

# Forecasting: principles and practice

**Rob J Hyndman**

2.3 Stationarity and differencing

# Outline

- 1 Stationarity**
- 2 Differencing
- 3 Unit root tests
- 4 Lab session 10
- 5 Backshift notation

# Stationarity

## Definition

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

A stationary series is:

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

# Stationarity

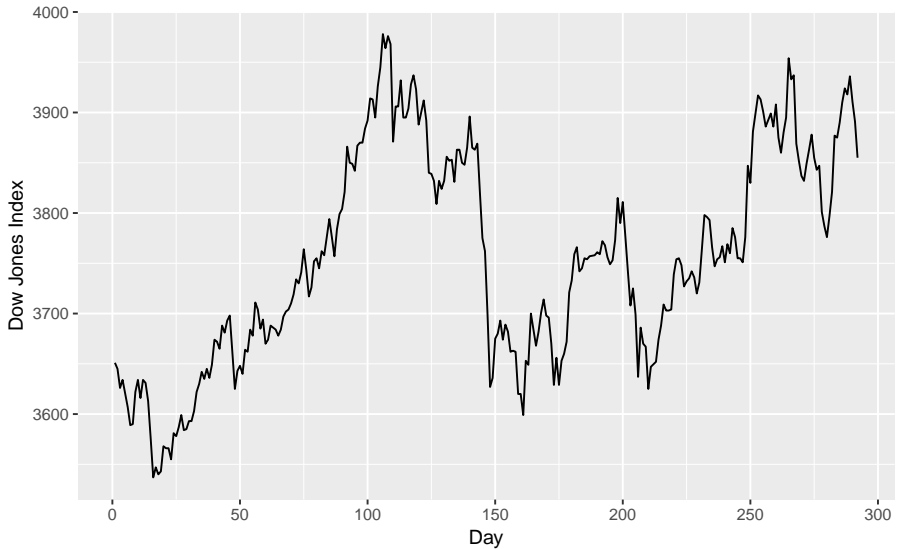
## Definition

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

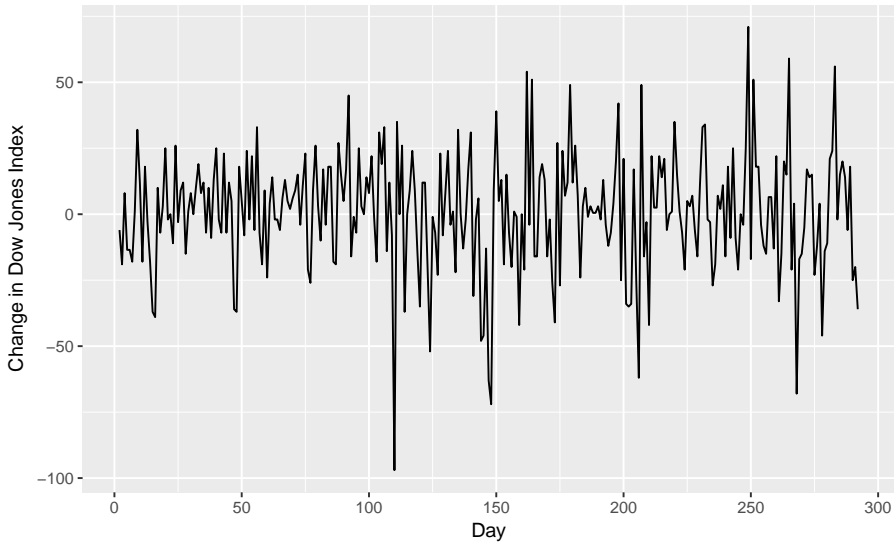
A **stationary series** is:

