CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff Business School
Ysgol Busnes Caerdydd

# Forecasting in R

## ARIMA models

Bahman Rostami-Tabar

# Outline

# Outline

# Learning objectives

- Describe model building strategy for ARIMA models
- Explain criteria for best model selection
- Produce forecast using ARIMA models

# Outline

# Exponential smoothing vs ARIMA models

- Exponential smoothing models were based on a description of trend and seasonality in the data,
- ARIMA models aim to describe the autocorrelations in the data.
- Exponential smoothing and ARIMA models are the two most widely-used approaches to time series forecasting

# ARIMA models

**Autoregressive Integrated Moving Average models**

**AR**: autoregressive (lagged observations as inputs)

**I**: integrated (differencing to make series stationary)

**MA**: moving average (lagged errors as inputs)

# What does ARIMA account for?

- Previous observations
- Rate of change in the previous observations
- Error term in the previous observations
- Perform weel for short term horizons

# ARIMA model

- Combine ARMA model with **differencing**.
- $(1 - B)^d y_t$ follows an ARMA model.

**Autoregressive Moving Average(ARMA) models:**

$$\begin{aligned} y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t. \end{aligned}$$

# ARIMA model

- Combine ARMA model with **differencing**.
- $(1-B)^d y_t$ follows an ARMA model.

**Autoregressive Moving Average(ARMA) models:**

$$\begin{aligned} y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t. \end{aligned}$$

**ARIMA($p, d, q$) model**

| | |
|---|---|
| AR: | $p$ = number of preceding/lagged $y$ values |
| I: | $d$ = number of times series have to be "differenced" |
| MA: | $q$ = number of preceding/lagged values for the error term . |

# Stationarity

ARIMA models are stationary

**Definition**

If $\{y_t\}$ is a stationary time series, then for all $s$, the distribution of $(y_t, \dots, y_{t+s})$ does not depend on $t$.

# Stationarity

ARIMA models are stationary

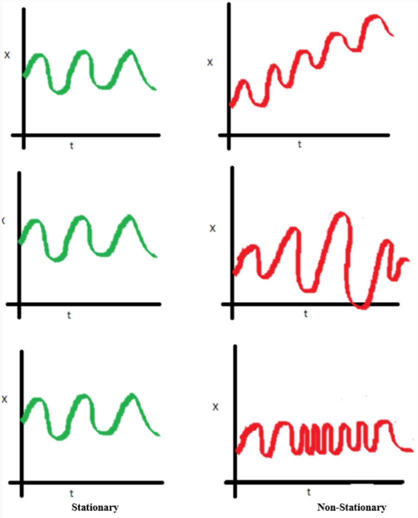
**Definition**

If $\{y_t\}$ is a stationary time series, then for all $s$, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on $t$.

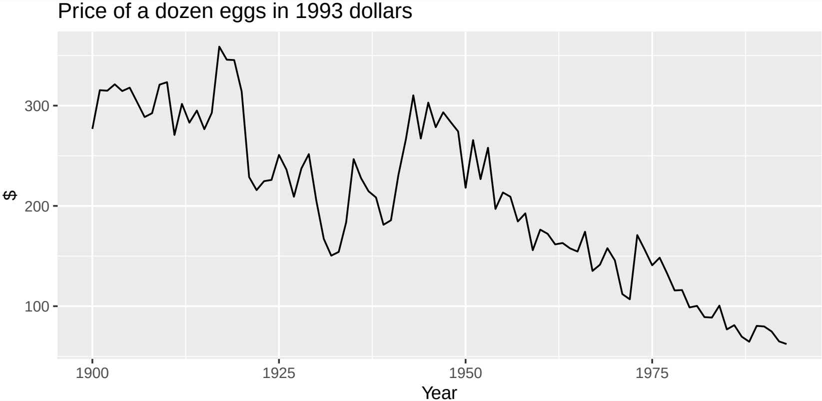
A **stationary series** is:

