MONASH University

MONASH BUSINESS SCHOOL

ETW3420: Principles of Forecasting and Applications

Topic 6: ARIMA Models

Dr. Jason Ng

Outline

Introduction

- ARIMA models provide another 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.
- Before we introduce ARIMA models, we must first discuss the concept of stationarity and the technique of differencing time series.

Outline

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 $(y_t, \ldots, y_{t+s})$ does not depend on t.

A **stationary series** is:

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

Stationarity

Statistical Properties of Stationary Series

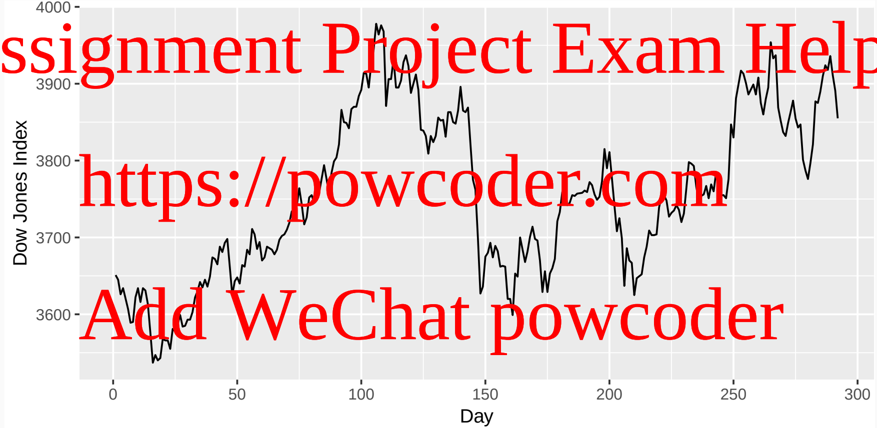
Statistically, a time series $\{y_t\}$ is stationary if:

$$E(y_t) = \mu < \infty \text{ for all } t$$

$$Var(y_t) = \gamma_0 < \infty \text{ for all } t$$

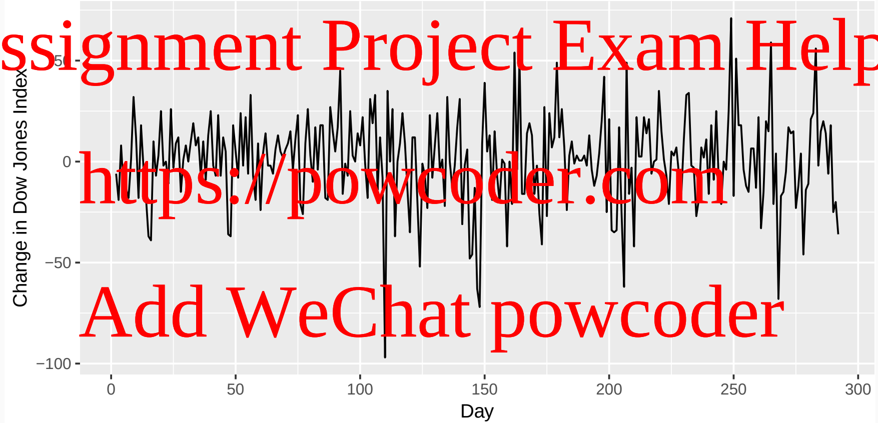
$$cov(y_t, y_{t-j}) = \gamma_j < \infty \text{ for all } t$$

Stationary?



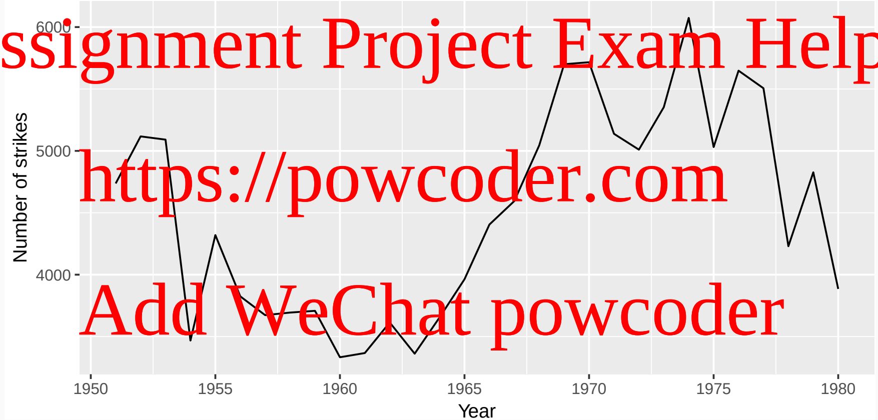
4000
3900
3800
3700
3600
Dow Jones Index
0
50
100
150
200
250
300
Day

Stationary?

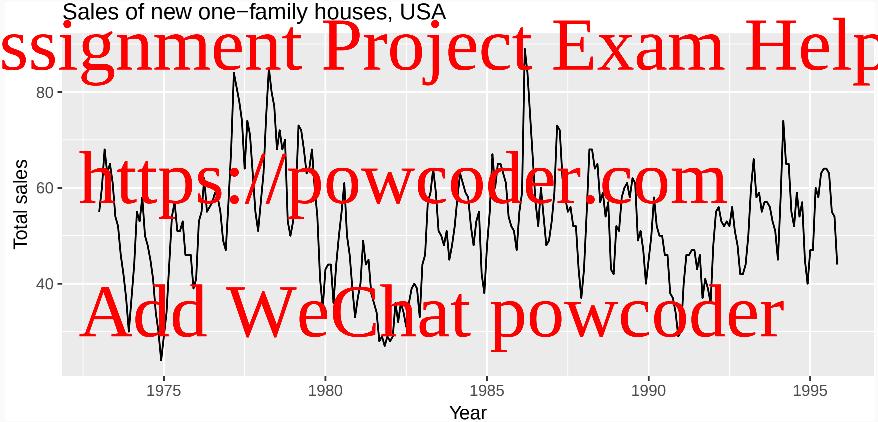


Change in Dow Jones Index
50
0
−50
−100
0
50
100
150
200
250
300
Day

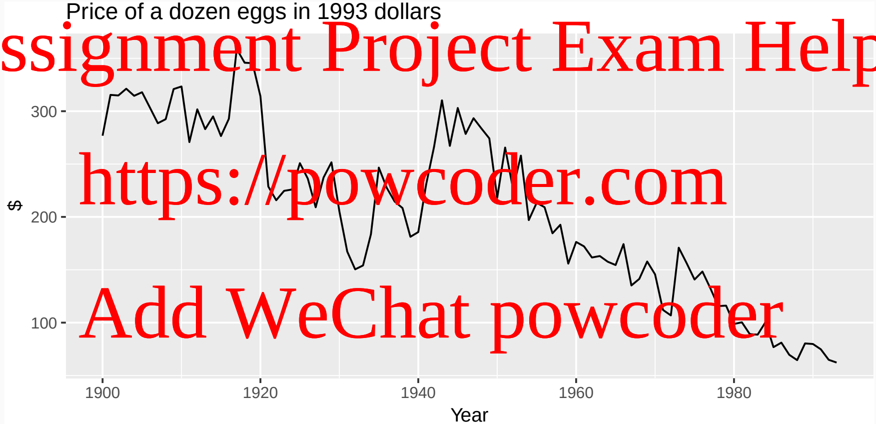
Stationary?


6000
5000
4000
Number of strikes
1950
1955
1960
1965
1970
1975
1980
Year

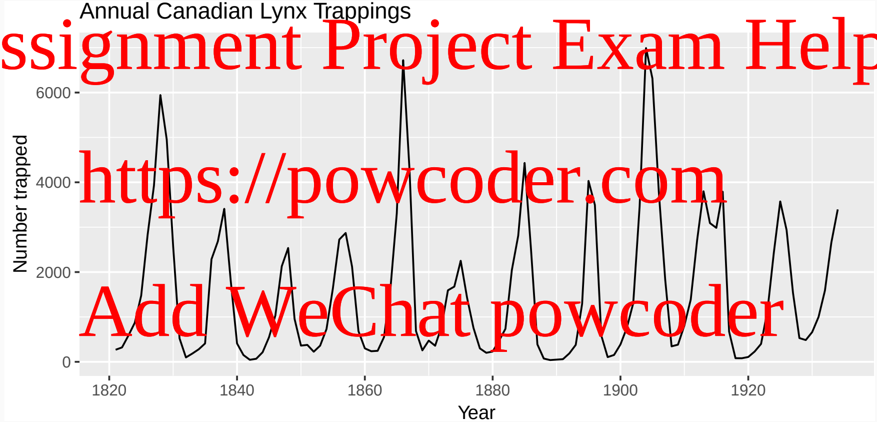
Stationary?


Sales of new one−family houses, USA
Total sales
80
60
40
1975
1980
1985
1990
1995
Year

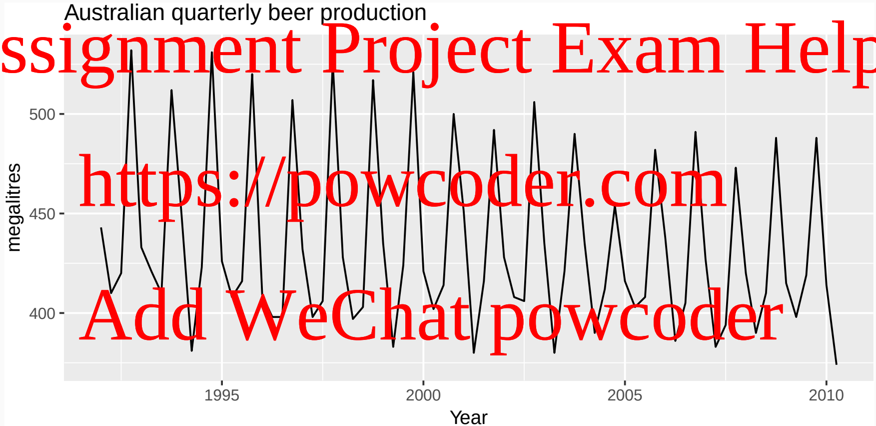
Stationary?


Price of a dozen eggs in 1993 dollars
300
200
100
$
1900
1920
1940
1960
1980
Year

Stationary?


Annual Canadian Lynx Trappings
Number trapped
6000
4000
2000
0
1820
1840
1860
1880
1900
1920
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 $(y_t, \ldots, y_{t+s})$ does not depend on t.

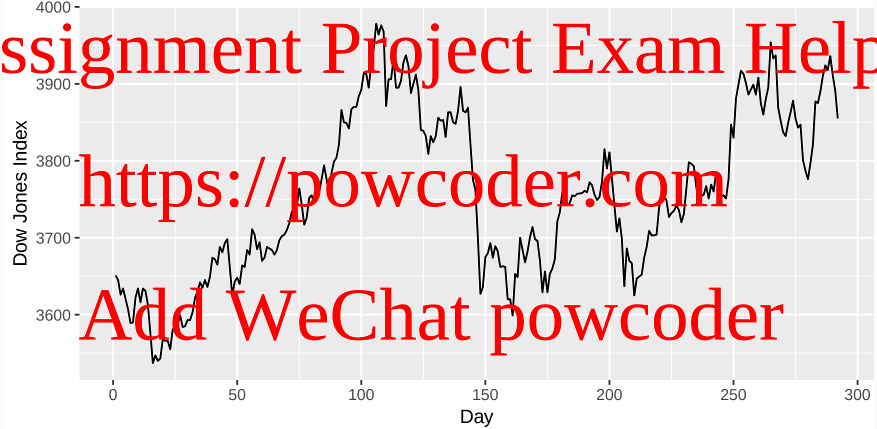
- Transformations help to **stabilize the variance**.
- For ARIMA modelling, we also need to **stabilize the mean**.

