MONASH University

MONASH BUSINESS SCHOOL

Forecasting: principles and practice

Rob J Hyndman

2.4 Non-seasonal ARIMA models

Outline

1 Autoregressive models

2 Moving average models

3 Non-seasonal ARIMA models

4 Partial autocorrelations

5 Estimation and order selection

6 ARIMA modelling in R

7 Forecasting

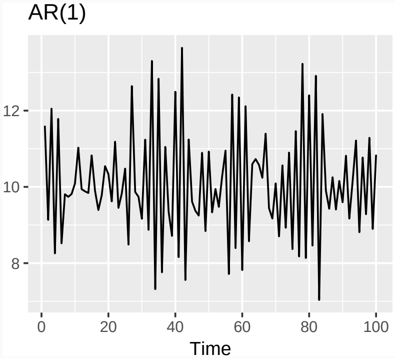
8 Lab session 11

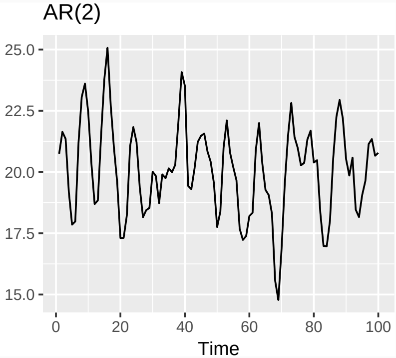
Autoregressive models

Autoregressive (AR) models:

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

where e_t is white noise. This is a multiple regression with **lagged values** of y_t as predictors.

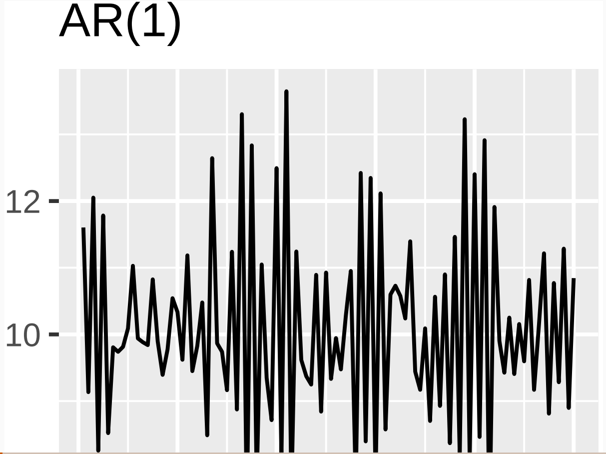




AR(1) model

$$y_t = 2 - 0.8y_{t-1} + e_t$$

$e_t \sim N(0, 1), \quad T = 100.$



AR(1) model

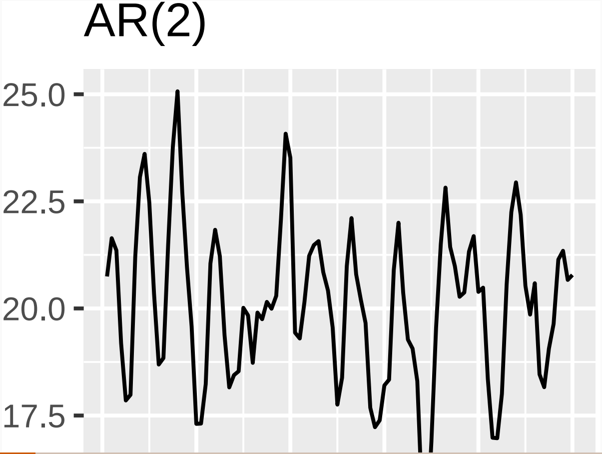
$$y_t = c + \phi_1 y_{t-1} + e_t$$

- When $\phi_1 = 0$, y_t is **equivalent to WN**
- When $\phi_1 = 1$ and $c = 0$, y_t is **equivalent to a RW**
- When $\phi_1 = 1$ and $c \neq 0$, y_t is **equivalent to a RW with drift**
- When $\phi_1 < 0$, y_t tends to **oscillate between positive and negative values**.

AR(2) model

$$y_t = 8 + 1.3y_{t-1} - 0.7y_{t-2} + e_t$$

$e_t \sim N(0, 1), \qquad T = 100.$



Stationarity conditions

We normally restrict autoregressive models to stationary data, and then some constraints on the values of the parameters are required.

