MONASH University

Rob J Hyndman

Forecasting: Principles and Practice

6. Transformations, stationarity, differencing

OTexts.com/fpp/2/4/

OTexts.com/fpp/8/1/

Outline

Transformations to stabilize the variance

If the data show different variation at different levels of the series, then a transformation can be useful.

Denote original observations as $y_1, \dots, y_n$ and transformed observations as $w_1, \dots, w_n$.

Mathematical transformations for stabilizing variation

Square root	$w_t = \sqrt{y_t}$	$\downarrow$
Cube root	$w_t = \sqrt[3]{y_t}$	Increasing
Logarithm	$w_t = \log(y_t)$	strength

Logarithms, in particular, are useful because they are more interpretable: changes in a log value are **relative (percent) changes on the original scale.**

Transformations to stabilize the variance

If the data show different variation at different levels of the series, then a transformation can be useful.

Denote original observations as $y_1, \dots, y_n$ and transformed observations as $w_1, \dots, w_n$.

Mathematical transformations for stabilizing variation

Square root	$w_t = \sqrt{y_t}$	$\downarrow$
Cube root	$w_t = \sqrt[3]{y_t}$	Increasing
Logarithm	$w_t = \log(y_t)$	strength

Logarithms, in particular, are useful because they are more interpretable: changes in a log value are **relative (percent) changes on the original scale.**

Transformations to stabilize the variance

If the data show different variation at different levels of the series, then a transformation can be useful.
Denote original observations as $y_1, \dots, y_n$ and transformed observations as $w_1, \dots, w_n$.

Mathematical transformations for stabilizing variation

Square root	$w_t = \sqrt{y_t}$	$\downarrow$
Cube root	$w_t = \sqrt[3]{y_t}$	Increasing
Logarithm	$w_t = \log(y_t)$	strength

Logarithms, in particular, are useful because they are more interpretable: changes in a log value are **relative (percent) changes on the original scale.**

Transformations to stabilize the variance

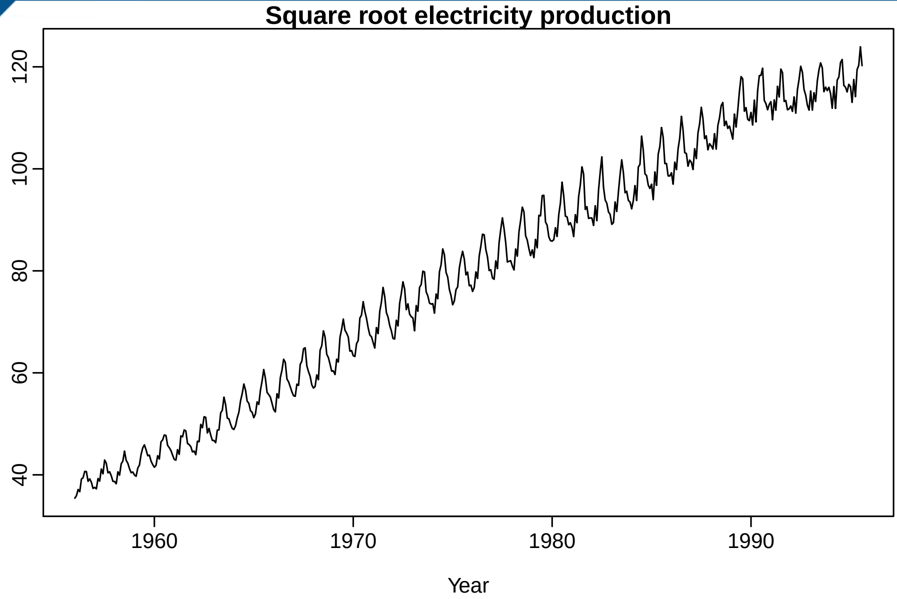
If the data show different variation at different levels of the series, then a transformation can be useful.
Denote original observations as $y_1, \dots, y_n$ and transformed observations as $w_1, \dots, w_n$.

