MONASH University

MONASH BUSINESS SCHOOL

Forecasting: principles and practice

Rob J Hyndman

2.4 Non-seasonal ARIMA models

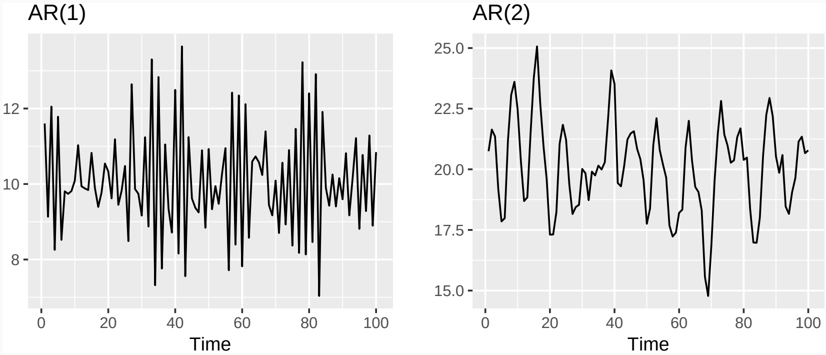
Outline

Autoregressive models

Autoregressive (AR) models:

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

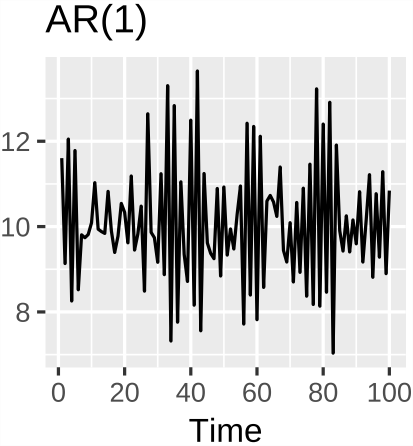
where ε_t is white noise. This is a multiple regression with **lagged values** of y_t as predictors.



AR(1) model

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

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



AR(1) model

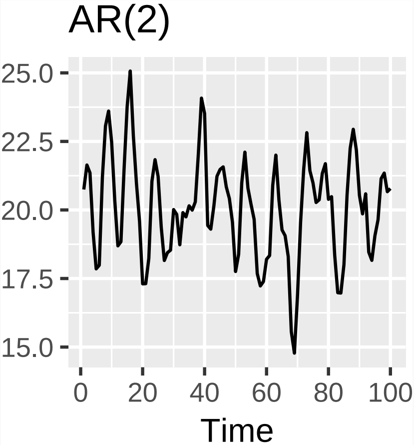
$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t$$

- 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

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

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



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 $p \geq 3$.

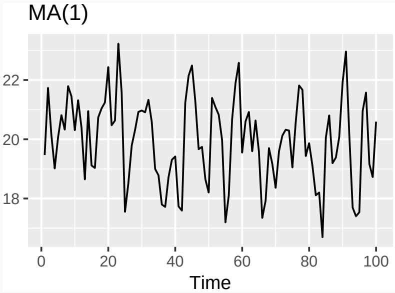
Outline

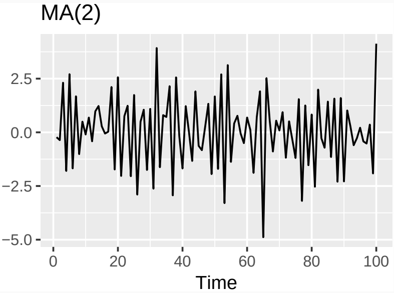
Moving Average (MA) models

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

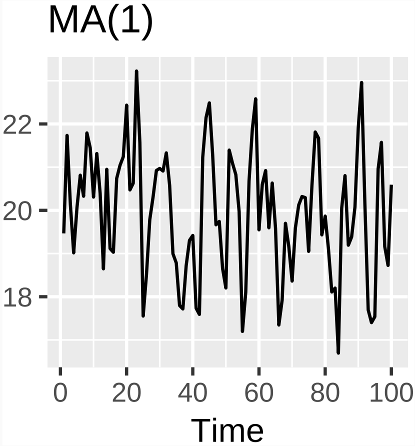




MA(1) model

$$y_t = 20 + \varepsilon_t + 0.8\varepsilon_{t-1}$$

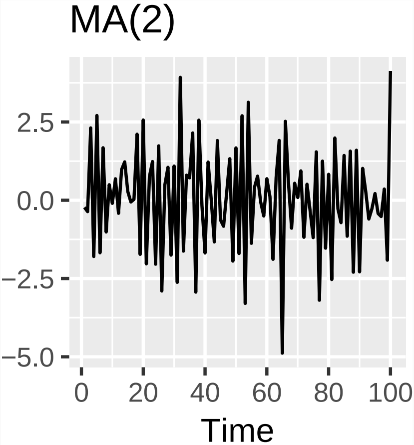
$$\varepsilon_t \sim N(0, 1), \quad T = 100.$$