Non-stationarity in the mean

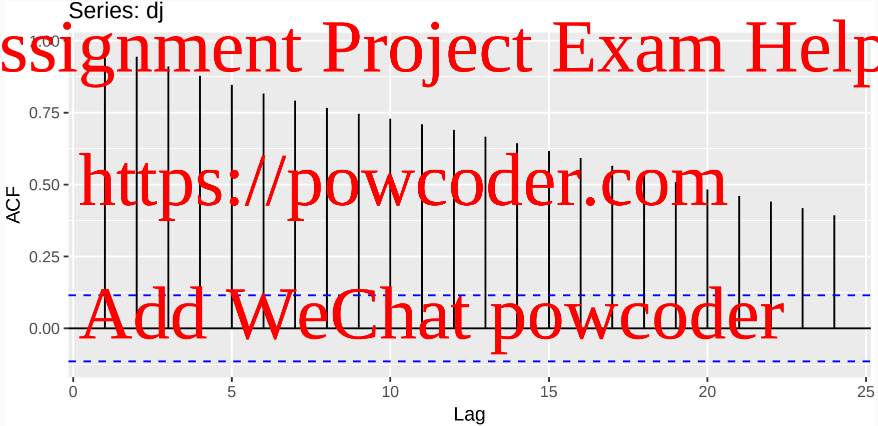
Identifying non-stationary series

- 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.

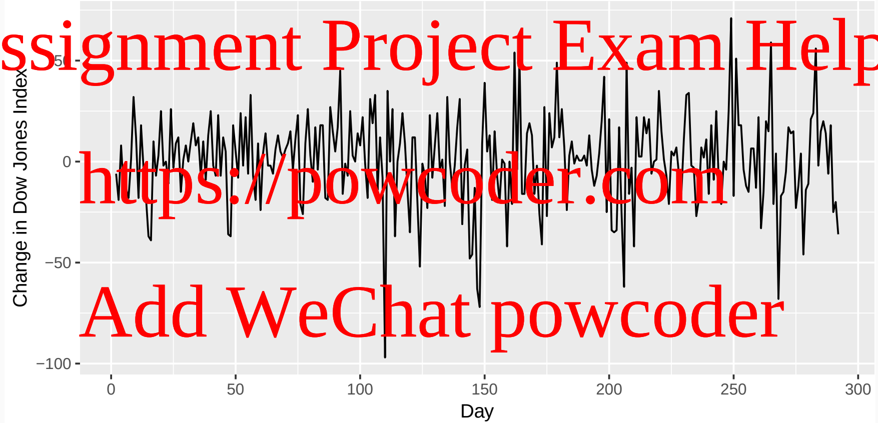
Example: Dow-Jones index



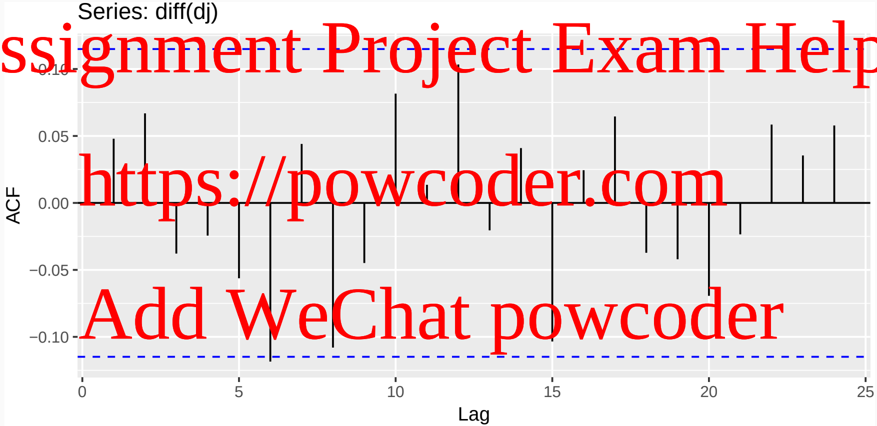
Example: Dow-Jones index



Example: Dow-Jones index



Example: Dow-Jones index



Differencing

- Differencing helps to **stabilize the mean**.
- The differenced series is the *change* between each observation in the original series: $y'_t = y_t - y_{t-1}$.
- The differenced series will have only $T - 1$ values since it is not possible to calculate a difference y'_1 for the first observation.

Second-order differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

Second-order differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

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

Second-order differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

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

- y''_t will have $T-2$ values.
- In practice, it is almost never necessary to go beyond second-order differences.

Seasonal differencing

A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

Seasonal differencing

A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

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

where m = number of seasons.

Seasonal differencing

A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

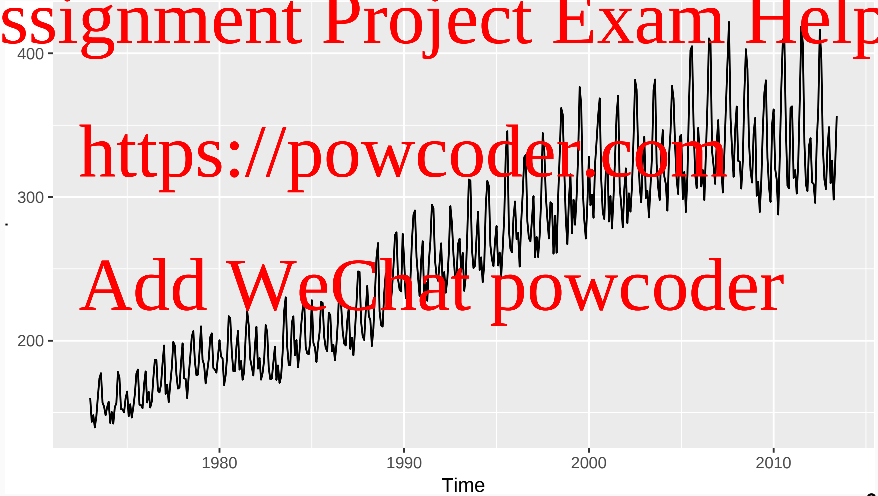
$$y'_t = y_t - y_{t-m}$$

where m = number of seasons.

- For monthly data $m = 12$.
- For quarterly data $m = 4$.

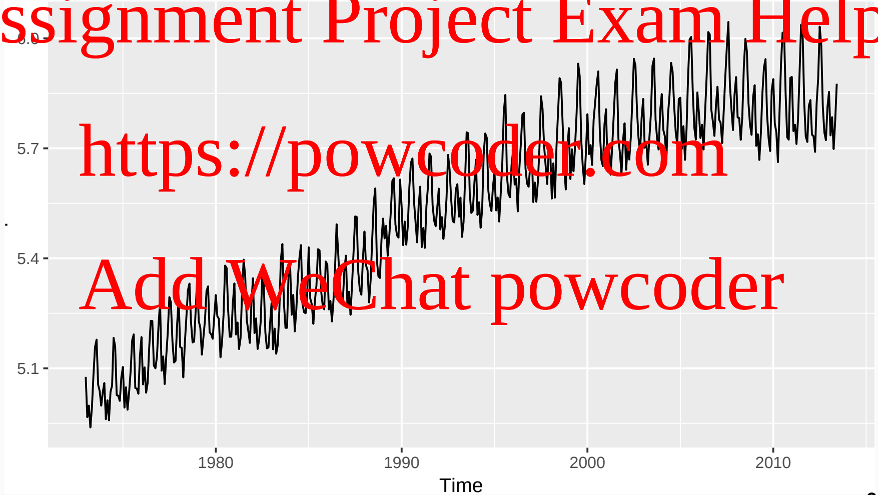
Electricity production

```
usmelec %>% autoplot()
```



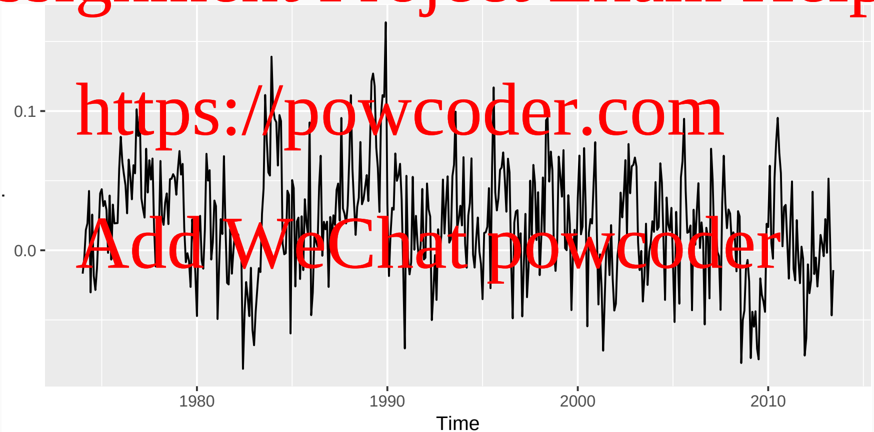
Electricity production

```r
usmelec %>% log() %>% autoplot()
```

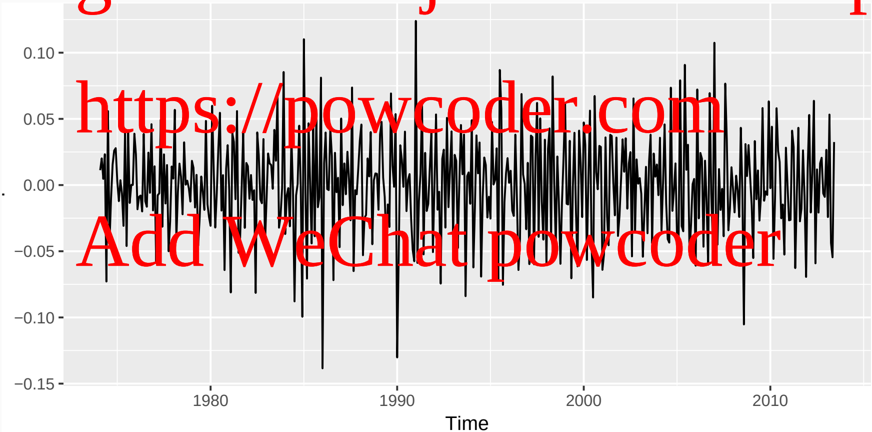
Electricity production

```
usmelec %>% log() %>% diff(lag=12) %>%
  autoplot()
```

Electricity production

```
usmelec %>% log() %>% diff(lag=12) %>%
  diff(lag=1) %>% autoplot()
```

Electricity production

- Seasonally differenced series is closer to being stationary.
- Remaining non-stationarity can be removed with further first difference.

If $y'_t = y_t - y_{t-12}$ denotes seasonally differenced series, then twice-differenced series is given by:

$$\begin{aligned} y^*_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-12}) - (y_{t-1} - y_{t-13}) \\ &= y_t - y_{t-1} - y_{t-12} + y_{t-13} \,. \end{aligned}$$

Seasonal differencing

When both seasonal and first differences are applied. . .

Seasonal differencing

When both seasonal and first differences are applied. . .

- it makes no difference which is done first—the result will be the same.
- If seasonality is strong, we recommend that seasonal differencing be done first because sometimes the resulting series will be stationary and there will be no need for further first difference.

Seasonal differencing

When both seasonal and first differences are applied. . .

- it makes no difference which is done first—the result will be the same.
- If seasonality is strong, we recommend that seasonal differencing be done first because sometimes the resulting series will be stationary and there will be no need for further first difference.

It is important that if differencing is used, the differences are interpretable.

Interpretation of differencing

- first differences are the change between **one observation and the next**;

- seasonal differences are the change between **one year to the next**.

Interpretation of differencing

- first differences are the change between **one observation and the next**;

- seasonal differences are the change between **one year to the next**.

But taking lag 3 differences for yearly data, for example, results in a model which cannot be sensibly interpreted.

Unit root tests

Statistical tests to determine the required order of differencing.

1. Augmented Dickey Fuller test: null hypothesis is that the data are non-stationary.
2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: null hypothesis is that the data are stationary. (Will use this for our analysis)
3. Other tests available for seasonal data.

KPSS test

```
library(urca)
summary(ur.kpss(goog))
```



```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 7 lags.
##
## Value of test-statistic is: 10.7223
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```


- Test statistic is much larger than the 1% critical value, indicating that the null hypothesis is rejected, i.e. data are not stationary.

KPSS test

```r
goog %>% diff() %>% ur.kpss() %>% summary()
```



```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 7 lags.
##
## Value of test-statistic is: 0.0324
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```


- Conclude that the differenced data are stationary.