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

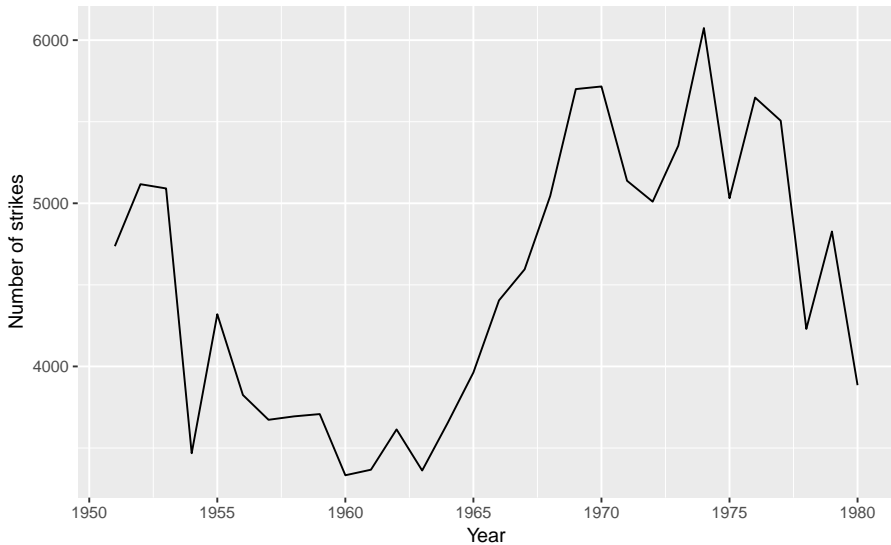
# Stationary?