MA(2) model

$$y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\varepsilon_{t-2}$$

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



Invertibility

- Invertible models have property that distant past has negligible effect on forecasts. Requires consraints on MA parameters.

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

- Invertible models have property that distant past has negligible effect on forecasts. Requires consraints on MA parameters.

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.

- 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 $q \geq 3$.

Outline

ARIMA models

Autoregressive Moving Average models:

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

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:

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

- 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)
- MA(q): ARIMA(0,0,q)

Backshift notation for ARIMA

- ARMA model:

$$y_t = c + \phi_1 B y_t + \cdots + \phi_p B^p y_t + \varepsilon_t + \theta_1 B \varepsilon_t + \cdots + \theta_q B^q \varepsilon_t$$

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

- ARIMA(1,1,1) model:

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

Backshift notation for ARIMA

- ARMA model:

$$y_t = c + \phi_1 B y_t + \cdots + \phi_p B^p y_t + \varepsilon_t + \theta_1 B \varepsilon_t + \cdots + \theta_q B^q \varepsilon_t$$

$$\text{or} \quad (1 - \phi_1 B - \cdots - \phi_p B^p) y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

- ARIMA(1,1,1) model:

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

Written out:

$$y_t = c + y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

R model

Intercept form

$$(1 - \phi_1 B - \cdots - \phi_p B^p) y'_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

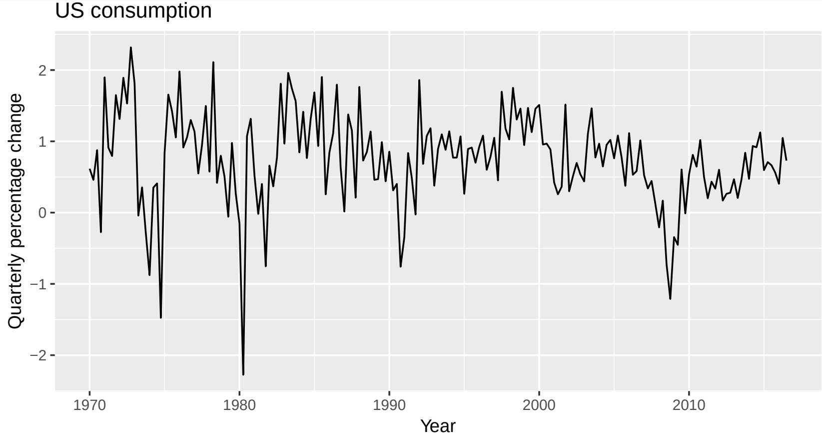
Mean form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(y'_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

- $y'_t = (1 - B)^d y_t$
- μ is the mean of y'_t.
- $c = \mu(1 - \phi_1 - \cdots - \phi_p)$.
- R uses mean form.

US personal consumption

```r
autoplot(uschange[,"Consumption"]) +
  xlab("Year") + ylab("Quarterly percentage change") +
  ggtitle("US consumption")
```



US personal consumption

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

```
## Series: uschange[, "Consumption"]
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ma1    ma2   mean
##        1.391  -0.581  -1.180  0.558  0.746
## s.e.   0.255   0.208   0.238  0.140  0.084
##
## sigma^2 estimated as 0.351:  log likelihood=-165.1
## AIC=342.3   AICc=342.8   BIC=361.7
```

US personal consumption

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

```
## Series: uschange[, "Consumption"]
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##          ar1     ar2     ma1    ma2   mean
##        1.391  -0.581  -1.180  0.558  0.746
## s.e.   0.255   0.208   0.238  0.140  0.084
##
## sigma^2 estimated as 0.351:  log likelihood=-165.1
## AIC=342.3   AICc=342.8   BIC=361.7
```

