Time series analysis:
ARIMA models
Exponential Smoothing

Forecasting accuracy

Let us define a forecasting error $e_t = Y_t - F_t$.
We may then define some forecasting accuracy measures:
Mean Error, Mean Absolute Error, Mean Squared Error

$$\mathsf{ME} = \frac{1}{n}\sum_{t=1}^{n} e_t$$

$$\mathsf{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

$$\mathsf{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2.$$

Forecasting accuracy

The value of ME, MAE, MSE depend on the scale of data.
This makes difficult to compare different models.
We may define the percentage error and related measures.

$$\mathsf{PE}_t = \frac{Y_t - F_t}{Y_t} 100$$

$$\mathsf{MPE} = \frac{1}{n} \sum_{t=1}^{n} \mathsf{PE}_t$$

$$\mathsf{MAPE} = \frac{1}{n} \sum_{t=1}^{n} |\mathsf{PE}_t|$$

Short-term forecasting: simple exponential smoothing

If $\alpha = 0$ then $F_{t+1} = F_t \rightarrow$ Naive Model

The simple exponential smoothing is defined as

observed data at time t

forecast at time t

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$

and

where α is a constant term taking values between 0 e 1.
The new forecast F_{t+1} is the old forecast F_t with an adjustment.

Short-term forecasting: simple exponential smoothing

If $\alpha = 1$ $F_{t+1} = Y_t$

An equivalent way to express the simple exponential smoothing is

$$F_{t+1} = \alpha Y_t + (1 - \alpha) F_t \quad \text{— } 1$$

The new forecast F_{t+1} is a weighted average of the last observation, Y_t, and the last forecast, F_t.

$F_t = \alpha Y_{t-1} + (1-\alpha) F_{t-1}$

Sub F_t in 1

$F_{t+1} = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + (1-\alpha)^2 F_{t-1}$

Short-term forecasting: simple exponential smoothing

Why exponential smoothing?

$$\begin{aligned} F_{t+1} &= \alpha Y_t + (1-\alpha)[\alpha Y_{t-1} + (1-\alpha)F_{t-1}] \\ &= \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + (1-\alpha)^2 F_{t-1} \end{aligned}$$

so that we obtain

$$\begin{aligned} F_{t+1} &= \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} \\ &\quad + \alpha(1-\alpha)^3 Y_{t-3} + \ldots + \alpha(1-\alpha)^{t-1} Y_1 + (1-\alpha)^t F_1 \end{aligned}$$

coefficients are decaying exponential

Short-term forecasting: simple exponential smoothing

Initialization of the process

$$F_2 = \alpha Y_1 + (1 - \alpha) F_1$$

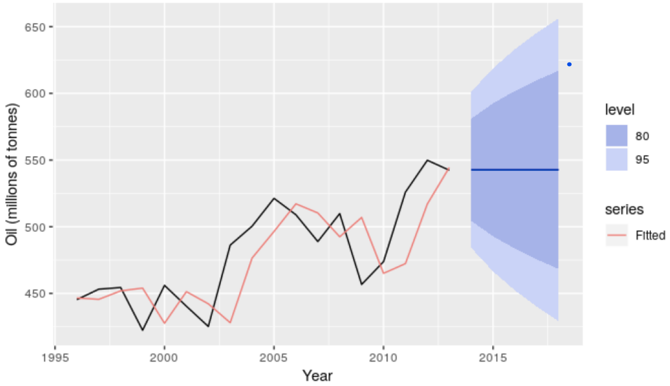
Since F_1 is not available, typically we use the first observation, $Y_1 = F_1$.

$$F_2 = Y_1$$

Short-term forecasting: simple exponential smoothing

- A crucial point in exponential smoothing concerns choosing a suitable value for α.
- A higher value for α is more sensitive to a change in the data structure, while a lower value generates a 'flat' forecast.
- A suitable selection for α is that minimizing the MSE.

Example



Oil production in Saudi Arabia: forecasting with $\hat{\alpha} = 0.83$

simple model with just one parameter α.

Short-term forecasting: Holt's exponential smoothing

- The Holt's linear trend method is a useful extension to allow the forecasting of data with a trend.
- This method involves a forecast equation and two smoothing equations (one for the level and one for the trend)

double smoothing

$$
\begin{aligned}
L_t &= \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \\
b_t &= \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \\
F_{t+m} &= L_t + b_t m
\end{aligned}
$$

$m=1 \quad F_{t+1} = L_t + b_t$

$m=2 \quad F_{t+2} = L_t + b_t \times 2$

straight lines

L_t denotes an estimate of the level of the series at time t and b_t an estimate of the slope t.