Automatic selection of first difference

This process of using a sequence of KPSS tests to determine the appropriate number of first differences is carried out by the function `ndiffs()`.


```
ndiffs(goog)
```



```
## [1] 1
```

Automatic selection of seasonal difference

- STL decomposition: $y_t = T_t + S_t + R_t$
- Seasonal strength $F_s = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right)$
- If $F_s > 0.64$, do one seasonal difference. Otherwise, no seasonal difference required.
- Use the `nsdiffs()` function to test.

Automatic selection of seasonal difference

- STL decomposition: $y_t = T_t + S_t + R_t$
- Seasonal strength $F_s = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t+R_t)}\right)$
- If $F_s > 0.64$, do one seasonal difference. Otherwise, no seasonal difference required.
- Use the `nsdiffs()` function to test.


```
usmelec %>% log() %>% nsdiffs()
```

```
## [1] 1
```

```
usmelec %>% log() %>% diff(lag=12) %>% ndiffs()
```

```
## [1] 1
```

- These functions suggest performing both a seasonal and first difference.

Self-practice

For the `visitors` series, find an appropriate differencing (after transformation if necessary) to obtain stationary data.

Backshift notation

A very useful notational device is the backward shift operator, B, which is used as follows:

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

Backshift notation

A very useful notational device is the backward shift operator, B, which is used as follows:

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

In other words, B, operating on y_t, has the effect of **shifting the data back one period**.

Backshift notation

A very useful notational device is the backward shift operator, B, which is used as follows:

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

In other words, B, operating on y_t, has the effect of **shifting the data back one period**. Two applications of B to y_t **shifts the data back two periods**:

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

Backshift notation

A very useful notational device is the backward shift operator, B, which is used as follows:

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

In other words, B, operating on y_t, has the effect of **shifting the data back one period**. Two applications of B to y_t **shifts the data back two periods**:

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

For monthly data, if we wish to shift attention to "the same month last year," then B^{12} is used, and the notation is $B^{12}y_t = y_{t-12}$.

Backshift notation

The backward shift operator is convenient for describing the process of *differencing*.

Backshift notation

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

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t .$$

Backshift notation

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

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t .$$

Note that a first difference is represented by $(1 - B)$.

Backshift notation

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

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t \; .$$

Note that a first difference is represented by $(1 - B)$.

Similarly, if second-order differences (i.e., first differences of first differences) have to be computed, then:

$$y''_t = y_t - 2y_{t-1} + y_{t-2} = (1 - B)^2 y_t \; .$$

Backshift notation

- Second-order difference is denoted $(1-B)^2$;
- *Second-order difference* is not the same as a *second difference*, which would be denoted $1-B^2$;

- In general, a dth-order difference can be written as
$$(1-B)^d y_t.$$

- A seasonal difference followed by a first difference can be written as
$$(1-B)(1-B^m)y_t\ .$$

Backshift notation

The "backshift" notation is convenient because the terms can be multiplied together to see the combined effect.

$$(1 - B)(1 - B^m)y_t = (1 - B - B^m + B^{m+1})y_t$$
$$= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}$$

Backshift notation

The "backshift" notation is convenient because the terms can be multiplied together to see the combined effect.

$$
\begin{aligned}
(1 - B)(1 - B^m)y_t &= (1 - B - B^m + B^{m+1})y_t \\
&= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}
\end{aligned}
$$

For monthly data, $m = 12$ and we obtain the same result as earlier.

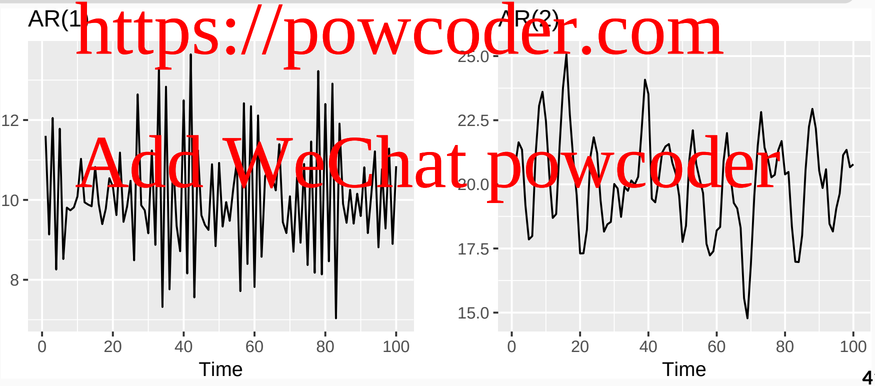
Outline

Autoregressive models

Autoregressive (AR) models:

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

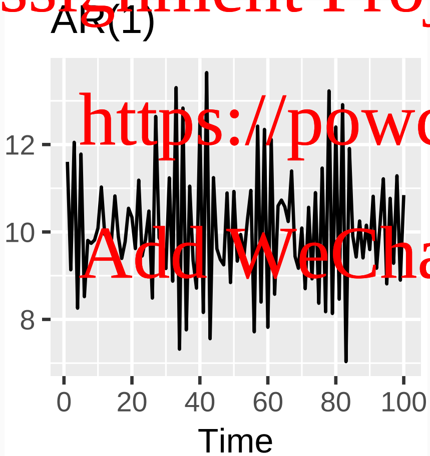
where ε_t is white noise. This is a multiple regression with **lagged values** of y_t as predictors.



AR(1) model

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

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



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

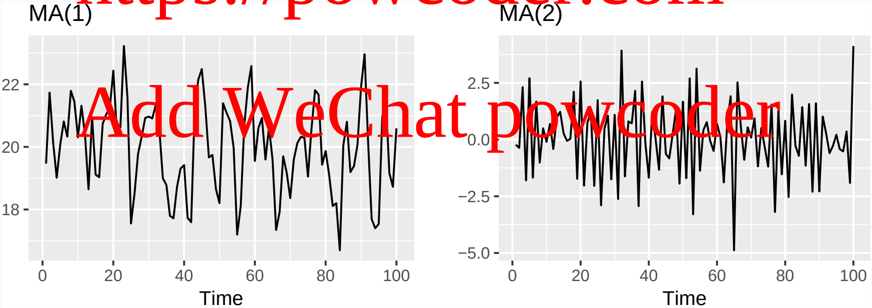
Complex roots of $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ lie outside the unit circle on the complex plane.

- 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 $p \geq 3$.
- Estimation software takes care of this.

Moving Average (MA) models:

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

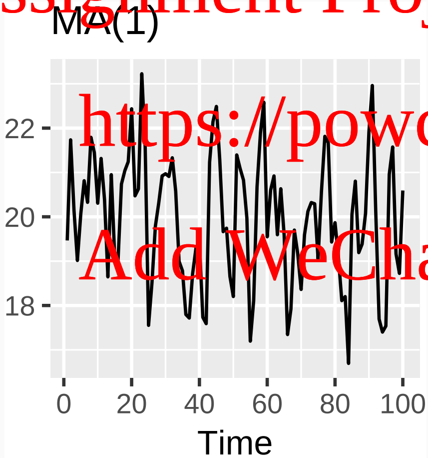
where ε_t is white noise. This is a multiple regression with **past *errors*** as predictors. *Don't confuse this with moving average smoothing!*



MA(1) model

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

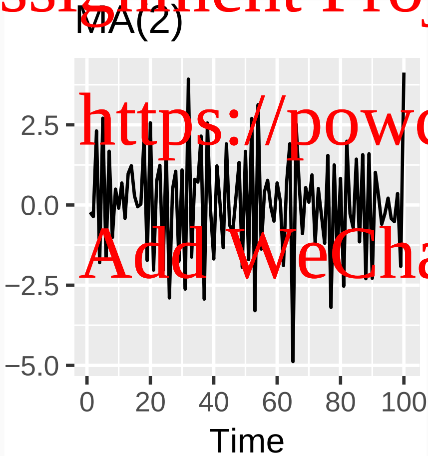
$\varepsilon_t \sim N(0,1), \quad T = 100.$



MA(2) model

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

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



MA(∞) models

It is possible to write any stationary AR(p) process as an MA(∞) process.

Example: AR(1)

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \varepsilon_t \\
&= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&\dots
\end{aligned}
$$

MA(∞) models

It is possible to write any stationary AR(p) process as an MA(∞) process.

Example: AR(1)

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \varepsilon_t \\
&= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&\dots
\end{aligned}
$$

Provided $-1 < \phi_1 < 1$:

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \cdots$$

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 $1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ 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 $q \geq 3$.
- Estimation software takes care of this.

ARMA models

Autoregressive Moving Average models:

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

ARMA models

Autoregressive Moving Average 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 AR coefficients ensure stationarity.
- Conditions on MA coefficients ensure invertibility.

ARIMA Models

Autoregressive Integrated Moving Average models:

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

- Combine ARMA model with **differencing**.
- $y'_t = (1 - B)^d y_t$ 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:

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

- ARIMA(1,1,1) model:

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

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 parameterisation

Intercept form

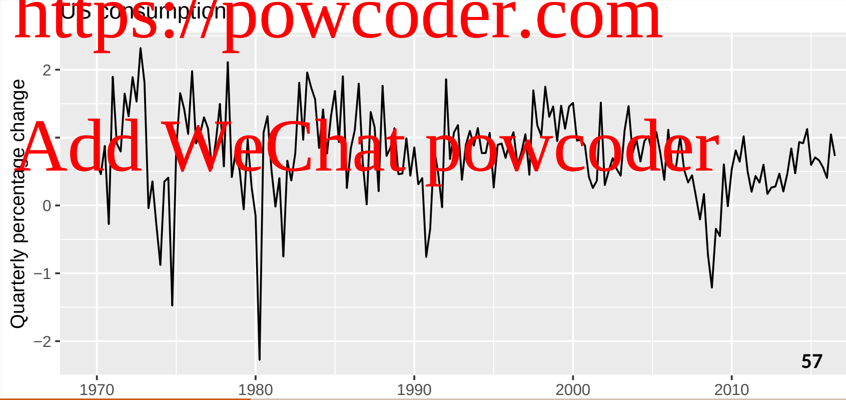
$$(1 - \phi_1 B - \cdots - \phi_p B^p) y'_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

Mean form (Parameterisation in R)

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

- $y'_t = (1 - B)^d y_t$
- μ is the mean of y'_t.
- Set $c = \mu(1 - \phi_1 - \cdots - \phi_p)$ to convert from mean form to intercept form

- Figure below shows quarterly percentage changes in US consumption expenditure.
- Although it is a quarterly series, there does NOT appear to be a seasonal pattern.
- Hence, fit a non-seasonal ARIMA model.