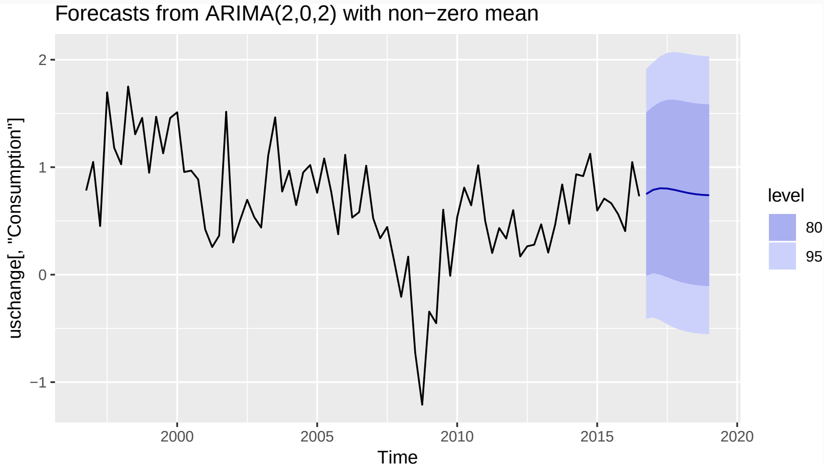
ARIMA(2,0,2) model:

$$y_t = c + 1.391y_{t-1} - 0.581y_{t-2} - 1.180\varepsilon_{t-1} + 0.558\varepsilon_{t-2} + \varepsilon_t,$$

where $c = 0.746 \times (1 - 1.391 + 0.581) = 0.142$ and $\varepsilon_t \sim N(0, 0.351)$.

US personal consumption

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



Understanding ARIMA models

Long-term forecasts

zero	$c = 0, d = 0$	
non-zero constant	$c = 0, d = 1$	$c \neq 0, d = 0$
linear	$c = 0, d = 2$	$c \neq 0, d = 1$
quadratic	$c = 0, d = 3$	$c \neq 0, d = 2$

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.

Understanding ARIMA models

Cyclic behaviour

- For cyclic forecasts, $p \geq 2$ 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

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:

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

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

Outline

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 KPSS test and seasonal strength measure.
- Select p, q by minimising AICc.
- Use stepwise search to traverse model space.

How does auto.arima() work?

Step 1: Select values of d and D.

Step 2: 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?

Step 1: Select values of d and D.

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

Step 3: Consider variations of current model:

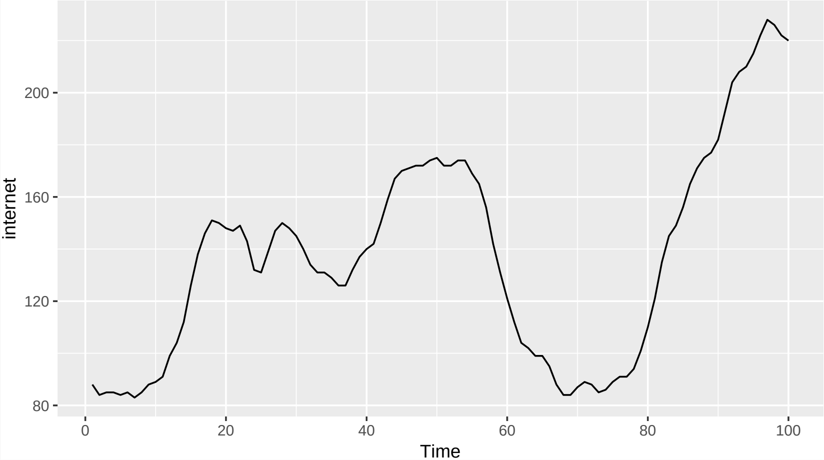
- vary one of $p, q,$ from current model by ± 1;
- p, q both vary from current model by ± 1;
- Include/exclude c from current model.

Model with lowest AICc becomes current model.

Repeat Step 3 until no lower AICc can be found.

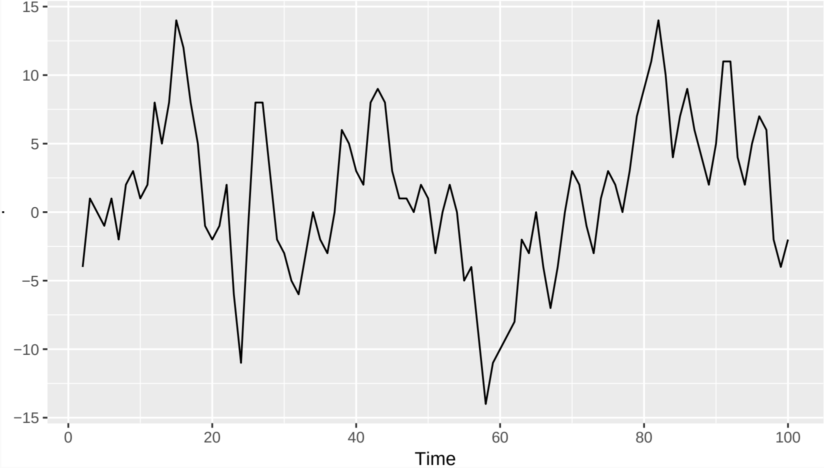
Choosing an ARIMA model

```
autoplot(internet)
```