Mathematical transformations for stabilizing variation

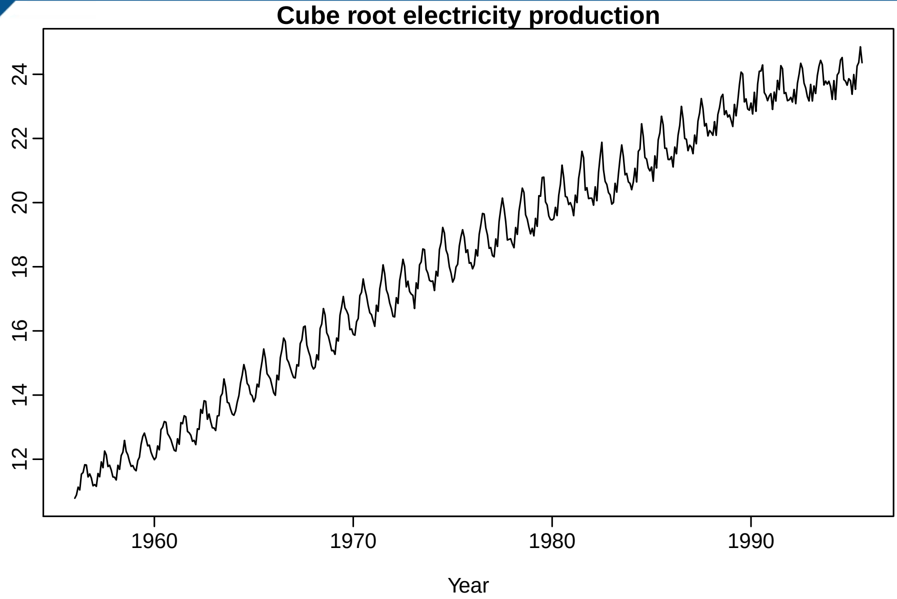
Square root	$w_t = \sqrt{y_t}$	$\downarrow$
Cube root	$w_t = \sqrt[3]{y_t}$	Increasing
Logarithm	$w_t = \log(y_t)$	strength

Logarithms, in particular, are useful because they are more interpretable: changes in a log value are **relative (percent) changes on the original scale.**

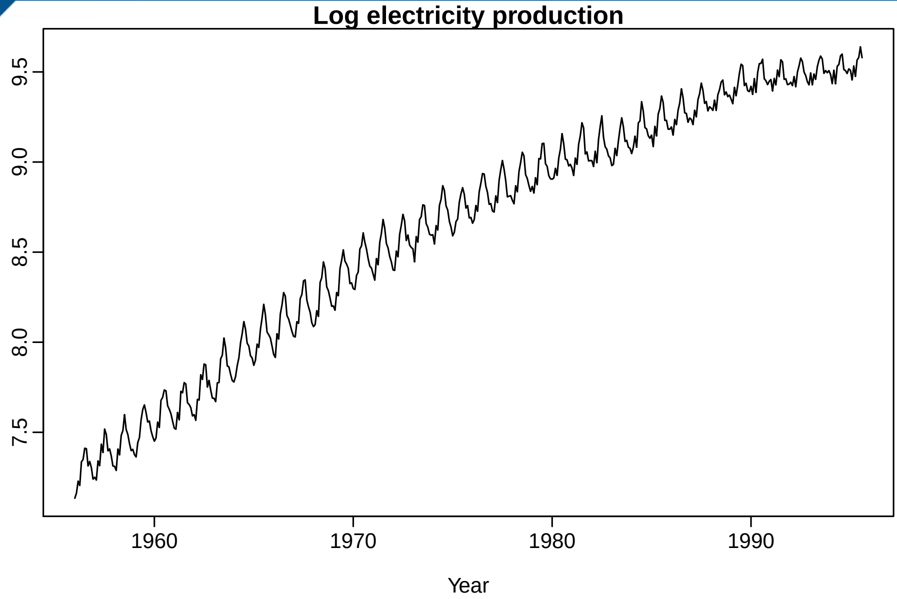
Transformations



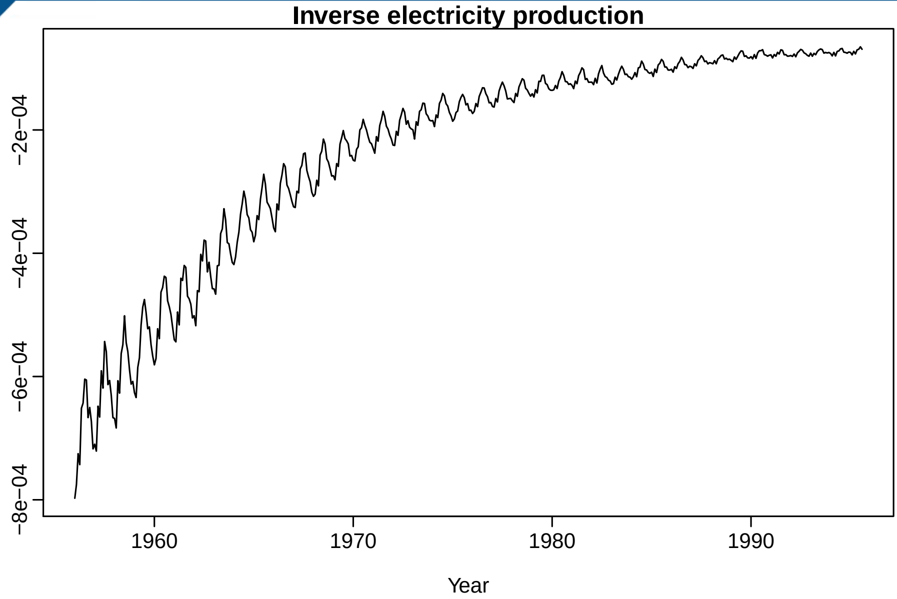
Transformations



Transformations



Transformations



Box-Cox transformations

Each of these transformations is close to a member of the family of **Box-Cox transformations**:

