Time Series Analysis 3
DS351

Holt's linear trend method

- Like exponential smoothing but with **trend** element.
- Use when there is no seasonallity.
- If there is seasonallity, use Holt-Winters instead (next model)

Holt's method

$$\begin{aligned}
\text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\
\text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Trend equation} \quad & b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1},
\end{aligned}$$

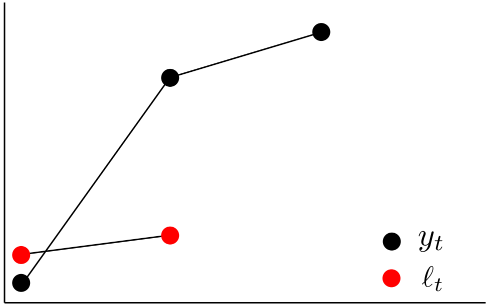
There are **4** parameters here: α, β, ℓ_0 and b_0.

Holt's linear trend method

Holt's method

$$\begin{aligned} \text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\ \text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ \text{Trend equation} \quad & b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}, \end{aligned}$$

- l_t is the level (estimate of y_t).
- b_t is the slope.
- Suppose that we have l_{t-1} and b_{t-1}.

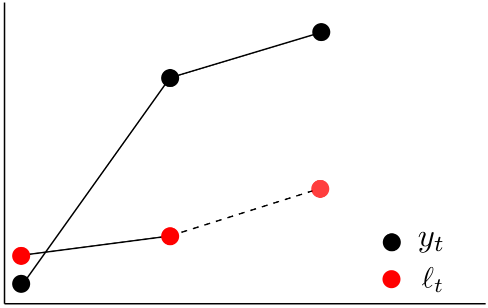


Holt's linear trend method

Holt's method

$$\begin{aligned} \text{Forecast equation} \qquad \hat{y}_{t+h|t} &= \ell_t + hb_t \\ \text{Level equation} \qquad \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ \text{Trend equation} \qquad b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}, \end{aligned}$$

- l_t is the "average" between y_t and $l_{t-1} + b_{t-1}$.
- Find $l_{t-1} + b_{t-1}$.

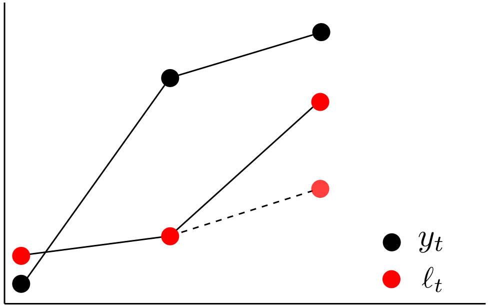


Holt's linear trend method

Holt's method

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t + hb_t$
Level equation	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$
Trend equation	$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1},$

- l_t is the "average" between y_t and $l_{t-1} + b_{t-1}$.
- Find $l_{t-1} + b_{t-1}$.
- Then find l_t.

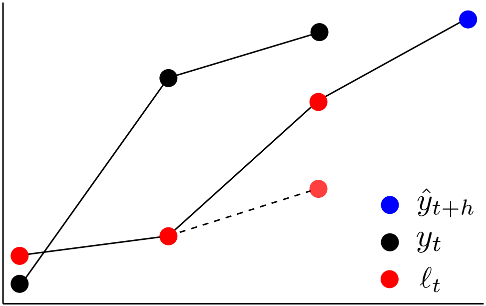


Holt's linear trend method

Holt's method

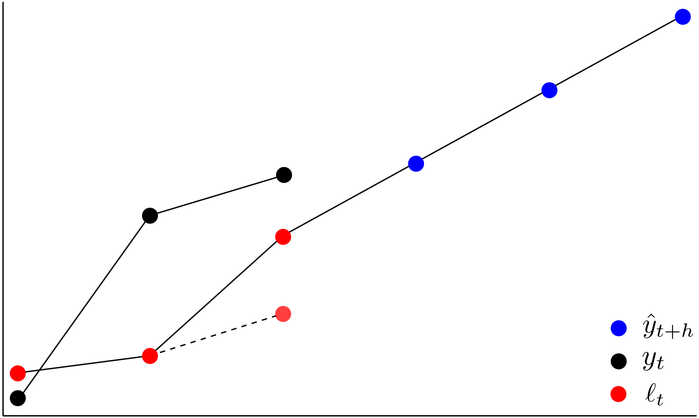
$$
\begin{aligned}
\text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\
\text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Trend equation} \quad & b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1},
\end{aligned}
$$

- b_t is the "average" between $l_t - l_{t-1}$ and b_{t-1}.
- Start the first forecast $\hat{y}_{t+1|t}$.