- roughly horizontal
- constant variance
- no patterns predictable in the long-term

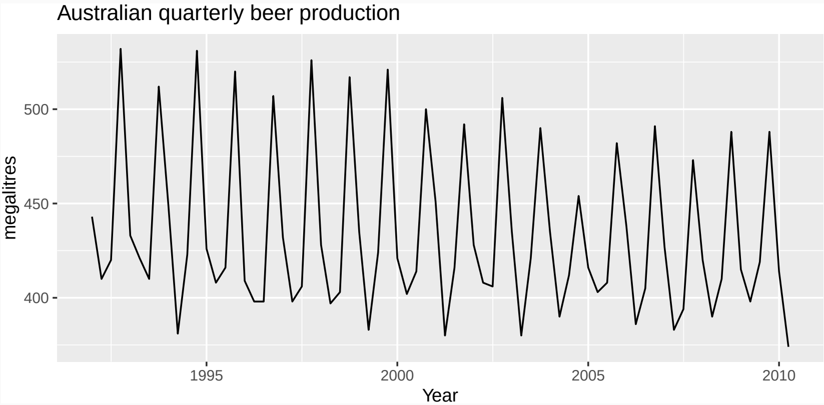
# Stationarity vs. Non-Stationarity



# Stationary?


Price of a dozen eggs in 1993 dollars
300
200
100
$
1900
1925
1950
1975
Year

# Stationary?



Australian quarterly beer production
megalitres
500
450
400
1995
2000
2005
2010
Year

# Stationarity

**Definition**

If $\{y_t\}$ is a stationary time series, then for all $s$, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on $t$.

# Stationarity

**Definition**

If $\{y_t\}$ is a stationary time series, then for all $s$, the distribution of $(y_t, \dots, y_{t+s})$ does not depend on $t$.

Transformations help to **stabilize the variance**.

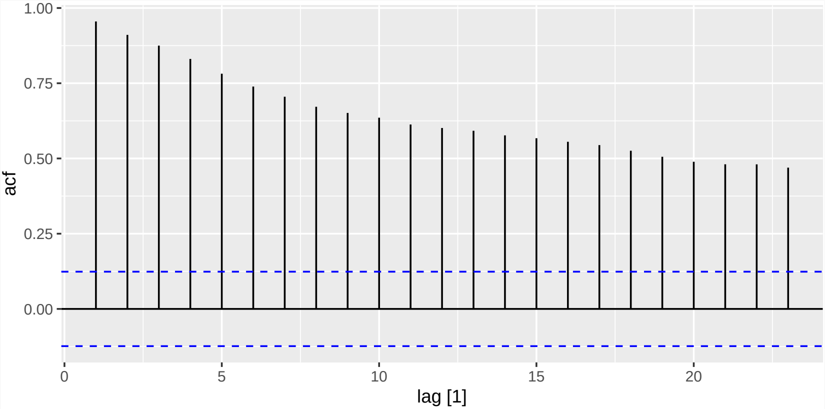
For ARIMA modelling, we also need to **stabilize the mean**.

# Is your data stationarity?

## Identifying non-stationary series

- Time plot.
- The ACF of stationary data drops to zero relatively quickly
- The ACF of non-stationary data decreases slowly.
- For non-stationary data, the value of $r_1$ is often large and positive.

# Example: Google stock price



# Unit root tests

- One way to determine more objectively whether data is non stationary to use a unit root test
- These are statistical hypothesis tests of stationarity that are designed for determining whether differencing is required

**Statistical tests to determine the required order of differencing.**

- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: null hypothesis is that the data are stationary

# KPSS test

```r
google_2018 %>%
  features(Close, unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Symbol kpss_stat kpss_pvalue
##   <chr>      <dbl>       <dbl>
## 1 GOOG       0.573      0.0252
```

# Differencing

- Differencing helps to **stabilize the mean**.
- The differenced series is the *change* between each observation in the original series.
- Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time.
- In practice, it is almost never necessary to go beyond second-order differences.