General condition for stationarity

Complex roots of $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ lie outside the unit circle on the complex plane.

Stationarity conditions

We normally restrict autoregressive models to stationary data, and then some constraints on the values of the parameters are required.

General condition for stationarity

Complex roots of $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ lie outside the unit circle on the complex plane.

- For $p = 1$: $-1 < \phi_1 < 1$.
- For $p = 2$:
  $-1 < \phi_2 < 1 \qquad \phi_2 + \phi_1 < 1 \qquad \phi_2 - \phi_1 < 1.$
- More complicated conditions hold for $p \geq 3$.
- Estimation software takes care of this.

Outline

1 Autoregressive models

2 Moving average models

3 Non-seasonal ARIMA models

4 Partial autocorrelations

5 Estimation and order selection

6 ARIMA modelling in R

7 Forecasting

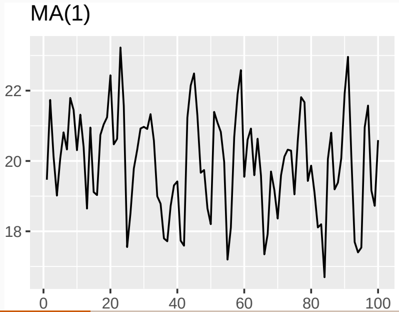
8 Lab session 11

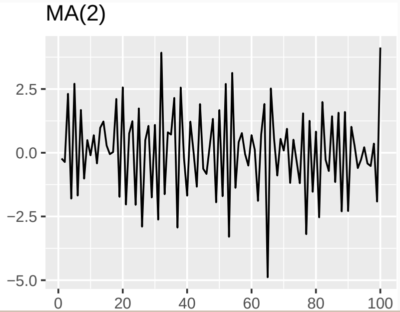
Moving Average (MA) models

Moving Average (MA) models:

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

where e_t is white noise. This is a multiple regression with ***past errors*** as predictors. *Don't confuse this with moving average smoothing!*

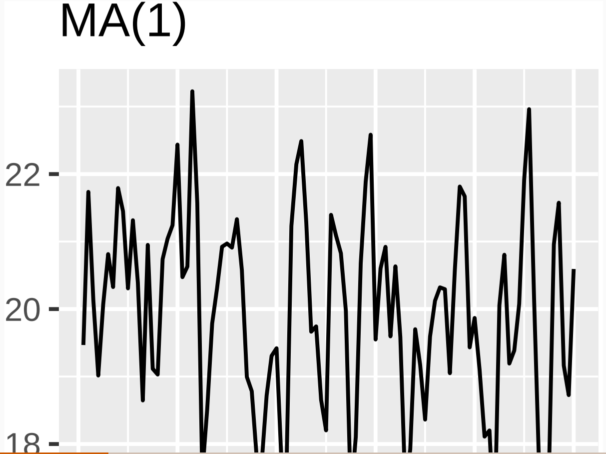




MA(1) model

$$y_t = 20 + e_t + 0.8e_{t-1}$$

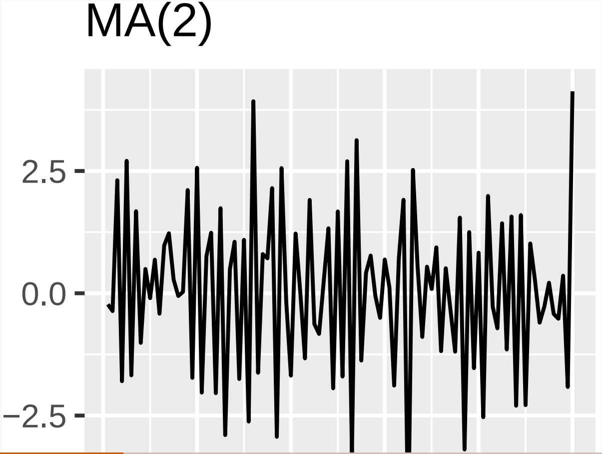
$e_t \sim N(0, 1), \quad T = 100.$



MA(2) model

$$y_t = e_t - e_{t-1} + 0.8e_{t-2}$$

$e_t \sim N(0, 1), \quad T = 100.$



Invertibility

- Any MA(q) process can be written as an AR(∞) process if we impose some constraints on the MA parameters.
- Then the MA model is called "invertible".
- Invertible models have some mathematical properties that make them easier to use in practice.
- Invertibility of an ARIMA model is equivalent to forecastability of an ETS model.

