Forecasting with SARIMA Models

Maria Eduarda Silva

FEP.UP

Forecasting is estimating how the sequence of observations will continue into the future

Types of data:

- Time series data collected at regular intervals over time
- Cross-sectional data are for a single point in time

Time series models

Time series models use only information on the variable to be forecast

$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}, \ldots, error)$

where t is time and y_t is the quantity of interest at time t like: sales, electricity demand.

ARIMA models and exponential smoothing

- Useful when predictor variables not known or measured
- Useful if prediction of predictor variables difficult
- Does not lead to understanding of the system

Cross-sectional models

Cross-sectional models assume that variable to be forecast is affected by one or more **predictor variables**

$y = f(x_1, x_2, \ldots, error)$

where $x_1, x_2, \ldots$ are variables such as current temperature, GDP, population, time of the day, day of the week, etc

regression models

Mixed models

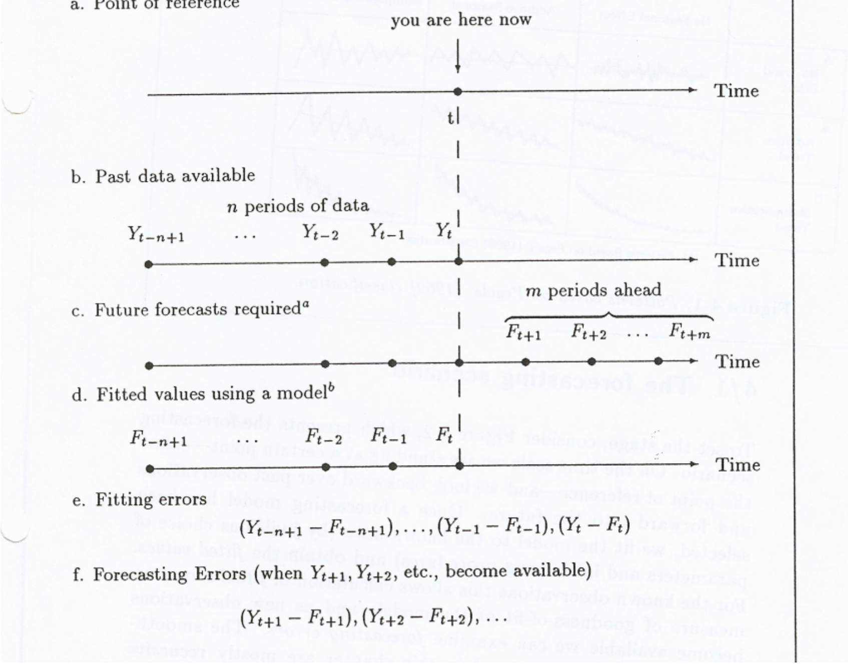
$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}, \ldots, x_1, x_2, \ldots, error)$
dynamic regression models, panel data models, longitudinal models, transfer function models.

Statistical Forecasting

- Thing to be forecasted: a random variable y
- Forecast distribution: if $\mathcal{F}$ represents all the observations, then $y|\mathcal{F}$ means *the random variable y given what we know in $\mathcal{F}$*
- The point forecast is the mean (or median) of $y|\mathcal{F}$
- The forecast variance is $\mathrm{var}(y|\mathcal{F})$
- A prediciton interval or interval forecast is a range of values of y with high confidence

Prediction and Forecasting

- Forecasting method: algorithm which produces a point forecast
- Statistical model: is a Data Generation Process may be used to
  - construct a probability distribution for y_{n+h} (from where one can obtain a point forecast)
  - construct confidence intervals for the forecasts



a. Point of reference
you are here now
Time
t
b. Past data available
n periods of data
Y_{t-n+1} $\dots$ Y_{t-2} Y_{t-1} Y_t
Time
c. Future forecasts required[a]
m periods ahead
F_{t+1} F_{t+2} $\dots$ F_{t+m}
Time
d. Fitted values using a model[b]
F_{t-n+1} $\dots$ F_{t-2} F_{t-1} F_t
Time
e. Fitting errors
$(Y_{t-n+1} - F_{t-n+1}), \dots, (Y_{t-1} - F_{t-1}), (Y_t - F_t)$
f. Forecasting Errors (when Y_{t+1}, Y_{t+2}, etc., become available)
$(Y_{t+1} - F_{t+1}), (Y_{t+2} - F_{t+2}), \dots$