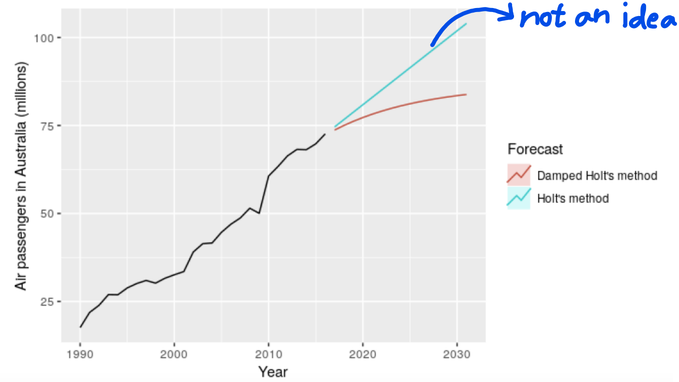
- This exponential smoothing is a double smoothing.
- The forecast function is no longer flat but trending.
- The m-step-ahead forecast is equal to the last estimated level (L_t) plus times the last estimated trend value. (b_t)
- Hence the forecasts are a linear function of m.

Short-term forecasting: Damped trend methods

- The forecasts generated by Holt's linear method display a constant trend (increasing or decreasing) indefinitely into the future.
- Empirical evidence indicates that these methods tend to over-forecast, especially for longer forecast horizons.
- A useful extension includes a damping parameter $0 < \phi < 1$

$$
\begin{aligned}
L_t &= \alpha Y_t + (1-\alpha)(L_{t-1} + \phi b_{t-1}) \\
b_t &= \beta(L_t - L_{t-1}) + (1-\beta)\phi b_{t-1} \\
F_{t+m} &= L_t + (\phi + \phi^2 + \cdots + \phi^h) b_t
\end{aligned}
$$

Example



not an ideal situation

Airline passengers:
Holt's and Dampened Holt's exponential smoothing $\phi = 0.90$

Holt's winters considers both trend and seasonality

ARIMA models: introduction

Chapter 8 of principles of forecasting – Robhyndman.

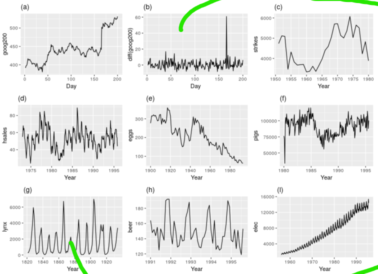
- ARIMA models provide a typical approach to time series forecasting.
- Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting, and provide complementary approaches to the problem.
- While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

Stationarity and differencing

A stationary time series is one whose properties do not depend on the time at which the series is observed.

Thus, time series with trends, or with seasonality, are not stationary.

stationary



Which of these series are stationary? (a) Google stock price for 200 consecutive days; (b) Daily change in the Google stock price for 200 consecutive days; (c) Annual number of strikes in the US; (d) Monthly sales of new one-family houses sold in the US; (e) Annual price of a dozen eggs in the US (constant dollars); (f) Monthly total of pigs slaughtered in Victoria, Australia; (g) Annual total of lynx trapped in the McKenzie River district of north-west Canada; (h) Monthly Australian beer production; (i) Monthly Australian electricity production.

Differencing

applied to eliminate or reduce trend or seasonality of time series

- Differencing can help stabilise the mean of a time series by removing changes in the level of a time series, and therefore eliminating (or reducing) trend and seasonality.

$$y'_t = y_t - y_{t-1}.$$

- Seasonal differencing (for monthly data)

$$y'_t = y_t - y_{t-12}.$$

because of monthly data
Ex: 4 in case of quatenly data.

- A further differencing may be performed

$$y^*_t = y'_t - y'_{t-1} = (y_t - y_{t-12}) - (y_{t-1} - y_{t-13}).$$

Backshift notation

The backward shift operator B is a useful notational device when working with time series lags:

$$By_t = y_{t-1}.$$

In other words, B has the effect of shifting the data back one period.
Two applications of B shifts the data back two periods

$$B(By_t) = B^2 y_t = y_{t-2}.$$

Backshift notation

The backward shift operator is convenient for describing the process of differencing. A first difference can be written as

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t \rightarrow (1-B)^{*} y_t$$

Similarly, if second-order differences have to be computed, then

$$\begin{aligned} y''_t &= (y'_t - y'_{t-1}) \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \\ &= (1 - 2B + B^2)y_t \\ &= (1-B)^2 y_t \end{aligned}$$

Autoregressive models

- In a multiple regression model, we forecast the variable of interest using a linear combination of predictors.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \varepsilon.$$

- In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + \varepsilon_t$$

$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \longrightarrow AR(1)$ $\quad$ $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t \rightarrow AR(2)$

- We refer to this as an $AR(p)$, an *autoregressive* model of order p.
- This is like a multiple regression but with lagged values of y_t as predictors.

Moving-average models

Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model.

$$y_t = \beta_0 + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + ... + \beta_q \varepsilon_{t-q} + \varepsilon_t$$

We refer to this as an $MA(q)$, a Moving Average of order q.