# Number of differencing

```r
google_2018 %>%
  features(Close, unitroot_ndiffs)
```

```
## # A tibble: 1 x 2
##   Symbol ndiffs
##   <chr>   <int>
## 1 GOOG        1
```

```r
#seasonal differencing
#features(Close, unitroot_nsdiffs)
```

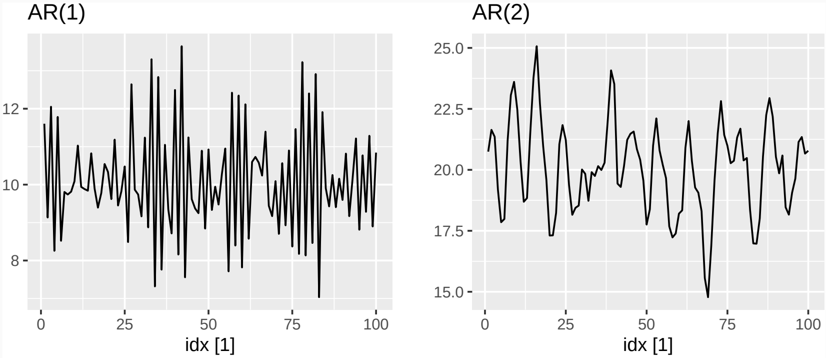
# Outline

# Autoregressive models

## Autoregressive (AR) models:

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

where $\varepsilon_t$ is white noise. We use **lagged values** of $y_t$ as predictors.



# Autoregressive models

- In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable.
- Where $c$ is a constant and $e_t$ i.i.d. (white noise) random variable with zero mean and known variance, $\sigma^2$.
- Changing the parameters $\phi_1, \phi_2, \ldots, \phi_p$ results in different time series patterns.

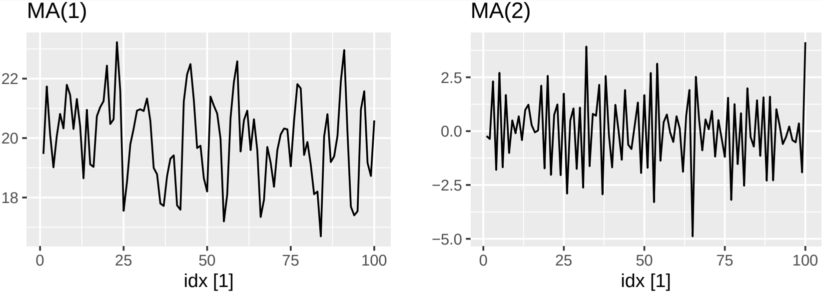
# Moving Average (MA) models

## Moving Average (MA) models:

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

where $\varepsilon_t$ is white noise.

We use ***past errors*** as predictors. *Don't confuse this with moving average smoothing!*



# Moving Average (MA) models

- We forecast the variable of interest using a linear combination of **past *errors***
- $c$ is a constant and $e_t$ i.i.d. (white noise) random variable with zero mean and known variance, $\sigma^2$.
- Changing the parameters $\theta_1, \theta_2, \ldots, \theta_q$ results in different time series patterns.

# ARMA(p,q) models

**Autoregressive Moving Average(ARMA) models:**

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

- Predictors include both **lagged values of $y_t$ and lagged errors.**
- Conditions on coefficients ensure stationarity.
- Conditions on coefficients ensure invertibility.

# Outline

# Select order of p,d,q

- Once you have a stationary time series, the next step is to select the appropriate ARIMA model. - Number of differencing determine *d*
- This means finding the most appropriate values for *p* and *q* in the $ ARIMA($p, d, q$) model.
- To do so, you need to examine the Autocorrelation and Partial Autocorrelation of the stationary time series.