Strategy

training set test

N T

- Split the series into (i) training set (ii) test set
- Choose a method/model
- Start the forecasting procedure /Estimate the model using the training set
- Produce forecasts for the test set to obtain the forecasting accuracy measures
- Optimize parameters or model
- Obtain forecasts

Definitions and notations

- Aim: given a time series with n observations, $y_1, y_2, \ldots, y_n$ forecast $y_{n+1}, y_{n+2}, \ldots, y_{n+h}$

- n is the origin, $n+h$ is the forecast horizon and h is the number of steps-ahead

- $\hat{y}_n(h)$ denotes de (point) forecast of y_{n+h}, $h = 1, 2, \ldots$

- $\hat{y}_n(h)$ is a function of $\underline{y} = (y_1, y_2, \ldots, y_n)$, $\hat{y}_n(h) = g(\underline{y})$

- Forecast error at i steps-ahead $e_n(i) = y_{n+i} - \hat{y}_n(i)$

- Criterion to compute $\hat{y}_n(h)$: minimize mean squared error $\sum_{i=1}^{h} (\hat{y}_n(i) - y_{n+i})^2$

Criterion: minimum squared error

c: $\boxed{E(X-c)^2}$ minimize $\boxed{E(X-\mu)^2}$

- If X is a r.v. with $E(X) = \mu$ and $V(X) = \sigma^2$ then $E(X-c)^2$ is minmum for $c = \mu$
- If Y is a r.v. and $h(X)$is a function of X then $E(Y - h(X))^2$ is minimum for $h(X) = E(Y|X)$

  $E(Y - E(Y|X))^2$ is minimum
- Then

$$\hat{y}_n(h) = \mathrm{E}(y_{n+h}|y_1, y_2, \ldots, y_n)$$

$\hat{Y}_n(1) = E(Y_{n+1}|Y_n, Y_{n-1}, \ldots, Y_1)$

$f_{Y_{n+1}|Y_n \ldots Y_1}(y)$

$Y_{n+1} \approx \sum_{j=0}^{n-1} \alpha_j Y_{n-j}$

Y_{n+h} $\hat{Y}_n(h) = E(Y_{n+h}|Y_n, Y_{n-1}, \ldots, Y_1)$

Criteria to evaluate forecasts

$e_n(h) = y_{n+h} - \hat{y}_n(h)$

- Scale dependent measures
  - Mean Error, **ME**, $(1/h)\sum_{i=1}^{h} e_n(i)$
  - Mean Absolute Error, **MAE**, $(1/h)\sum_{i=1}^{h} |e_n(i)|$
  - Mean Squared Error, **MSE**, $(1/h)\sum_{i=1}^{h}(e_n(i))^2$

- Relative measures: Mean Absolute Percentual Error, **MAPE**, $(1/h)\sum_{i=1}^{h} |e_n(i)/y_{n+i}| \times 100$

  $\frac{1}{h}\sum_{i=1}^{h} \frac{|y_{n+i} - \hat{y}_n(i)|}{y_{n+i}} \times 100$

- Scale independent measures: Mean Absolute Scaled Error, **MASE** (Hyndman and Koehler, 2006)

  $\hat{y}_n(1) = y_n$

  Define the scaled error as:

$$q_n(i) = \frac{e_n(i)}{1/(n-1)\sum_{i=2}^{n} \underbrace{|y_i - y_{i-1}|}_{\text{forecast error from naive forecast}}}$$

  MASE=mean($|q_n(i)|$)

Cont.