Invertibility

General condition for invertibility

Complex roots of $1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ lie outside the unit circle on the complex plane.

Invertibility

General condition for invertibility

Complex roots of $1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$ lie outside the unit circle on the complex plane.

- For $q = 1$: $-1 < \theta_1 < 1$.
- For $q = 2$:
  $-1 < \theta_2 < 1 \qquad \theta_2 + \theta_1 > -1 \qquad \theta_1 - \theta_2 < 1.$
- More complicated conditions hold for $\{q \geq 3.\}$
- Estimation software takes care of this.

Outline

1 Autoregressive models

2 Moving average models

3 Non-seasonal ARIMA models

4 Partial autocorrelations

5 Estimation and order selection

6 ARIMA modelling in R

7 Forecasting

8 Lab session 11

ARIMA models

Autoregressive Moving Average models:

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

ARIMA models

Autoregressive Moving Average models:

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

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

ARIMA models

Autoregressive Moving Average models:

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

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

Autoregressive Integrated Moving Average models

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

ARIMA models

Autoregressive Integrated Moving Average models

ARIMA(p, d, q) model

AR:	p = order of the autoregressive part
I:	d = degree of first differencing involved
MA:	q = order of the moving average part.

- White noise model: ARIMA(0,0,0)
- Random walk: ARIMA(0,1,0) with no constant
- Random walk with drift: ARIMA(0,1,0) with const.
- AR(p): ARIMA(p,0,0)

Backshift notation for ARIMA

- ARMA model:

$$
\begin{aligned}
& y_t = c + \phi_1 B y_t + \cdots + \phi_p B^p y_t + e_t + \theta_1 B e_t + \cdots + \theta_q B^q e \\
\text{or} \quad & (1 - \phi_1 B - \cdots - \phi_p B^p) y_t = c + (1 + \theta_1 B + \cdots + \theta_q B
\end{aligned}
$$

- ARIMA(1,1,1) model:

$$
\underset{\substack{\uparrow \\ \text{AR(1)}}}{(1 - \phi_1 B)} \quad \underset{\substack{\uparrow \\ \text{First} \\ \text{difference}}}{(1 - B) y_t} \quad = \quad c + \underset{\substack{\uparrow \\ \text{MA(1)}}}{(1 + \theta_1 B) e_t}
$$

Backshift notation for ARIMA

- ARMA model:

$$
\begin{aligned}
y_t &= c + \phi_1 B y_t + \cdots + \phi_p B^p y_t + e_t + \theta_1 B e_t + \cdots + \theta_q B^q e \\
\text{or} \quad (1 - \phi_1 B - \cdots - \phi_p B^p) y_t &= c + (1 + \theta_1 B + \cdots + \theta_q B
\end{aligned}
$$

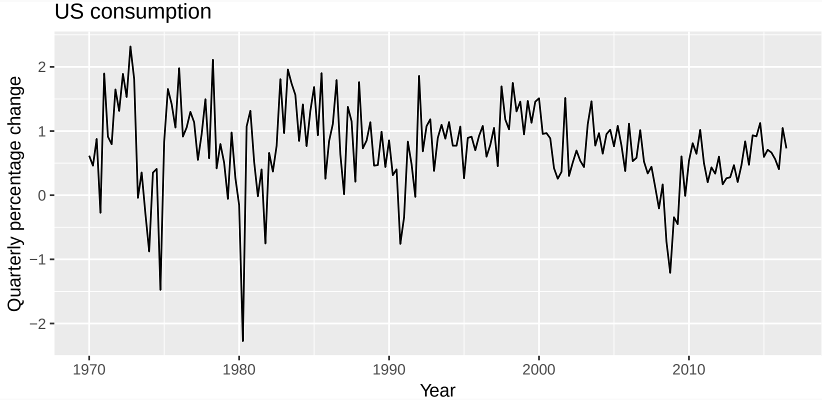
- ARIMA(1,1,1) model:

$$
\underset{\substack{\uparrow \\ \text{AR(1)}}}{(1 - \phi_1 B)} \quad \underset{\substack{\uparrow \\ \text{First} \\ \text{difference}}}{(1 - B) y_t} \quad = \quad c + \underset{\substack{\uparrow \\ \text{MA(1)}}}{(1 + \theta_1 B) e_t}
$$