$$w_t = \begin{cases} \log(y_t), & \lambda = 0; \\ (y_t^\lambda - 1)/\lambda, & \lambda \neq 0. \end{cases}$$

- $\lambda = 1$: (No substantive transformation)
- $\lambda = \frac{1}{2}$: (Square root plus linear transformation)
- $\lambda = 0$: (Natural logarithm)
- $\lambda = -1$: (Inverse plus 1)

Box-Cox transformations

Each of these transformations is close to a member of the family of **Box-Cox transformations**:

$$w_t = \begin{cases} \log(y_t), & \lambda = 0; \\ (y_t^\lambda - 1)/\lambda, & \lambda \neq 0. \end{cases}$$

- $\lambda = 1$: (No substantive transformation)
- $\lambda = \frac{1}{2}$: (Square root plus linear transformation)
- $\lambda = 0$: (Natural logarithm)
- $\lambda = -1$: (Inverse plus 1)

Box-Cox transformations

Each of these transformations is close to a member of the family of **Box-Cox transformations**:

$$w_t = \begin{cases} \log(y_t), & \lambda = 0; \\ (y_t^\lambda - 1)/\lambda, & \lambda \neq 0. \end{cases}$$

- $\lambda = 1$: (No substantive transformation)
- $\lambda = \frac{1}{2}$: (Square root plus linear transformation)
- $\lambda = 0$: (Natural logarithm)
- $\lambda = -1$: (Inverse plus 1)

Box-Cox transformations

Each of these transformations is close to a member of the family of **Box-Cox transformations**:

$$w_t = \begin{cases} \log(y_t), & \lambda = 0; \\ (y_t^\lambda - 1)/\lambda, & \lambda \neq 0. \end{cases}$$

- $\lambda = 1$: (No substantive transformation)
- $\lambda = \frac{1}{2}$: (Square root plus linear transformation)
- $\lambda = 0$: (Natural logarithm)
- $\lambda = -1$: (Inverse plus 1)

Box-Cox transformations

Each of these transformations is close to a member of the family of **Box-Cox transformations**:

$$w_t = \begin{cases} \log(y_t), & \lambda = 0; \\ (y_t^\lambda - 1)/\lambda, & \lambda \neq 0. \end{cases}$$

- $\lambda = 1$: (No substantive transformation)
- $\lambda = \frac{1}{2}$: (Square root plus linear transformation)
- $\lambda = 0$: (Natural logarithm)
- $\lambda = -1$: (Inverse plus 1)

Box-Cox transformations

Box-Cox transformations

- y_t^λ for λ close to zero behaves like logs.
- If some $y_t = 0$, then must have $\lambda > 0$
- if some $y_t < 0$, no power transformation is possible unless all y_t adjusted by **adding a constant to all values**.
- Choose a simple value of λ. It makes explanation easier.
- Results are relatively insensitive to value of λ
- Often no transformation ($\lambda = 1$) needed.
- Transformation often makes little difference to forecasts but has large effect on PI.
- Choosing $\lambda = 0$ is a simple way to force forecasts to be positive

Box-Cox transformations

- y_t^λ for λ close to zero behaves like logs.
- If some $y_t = 0$, then must have $\lambda > 0$
- if some $y_t < 0$, no power transformation is possible unless all y_t adjusted by **adding a constant to all values**.
- Choose a simple value of λ. It makes explanation easier.
- Results are relatively insensitive to value of λ
- Often no transformation ($\lambda = 1$) needed.
- Transformation often makes little difference to forecasts but has large effect on PI.
- Choosing $\lambda = 0$ is a simple way to force forecasts to be positive

Box-Cox transformations

- y_t^λ for λ close to zero behaves like logs.
- If some $y_t = 0$, then must have $\lambda > 0$
- if some $y_t < 0$, no power transformation is possible unless all y_t adjusted by **adding a constant to all values**.
- Choose a simple value of λ. It makes explanation easier.
- Results are relatively insensitive to value of λ
- Often no transformation ($\lambda = 1$) needed.
- Transformation often makes little difference to forecasts but has large effect on PI.
- Choosing $\lambda = 0$ is a simple way to force forecasts to be positive