Holt's linear trend method

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$



The forecast is a linear function of h.

Air passengers data

Year	Time	Observation	Level	Slope	Forecast
	t	y_t	ℓ_t	b_t	$y_{t\|t1}$
1989	0		15.57	2.102	
1990	1	17.55	17.57	2.102	17.67
1991	2	21.86	21.49	2.102	19.68
1992	3	23.89	23.84	2.102	23.59
1993	4	26.93	26.76	2.102	25.94
⋮	⋮	⋮	⋮	⋮	⋮
2016	27	72.60	72.50	2.102	72.02
	h				$\hat{y}_{t+h\|t}$
	1				74.60
	2				76.70
	3				78.80
	4				80.91
	5				83.01

Damped Holt's method

- Linear trend is not realistic in many situations.
- Examples: Total factory output with a fixed number of machine.

Damped Holt's method (Gardner & McKenzie, 1985)

Fix $0 \leq \phi \leq 1$

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}.
\end{aligned}
$$

Damped Holt's method

- Linear trend is not realistic in many situations.
- Examples: Total factory output with a fixed number of machine.

Damped Holt's method (Gardner & McKenzie, 1985)

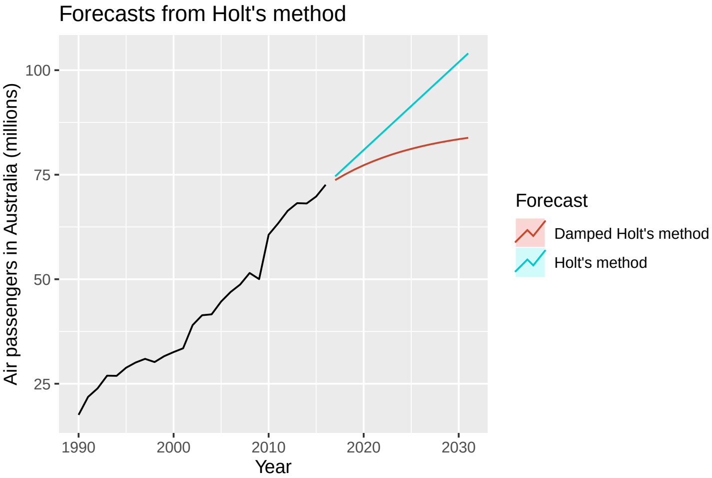
Fix $0 \leq \phi \leq 1$

$$\begin{aligned} \hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t \\ \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}. \end{aligned}$$

- $\phi = 1 \rightarrow$ Holt's method
- $\phi = 0 \rightarrow$ forecast with a constant
- In practice, $\phi \geq 0.8$.

Air passengers data

$$\phi = 0.9$$



Holt-Winters' seasonal method

- Use this method when there is seasonality.

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{\text{t-}} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}
$$

- Basically Holt's method + seasonality.
- m is the frequency of seasonality e.g. $m = 12$ for monthly data.
- l_t is the "average" between **observation with seasonality removed** $y_t - s_{t-m}$ and $l_{t-1} + b_{t-1}$.
- s_t is the "average" between **observation with level and trend removed** $y_t - l_{t-1} - b_{t-1}$ and the value of previous season s_{t-m}.

Holt-Winters' seasonal method

Holt-Winters' method

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t-} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}$$

- $t-$ in s_{t-} is the latest time in the sample that has the same seasonal index as $t+h$.
- For example, if $t =$ January, 2019 and $t+h =$ March, 2019 then $t- =$ March, 2018.
- There are a lot of parameters now: α, β, γ, ℓ_0, b_0, s_{-m+1}, $s_{-m+2}, \ldots, s_0$.