# Stationary?

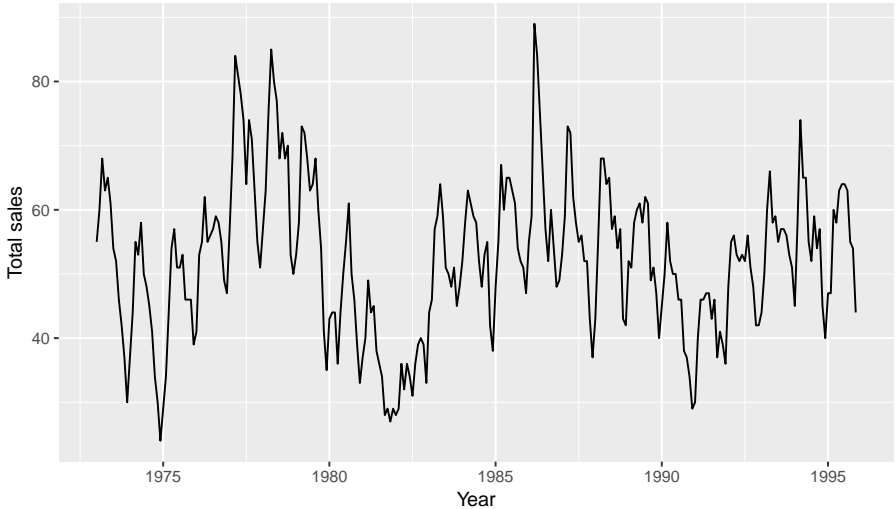


# Stationary?



# Stationary?

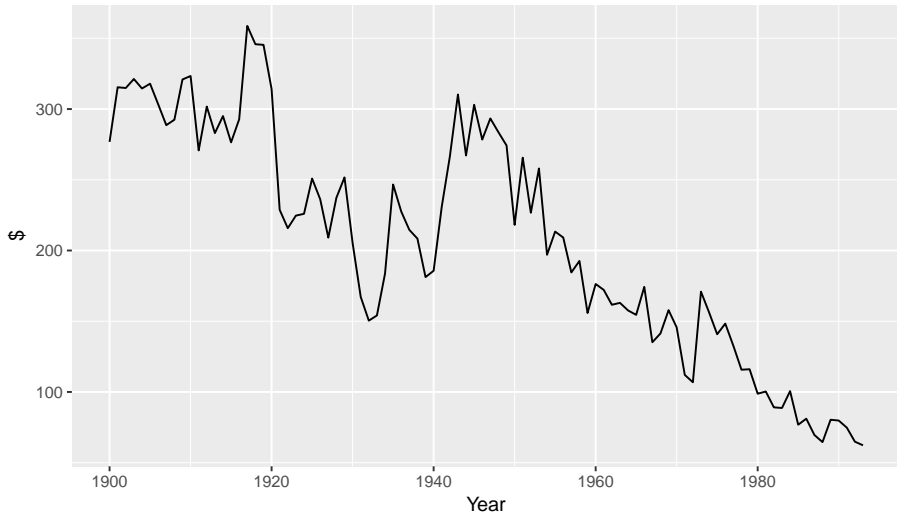
Sales of new one-family houses, USA





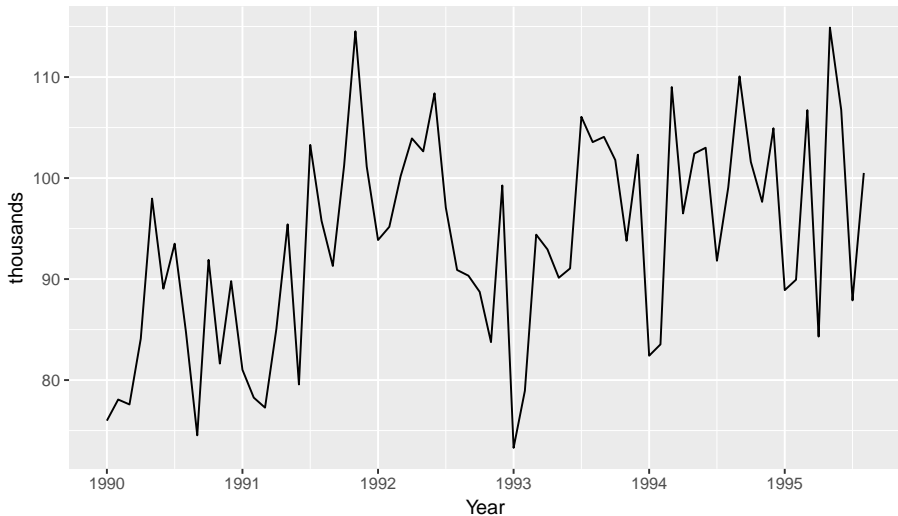
# Stationary?

Price of a dozen eggs in 1993 dollars



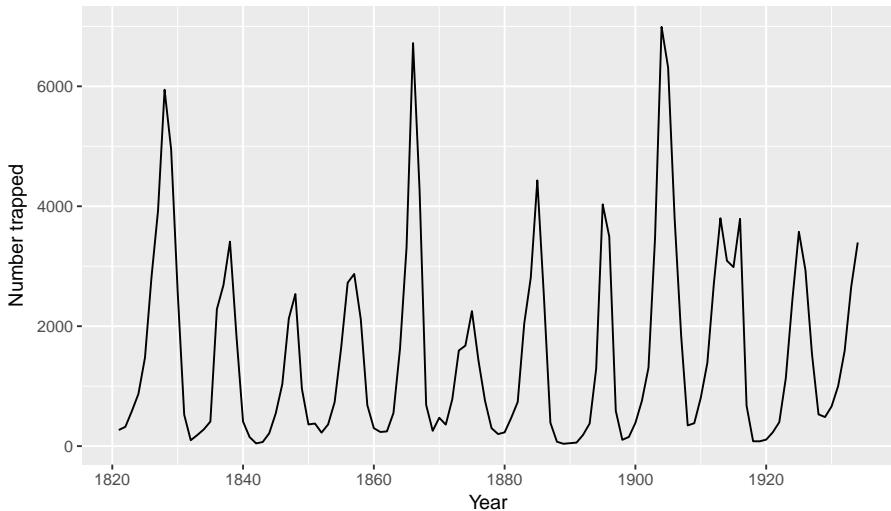
# Stationary?

Number of pigs slaughtered in Victoria



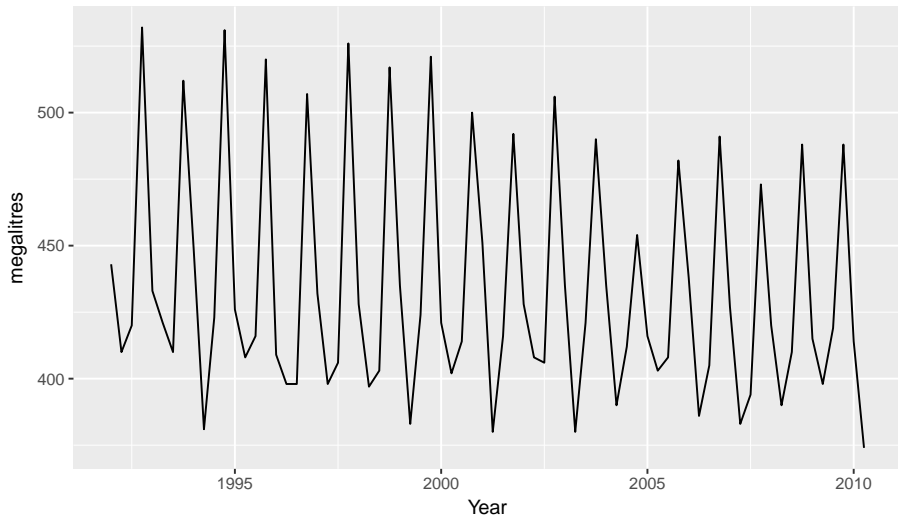
# Stationary?