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
##       1.3908  -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"]))
```



```
## 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
```


“‘

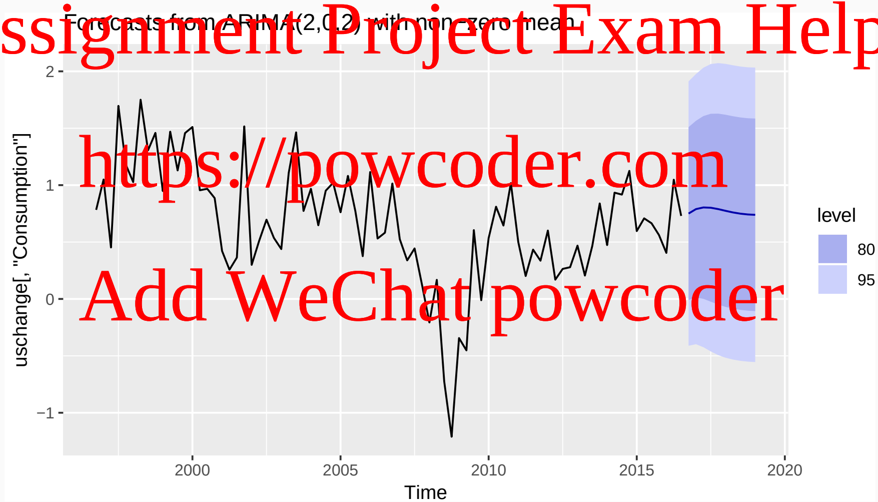
ARIMA(2,0,2) model:

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

where $c = 0.746 \times (1 - 1.391 + 0.581) = 0.142$ and ε_t is white noise with a standard deviation of $0.593 = \sqrt{0.351}$.

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 determined by the last few observations.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a straight line with intercept and slope determined by the last few observations.
- 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 with slope equal to the mean of the differenced data.
- If $c \neq 0$ and $d = 2$, the long-term forecasts will follow a quadratic trend.

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, $p \geq 2$ 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

Model Identification

- Question: How do we identify an ARIMA model? That is, how do we identify the values of p, d, q?

Model Identification

- Question: How do we identify an ARIMA model? That is, how do we identify the values of p, d, q?

- Rely on the sample ACF and sample PACF plots to determine appropriate values for p and q.
- To do this, will need to know the theoretical ACF and PACF of some common AR and MA models.

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.

α_k = kth partial autocorrelation coefficient

= equal to the estimate of ϕ_k in regression:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_k y_{t-k}.$$

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.

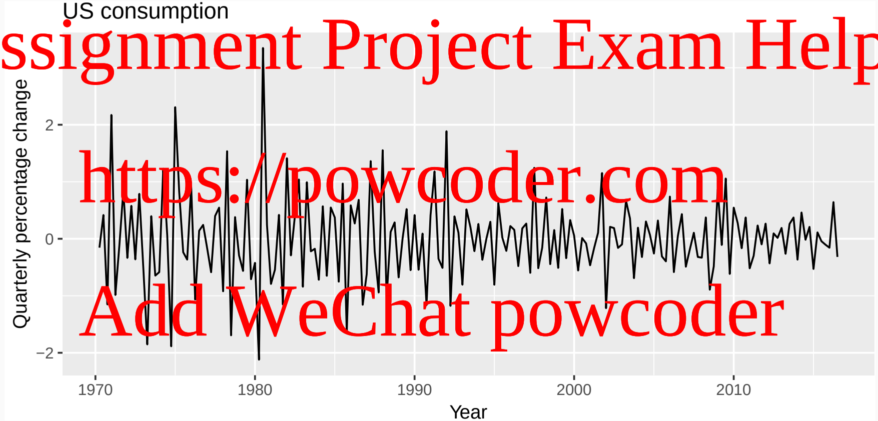
$$\alpha_k = k\text{th partial autocorrelation coefficient}$$

$$= \text{equal to the estimate of } \phi_k \text{ in regression:}$$

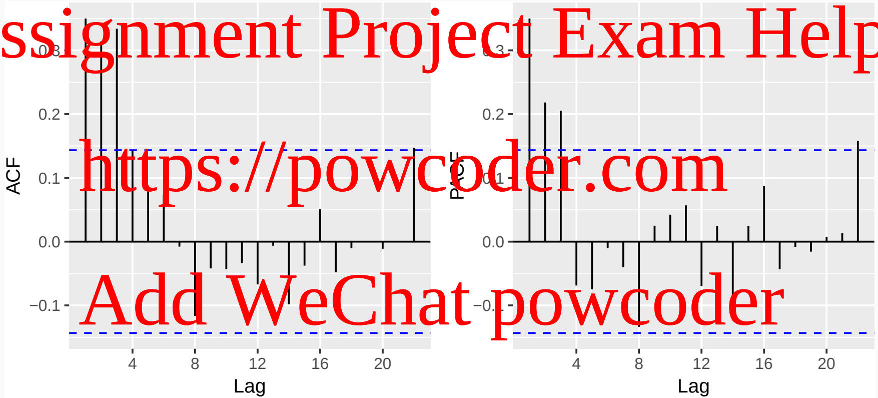
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_k y_{t-k}.$$

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

Example: US consumption



Example: US consumption



AR(1): ACF and PACF interpretation

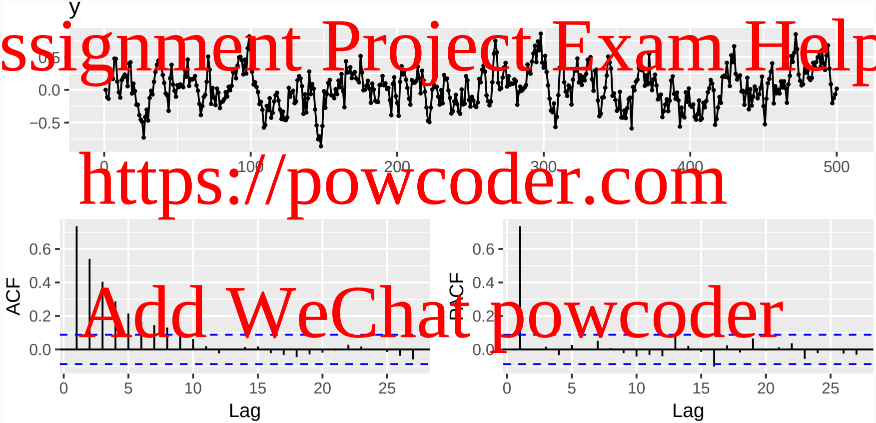
AR(1)

$$\rho_k = \phi_1^k \qquad \text{for } k = 1, 2, \ldots;$$
$$\alpha_1 = \phi_1 \qquad \alpha_k = 0 \qquad \text{for } k = 2, 3, \ldots.$$

So we have an AR(1) model when

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

AR(1): Simulation Example



AR(p): ACF and PACF interpretation

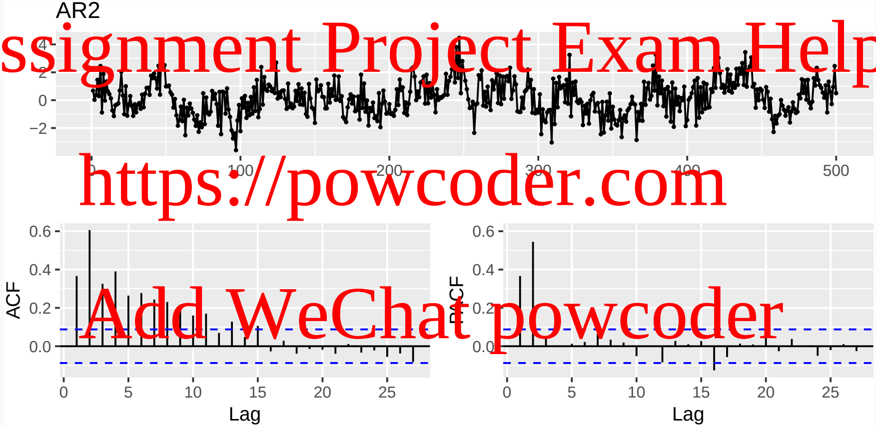
AR(p)

- ACF dies out in an exponential or damped sine-wave manner
- PACF has all zero spikes beyond the pth spike

So we have an AR(p) model when

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

AR(2): Simulation Example



MA(1): ACF and PACF interpretation

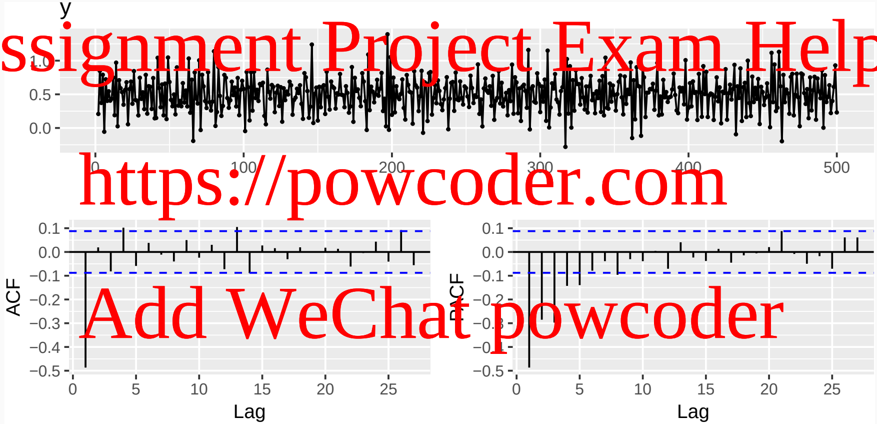
MA(1)

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

So we have an MA(1) model when

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

MA(1): Simulation Experiment



MA(q): ACF and PACF interpretation

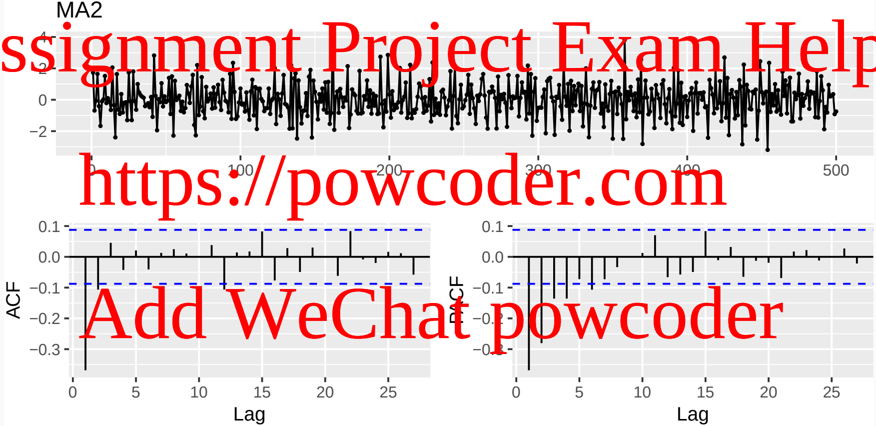
MA(q)

- PACF dies out in an exponential or damped sine-wave manner
- ACF has all zero spikes beyond the qth spike

So we have an MA(q) model when

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

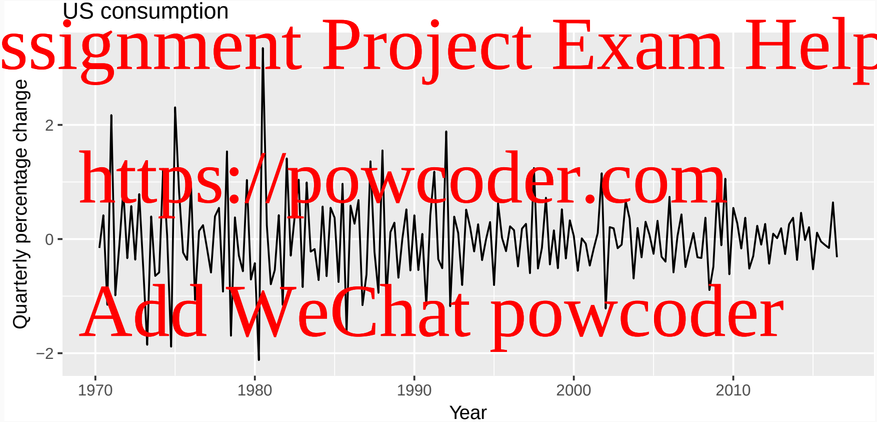
MA(2): Simulation Example



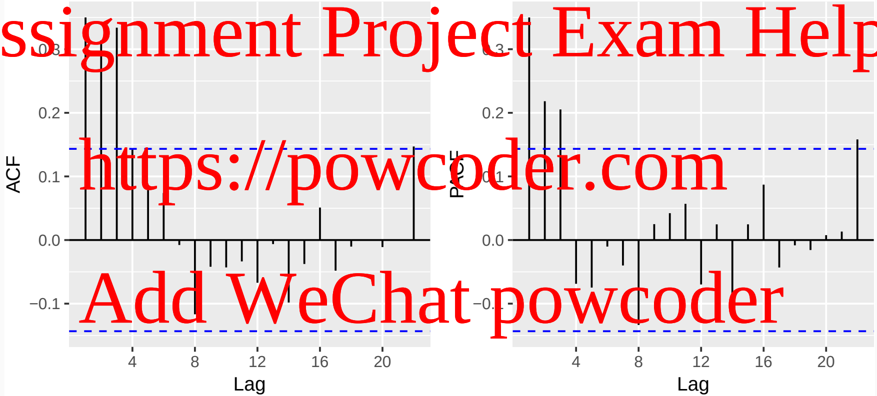
Model Identification: Some caveats

- If the data are from an ARIMA($p, d, 0$) or ARIMA($0, d, q$) model, then the ACF and PACF plots can be helpful in determining the value of p or q.
- If both p and q are positive, then the plots do not help in finding suitable values of p and q.
- It is common practice to identify two sets of models: 1 set of models consider $p = 0$ and $q > 0$ while the other set of models consider $p > 0$ and $q = 0$. The models can then be modified subsequently to have both p and $q > 0$.
- This step of model identification is largely a trial and error process. Known as Box-Jenkins methodology.

Example: US consumption



Example: US consumption



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, $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$.

- R uses **maximum likelihood estimation (MLE)** to estimate the ARIMA model. This technique finds the values of the parameters which **maximize the probability** of obtaining the data we have observed.

- For ARIMA models, 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.
- Different software will give different estimates due to different optimization algorithms.
- In practice, R will try to maximise the *log likelihood* of the

- The AIC is useful for determining the order of an ARIMA model.

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

- The AIC is useful for determining the order of an ARIMA model.

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

- The AIC is useful for determining the order of an ARIMA model.

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

- The AIC is useful for determining the order of an ARIMA model.

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

1 Introduction

2 Stationarity and differencing

3 Non-seasonal ARIMA models

4 Model identification

5 Estimation and order selection

6 ARIMA modelling in R

7 Forecasting

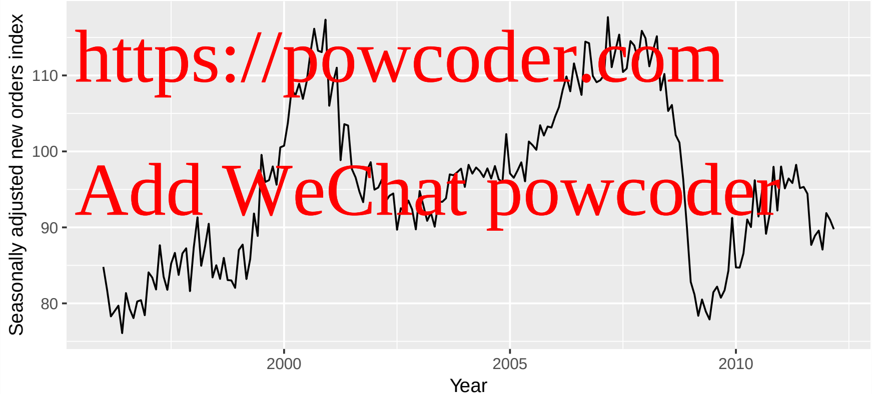
8 Seasonal ARIMA models

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.

Seasonally adjusted electrical equipment

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

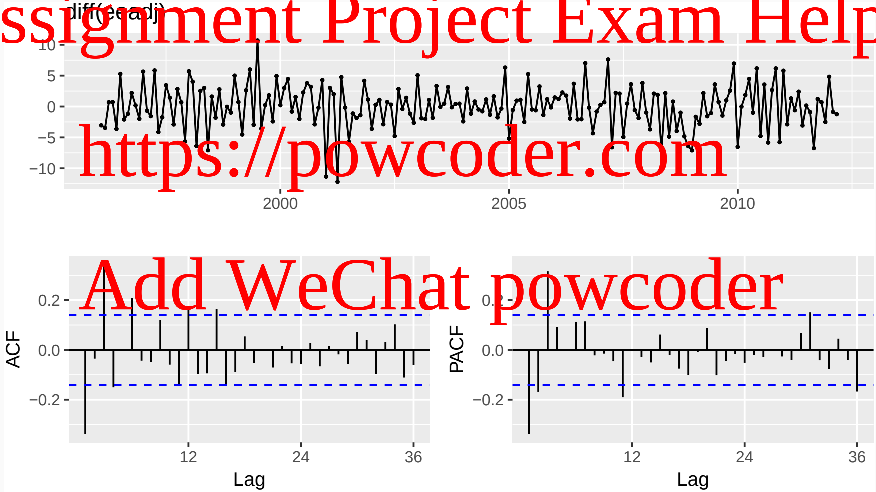



Seasonally adjusted electrical equipment

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
ggtsdisplay(diff(eeadj))
```