# Partial autocorrelations

**Partial autocorrelations** measure relationship between $y_t$ and $y_{t-k}$, when the effects of other time lags — $1, 2, 3, \dots, k-1$ — are removed.

# Partial autocorrelations

**Partial autocorrelations** measure relationship between $y_t$ and $y_{t-k}$, when the effects of other time lags — $1, 2, 3, \dots, k-1$ — are removed.

$$\begin{aligned} \alpha_k &= k\text{th partial autocorrelation coefficient} \\ &= \text{equal to the estimate of } \phi_k \text{ in regression:} \\ & \quad y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_k y_{t-k}. \end{aligned}$$

# Partial autocorrelations

**Partial autocorrelations** measure relationship between $y_t$ and $y_{t-k}$, when the effects of other time lags — $1, 2, 3, \dots, k-1$ — are removed.

$$\begin{aligned} \alpha_k &= k\text{th partial autocorrelation coefficient} \\ &= \text{equal to the estimate of } \phi_k \text{ in regression:} \\ &\qquad y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_k y_{t-k}. \end{aligned}$$

- Varying number of terms on RHS gives $\alpha_k$ for different values of $k$.
- There are more efficient ways of calculating $\alpha_k$.
- $\alpha_1 = \rho_1$
- same critical values of $\pm 1.96/\sqrt{T}$ as for ACF.

# ACF and PACF interpretation

**AR(1)**

$$\rho_k = \phi_1^k \qquad \text{for } k = 1, 2, \dots;$$

$$\alpha_1 = \phi_1 \qquad \alpha_k = 0 \qquad \text{for } k = 2, 3, \dots.$$

So we have an AR(1) model when

- autocorrelations exponentially decay
- there is a single significant partial autocorrelation.

# ACF and PACF interpretation

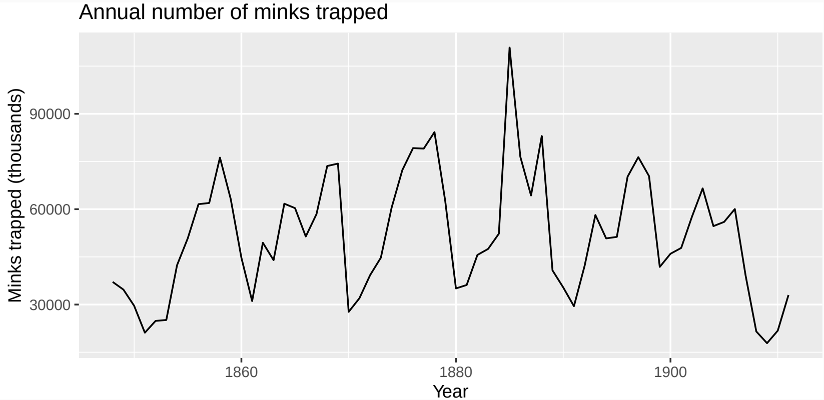
**MA(1)**

$$\rho_1 = \theta_1 \qquad \rho_k = 0 \qquad \text{for } k = 2, 3, \ldots;$$
$$\alpha_k = -(-\theta_1)^k$$

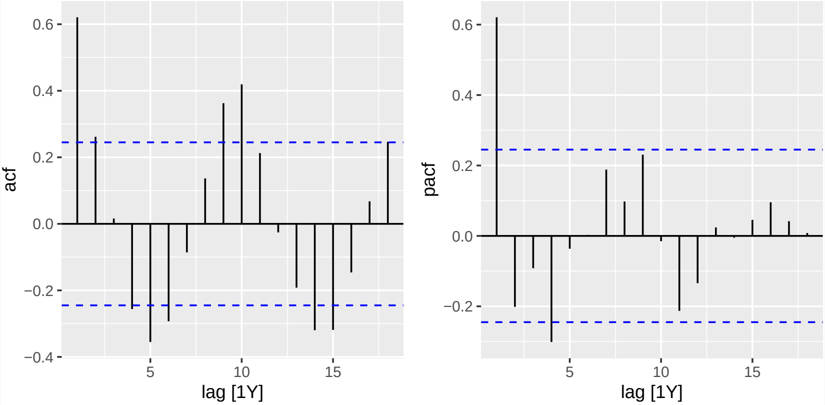
So we have an MA(1) model when

- the PACF is exponentially decaying and
- there is a single significant spike in ACF