Annual Canadian Lynx Trappings



# Stationary?

Australian quarterly beer production



# Stationarity

## Definition

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

Transformations help to **stabilize the variance**.

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

# Stationarity

## Definition

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

Transformations help to **stabilize the variance**.

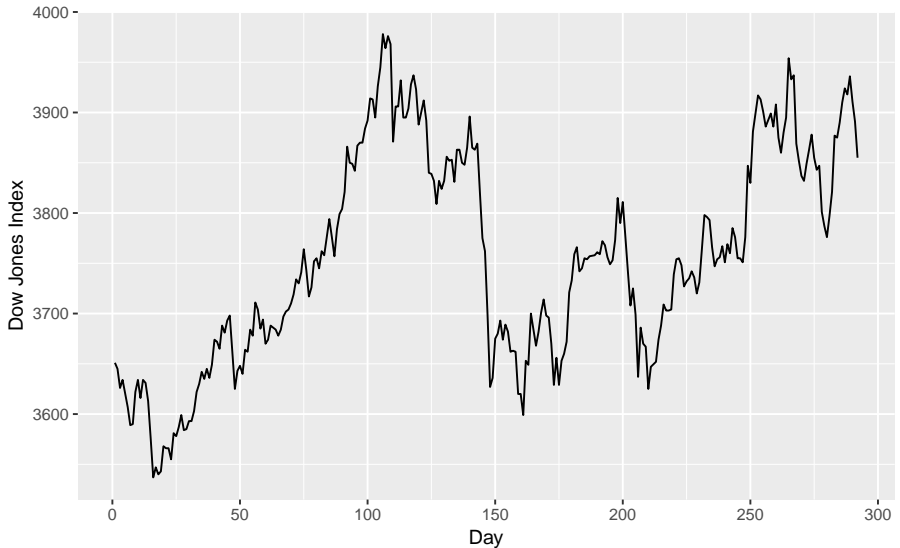
For ARIMA modelling, we also need to **stabilize the mean**.

# Non-stationarity in the mean

## Identifying non-stationary series

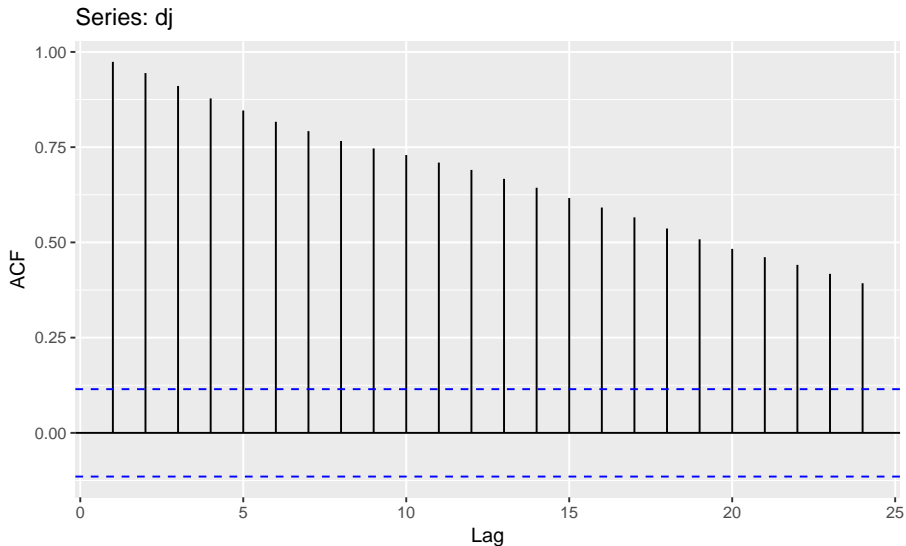
- time plot.
- The ACF of stationary data drops to zero relatively quickly
- The ACF of non-stationary data decreases slowly.
- For non-stationary data, the value of  $r_1$  is often large and positive.