Seasonally adjusted electrical equipment

4. PACF is suggestive of AR(3). So initial candidate model is ARIMA(3,1,0). No other obvious candidates.

5. Fit ARIMA(3,1,0) model along with variations: ARIMA(4,1,0), ARIMA(2,1,0), ARIMA(3,1,1), etc. ARIMA(3,1,1) has smallest AICc value.

Seasonally adjusted electrical equipment


```
(fit <- Arima(eeadj, order=c(3,1,1)))
```

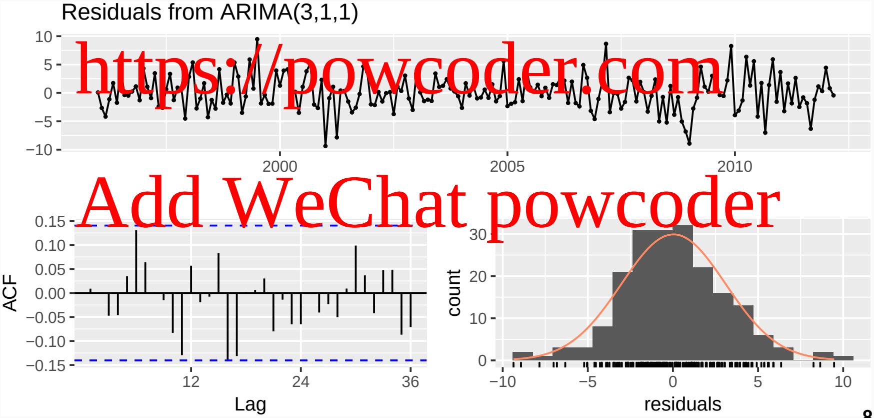


```
## Series: eeadj
## ARIMA(3,1,1)
##
## Coefficients:
##          ar1     ar2     ar3      ma1
##       0.0044  0.0916  0.3698  -0.3921
## s.e.  0.2201  0.0984  0.0669   0.2426
##
## sigma^2 estimated as 9.577:  log likelihood=-492.69
## AIC=995.38   AICc=995.7   BIC=1011.72
```

Seasonally adjusted electrical equipment

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


```
checkresiduals(fit)
```

##

Seasonally adjusted electrical equipment


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,1,1)
## Q* = 24.034, df = 20, p-value = 0.2409
##
## Model df: 4.   Total lags used: 24
```

Seasonally adjusted electrical equipment

```
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.

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.

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?

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

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:

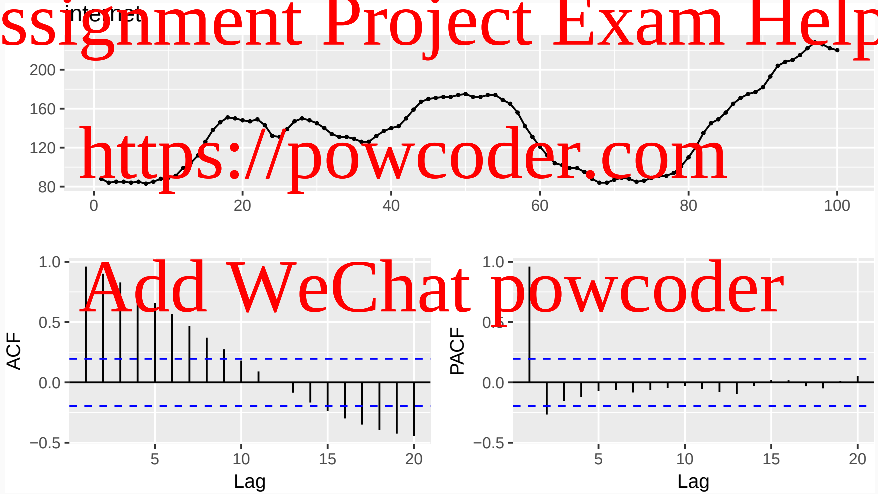
- vary one of $p, q,$ from current model by ± 1;
- p, q both vary from current model by ± 1;
- Include/exclude c from current model.

Model with lowest AICc becomes current model.

Repeat Step 2 until no lower AICc can be found.

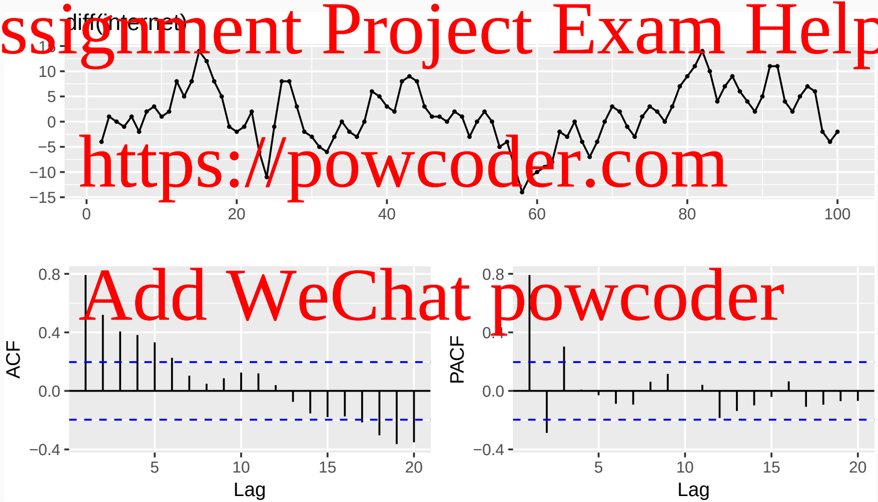
Choosing your own model

```
ggtsdisplay(internet)
```



Choosing your own model

```
ggtsdisplay(diff(internet))
```



Choosing your own model

```
(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
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

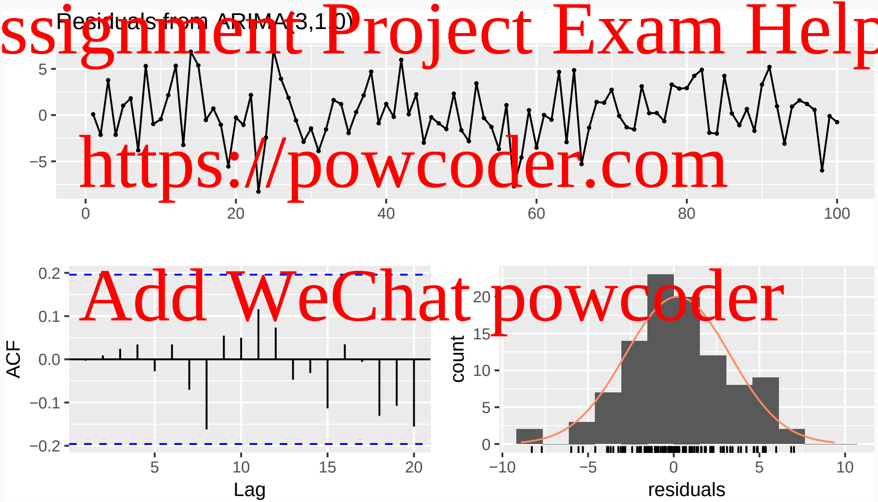
```r
auto.arima(internet, stepwise=FALSE,
  approximation=FALSE)
```