Holt-Winters' multiplicative method

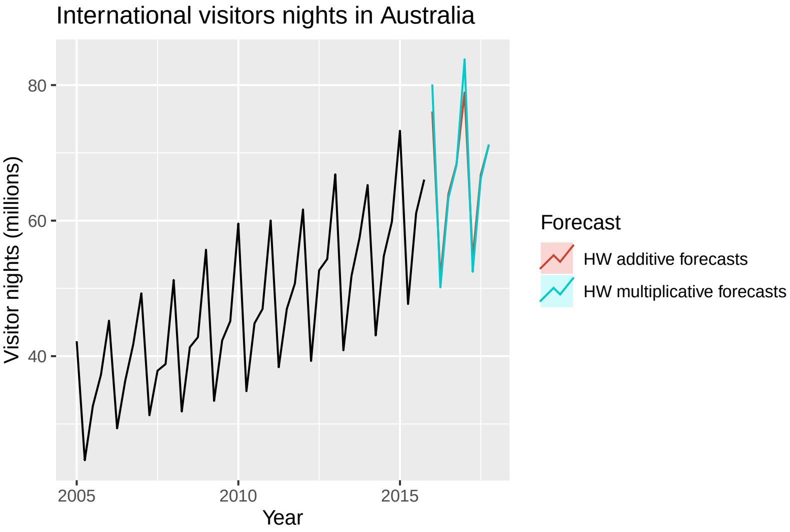
We can replace

$$\text{add by } s_t \to \text{multiply by} s_t$$
$$\text{subtract by } s_t \to \text{divide by} s_t.$$

Holt-Winters' multiplicative method

$$\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t-} \\
\ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}.
\end{aligned}$$

International visitors nights in Australia



Forecasts using Holt-Winters' method

	t	y_t	ℓ_t	b_t	s_t	y_t
2004 Q1	-3				9.70	
2004 Q2	-2				-9.31	
2004 Q3	-1				-1.69	
2004 Q4	0		32.26	0.70	1.31	
2005 Q1	1	42.21	32.82	0.70	9.50	42.66
2005 Q2	2	24.65	33.66	0.70	-9.13	24.21
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2015 Q4	44	66.06	63.22	0.70	2.35	64.22
	h					$y_{t+h\|t}$
2016 Q1	1					76.10
2016 Q2	2					51.60
2016 Q3	3					63.97
2016 Q4	4					68.37
2017 Q1	5					78.90
2017 Q2	6					54.41

AutoRegressive Integrated Moving Average (ARIMA)

Backshift operator

- This is where differencing is important.
- Backshift operator B helps us with differencing.

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

Backshift operator

- This is where differencing is important.
- Backshift operator B helps us with differencing.

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

Thus the first difference can be written as

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

Backshift operator

- This is where differencing is important.
- Backshift operator B helps us with differencing.

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

Thus the first difference can be written as

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

The second difference is

$$y''_t =$$

Backshift operator

- This is where differencing is important.
- Backshift operator B helps us with differencing.

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

Thus the first difference can be written as

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

The second difference is

$$y''_t =$$

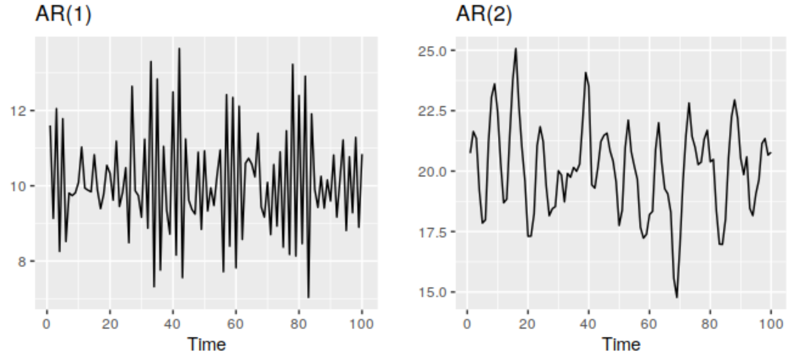
In general, the d-th order difference is

Autoregressive model

An autoregressive model of order p, $\mathbf{AR}(\boldsymbol{p})$, is

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

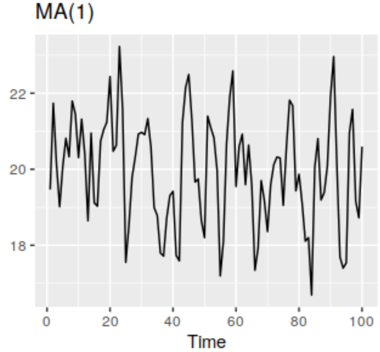
which is a multiple linear regression with $y_{t-p}, \ldots, y_{t-1}$ as predictors and y_t as the responses variable.