# Example: Mink trapping



# Example: Mink trapping



# Maximum likelihood estimation

Having identified the model order, we need to estimate the parameters $c$, $\phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$.

# Maximum likelihood estimation

Having identified the model order, we need to estimate the parameters $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$.

- MLE is very similar to least squares estimation obtained by minimizing

$$\sum_{t-1}^{T} e_t^2$$

# Information criteria

**Akaike's Information Criterion (AIC):**

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

where $L$ is the likelihood of the data,

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

# Information criteria

## Akaike's Information Criterion (AIC):

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

where $L$ is the likelihood of the data,

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

## Corrected AIC:

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

# Information criteria

**Akaike's Information Criterion (AIC):**

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

where $L$ is the likelihood of the data,

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

**Corrected AIC:**

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

**Bayesian Information Criterion:**

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

# Information criteria

**Akaike's Information Criterion (AIC):**

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

where $L$ is the likelihood of the data,
$k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.

**Corrected AIC:**

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

**Bayesian Information Criterion:**

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

Good models are obtained by minimizing either the AIC, AICc or BIC. Our preference is to use the AICc.

# Outline

# How does ARIMA() work?

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left[1 + \frac{(p+q+k+2)}{T-p-q-k-2}\right].$$

where $L$ is the maximised likelihood fitted to the *differenced* data, $k=1$ if $c \neq 0$ and $k=0$ otherwise.

# How does ARIMA() work?

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left[1 + \frac{(p+q+k+2)}{T-p-q-k-2}\right].$$

where $L$ is the maximised likelihood fitted to the *differenced* data, $k=1$ if $c\neq 0$ and $k=0$ otherwise.

**Step1:** Select current model (with smallest AICc) from:
ARIMA$(2,d,2)$
ARIMA$(0,d,0)$
ARIMA$(1,d,0)$
ARIMA$(0,d,1)$

# How does ARIMA() work?

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left[1 + \frac{(p+q+k+2)}{T-p-q-k-2}\right].$$

where $L$ is the maximised likelihood fitted to the *differenced* data, $k = 1$ if $c \neq 0$ and $k = 0$ otherwise.

**Step1:** Select current model (with smallest AICc) from:
ARIMA$(2, d, 2)$
ARIMA$(0, d, 0)$
ARIMA$(1, d, 0)$
ARIMA$(0, d, 1)$

**Step 2:** Consider variations of current model:

- vary one of $p, q$, from current model by $\pm 1$;
- $p, q$ both vary from current model by $\pm 1$;
- Include/exclude $c$ from current model.

Model with lowest AICc becomes current model.

**Repeat Step 2 until no lower AICc can be found.**

# Modelling procedure with `ARIMA`

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance.
3. If the data are non-stationary: take first differences of the data until the data are stationary.
4. Examine the ACF/PACF: Is an AR($p$) or MA($q$) model appropriate?
5. Try your chosen model(s), and use the AICc to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

# Automatic modelling procedure with `ARIMA`

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance.
3. Use `ARIMA` to automatically select a model.
4. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals. If they do not look like white noise, try a modified model.
5. Once the residuals look like white noise, calculate forecasts.

# Outline

# Point forecasts

1. Rearrange ARIMA equation so $y_t$ is on LHS.
2. Rewrite equation by replacing $t$ by $T + h$.
3. On RHS, replace future observations by their forecasts, future errors by zero, and past errors by corresponding residuals.