Theil´s U-statistic

$$U_1 = \frac{\sqrt{\frac{1}{h}\sum_{i=1}^{h} e_n(i)^2}}{\sqrt{\frac{1}{h}\sum_{i=1}^{h} y_{n+i}^2} + \sqrt{\frac{1}{h}\sum_{i=1}^{h} \hat{y}_n^2(i)}}$$

takes values between 0 and 1 and values near 0 indicate an higher precision in the prediction.

Y_t AR(1) $\quad Y_t = a Y_{t-1} + e_t \quad e_t \sim N(0, \sigma_e^2)$

$Y_1, \dots, Y_T \quad \widehat{Y_{T+1}}\ ? \quad Y_t \sim N\left(0, \frac{\sigma_e^2}{1+a^2}\right)$

$$\widehat{Y_{T+1}} = \hat{Y}_T(1) = E\left(Y_{T+1} \mid Y_T, Y_{T-1}, \dots, Y_1\right)$$
$$= E\left(a Y_T + e_{T+1} \mid Y_T, Y_{T-1}, \dots, Y_1\right)$$
$$= a E(Y_T \mid Y_T, Y_{T-1}, \dots, Y_1) + E\left(e_{T+1} \mid Y_T, Y_{T-1} \dots Y_1\right)$$
$$= a Y_T + 0 = a Y_T$$

$$e_T(1) = Y_{T+1} - \hat{Y}_T(1) = Y_{T+1} - a Y_T = e_{T+1}$$
$$E\left(e_T(1)\right) = E\left(e_{T+1}\right) = 0$$
$$V\left(e_T(1)\right) = V\left(e_{T+1}\right) = \sigma_e^2$$
$$\underset{\substack{\shortparallel \\ a Y_T}}{\hat{Y}_T(1)} \sim N\left(0, V(\hat{Y}_T(1)\right)$$

$\widehat{Y_{T+2}}$? $\hat{Y}_T(2) = E(Y_{T+2} \mid Y_T, Y_{T-1}, \dots, Y_1)$

$= E(a Y_{T+1} + e_{T+2} \mid Y_T, \dots, Y_1)$

$= a E(Y_{T+1} \mid Y_T \dots Y_1) + E(e_{T+2} \mid Y_T, \dots, Y_1)$

$= a \hat{Y}_T(1) + 0$

$= a^2 Y_T$

$\hat{Y}_T(h) = E(Y_{T+h} \mid Y_T, \dots, Y_1) = E(a Y_{T+h-1} + e_{T+h} \mid Y_T, \dots, Y_1)$

$= a \hat{Y}_T(h-1) = \underset{|a|<1}{a^h Y_T} \xrightarrow[h \to \infty]{} 0$ as h increases the AR(1) forecast tends to the mean of process

$\hat{e}_T(2) = Y_{T+2} - \hat{Y}_T(2)$

$= a Y_{T+1} + e_{T+2} - a \hat{Y}_T(1) = a(Y_{T+1} - \hat{Y}_T(1)) + e_{T+2}$

$= a \hat{e}_T(1) + e_{T+2} = a e_{T+1} + e_{T+2}$

$E(\hat{e}_T(2)) = 0 \Rightarrow E(\hat{Y}_T(2)) = Y_{T+2}$ ✓

• One-step-ahead forecast errors $e_T(1) = e_{T+1}$ are independent

$e_{T-2}(1) = e_{T-1}$ $\quad$ $e_{T-1}(1) = e_T$ $\quad$ $e_T(1) = e_{T+1}$

$e_T(1)$ $\quad$ $e_T(2)$ are not indep

$$e_T(1) = e_{T+1} \qquad e_T(2) = a\, e_{T+1} + e_{T+2}$$

Forecasting with AR(1)

