beta_1} Z_{t-1}.$$  

Forecasting ARIMA($p, d, q$) for an ARIMA($p, d, q$) process is explicitly calculated recursively as

$$\hat{X}_{t+h} = \sum_{i=1}^{p+d} \alpha_i^* \hat{X}_{t+h-i} + \sum_{i=1}^{q} \beta_i Z_{t+h-i},$$

where $\hat{X}_{t+h-i}$ are the observed values of $X_{t+h-i}$ for $i \geq h$ and the Z’s that have not yet happened are replaced by zeroes.

Example 2.

$$\hat{X}_{t+1} = \sum_{i=1}^{2} \alpha_i^* \hat{X}_{t+1-i} + \sum_{i=1}^{1} \beta_i Z_{t+1-i}$$

$$= \alpha_1^* \hat{X}_t + \alpha_2^* X_{t-1} + \beta_1 Z_t$$

$$= 1.8X_t - 0.8X_{t-1} + 0.3Z_t.$$  

Forecasting ARIMA($p, d, q$)  

Example 3.

$$\hat{X}_{t+2} = \sum_{i=1}^{2} \alpha_i^* \hat{X}_{t+2-i} + \sum_{i=1}^{1} \beta_i Z_{t+2-i}$$

$$= \alpha_1^* \hat{X}_{t+1} + \alpha_2^* X_t + \beta_1 Z_{t+1}$$

$$= 1.8\hat{X}_{t+1} - 0.8X_t + 0$$

$$= 1.8(1.8X_t - 0.8X_{t-1} + 0.3Z_t) - 0.8X_t.$$  

$$= 2.44X_t - 0.8X_{t-1} + 0.3Z_t.$$