Start with $h = 1$. Repeat for $h = 2, 3, \ldots$.

# Prediction intervals

## 95% prediction interval

$$\hat{y}_{T+h|T} \pm 1.96\sqrt{v_{T+h|T}}$$

where $v_{T+h|T}$ is estimated forecast variance.

- Multi-step prediction intervals for ARIMA(0,0,$q$):

$$y_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

$$v_{T|T+h} = \hat{\sigma}^2 \left[1 + \sum_{i=1}^{h-1} \theta_i^2\right], \qquad \text{for } h = 2, 3, \dots .$$

# Outline

# Seasonal ARIMA models

| ARIMA | $\underbrace{(p, d, q)}$ | $\underbrace{(P, D, Q)_m}$ |
|---|---|---|
| | $\uparrow$ <br> Non-seasonal part of the model | $\uparrow$ <br> Seasonal part of of the model |

where $m$ = number of observations per year.

# Seasonal ARIMA models

The seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF.

**ARIMA(0,0,0)(0,0,1)$_{12}$ will show:**

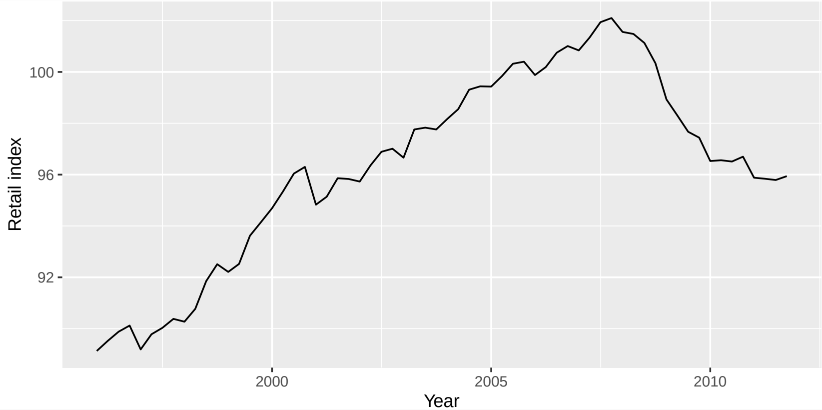
- a spike at lag 12 in the ACF but no other significant spikes.
- The PACF will show exponential decay in the seasonal lags; that is, at lags 12, 24, 36, . . . .

**ARIMA(0,0,0)(1,0,0)$_{12}$ will show:**

- exponential decay in the seasonal lags of the ACF
- a single significant spike at lag 12 in the PACF.

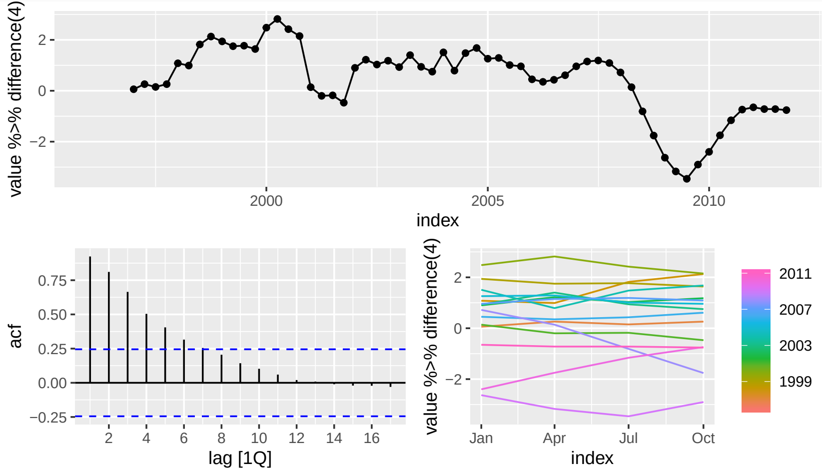
# European quarterly retail trade

```r
eu_retail %>% autoplot(value) +
  xlab("Year") + ylab("Retail index")
```