- $y_t = ay_{t-1} + e_t$, $|a| < 1$ $e_t \sim N(0, \sigma^2)$
- Forecast y_{n+h} given $y_1, \ldots, y_n$
- Note:

$$
\begin{aligned}
y_{n+1} &= ay_n + e_{n+1} \\
y_{n+2} &= ay_{n+1} + e_{n+2} = a(ay_n + e_{n+1}) + e_{n+2} \\
&= a^2 y_n + ae_{n+1} + e_{n+2} \\
y_{n+3} &= ay_{n+2} + e_{n+3} = a(a^2 y_n + ae_{n+1} + e_{n+2}) + e_{n+3} \\
&= a^3 y_n + a^2 e_{n+1} + ae_{n+2} + e_{n+3} \\
\vdots &\quad \vdots \\
y_{n+h} &= a^h y_n + a^{h-1} e_{n+1} + \ldots + e_{n+h} \\
&= a^h y_n + \sum_{j=0}^{h-1} a^j e_{n+h-j}
\end{aligned}
$$

then h steps-ahead forecast is:

$$\begin{aligned}
\hat{y}_n(h) &= \mathrm{E}(y_{n+h}|y_n) \\
&= \mathrm{E}(a^h y_n + \sum_{j=0}^{h-1} a^j e_{n+h-j})|y_n \\
&= a^h y_n + \mathrm{E}(\sum_{j=0}^{h-1} a^j e_{n+h-j}|y_n) \\
&= a^h y_n + \sum_{j=0}^{h-1} a^j \mathrm{E}(e_{n+h-j}|y_n) \\
&= a^h y_n
\end{aligned}$$

Note that the forecast tends to the mean of the process!

Prediction error

- $\hat{e}_n(h) = (\hat{y}_n(h) - y_{n+h})$ h-steps ahead prediction with origin at n
- $$\begin{aligned} \hat{e}_n(h) &= y_{n+h} - \hat{y}_n(h) \\ &= a^h y_n - a^h y_n + \sum_{j=0}^{h-1} a^j e_{n+h-j} \\ &= \sum_{j=0}^{h-1} a^j e_{n+h-j} \end{aligned}$$

and the variance of the prediction error

- Note that $\mathrm{E}(\hat{e}_n(h)) = 0$
-

$$\begin{aligned}
\mathrm{var}(\hat{e}_n(h)) &= \mathrm{var}(\hat{y}_n(h) - y_{n+h}) \\
&= \mathrm{var}(\sum_{j=0}^{h-1} a^j e_{n+h-j})
\end{aligned}$$

but e_t are iid, then

-

$$\mathrm{var}(\hat{e}_n(h)) = \sigma_e^2 \sum_{j=0}^{h-1} a^{2j}$$

Interpretation

- For $h = 1$ $\hat{e}_n(1) = e_{n+1}$ that is the reason why e_t are called innovations: e_t represents the news or surprise at each time period, the quantity that is not predictable.
- The forecasts are unbiased: $\mathrm{E}(\hat{y}_n(h)) = y_{n+h}$
- The variance of the prediction error depends on: a, σ_e^2 and the horizon h
- The variance of the prediction error increases with h.
- The variance of the prediction error tends to the value $\sigma_e^2/(1-a^2)$ which is the same as the variance of the process.
- In practice a and σ_e^2 are replaced by its estimated values $\hat{a}$ and $\hat{\sigma}_e^2$, respectively.
- Assuming $e_t \sim N(0, \sigma_e^2)$ a $\gamma\%$ confidence interval is given by: $\hat{y}_n(h) \pm 2\sqrt{\frac{(1-\hat{a}^{2k})}{(1-\hat{a}^2)}\hat{\sigma}_e^2}$

Forecasting with AR(p): notation

- $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + e_t$ where
  - $e_t \sim N(0,1)$
  - $a_1, a_2, \ldots, a_p$ are such that the roots $z_1, z_2, ldots, z_p$ of the AR polinomial $\phi(z) = 1 - a_1 Z - a_2 Z^2 - \ldots - a_p Z^p$ are $|z_i| > 1$
- Having observed $y_1, \ldots, y_n$ forecast $y_{n+1}, y_{n+2}, \ldots, y_{n+k}$
- Let $\hat{y}_n(k) = \mathrm{E}(y_{n+k} | y_1, \ldots, y_n)$