Choosing an ARIMA model

```r
internet %>% diff() %>% autoplot()
```



Choosing an ARIMA model

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

```
## Series: internet
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1    ma1
##        0.650  0.526
## s.e.   0.084  0.090
##
## sigma^2 estimated as 10:  log likelihood=-254.2
## AIC=514.3   AICc=514.5   BIC=522.1
```

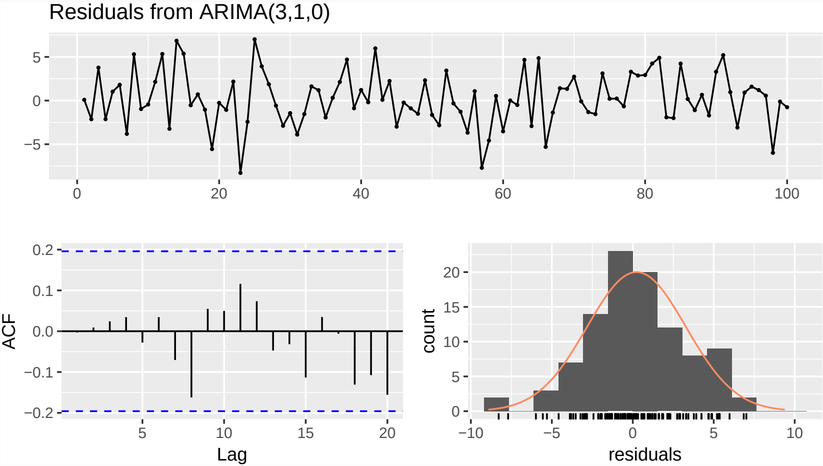
Choosing an ARIMA model

```r
(fit <- auto.arima(internet, stepwise=FALSE,
  approximation=FALSE))
```

```
## Series: internet
## ARIMA(3,1,0)
##
## Coefficients:
##          ar1     ar2    ar3
##        1.151  -0.661  0.341
## s.e.  0.095   0.135  0.094
##
## sigma^2 estimated as 9.66:  log likelihood=-252
## AIC=512   AICc=512.4   BIC=522.4
```