# Example: Dow-Jones index

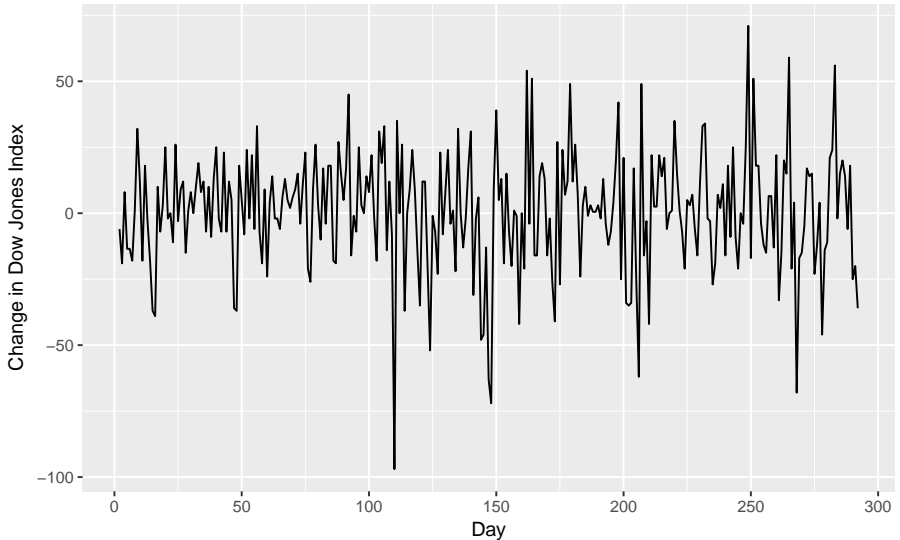




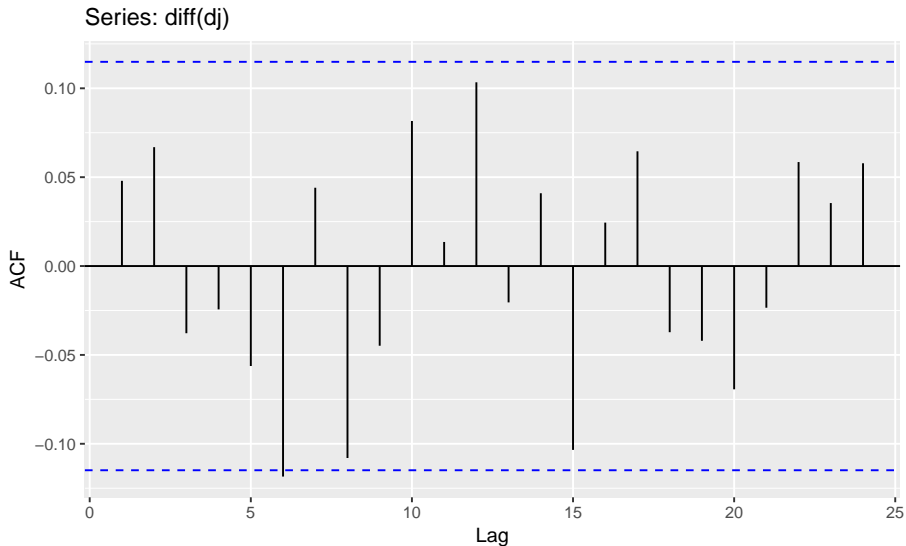
# Example: Dow-Jones index



# Example: Dow-Jones index



# Example: Dow-Jones index



# Outline

- 1 Stationarity
- 2 Differencing**
- 3 Unit root tests
- 4 Lab session 10
- 5 Backshift notation

# 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:

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

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

# 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.

# 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.

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

where  $m$  = number of seasons.