Forecast with AR(p): 1 step-ahead

$$
\begin{aligned}
\hat{y}_n(1) &= a_1 y_n + a_2 y_{n-1} + \ldots + a_p y_{n+1-p} \\
\\
\hat{e}_n(1) &= y_{n+1} - \hat{y}_n(1) \\
&= a_1 y_n + a_2 y_{n-1} + \ldots + a_p y_{n+1-p} + e_{n+1} \\
&\quad -\left(a_1 y_n + a_2 y_{n-1} + \ldots + a_p y_{n+1-p} + e_{n+1}\right) \\
&= e_{n+1} \\
\\
\mathrm{var}(\hat{e}_n(1)) &= \sigma_e^2
\end{aligned}
$$

Forecast with AR(p): 2 steps-ahead

$$
\begin{aligned}
\hat{y}_n(2) &= \mathrm{E}(a_1 y_{n+1} + a_2 y_n + \ldots + a_p y_{n+2-p} + e_{n+2} | y_1, \ldots, y_n) \\
&= a_1 \hat{y}_n(1) + a_2 y_n + \ldots + a_p y_{n+2-p}
\end{aligned}
$$

$$
\begin{aligned}
\hat{e}_n(2) &= y_{n+2} - \hat{y}_n(2) \\
&= a_1 y_{n+1} + a_2 y_n + \ldots + a_p y_{n+2-p} + e_{n+2} \\
&\quad -(a_1 \hat{y}_n(1) + a_2 y_n + \ldots + a_p y_{n+2-p}) \\
&= a_1 \hat{e}_n(1) + e_{n+2}
\end{aligned}
$$

$$
\mathrm{var}(\hat{e}_n(2)) = (1 + a_1^2)\sigma_e^2
$$

Then for k-steps ahead:

$$\begin{aligned}\hat{y}_n(h) &= \mathrm{E}(a_1 y_{n+h-1} + a_2 y_{n+h-2} + \ldots + a_p y_{n+h-p} + e_{n+h} | y_1, \ldots, y_n) \\ &= a_1 \hat{y}_n(h-1) + a_2 \hat{y}_n(h-2) + \ldots + a_p \hat{y}_n(h-p)\end{aligned}$$

where

$$\hat{y}_n(j) = \mathrm{E}(y_{n+j} | y_1, \ldots, y_n) = \begin{cases} y_{n+j}, & 1 \leq n+j \leq n, \\ 0, & n+j \leq 0 \end{cases}$$

and the prediction error and its variance are:

$$\begin{aligned}\hat{e}_n(h) &= y_{n+h} - \hat{y}_n(h) \\ &= a_1\hat{e}_n(h-1) + a_2\hat{e}_n(h-2) + \ldots + a_p\hat{e}_n(h-p)\end{aligned}$$

where

$$\hat{e}_n(j) = \begin{cases} e_{n+j}, & 1 \leq n+j \leq n, \\ 0, & n+j \geq 0 \end{cases}$$

NOTE: for fixed n $\hat{e}_n(h_1)$ and $\hat{e}_n(h_2)$ $h_1 \neq h_2$ are correlated!!!
To compute $\mathrm{var}(\hat{e}_n(h))$, $\hat{e}_n(h)$ is written as a function of $e_{n+1}, \ldots, e_{n+h}$

Example: MA(1)

$y_1 \cdots y_n \quad y_{n+h}?$

$\hat{y}_n(h) = E(y_{n+h} \mid y_n \cdots y_1)$

Forecast $h = 2$ steps-ahead $X_t = 0.4e_{t-1} + e_t$, $\sigma_e^2 = 1.5$ Note that