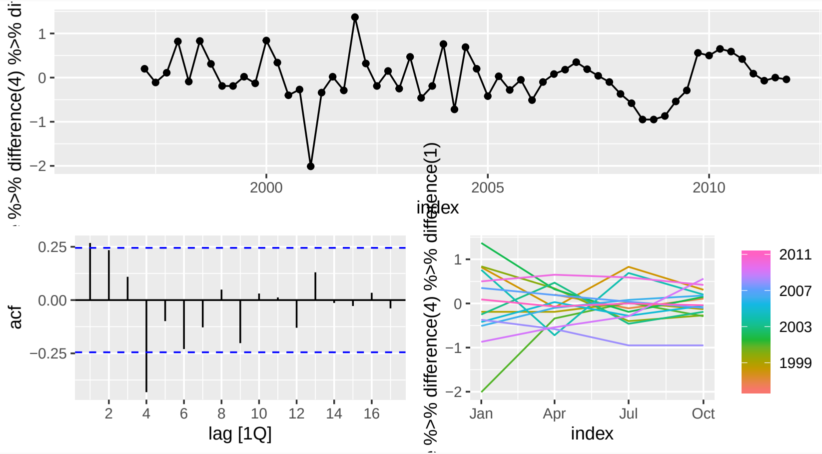
# European quarterly retail trade

```r
eu_retail %>% gg_tsdisplay(
  value %>% difference(4))
```



# European quarterly retail trade

```r
eu_retail %>% gg_tsdisplay(
  value %>% difference(4) %>% difference(1))
```

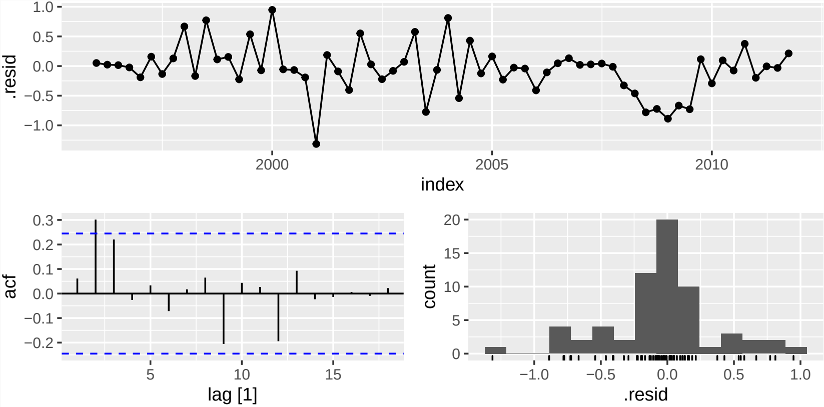


# European quarterly retail trade

- $d = 1$ and $D = 1$ seems necessary.
- Significant spike at lag 1 in ACF suggests non-seasonal MA(1) component.
- Significant spike at lag 4 in ACF suggests seasonal MA(1) component.
- Initial candidate model: ARIMA(0,1,1)(0,1,1)$_4$.
- We could also have started with ARIMA(1,1,0)(1,1,0)$_4$.

# European quarterly retail trade

```
fit <- eu_retail %>%
  model(arima = ARIMA(value ~ pdq(0,1,1) + PDQ(0,1,1)))
fit %>% gg_tsresiduals()
```



# European quarterly retail trade

```r
augment(fit) %>%
  features(.resid, ljung_box, lag = 8, dof = 2)
```

```
## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
##   <chr>    <dbl>     <dbl>
## 1 arima     10.7    0.0997
```

# Outline

# Lab Session 8

For the daily A&E data :

- Fit a suitable ARIMA model.
- Produce forecasts of your fitted models for 42 days.
- Check residuals
- Check the forecasts. Do they look reasonable?