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

# 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  $m = 4$ .

# Seasonal differencing

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

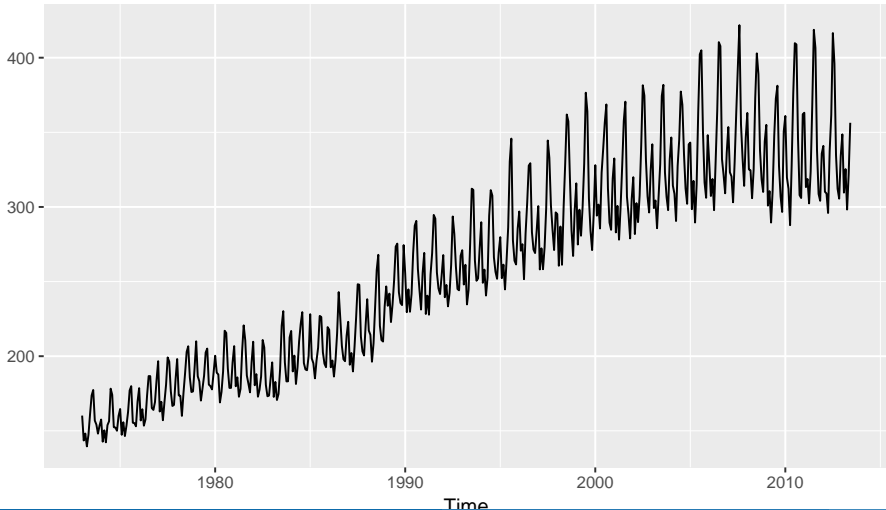
$$y'_t = y_t - y_{t-m}$$

where  $m$  = number of seasons.