Written out:

$$
y_t = c + y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + \theta_1 e_{t-1} + e_t
$$

US personal consumption



US personal consumption

```r
(fit <- auto.arima(uschange[,"Consumption"],
    seasonal=FALSE))
```

```
## Series: uschange[, "Consumption"]
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ma1     ma2    mean
##       1.3908  -0.5813  -1.1800  0.5584  0.7463
## s.e.  0.2553   0.2078   0.2381  0.1403  0.0845
##
## sigma^2 estimated as 0.3511:  log likelihood=-165.14
## AIC=342.28   AICc=342.75   BIC=361.67
```

US personal consumption

```r
(fit <- auto.arima(uschange[,"Consumption"],
    seasonal=FALSE))
```

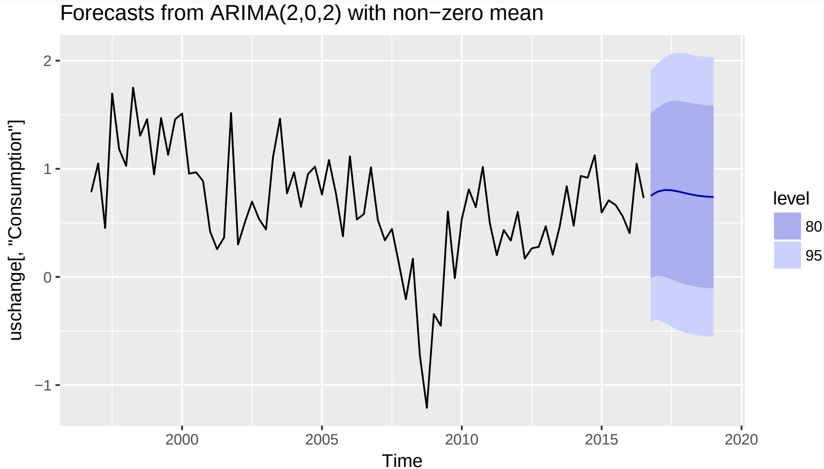
```
## Series: uschange[, "Consumption"]
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ma1     ma2    mean
##       1.3908  -0.5813  -1.1800  0.5584  0.7463
## s.e.  0.2553   0.2078   0.2381  0.1403  0.0845
##
## sigma^2 estimated as 0.3511:  log likelihood=-165.14
## AIC=342.28   AICc=342.75   BIC=361.67
```

ARIMA(0,0,3) or MA(3) model:

$$y_t = 0.756 + e_t + 0.254e_{t-1} + 0.226e_{t-2} + 0.269e_{t-3},$$

US personal consumption

```
fit %>% forecast(h=10) %>% autoplot(include=80
```



Understanding ARIMA models

- If $c = 0$ and $d = 0$, the long-term forecasts will go to zero.
- If $c = 0$ and $d = 1$, the long-term forecasts will go to a non-zero constant.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 0$, the long-term forecasts will go to the mean of the data.
- If $c \neq 0$ and $d = 1$, the long-term forecasts will follow a straight line.

Understanding ARIMA models

Forecast variance and d

- The higher the value of d, the more rapidly the prediction intervals increase in size.
- For $d = 0$, the long-term forecast standard deviation will go to the standard deviation of the historical data.

Cyclic behaviour

- For cyclic forecasts, $p > 2$ and some restrictions on coefficients are required.
- If $p = 2$, we need $\phi_1^2 + 4\phi_2 < 0$. Then average cycle of length

$$(2\pi)/\left[\text{arc cos}(-\phi_1(1-\phi_2)/(4\phi_2))\right].$$

Outline

1 Autoregressive models

2 Moving average models

3 Non-seasonal ARIMA models

4 Partial autocorrelations

5 Estimation and order selection

6 ARIMA modelling in R

7 Forecasting

8 Lab session 11

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.