```
## 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
checkresiduals(fit)
```

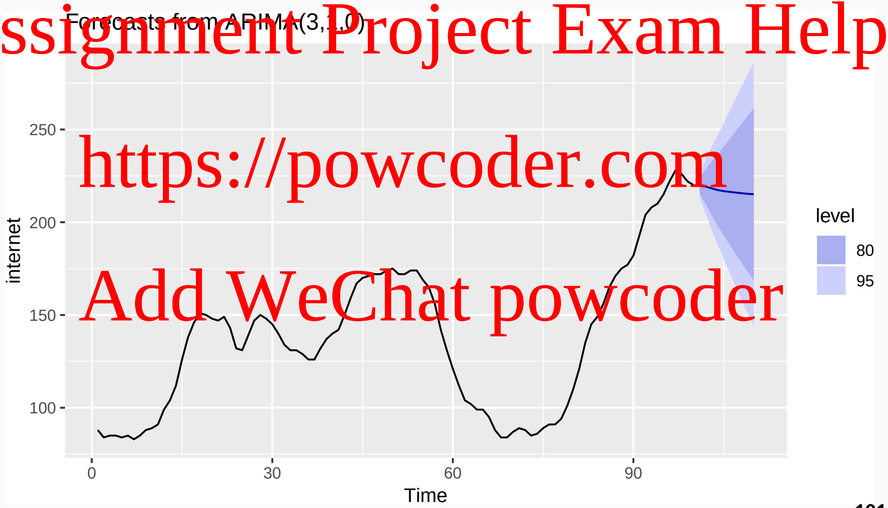



Choosing your own model

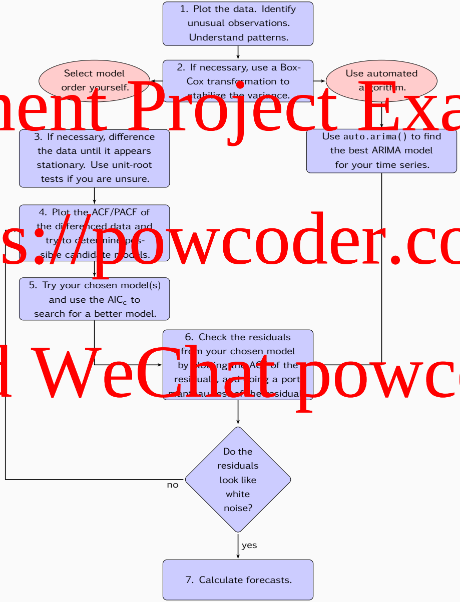

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,1,0)
## Q* = 4.4913, df = 7, p-value = 0.7218
##
## Model df: 3.   Total lags used: 10
```

Choosing your own model

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

Modelling procedure



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 $h = 2, 3, \ldots$.

Point forecasts

ARIMA(3,1,1) forecasts: Step 1

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B) y_t = (1 + \theta_1 B)\varepsilon_t,$$

Point forecasts

ARIMA(3,1,1) forecasts: Step 1

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B) y_t = (1 + \theta_1 B)\varepsilon_t,$$

$$\left[1 - (1 + \phi_1)B + (\phi_1 - \phi_2)B^2 + (\phi_2 - \phi_3)B^3 + \phi_3 B^4\right] y_t = (1 + \theta_1 B)\varepsilon_t,$$

Point forecasts

ARIMA(3,1,1) forecasts: Step 1

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)y_t = (1 + \theta_1 B)\varepsilon_t,$$

$$\left[1 - (1 + \phi_1)B + (\phi_1 - \phi_2)B^2 + (\phi_2 - \phi_3)B^3 + \phi_3 B^4\right] y_t = (1 + \theta_1 B)\varepsilon_t,$$

$$y_t - (1 + \phi_1)y_{t-1} + (\phi_1 - \phi_2)y_{t-2} + (\phi_2 - \phi_3)y_{t-3} + \phi_3 y_{t-4} = \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

ARIMA(3,1,1) forecasts: Step 1

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B) y_t = (1 + \theta_1 B)\varepsilon_t,$$

$$\left[1 - (1 + \phi_1)B + (\phi_1 - \phi_2)B^2 + (\phi_2 - \phi_3)B^3 + \phi_3 B^4\right] y_t = (1 + \theta_1 B)\varepsilon_t,$$

$$y_t - (1 + \phi_1)y_{t-1} + (\phi_1 - \phi_2)y_{t-2} + (\phi_2 - \phi_3)y_{t-3} + \phi_3 y_{t-4} = \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

$$y_t = (1 + \phi_1)y_{t-1} - (\phi_1 - \phi_2)y_{t-2} - (\phi_2 - \phi_3)y_{t-3} - \phi_3 y_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

Point forecasts (h=1)

$$y_t = (1 + \phi_1)y_{t-1} - (\phi_1 - \phi_2)y_{t-2} - (\phi_2 - \phi_3)y_{t-3}$$
$$- \phi_3 y_{t-4} + \varepsilon_t + \theta_1\varepsilon_{t-1}$$

Point forecasts (h=1)

$$
\begin{aligned}
y_t = (1 + \phi_1)y_{t-1} - (\phi_1 - \phi_2)y_{t-2} - (\phi_2 - \phi_3)y_{t-3} \\
- \phi_3 y_{t-4} + \varepsilon_t + \theta_1\varepsilon_{t-1}
\end{aligned}
$$

ARIMA(3,1,1) forecasts: Step 2

$$
\begin{aligned}
y_{T+1} = (1 + \phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2} \\
- \phi_3 y_{T-3} + \varepsilon_{T+1} + \theta_1\varepsilon_T.
\end{aligned}
$$

Point forecasts (h=1)

$$y_t = (1+\phi_1)y_{t-1} - (\phi_1 - \phi_2)y_{t-2} - (\phi_2 - \phi_3)y_{t-3}$$
$$- \phi_3 y_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ARIMA(3,1,1) forecasts: Step 2

$$y_{T+1} = (1+\phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2}$$
$$- \phi_3 y_{T-3} + \varepsilon_{T+1} + \theta_1 \varepsilon_T.$$

ARIMA(3,1,1) forecasts: Step 3

$$\hat{y}_{T+1|T} = (1+\phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2}$$
$$- \phi_3 y_{T-3} + \theta_1 e_T.$$

Point forecasts (h=2)

$$y_t = (1 + \phi_1)y_{t-1} - (\phi_1 - \phi_2)y_{t-2} - (\phi_2 - \phi_3)y_{t-3}$$
$$- \phi_3 y_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Point forecasts (h=2)

$$\begin{aligned} y_t = (1+\phi_1)y_{t-1} - (\phi_1-\phi_2)y_{t-2} - (\phi_2-\phi_3)y_{t-3} \\ - \phi_3 y_{t-4} + \varepsilon_t + \theta_1\varepsilon_{t-1} \end{aligned}$$

ARIMA(3,1,1) forecasts: Step 2

$$\begin{aligned} y_{T+2} = (1+\phi_1)y_{T+1} - (\phi_1-\phi_2)y_{T} - (\phi_2-\phi_3)y_{T-1} \\ - \phi_3 y_{T-2} + \varepsilon_{T+2} + \theta_1\varepsilon_{T+1}. \end{aligned}$$

Point forecasts (h=2)

$$y_t = (1+\phi_1)y_{t-1} - (\phi_1 - \phi_2)y_{t-2} - (\phi_2 - \phi_3)y_{t-3} - \phi_3 y_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ARIMA(3,1,1) forecasts: Step 2

$$y_{T+2} = (1+\phi_1)y_{T+1} - (\phi_1 - \phi_2)y_T - (\phi_2 - \phi_3)y_{T-1} - \phi_3 y_{T-2} + \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1}.$$

ARIMA(3,1,1) forecasts: Step 3

$$\hat{y}_{T+2|T} = (1+\phi_1)\hat{y}_{T+1|T} - (\phi_1 - \phi_2)y_T - (\phi_2 - \phi_3)y_{T-1} - \phi_3 y_{T-2}.$$

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):

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

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

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):

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

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

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):

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

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

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

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

Self-Practice

For the `usgdp` data:

- if necessary, find a suitable Box-Cox transformation for the data;
- fit a suitable ARIMA model to the transformed data using `auto.arima()`;
- check the residual diagnostics;
- produce forecasts of your fitted model. Do the forecasts look reasonable?

Outline

1. Introduction
2. Stationarity and differencing
3. Non-seasonal ARIMA models
4. Model identification
5. Estimation and order selection
6. ARIMA modelling in R
7. Forecasting
8. **Seasonal ARIMA models**

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

where m = number of observations per year.

Seasonal ARIMA models

E.g., ARIMA$(1,1,1)(1,1,1)_4$ model (without constant)

Seasonal ARIMA models

E.g., ARIMA(1, 1, 1)(1, 1, 1)$_4$ model (without constant)

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

Seasonal ARIMA models

E.g., ARIMA$(1, 1, 1)(1, 1, 1)_4$ model (without constant)

$$\underbrace{(1-\phi_1 B)}_{\text{Non-seasonal AR(1)}}\underbrace{(1-\Phi_1 B^4)}_{\text{Seasonal AR(1)}}\underbrace{(1-B)}_{\text{Non-seasonal difference}}\underbrace{(1-B^4)}_{\text{Seasonal difference}}y_t = \underbrace{(1+\theta_1 B)}_{\text{Non-seasonal MA(1)}}\underbrace{(1+\Theta_1 B^4)}_{\text{Seasonal MA(1)}}\varepsilon_t.$$

E.g., ARIMA$(1,1,1)(1,1,1)_4$ model (without constant)

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

All the factors can be multiplied out and the general model written as follows:

$$\begin{aligned} y_t = {} & (1+\phi_1)y_{t-1} - \phi_1 y_{t-2} + (1+\Phi_1)y_{t-4} \\ & - (1+\phi_1+\Phi_1+\phi_1\Phi_1)y_{t-5} + (\phi_1+\phi_1\Phi_1)y_{t-6} \\ & - \Phi_1 y_{t-8} + (\Phi_1+\phi_1\Phi_1)y_{t-9} - \phi_1\Phi_1 y_{t-10} \\ & + \varepsilon_t + \theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-4} + \theta_1\Theta_1\varepsilon_{t-5}. \end{aligned}$$

Common ARIMA models

The US Census Bureau uses the following models most often:

ARIMA(0,1,1)(0,1,1)$_m$	with log transformation
ARIMA(0,1,2)(0,1,1)$_m$	with log transformation
ARIMA(2,1,0)(0,1,1)$_m$	with log transformation
ARIMA(0,2,2)(0,1,1)$_m$	with log transformation
ARIMA(2,1,2)(0,1,1)$_m$	with no transformation

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:

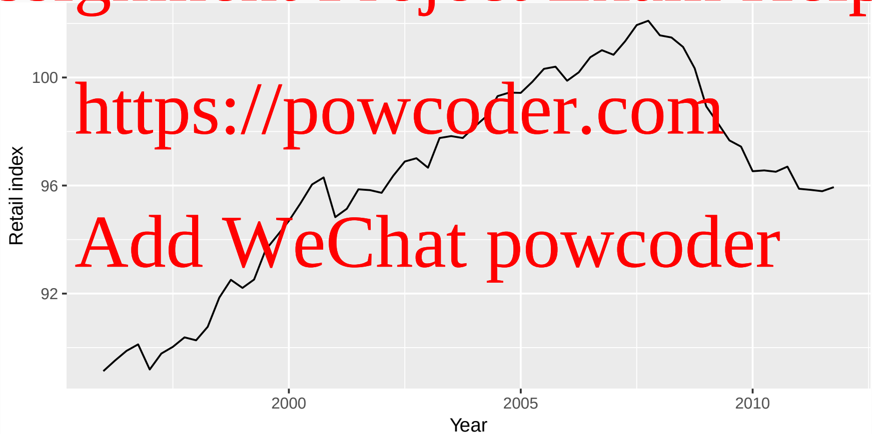
- a spike at lag 12 in the ACF but no other significant spikes.
- The PACF will show exponential decay in the seasonal lags; that is, at lags 12, 24, 36, ...

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.