AR I MA

Auto regressive

Differentiat- ion.

Moving average

ARIMA models

- If we combine differencing with autoregression and a moving average model, we obtain a non-seasonal ARIMA model.
- ARIMA is an acronym for AutoRegressive Integrated Moving Average, ARIMA (p, d, q) where p refers to the AR part, q refers to the MA part and d is the degree of first differencing involved.
- Notice that a White Noise model $y_t = c + \varepsilon_t$ is an ARIMA$(0, 0, 0)$, while a *Random Walk* $y_t = y_{t-1} + \varepsilon_t$, is an ARIMA (0,1,0).

ARMA(p,q) models

An ARMA (p, q) may be expressed as

$$y_t = c + \phi_1 y_{t-1} + ... + \phi_q y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$$

or, by using backshift notation,

$$(1 - \phi_1 B - ... - \phi_p B^p) y_t = c + (1 - \theta_1 B - ... - \theta_q B^q) \varepsilon_t$$

ARIMA(p,d,q) models

Presentations: Jan9 - Jan13 - 4:30pm
Dec 12th - working groups meeting (not mandatory) - last day of lecture

If an ARMA (p, q) model is non stationary, we obtain an ARIMA (p, d, q) model. The simplest case, ARIMA (1,1,1), is defined as

$$(1 - \phi_1 B)(1 - B)y_t = c + (1 - \theta_1 B)\varepsilon_t$$

The general form of an ARIMA (p, d, q) may produce a great variety of ACF and PACF.

p - Auto regressive part - describes the past output behavior
q - Moving average part - describes weighted average of past error terms

Stationary series - does not depend on time (i.e no trend and seasonality)

AR(1) ------> MA(inf)
MA(1) ------> AR(inf)

Differentiation is used to eliminate trend and seasonality of the time-series

ARIMA(p,d,q) and seasonality - SARIMA

A further extension to ARMA models concerns seasonality.
An ARIMA model with seasonal components is an ARIMA $(p, d, q)(P, D, Q)_s$, where (p, d, q) indicates the non-seasonal part of the model, while (P, D, Q) indicates the seasonal part of order s.
The ARIMA model $(1, 1, 1)(1, 1, 1)_4$ is

Quarterly Data

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 - \theta_1 B)(1 - \Theta_1 B^4)\varepsilon_t$$

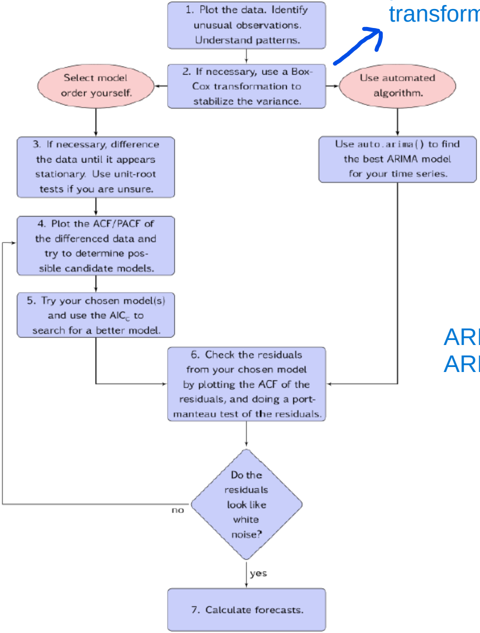
MA

Differentiation

AR

Model selection

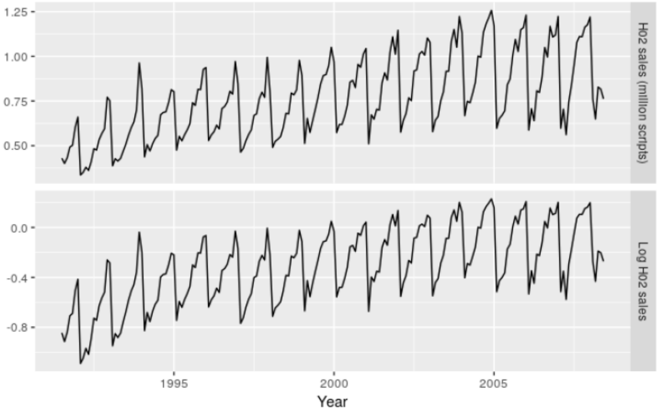
we can also perform log transformation



- selects best model based on AIC

ARMA(15, 12) - not ok
ARMA(3, 0, 2) - ok

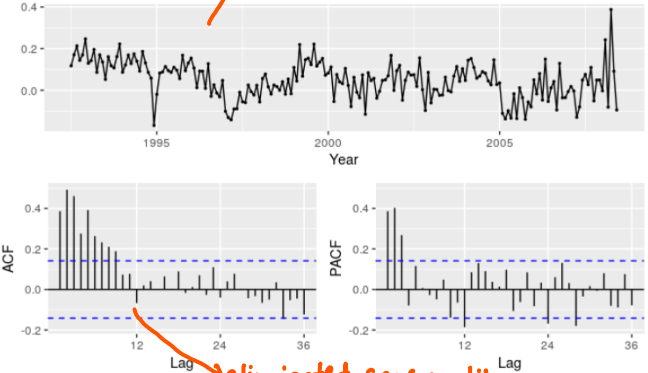
Example



A drug's sales (July 1991- June 2008) What are the main features of the series?