$\sum_{j=0}^{\infty} \pi_j y_{t-j} = e_t$

$Y_t = e_t + b\, e_{t-1}$ $\quad e_t$ iid $\quad |b|<1 \quad Y_1 \cdots Y_n$

$$\hat{Y}_n(1) = E\left(Y_{n+1} \mid Y_1 \cdots Y_n\right)$$

$$= E\left(e_{n+1} + b\, e_n \mid Y_1 \cdots Y_n\right) =$$

$$\sum |\pi_i| < \infty \qquad \pi_i \xrightarrow[i\to\infty]{} 0$$

$$\sum_{i=0}^{\infty} \pi_i Y_{t-i} = e_t$$

$$Y_t + \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \cdots = e_t$$

$$\hat{Y}_n(1) = E\left(e_{n+1} - \sum_{i=1}^{\infty} \pi_i Y_{n+1-i} \mid Y_1 \cdots Y_n\right)$$

$$= 0 - E\left(\sum_{i=1}^{\infty} \pi_i Y_{t-i} \mid Y_1 \cdots Y_n\right) \approx \sum_{i=1}^{n} \pi_i Y_{t-i}$$

$$Y_{n+1} = -\pi_1 Y_n - \pi_2 Y_{n-1} - \cdots + e_{n+1}$$

$$\hat{Y}_n(1) = -\pi_1 Y_n - \pi_2 Y_{n-1} - \cdots$$

AR(1)

$$X_t = a X_{t-1} + e_t$$

$$X_1 \cdots X_n$$

$$\hat{X}_n(1) = E\left(X_{n+1} \mid X_n \cdots X_1\right)$$

$$= E\left(a X_n + e_{n+1} \mid X_n \cdots X_1\right)$$

$$= a X_n + \underbrace{E\left(e_{n+1} \mid X_n, \ldots, X_1\right)}_{0}$$

$$= a X_n$$

$$e_n(1) = Y_{n+1} - \hat{Y}_n(1)$$
$$= \sum_{i=0}^{\infty} \pi_i Y_{n+1-i} - \sum_{i=0}^{\infty} \pi_i \cdot Y_{n+1-i} = e_{n+1}$$

$$\hat{X}_n(2) = E\left(Y_{n+2} \,\middle|\, Y_n, Y_{n-1}, \dots, Y_1 \right)$$
$$= E\left(e_{n+2} - \pi_1 Y_{n+1} - \pi_2 Y_n - \dots \,\middle|\, Y_n, Y_{n-1}, \dots Y_1 \right)$$
$$= 0 - \pi_1 \hat{Y}_n(1) - \pi_2 Y_n - \dots \pi_n Y_1$$

```
ARMAtoMA(ar=-0.4,lag.max=50)
 [1] -4.000000e-01  1.600000e-01 -6.400000e-02  2.560000e-02 -1.024000e-02
 [9] -2.621440e-04  1.048576e-04 -4.194304e-05  1.677722e-05 -6.710886e-06
[17] -1.717987e-07  6.871948e-08 -2.748779e-08  1.099512e-08 -4.398047e-09
[25] -1.125900e-10  4.503600e-11 -1.801440e-11  7.205759e-12 -2.882304e-12
[33] -7.378698e-14  2.951479e-14 -1.180592e-14  4.722366e-15 -1.888947e-15
[41] -4.835703e-17  1.934281e-17 -7.737125e-18  3.094850e-18 -1.237940e-18
[49] -3.169127e-20  1.267651e-20
```

```
> ARMAtoMA(ar=-0.3,ma=-0.4,lag.max=50)
 [1] -7.000000e-01  2.100000e-01 -6.300000e-02  1.890000e-02 -5.670000e-03
 [9] -4.592700e-05  1.377810e-05 -4.133430e-06  1.240029e-06 -3.720087e-07
[17] -3.013270e-09  9.039811e-10 -2.711943e-10  8.135830e-11 -2.440749e-11
[25] -1.977007e-13  5.931020e-14 -1.779306e-14  5.337918e-15 -1.601375e-15
[33] -1.297114e-17  3.891342e-18 -1.167403e-18  3.502208e-19 -1.050662e-19
[41] -8.510366e-22  2.553110e-22 -7.659329e-23  2.297799e-23 -6.893396e-24
[49] -5.583651e-26  1.675095e-26
```