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

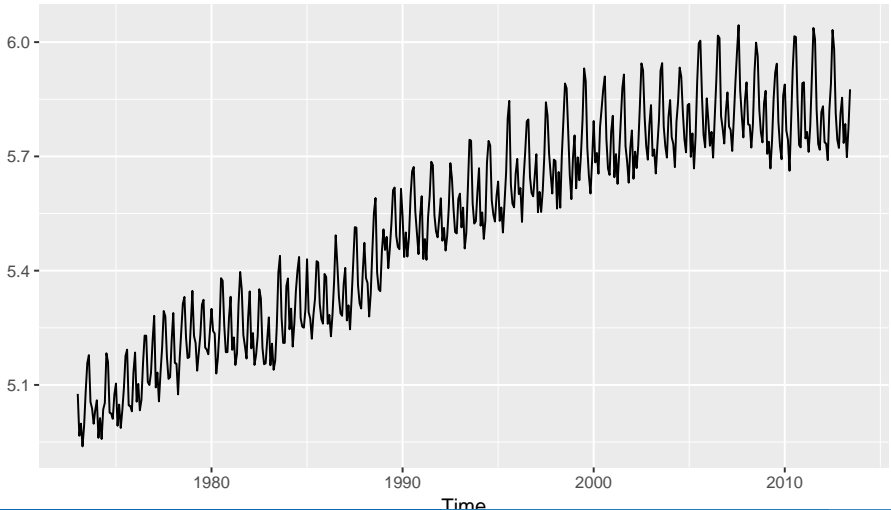
# Electricity production

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



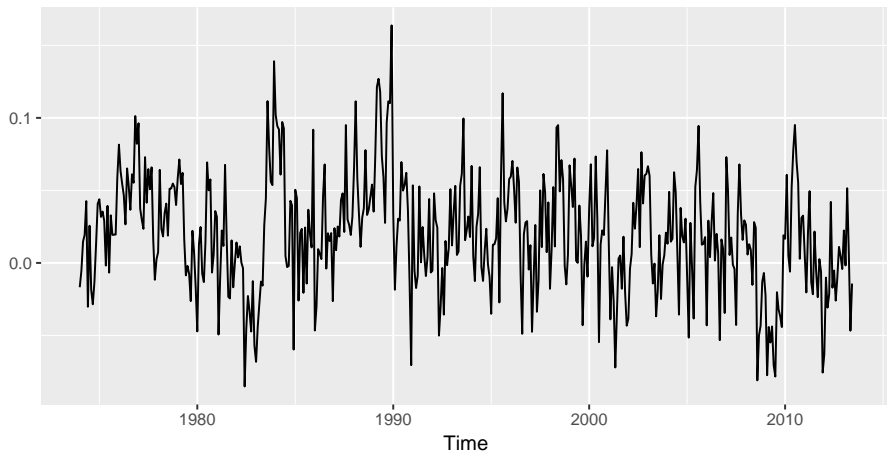
# Electricity production

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



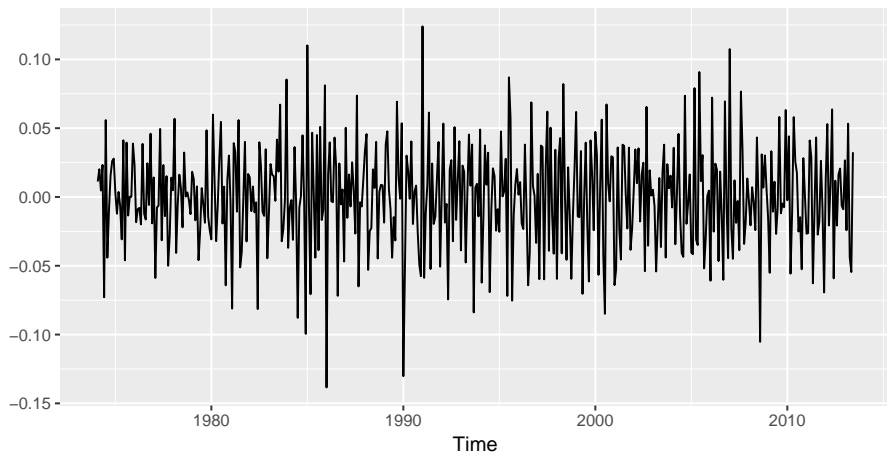
# Electricity production

```
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

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



# Seasonal differencing

When both seasonal and first differences are applied...

- 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.

# 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.

# 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**.

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

# 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.

# Outline

- 1 Stationarity
- 2 Differencing
- 3 Unit root tests**
- 4 Lab session 10
- 5 Backshift notation

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

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

# Dickey-Fuller test

## Test for “unit root”

- Estimate regression model

$$y'_t = \phi y_{t-1} + b_1 y'_{t-1} + b_2 y'_{t-2} + \dots + b_k y'_{t-k}$$

where  $y'_t$  denotes differenced series  $y_t - y_{t-1}$ .

- Number of lagged terms,  $k$ , is usually set to be about 3.
- If original series,  $y_t$ , needs differencing,  $\hat{\phi} \approx 0$ .
- If  $y_t$  is already stationary,  $\hat{\phi} < 0$ .
- In R: Use `tseries::adf.test()`.



# Dickey-Fuller test in R

```
tseries::adf.test(x,  
  alternative = c("stationary", "explosive"),  
  k = trunc((length(x)-1)^(1/3)))
```

- $k = \lfloor T - 1 \rfloor^{1/3}$

- Set alternative = stationary.

```
tseries::adf.test(dj)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: dj  
## Dickey-Fuller = -1.9872, Lag order = 6, p-value = 0.5816  
## alternative hypothesis: stationary
```

# Dickey-Fuller test in R

```
tseries::adf.test(x,  
  alternative = c("stationary", "explosive"),  
  k = trunc((length(x)-1)^(1/3)))
```

- $k = \lfloor T - 1 \rfloor^{1/3}$
- Set `alternative = stationary`.

```
tseries::adf.test(dj)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: dj  
## Dickey-Fuller = -1.9872, Lag order = 6, p-value = 0.5816  
## alternative hypothesis: stationary
```

# Dickey-Fuller test in R

```
tseries::adf.test(x,  
  alternative = c("stationary", "explosive"),  
  k = trunc((length(x)-1)^(1/3)))
```

- $k = \lfloor T - 1 \rfloor^{1/3}$
- Set alternative = stationary.

```
tseries::adf.test(dj)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: dj  
## Dickey-Fuller = -1.9872, Lag order = 6, p-value = 0.5816  
## alternative hypothesis: stationary
```

# How many differences?

```
ndiffs(x)  
nsdiffs(x)
```

```
ndiffs(dj)
```

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

```
nsdiffs(hsales)
```

```
## [1] 0
```

# Outline

- 1 Stationarity
- 2 Differencing
- 3 Unit root tests
- 4 Lab session 10**
- 5 Backshift notation

# Lab Session 10

# Outline

- 1 Stationarity
- 2 Differencing
- 3 Unit root tests
- 4 Lab session 10
- 5 Backshift notation**

# 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**:

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

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

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:

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

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

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**:

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

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

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**:

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

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*. 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

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

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

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  $d$ th-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.

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

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

# 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.