Choosing an ARIMA model

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



##

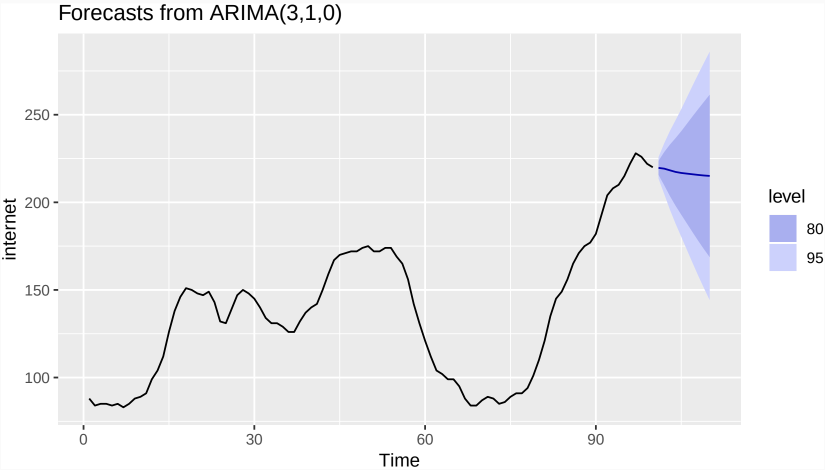
Choosing an ARIMA model

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

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,1,0)
## Q* = 4.5, df = 7, p-value = 0.7
##
## Model df: 3.   Total lags used: 10
```

Choosing an ARIMA model

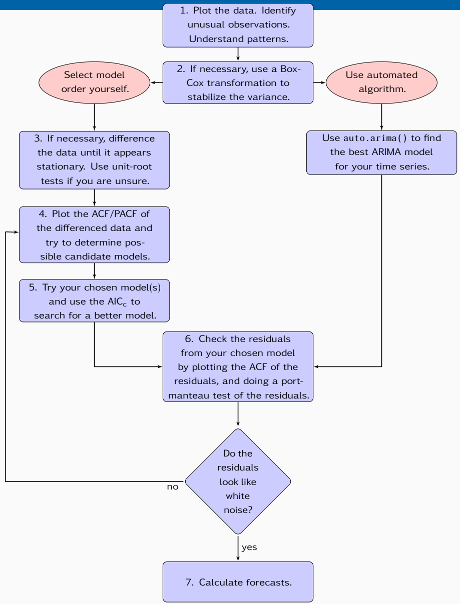
```
fit %>% forecast() %>% autoplot()
```



Modelling procedure with `auto.arima`

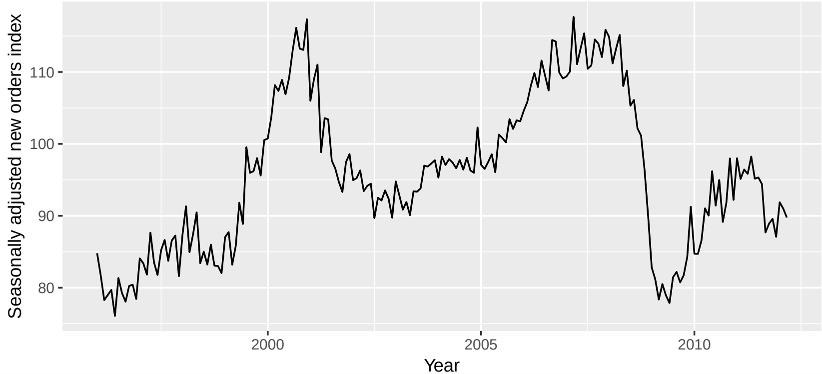
1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance.
3. Use `auto.arima` to select a model.
4. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test 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.

Modelling procedure



Seasonally adjusted electrical equipment

```r
eeadj <- seasadj(stl(elecequip, s.window="periodic"
autoplot(eeadj) + xlab("Year") +
  ylab("Seasonally adjusted new orders index")
```

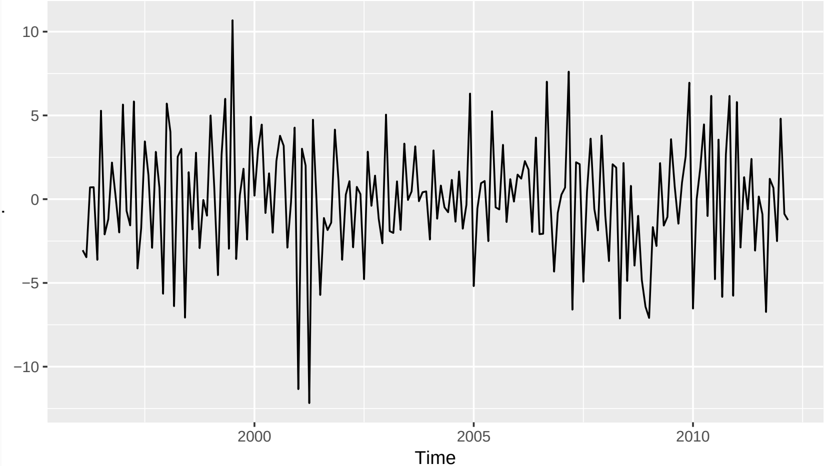


Seasonally adjusted electrical equipment

1. Time plot shows sudden changes, particularly big drop in 2008/2009 due to global economic environment. Otherwise nothing unusual and no need for data adjustments.
2. No evidence of changing variance, so no Box-Cox transformation.
3. Data are clearly non-stationary, so we take first differences.

Seasonally adjusted electrical equipment

```r
eeadj %>% diff() %>% autoplot()
```



Seasonally adjusted electrical equipment

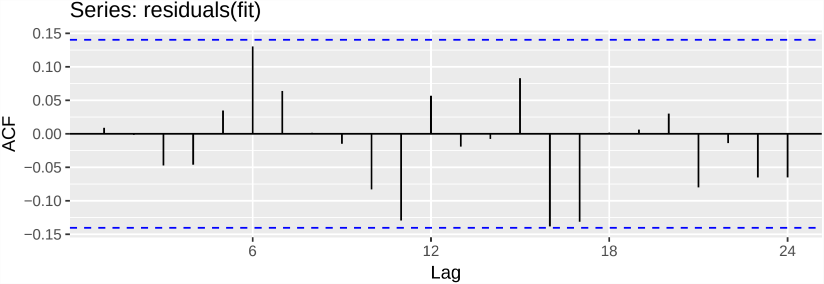
```r
fit <- auto.arima(eeadj, stepwise=FALSE, approximation=FALSE)
summary(fit)
```