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

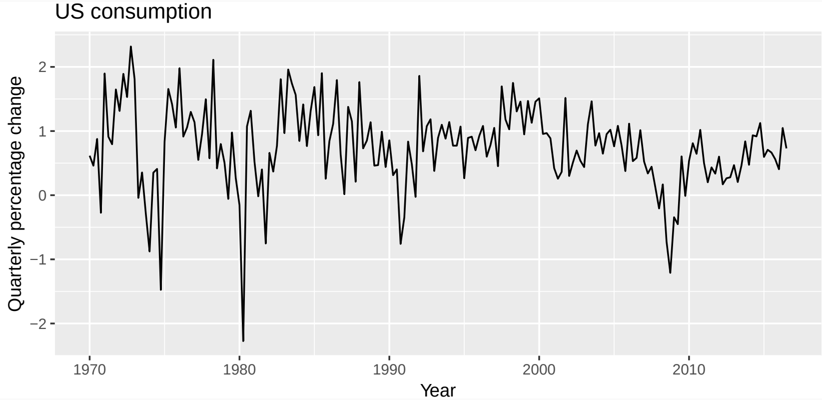
Partial autocorrelations

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.

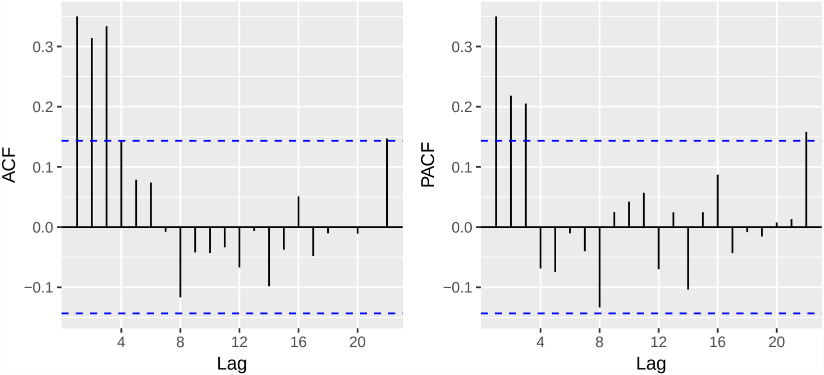
$$\begin{aligned}\alpha_k &= k\text{th partial autocorrelation coefficient}\\ &= \text{equal to the estimate of } b_k \text{ in regression:}\\ &\qquad y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_k y_{t-k}.\end{aligned}$$

- Varying number of terms on RHS gives α_k for different values of k.
- There are more efficient ways of calculating α_k.
- $\alpha_1 = \rho_1$

Example: US consumption



Example: US consumption



ACF and PACF interpretation

ARIMA(p,d,0) model if ACF and PACF plots of differenced data show:

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

ACF and PACF interpretation

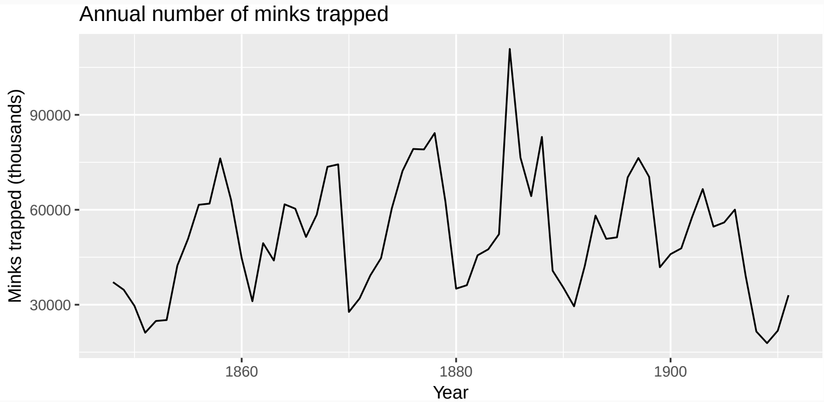
ARIMA(p,d,0) model if ACF and PACF plots of differenced data show:

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

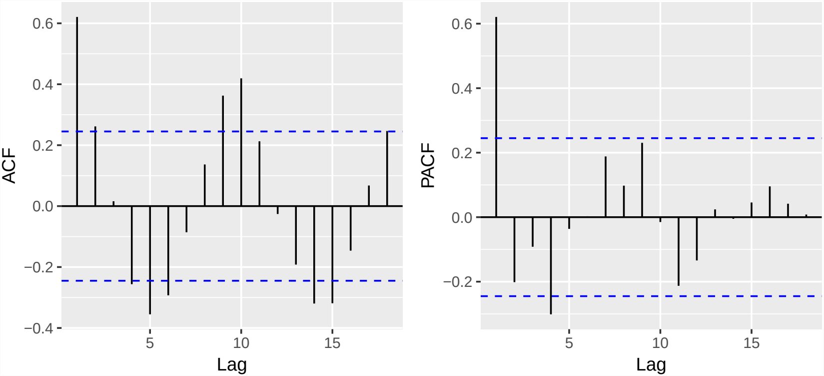
ARIMA(0,d,q) model if ACF and PACF plots of differenced data show:

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

Example: Mink trapping



Example: Mink trapping



Outline

1 Autoregressive models

2 Moving average models

3 Non-seasonal ARIMA models

4 Partial autocorrelations

5 **Estimation and order selection**

6 ARIMA modelling in R

7 Forecasting

8 Lab session 11

Maximum likelihood estimation

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

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
$$\sum_{t-1}^{T} e_t^2.$$
- The `Arima()` command allows CLS or MLE estimation.
- Non-linear optimization must be used in either case.
- Different software will give different estimates.

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:

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

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:

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

Outline

1 Autoregressive models

2 Moving average models

3 Non-seasonal ARIMA models

4 Partial autocorrelations

5 Estimation and order selection

6 **ARIMA modelling in R**

7 Forecasting

8 Lab session 11

How does auto.arima() work?

A non-seasonal ARIMA process

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t$$

Need to select appropriate orders: p, q, d

Hyndman and Khandakar (JSS, 2008) algorithm:

- Select no. differences d and D via unit root tests.
- Select p, q by minimising AICc.
- Use stepwise search to traverse model space.

How does auto.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 auto.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 auto.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.

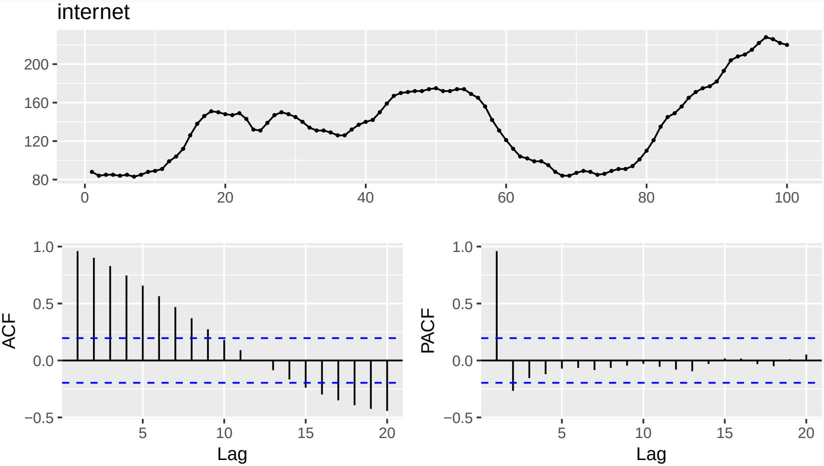
Step1: Select current model (with smallest AICc) from:
ARIMA$(2,d,2)$
ARIMA$(0,d,0)$
ARIMA$(1,d,0)$
ARIMA$(0,d,1)$

Step 2: Consider variations of current model:

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

Choosing your own model

```r
ggtsdisplay(internet)
```



Choosing your own model

```r
tseries::adf.test(internet)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  internet
## Dickey-Fuller = -2.6421, Lag order = 4, p-value = 0.3107
## alternative hypothesis: stationary
```

```r
tseries::kpss.test(internet)
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  internet
## KPSS Level = 0.72197, Truncation lag parameter = 2, p-
value =
## 0.01155
```