ARMA(1,1)- forecasts based on the infinite past

- $y_t = ay_{t-1} + be_{t-1} + e_t$ where $|a| < 1$ e $|b| < 1$
- Having observed $y_1, \ldots, y_n$ forecast $y_{n+1}, y_{n+2}, \ldots, y_{n+h}$
- $\hat{y}_n(1) = a\mathrm{E}(y_n|X1, \ldots, y_n) + b\mathrm{E}(e_n|X1, \ldots, y_n)$
- Note that
  - $y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$
  - $\sum_{j=0}^{\infty} \pi_j y_{t-j} = e_t$

  then the knowledge of $y_n, y_{n-1}, \ldots$ is equivalent to the knowledge $e_n, e_{n-1}, \ldots$.

cont

$$\hat{y}_n(1) = E(y_{n+1} \mid y_n \cdots y_1)$$
$$= E(a y_n + b e_n + e_{n+1} \mid y_n \cdots y_1)$$
$$= a y_n + b e_n + 0$$

$$e_t(1) = X_{t+1} - \hat{X}_t(1) = e_{t+1}$$
$$\Downarrow$$
$$e_n = X_n - \hat{X}_{n-1}(1) = e_{n-1}(1)$$

Thus

$$\begin{aligned}
\hat{y}_n(1) &= a y_n + b e_n && (1)\\
\hat{y}_n(h) &= a\hat{y}_n(h-1),\ \ h \geq 2 && (2)\\
\hat{e}_n(1) &= e_{n+1} && (3)\\
\mathrm{var}(\hat{e}_n(1)) &= \sigma_e^2 && (4)\\
\hat{e}_n(k) &= a\hat{e}_n(k-1) + b e_{n+k-1} + e_{n+k} && (5)
\end{aligned}$$

When computing $\mathrm{var}(\hat{e}_n(k))$ one must keep in mind that with n fixed the

prediction errors are dependent.

ARMA(1,1)- forecasts based on $y_1, \ldots, y_n$

In fact we have observed only $y_1, \ldots, y_n$ thus let

$$\mathrm{E}(e_k | X1, \ldots, y_n) = \hat{e}_n(-n+k),\ 1 \leq k \leq n$$

Since $e_t = y_t - ay_{t-1} - be_{t-1}$, we have:

$$\begin{aligned}
\hat{e}_n(-n+1) &= 0 \\
\hat{e}_n(-n+2) &= y_2 - ay_1 \\
\hat{e}_n(-n+3) &= y_3 - ay_2 - b\hat{e}_n(-n+2) \\
\vdots &\quad \vdots \\
\hat{e}_n(0) &= y_n - ay_{n-1} - b\hat{e}_n(1)
\end{aligned}$$

cont.

As such

$$\begin{aligned}\hat{X}_n(1) &= a y_n + b \hat{e}_n(0)\\ \hat{X}_n(h) &= a \hat{X}_n(h-1),\ \ h \geq 2\end{aligned}$$

$\mathrm{var}(\hat{e}_n(h)),\ h \geq 1$ is computed using (5)

Forecasting with ARMA(p, q)

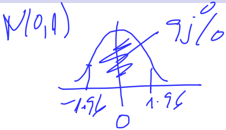
$$\begin{aligned}\hat{y}_n(h) &= a_1\hat{y}_n(h-1) + \ldots + a_p\hat{y}_n(h-p) + \\ &\quad b_1\mathrm{E}(e_{n+h-1}|y_n,\ldots) + \ldots + b_q\mathrm{E}(e_{n+h-q}|y_n,\ldots)\end{aligned}$$

where

$$\mathrm{E}(e_{n+k}|y_n,\ldots) = \hat{e}_n(k) = \begin{cases} e_{n+k}, & 1 \leq n+k \leq n, \\ 0, & n+j \geq 0 \end{cases}$$

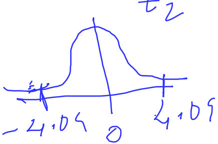
and $\hat{X}_n(k) = y_{n-k}$ if $-(p-1) \leq k \leq 0$
For $h > q$the forecasts are in fact obtained from the AR component of the ARMA model.