```
## Series: eeadj
## ARIMA(3,1,1)
##
## Coefficients:
##         ar1    ar2    ar3     ma1
##       0.004  0.092  0.370  -0.392
## s.e.  0.220  0.098  0.067   0.243
##
## sigma^2 estimated as 9.58:  log likelihood=-492.7
## AIC=995.4   AICc=995.7   BIC=1012
##
## Training set error measures:
##                     ME  RMSE   MAE      MPE  MAPE   MASE
## Training set 0.03288 3.055 2.357 -0.00647 2.482 0.2884
##                     ACF1
```

Seasonally adjusted electrical equipment

6. ACF plot of residuals from ARIMA(3,1,1) model look like white noise.

```
ggAcf(residuals(fit))
```



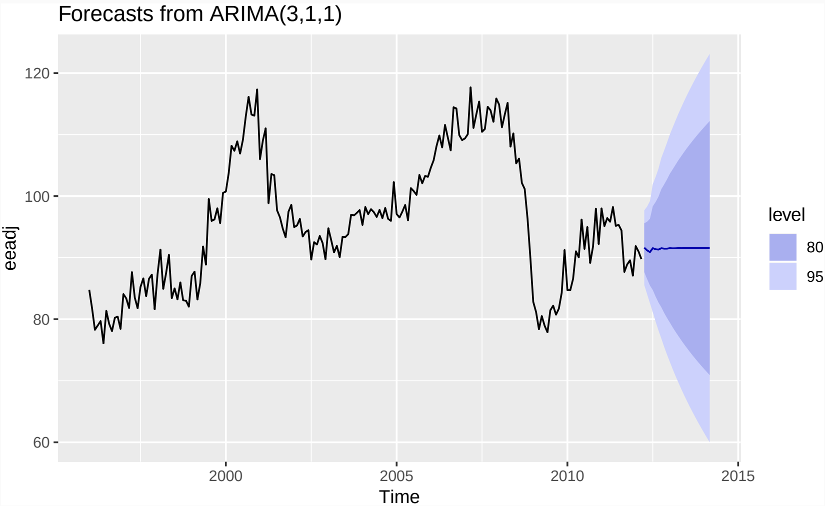
Seasonally adjusted electrical equipment

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

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,1,1)
## Q* = 24, df = 20, p-value = 0.2
##
## Model df: 4.   Total lags used: 24
```

Seasonally adjusted electrical equipment

```r
fit %>% forecast() %>% autoplot()
```

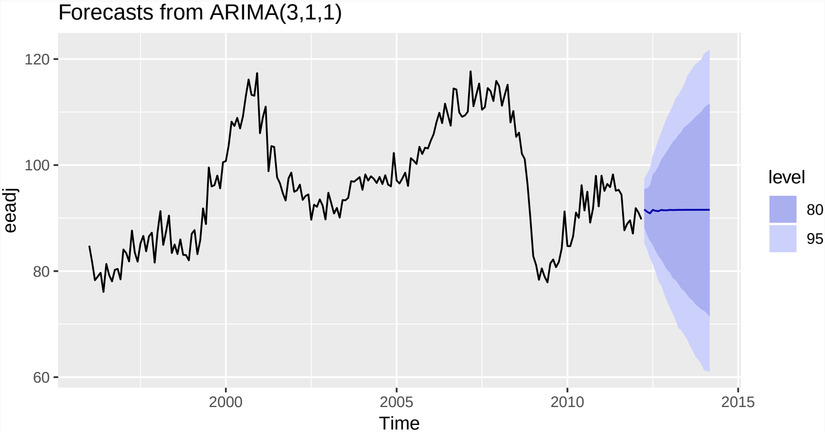


Prediction intervals

- Prediction intervals **increase in size with forecast horizon**.
- 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

Bootstrapped prediction intervals

```r
fit %>% forecast(bootstrap=TRUE) %>% autoplot()
```



- No assumption of normally distributed residuals.

Outline

Lab Session 15