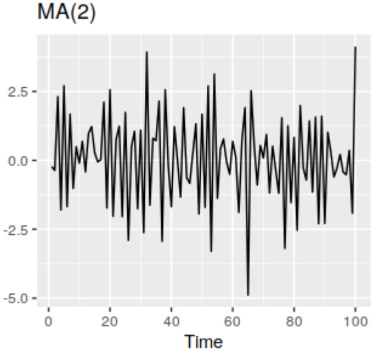


Moving average model

A moving average model, **MA(q)**, uses **past errors** to forecast.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}.$$





ARIMA model

AutoRegressive Integrated Moving Average (ARIMA) is a combination of both AR and MA models:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (1)$$

where y'_t is the differenced series. Thus there are a total of $p + q$ predictors. We call this an **ARIMA**$(\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{q})$ model.

ARIMA model

AutoRegressive Integrated Moving Average (ARIMA) is a combination of both AR and MA models:

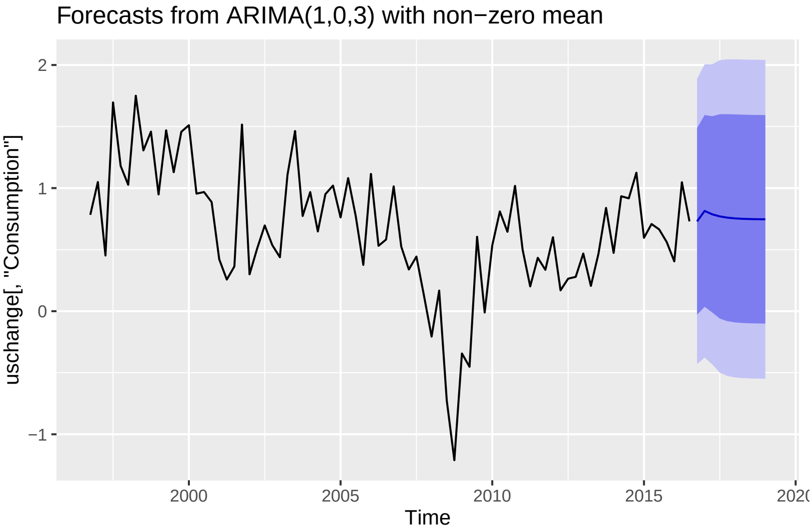
$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (1)$$

where y'_t is the differenced series. Thus there are a total of $p + q$ predictors. We call this an **ARIMA**$(\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{q})$ model.

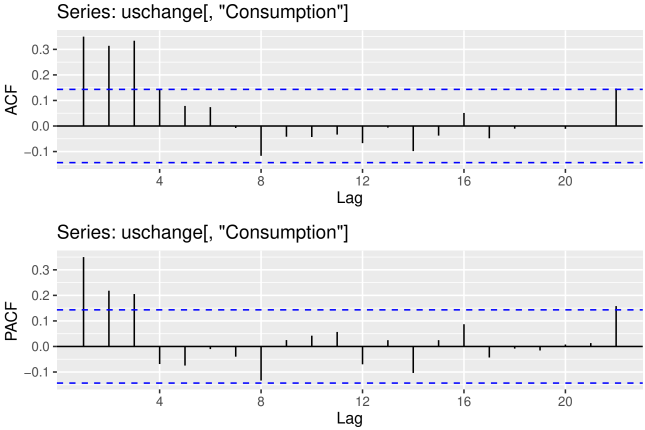
The value of c and d have the following effects on the long-term forecast.

$c = 0$	$d = 0$	forecasts go to zero.
$c = 0$	$d = 1$	forecasts go to a non-zero constant.
$c = 0$	$d = 2$	forecasts follow a straight line.
$c \neq 0$	$d = 0$	forecasts go to the mean of the data.
$c \neq 0$	$d = 1$	forecasts follow a straight line.
$c \neq 0$	$d = 2$	forecasts follow a quadratic trend.