Confidence Intervals for the forecasts



95%

Assuming $e_t \sim N(0, \sigma_e^2)$ a $\cancel{\gamma}$% confidence interval is given by:



$$\hat{y}_n(h) \pm 2\sqrt{\mathrm{var}(\hat{e}_n(h))}$$

$\boxed{\text{Bootstraping}}$

Equivalences

Simple exponential smoothing

- Forecasts equivalent to **ARIMA(0,1,1)**.
- Parameters: $\theta_1 = \alpha - 1$.

Holt's method

- Forecasts equivalent to **ARIMA(0,2,2)**.
- Parameters: $\theta_1 = \alpha + \beta - 2$ and $\theta_2 = 1 - \alpha$.

Damped Holt's method

- Forecasts equivalent to **ARIMA(1,1,2)**.
- Parameters: $\phi_1 = \phi$, $\theta_1 = \alpha + \phi\beta - 2$, $\theta_2 = (1 - \alpha)\phi$.

Holt-Winters' additive method

- Forecasts equivalent to **ARIMA(0,1,m+1)(0,1,0)$_\mathbf{m}$**.
- Parameter restrictions because ARIMA has $m + 1$ parameters whereas HW uses only three parameters.

Holt-Winters' multiplicative method

- No ARIMA equivalence

ETS
Models

Forecasting with SARIMA Models

- Expand the SARIMA equation so that y_t is on the left hand side and all other terms are on the right
- Rewrite the equation by replacing t by $T + h$
- On the right hand side of the equation, replace future observations by their forecasts, future errors by zero and past errors by the corresponding residuals

Begin with $h = 1$ and repeat the above steps are then repeated for $h = 2, 3, \ldots$ until all forecasts have been calculated.

SARIMA $(1,1,1) \times (0,1,1)_{12}$

$a_1 = 0.19$

$m_1 = -0.56$

$sma = -0.86$

$$(1-B^{12}) \times (1-B)(1 - 0{,}19B) X_t = (1-0{,}56B)(1-0{,}86B^{12}) e_t$$

$$(1 - B - B^{12} + B^{13})(1 - 0{.}19B) X_t = (1 - 0{.}86B^{12} - 0{.}56B + 0{,}56 \times 0{.}86 B^{13}) e_t$$

$$(1 - B - B^{12} + B^{13} - 0{\cdot}19B + 0{\cdot}19B^2 + 0{\cdot}19B^{13} - 0{\cdot}19B^{14}) X_t = \text{"}$$

$$(1 - 1{,}19B + 0{.}19B^2 - B^{12} + 1{,}19B^{13} - 0{\cdot}19B^{14}) X_t =$$

$$X_t = 1{.}19X_{t-1} - 0{\cdot}19X_{t-2} + X_{t-12} - 1{,}19X_{t-13} + 0{\cdot}19X_{t-14} + e_t +$$
$$- 0{,}56e_{t-1} - 0{\cdot}86e_{t-12} + 0{,}48e_{t-13}$$

$$\hat{X}_m(1) = E(* \mid Y_1, \ldots Y_m)$$
$$= 1{,}19X_m - 0{\cdot}19X_{m-1} + X_{m-11} - 1{.}19X_{m-12} + 0{\cdot}19X_{m-13}$$
$$+ 0 - 0{\cdot}56e_m - 0{\cdot}86e_{m-11} + 0{,}48e_{m-12}$$

$$\hat{e}_t = e_{t-1}(1)$$