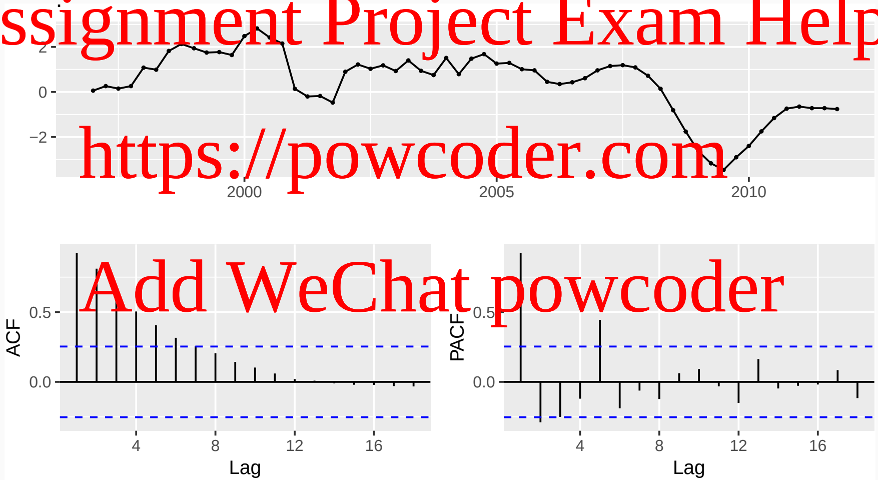
European quarterly retail trade

```r
autoplot(euretail) +
  xlab("Year") + ylab("Retail index")
```

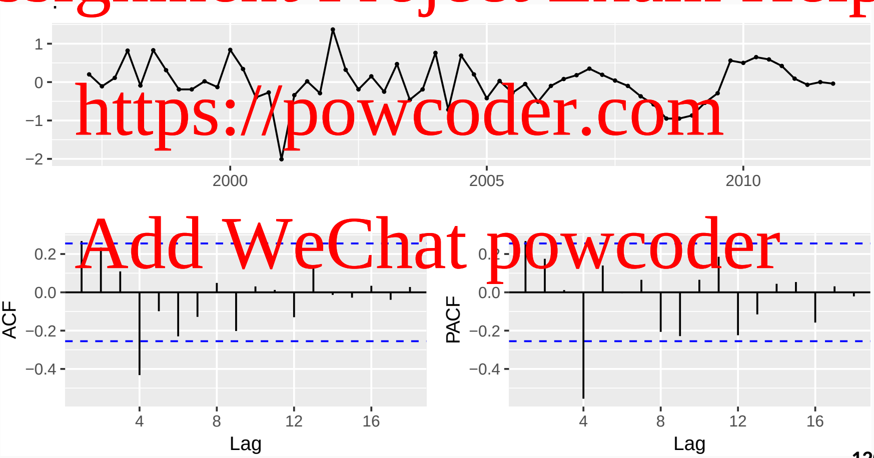
European quarterly retail trade

```r
euretail %>% diff(lag=4) %>% ggtsdisplay()
```



European quarterly retail trade

```
euretail %>% diff(lag=4) %>% diff() %>%
  ggtsdisplay()
```

European quarterly retail trade

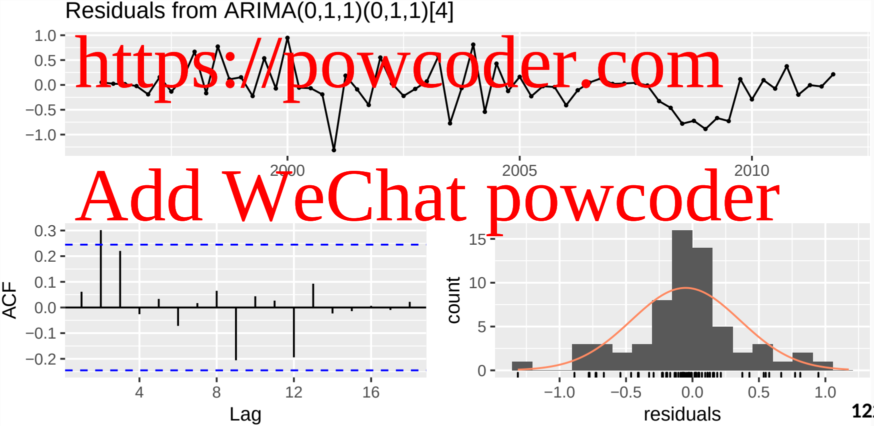
- $d = 1$ and $D = 1$ seems necessary.
- Significant spike at lag 1 in ACF suggests non-seasonal MA(1) component.

- 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 <- Arima(euretail, order=c(0,1,1),
  seasonal=c(0,1,1))
checkresiduals(fit)
```

European quarterly retail trade


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,3)(0,1,1)[4]
## Q* = 0.51128, df = 4, p-value = 0.9724
##
## Model df: 4.   Total lags used: 8
```

European quarterly retail trade

- ACF and PACF of residuals show significant spikes at lag 2, and maybe lag 3.
- AICc of ARIMA(0,1,2)(0,1,1)$_4$ model is 74.27.
- AICc of ARIMA(0,1,3)(0,1,1)$_4$ model is 68.39.

European quarterly retail trade

- ACF and PACF of residuals show significant spikes at lag 2, and maybe lag 3.
- AICc of ARIMA(0,1,2)(0,1,1)$_4$ model is 74.27.
- AICc of ARIMA(0,1,3)(0,1,1)$_4$ model is 68.39.

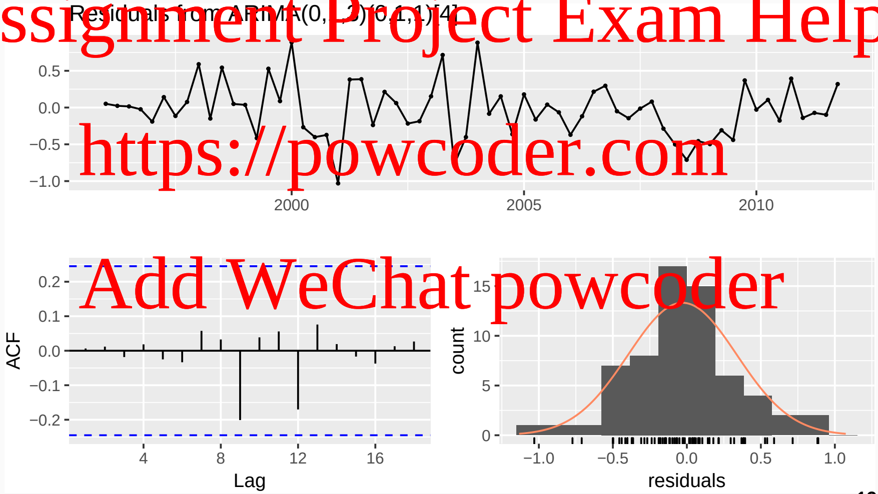

```
fit <- Arima(euretail, order=c(0,1,3),
  seasonal=c(0,1,1))
checkresiduals(fit)
```

European quarterly retail trade


```
## Series: euretail
## ARIMA(0,1,3)(0,1,1)[4]
##
## Coefficients:
##          ma1     ma2     ma3     sma1
##       0.2630  0.3694  0.4200  -0.6636
## s.e.  0.1237  0.1255  0.1294   0.1545
##
## sigma^2 estimated as 0.156:  log likelihood=-28.63
## AIC=67.26   AICc=68.39   BIC=77.65
```

European quarterly retail trade

```r
checkresiduals(fit)
```

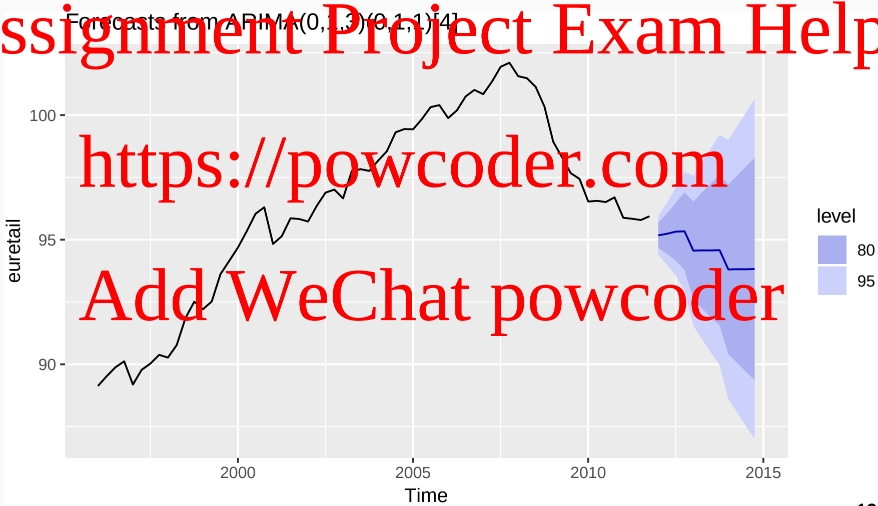



European quarterly retail trade


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,3)(0,1,1)[4]
## Q* = 0.51128, df = 4, p-value = 0.9724
##
## Model df: 4.   Total lags used: 8
```

European quarterly retail trade

```r
autoplot(forecast(fit, h=12))
```



European quarterly retail trade

```r
auto.arima(euretail)
```



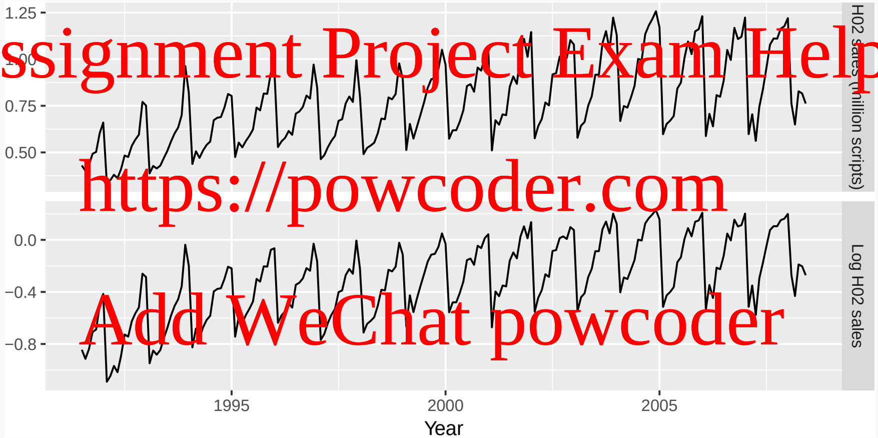
```
## Series: euretail
## ARIMA(1,1,2)(0,1,1)[4]
##
## Coefficients:
##          ar1      ma1     ma2     sma1
##       0.7362  -0.4663  0.2163  -0.8433
## s.e.  0.2243   0.1990  0.2101   0.1876
##
## sigma^2 estimated as 0.1587:  log likelihood=-29.62
## AIC=69.24   AICc=70.38   BIC=79.63
```

European quarterly retail trade

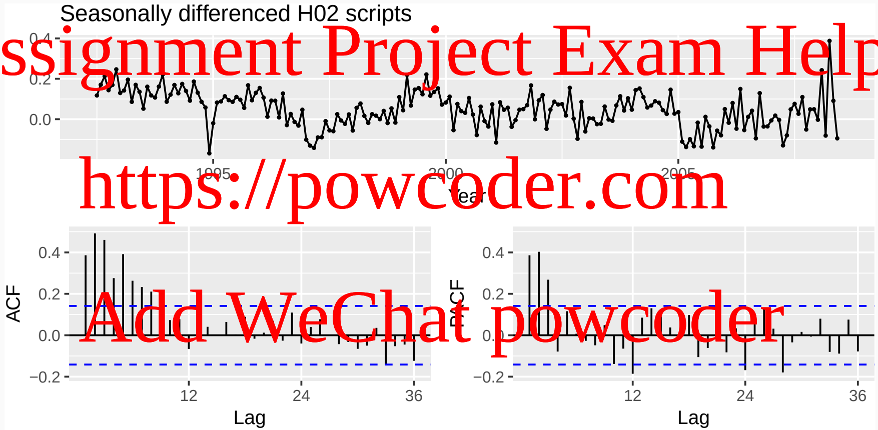
```r
auto.arima(euretail,
  stepwise=FALSE, approximation=FALSE)
```