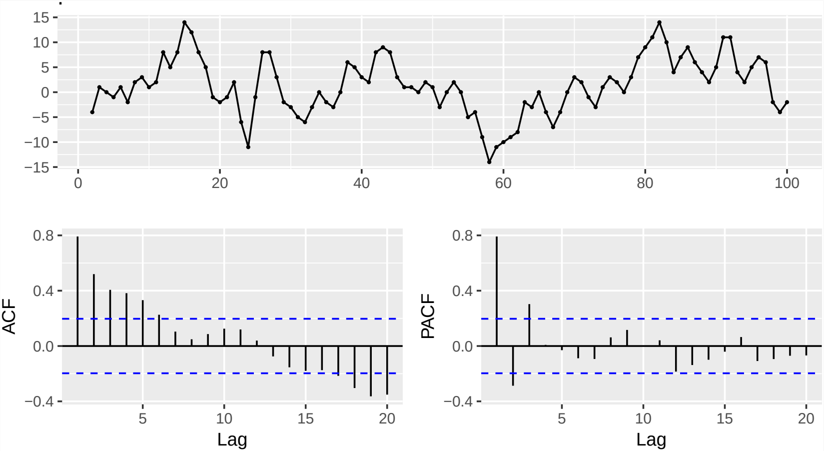
Choosing your own model

```
tseries::kpss.test(diff(internet))
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  diff(internet)
## KPSS Level = 0.26352, Truncation lag parameter = 2, p-
value = 0.1
```

Choosing your own model

```
internet %>% diff %>% ggtsdisplay
```



Choosing your own model

```r
(fit <- Arima(internet,order=c(3,1,0)))
```

```
## Series: internet
## ARIMA(3,1,0)
##
## Coefficients:
##          ar1      ar2     ar3
##       1.1513  -0.6612  0.3407
## s.e.  0.0950   0.1353  0.0941
##
## sigma^2 estimated as 9.656:  log likelihood=-
252
## AIC=511.99   AICc=512.42   BIC=522.37
```

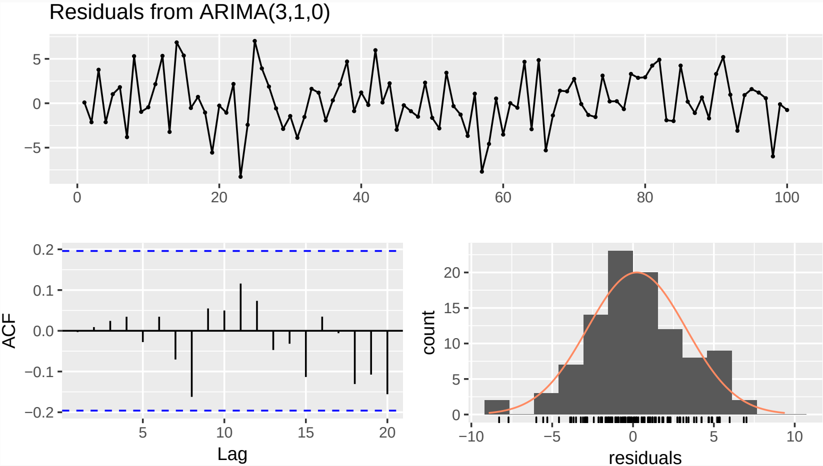
Choosing your own model

```r
(fit2 <- auto.arima(internet))
```

```
## Series: internet
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1     ma1
##       0.6504  0.5256
## s.e.  0.0842  0.0896
##
## sigma^2 estimated as 9.995:  log likelihood=-
254.15
## AIC=514.3   AICc=514.55   BIC=522.08
```

Choosing your own model

```
checkresiduals(fit, plot=TRUE)
```



#

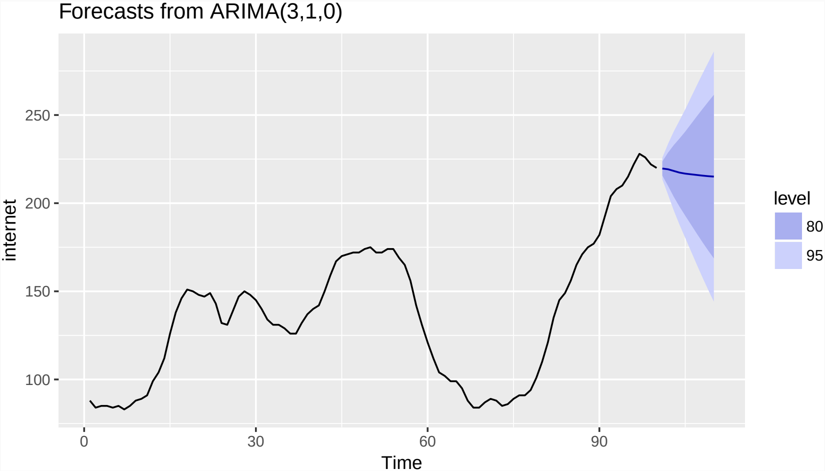
Choosing your own model

```r
checkresiduals(fit, plot=FALSE)
```