Example: Quarterly US consumption



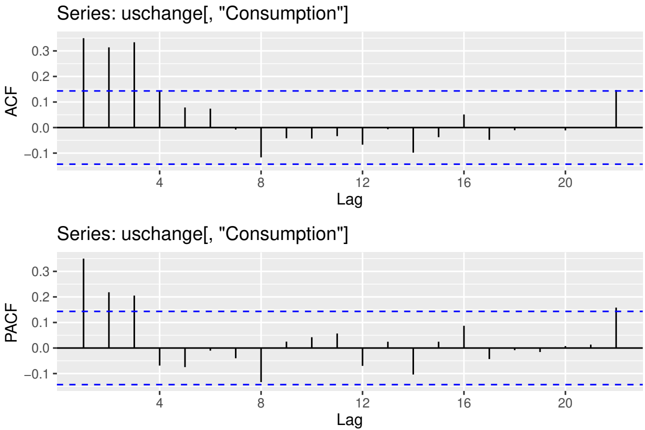
Find p and q from ACF and PACF plots



ACF and PACF plots of **differenced data** can help us find

- p in ARIMA$(p, d, 0)$
- q in ARIMA$(0, d, q)$
- Plots do not help for ARIMA(p, d, q) when $p, q > 0$.

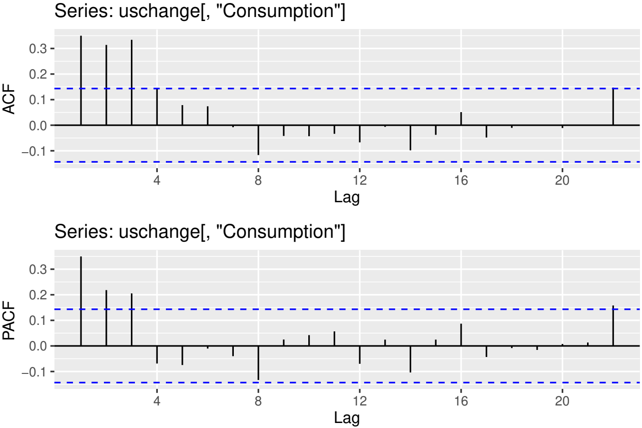
Find p and q from ACF and PACF plots



The data may follow an ARIMA$(p, d, 0)$ model if

- the ACF is exponentially decaying or sinusoidal
- there is a significant spike at lag p in the PACF, but none beyond lag p.

Find p and q from ACF and PACF plots



The data may follow an ARIMA$(0, d, q)$ model if

- the PACF is exponentially decaying or sinusoidal
- there is a significant spike at lag q in the ACF, but none beyond lag q.

Model selection

Our “score” of ARIMA model is based on the **likelihood**

$$L = \mathbb{P}(\text{data}|\text{model})$$

For example: We assume that a coin has probability 0.2 of turning head (the model is $\mathbb{P}(X = H) = 0.2$).

Model selection

Our “score” of ARIMA model is based on the **likelihood**

$$L = \mathbb{P}(\text{data}|\text{model})$$

For example: We assume that a coin has probability 0.2 of turning head (the model is $\mathbb{P}(X = H) = 0.2$).

If the results of five coin tosses are H, H, T, H, T, then the likelihood of the data given the model is

$$\mathbb{P}(H, H, T, H, T) = (0.2)^3(0.8)^2.$$

The likelihood would be higher if we assumed $\mathbb{P}(X = H) = 0.5$.

Model selection

Choose p and q, find parameters and compute $L = L(\text{data}|\text{model})$.

There are three "scores" that we can use

- Akaikes Information Criterion (AIC)

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1),$$

where $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.

Model selection

Choose p and q, find parameters and compute $L = L(\text{data}|\text{model})$.

There are three "scores" that we can use

- Akaikes Information Criterion (AIC)

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1),$$

  where $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.

- corrected AIC

$$\text{AICc} = \text{AIC} + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2},$$

  where T is the number of observations in the data.

Model selection

Choose p and q, find parameters and compute $L = L(\text{data}|\text{model})$.

There are three "scores" that we can use

- Akaikes Information Criterion (AIC)

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1),$$

  where $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.

- corrected AIC

$$\text{AICc} = \text{AIC} + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2},$$

  where T is the number of observations in the data.

- Bayesian Information Criterion

$$\text{BIC} = \text{AIC} + [\log(T) - 2](p + q + k + 1).$$

Model selection

These scores follow the same concepts.

- Better model has higher likelihood.
- But model with too many parameters (high $p + q$) tends to overfit and should be penalized.

We prefer AICc for ARIMA. The higher the score, the better.

Also check out seasonal ARIMA (not in scope of this course).