```
## Series: euretail
## ARIMA(0,1,3)(0,1,1)[4]
##
## Coefficients:
##          ma1     ma2     ma3     sma1
##       0.2630  0.3694  0.4200  -0.6636
## s.e.  0.1237  0.1255  0.1294   0.1545
##
## sigma^2 estimated as 0.156:  log likelihood=-28.63
## AIC=67.26   AICc=68.39   BIC=77.65
```

Cortecosteroid drug sales



Cortecosteroid drug sales



Cortecosteroid drug sales

- Choose $D = 1$ and $d = 0$.
- Spikes in PACF at lags 12 and 24 suggest seasonal AR(2) term.
- Spikes in PACF suggests possible non-seasonal AR(3) term.
- Initial candidate model: ARIMA(3,0,0)(2,1,0)$_{12}$.

Cortecosteroid drug sales

Model	AICc
ARIMA(3,0,1)(0,1,2)$_{12}$	-485.48
ARIMA(3,0,1)(1,1,1)$_{12}$	-484.25
ARIMA(3,0,1)(0,1,1)$_{12}$	-483.67
ARIMA(3,0,1)(2,1,0)$_{12}$	-476.31
ARIMA(3,0,0)(2,1,0)$_{12}$	-475.12
ARIMA(3,0,2)(2,1,0)$_{12}$	-474.88
ARIMA(3,0,1)(1,1,0)$_{12}$	-463.40

Cortecosteroid drug sales

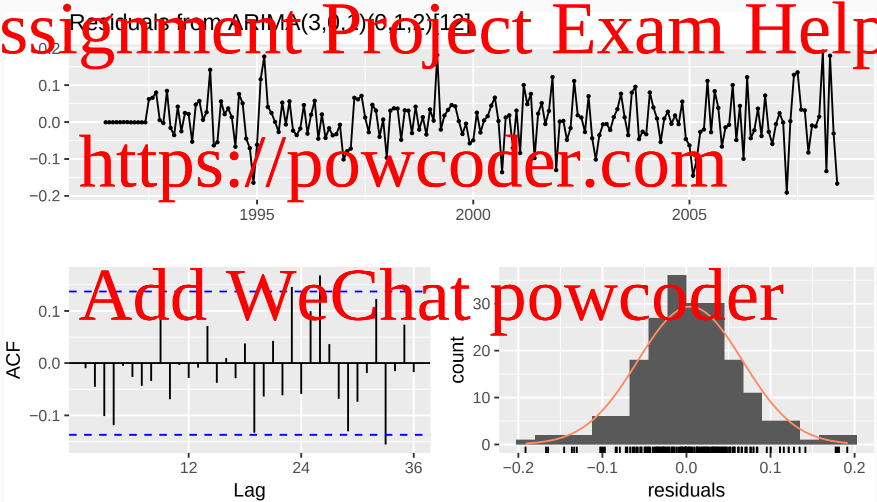
```r
(fit <- Arima(h02, order=c(3,0,1), seasonal=c(0,1,2),
  lambda=0))
```



```
## Series: h02
## ARIMA(3,0,1)(0,1,2)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ar1     ar2     ar3     ma1     sma1     sma2
##       -0.1603  0.5481  0.5678  0.3827  -0.5222  -0.1768
## s.e.   0.1636  0.0878  0.0942  0.1895   0.0861   0.0872
##
## sigma^2 estimated as 0.004278:  log likelihood=250.04
## AIC=-486.08   AICc=-485.48   BIC=-463.28
```

Cortecosteroid drug sales

```r
checkresiduals(fit, lag=36)
```

##

Cortecosteroid drug sales


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,0,1)(0,1,2)[12]
## Q* = 50.712, df = 30, p-value = 0.01045
##
## Model df: 6.   Total lags used: 36
```

Cortecosteroid drug sales

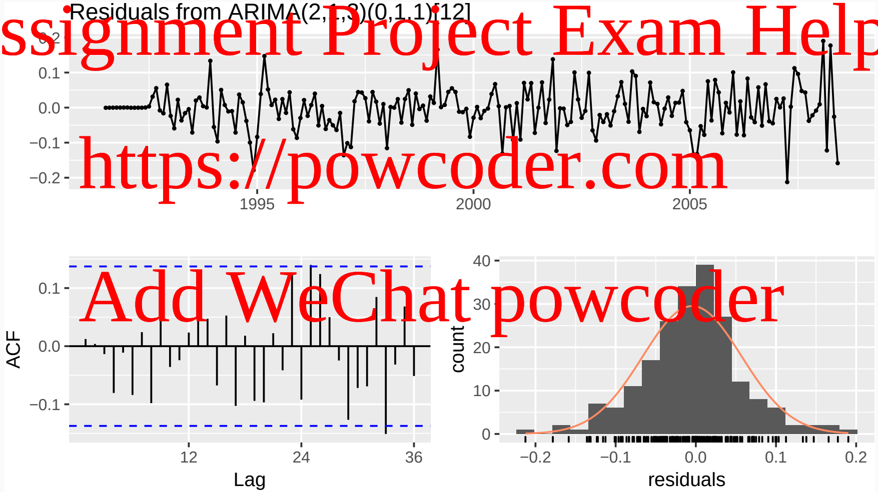
```
(fit <- auto.arima(h02, lambda=0))
```



```
## Series: h02
## ARIMA(2,1,3)(0,1,1)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ar1      ar2     ma1     ma2      ma3     sma1
##       -1.0194  -0.8351  0.1717  0.2578  -0.4206  -0.6528
## s.e.   0.1648   0.1203  0.2079  0.1177   0.1060   0.0657
##
## sigma^2 estimated as 0.004203:  log likelihood=250.8
## AIC=-487.6   AICc=-486.99   BIC=-464.83
```

Cortecosteroid drug sales

```r
checkresiduals(fit, lag=36)
```



##

Cortecosteroid drug sales


```
##
## 	Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,3)(0,1,1)[12]
## Q* = 46.149, df = 30, p-value = 0.03007
##
## Model df: 6.   Total lags used: 36
```

Cortecosteroid drug sales

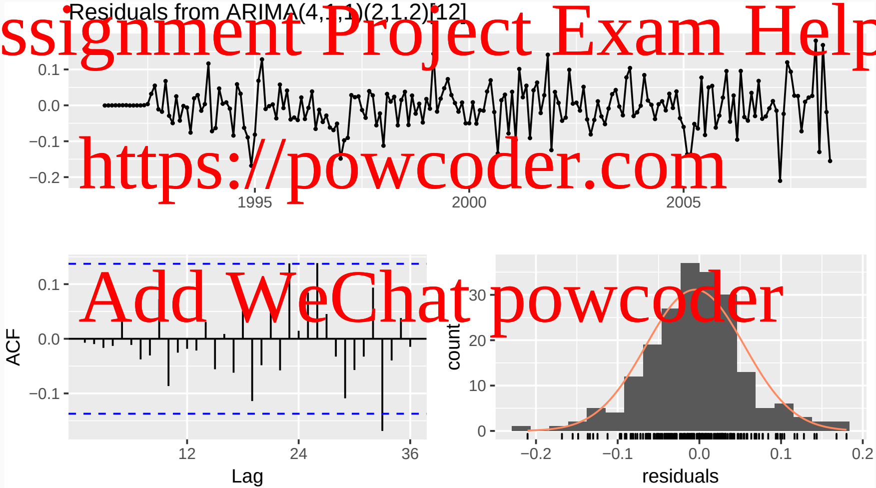
```r
(fit <- auto.arima(h02, lambda=0, max.order=9,
  stepwise=FALSE, approximation=FALSE))
```



```
## Series: h02
## ARIMA(4,1,1)(2,1,2)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##           ar1     ar2     ar3      ar4      ma1    sar1     sar2     sma1
##       -0.0425  0.2098  0.2017  -0.2273  -0.7424  0.6213  -0.3832  -1.2019
## s.e.   0.2167  0.1813  0.1144   0.0810   0.2074  0.2421   0.1185   0.2491
##         sma2
##       0.4959
## s.e.  0.2135
##
## sigma^2 estimated as 0.004049:  log likelihood=254.31
## AIC=-488.63   AICc=-487.4   BIC=-456.1
```

Cortecosteroid drug sales

```
checkresiduals(fit, lag=36)
```



##

Cortecosteroid drug sales


```
##
## 	Ljung-Box test
##
## data:  Residuals from ARIMA(4,1,1)(2,1,2)[12]
## Q* = 36.456, df = 27, p-value = 0.1057
##
## Model df: 9.   Total lags used: 36
```

Cortecosteroid drug sales

Training data: July 1991 to June 2006

Test data: July 2006–June 2008


```r
getrmse <- function(x,h,...)
{
  train.end <- time(x)[length(x)-h]
  test.start <- time(x)[length(x)-h+1]
  train <- window(x,end=train.end)
  test <- window(x,start=test.start)
  fit <- Arima(train,...)
  fc <- forecast(fit,h=h)
  return(accuracy(fc,test)[2,"RMSE"])
}
getrmse(h02,h=24,order=c(3,0,0),seasonal=c(2,1,0),lambda=0)
getrmse(h02,h=24,order=c(3,0,1),seasonal=c(2,1,0),lambda=0)
getrmse(h02,h=24,order=c(3,0,2),seasonal=c(2,1,0),lambda=0)
getrmse(h02,h=24,order=c(3,0,1),seasonal=c(1,1,0),lambda=0)
```

Cortecosteroid drug sales

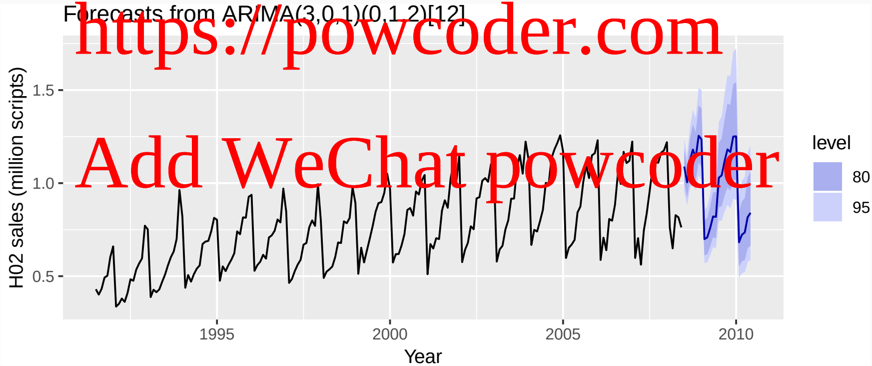
Model	RMSE
ARIMA(4,1,1)(2,1,2)[12]	0.0615
ARIMA(3,0,1)(0,1,2)[12]	0.0622
ARIMA(3,0,1)(1,1,1)[12]	0.0630
ARIMA(2,1,4)(0,1,1)[12]	0.0632
ARIMA(2,1,3)(0,1,1)[12]	0.0634
ARIMA(3,0,3)(0,1,1)[12]	0.0638
ARIMA(2,1,5)(0,1,1)[12]	0.0640
ARIMA(3,0,1)(0,1,1)[12]	0.0644
ARIMA(3,0,2)(0,1,1)[12]	0.0644
ARIMA(3,0,2)(2,1,0)[12]	0.0645
ARIMA(3,0,1)(2,1,0)[12]	0.0646
ARIMA(3,0,0)(2,1,0)[12]	0.0661

Cortecosteroid drug sales

- Models with lowest AICc values tend to give slightly better results than the other models.
- AICc comparisons must have the same orders of differencing. But RMSE test set comparisons can involve any models.
- Use the best model available, even if it does not pass all tests.

Cortecosteroid drug sales

```
fit <- Arima(h02, order=c(3,0,1), seasonal=c(0,1,2),
  lambda=0)
autoplot(forecast(fit)) +
  ylab("H02 sales (million scripts)") + xlab("Year")
```



Understanding Seasonal ARIMA models

- If $c = 0$ and $d + D = 0$, the long-term forecasts will go to zero.
- If $c = 0$ and $d + D = 1$, the long-term forecasts will go to a non-zero constant determined by the last few observations.
- If $c = 0$ and $d + D = 2$, the long-term forecasts will follow a straight line with intercept and slope determined by the last few observations.
- If $c \neq 0$ and $d + D = 0$, the long-term forecasts will go to the mean of the data.
- If $c \neq 0$ and $d + D = 1$, the long-term forecasts will follow a straight line with slope equal to the mean of the differenced data.
- If $c \neq 0$ and $d + D = 2$, the long-term forecasts will follow a quadratic trend.