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,1,0)
## Q* = 4.4913, df = 7, p-value = 0.7218
##
## Model df: 3.   Total lags used: 10
```

Choosing your own model

```
fit %>% forecast %>% autoplot
```



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.

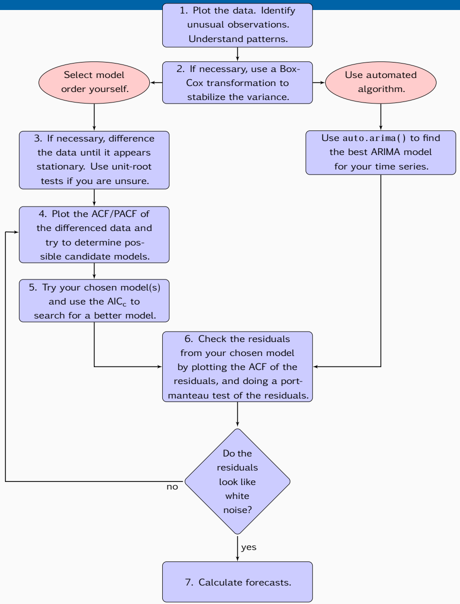
Modelling procedure with `auto.arima`

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 `auto.arima` to select a model.

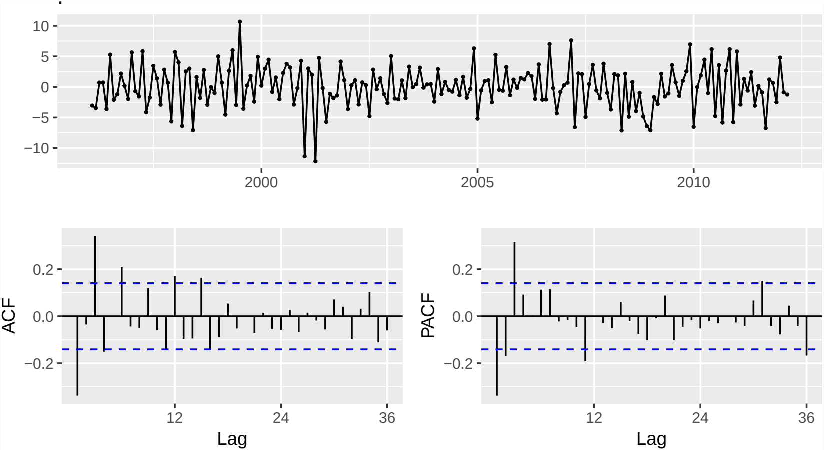
6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

Modelling procedure



Seasonally adjusted electrical equipment

```
eeadj %>% diff %>% ggtsdisplay
```



##Seasonally adjusted electrical equipment

Outline

1 Autoregressive models
2 Moving average models
3 Non-seasonal ARIMA models
4 Partial autocorrelations
5 Estimation and order selection
6 ARIMA modelling in R
7 Forecasting
8 Lab session 11

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.

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.

- $v_{T+1|T} = \hat{\sigma}^2$ for all ARIMA models regardless of parameters and orders.
- Multi-step prediction intervals for ARIMA(0,0,q):

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

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

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):

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

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

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):

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

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

- AR(1): Rewrite as MA(∞) and use above result.
- Other models beyond scope of this workshop.

Prediction intervals

- Prediction intervals **increase in size with forecast horizon**.
- Prediction intervals can be difficult to calculate by hand
- Calculations assume residuals are **uncorrelated** and **normally distributed**.
- Prediction intervals tend to be too narrow.
    - the uncertainty in the parameter estimates has not been accounted for.
    - the ARIMA model assumes historical patterns will not change during the forecast period.
    - the ARIMA model assumes uncorrelated future errors

Outline

1 Autoregressive models

2 Moving average models

3 Non-seasonal ARIMA models

4 Partial autocorrelations

5 Estimation and order selection

6 ARIMA modelling in R

7 Forecasting

8 **Lab session 11**

Lab Session 11