Example

obtained after seasonal differencing



eliminated seasonality

Seasonal differencing

PACF - calculates correlation between yt and yt-k and without considering corr of observations between the interval

Example

ARIMA(p, d, 0)

ACF ---> Exp Decaying
PACF ----> Significant spike until lag p

ARIMA(0, d, q)

- PACF - Exp Decaying
- ACF - significant spike until lag q

Model	AICc
$\text{ARIMA}(3,0,1)(0,1,2)_{12}$	−485.5
$\text{ARIMA}(3,0,1)(1,1,1)_{12}$	−484.2
$\text{ARIMA}(3,0,1)(0,1,1)_{12}$	−483.7
$\text{ARIMA}(3,0,1)(2,1,0)_{12}$	−476.3
$\text{ARIMA}(3,0,0)(2,1,0)_{12}$	−475.1
$\text{ARIMA}(3,0,2)(2,1,0)_{12}$	−474.9
$\text{ARIMA}(3,0,1)(1,1,0)_{12}$	−463.4

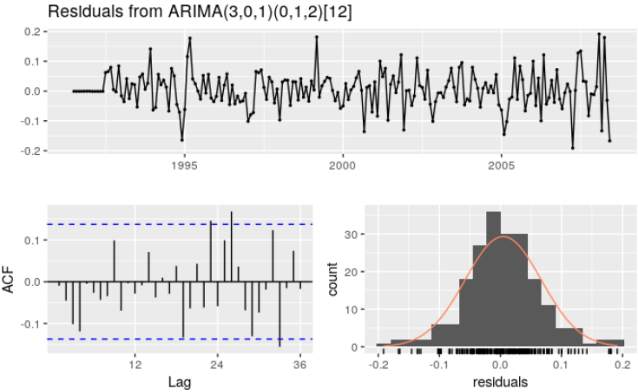
Different models have been estimated and compared on the basis of the AIC

$$AIC = -2\log(L) + 2k$$

No of parameters of the model

Example

ARIMAX = ARIMA(p,d,q) + Xt



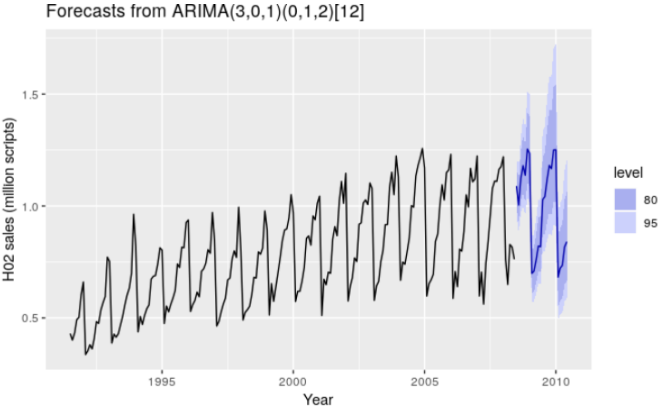
Are residuals white noise?

Example

Model	RMSE
ARIMA(3,0,1)(0,1,2)$_{12}$	0.0622
ARIMA(3,0,1)(1,1,1)$_{12}$	0.0630
ARIMA(2,1,4)(0,1,1)$_{12}$	0.0632
ARIMA(2,1,3)(0,1,1)$_{12}$	0.0634
ARIMA(3,0,3)(0,1,1)$_{12}$	0.0639
ARIMA(2,1,5)(0,1,1)$_{12}$	0.0640
ARIMA(3,0,1)(0,1,1)$_{12}$	0.0644
ARIMA(3,0,2)(0,1,1)$_{12}$	0.0644
ARIMA(3,0,2)(2,1,0)$_{12}$	0.0645
ARIMA(3,0,1)(2,1,0)$_{12}$	0.0646
ARIMA(4,0,2)(0,1,1)$_{12}$	0.0648
ARIMA(4,0,3)(0,1,1)$_{12}$	0.0648
ARIMA(3,0,0)(2,1,0)$_{12}$	0.0661
ARIMA(3,0,1)(1,1,0)$_{12}$	0.0679

Test set (july 2006–june 2008)
Auto.arima and model comparison by RMSE

Example



Forecast with the selected model