Box-Cox transformations

- y_t^λ for λ close to zero behaves like logs.
- If some $y_t = 0$, then must have $\lambda > 0$
- if some $y_t < 0$, no power transformation is possible unless all y_t adjusted by **adding a constant to all values**.
- Choose a simple value of λ. It makes explanation easier.
- Results are relatively insensitive to value of λ
- Often no transformation ($\lambda = 1$) needed.
- Transformation often makes little difference to forecasts but has large effect on PI.
- Choosing $\lambda = 0$ is a simple way to force forecasts to be positive

Box-Cox transformations

- y_t^λ for λ close to zero behaves like logs.
- If some $y_t = 0$, then must have $\lambda > 0$
- if some $y_t < 0$, no power transformation is possible unless all y_t adjusted by **adding a constant to all values**.
- Choose a simple value of λ. It makes explanation easier.
- Results are relatively insensitive to value of λ
- Often no transformation ($\lambda = 1$) needed.
- Transformation often makes little difference to forecasts but has large effect on PI.
- Choosing $\lambda = 0$ is a simple way to force forecasts to be positive

Box-Cox transformations

- y_t^λ for λ close to zero behaves like logs.
- If some $y_t = 0$, then must have $\lambda > 0$
- if some $y_t < 0$, no power transformation is possible unless all y_t adjusted by **adding a constant to all values**.
- Choose a simple value of λ. It makes explanation easier.
- Results are relatively insensitive to value of λ
- Often no transformation ($\lambda = 1$) needed.
- Transformation often makes little difference to forecasts but has large effect on PI.
- Choosing $\lambda = 0$ is a simple way to force forecasts to be positive

Box-Cox transformations

- y_t^λ for λ close to zero behaves like logs.
- If some $y_t = 0$, then must have $\lambda > 0$
- if some $y_t < 0$, no power transformation is possible unless all y_t adjusted by **adding a constant to all values**.
- Choose a simple value of λ. It makes explanation easier.
- Results are relatively insensitive to value of λ
- Often no transformation ($\lambda = 1$) needed.
- Transformation often makes little difference to forecasts but has large effect on PI.
- Choosing $\lambda = 0$ is a simple way to force forecasts to be positive

Box-Cox transformations

- y_t^λ for λ close to zero behaves like logs.
- If some $y_t = 0$, then must have $\lambda > 0$
- if some $y_t < 0$, no power transformation is possible unless all y_t adjusted by **adding a constant to all values**.
- Choose a simple value of λ. It makes explanation easier.
- Results are relatively insensitive to value of λ
- Often no transformation ($\lambda = 1$) needed.
- Transformation often makes little difference to forecasts but has large effect on PI.
- Choosing $\lambda = 0$ is a simple way to force forecasts to be positive

Back-transformation

We must reverse the transformation (or *back-transform*) to obtain forecasts on the original scale. The reverse Box-Cox transformations are given by

$$y_t = \begin{cases} \exp(w_t), & \lambda = 0; \\ (\lambda W_t + 1)^{1/\lambda}, & \lambda \neq 0. \end{cases}$$

```
plot(BoxCox(elec,lambda=1/3))
fit <- snaive(elec, lambda=1/3)
plot(fit)
plot(fit, include=120)
```

Back-transformation

We must reverse the transformation (or *back-transform*) to obtain forecasts on the original scale. The reverse Box-Cox transformations are given by

$$y_t = \begin{cases} \exp(w_t), & \lambda = 0; \\ (\lambda W_t + 1)^{1/\lambda}, & \lambda \neq 0. \end{cases}$$

```
plot(BoxCox(elec,lambda=1/3))
fit <- snaive(elec, lambda=1/3)
plot(fit)
plot(fit, include=120)
```

Automated Box-Cox transformations

```
BoxCox.lambda(elec)
```

- This attempts to balance the seasonal fluctuations and random variation across the series.
- Always check the results.
- A low value of λ can give extremely large prediction intervals.

Automated Box-Cox transformations

```
BoxCox.lambda(elec)
```

- This attempts to balance the seasonal fluctuations and random variation across the series.
- Always check the results.
- A low value of λ can give extremely large prediction intervals.

Automated Box-Cox transformations

```
BoxCox.lambda(elec)
```

- This attempts to balance the seasonal fluctuations and random variation across the series.
- Always check the results.
- A low value of λ can give extremely large prediction intervals.

Automated Box-Cox transformations

`BoxCox.lambda(elec)`

- This attempts to balance the seasonal fluctuations and random variation across the series.
- Always check the results.
- A low value of λ can give extremely large prediction intervals.

ETS and transformations

- A Box-Cox transformation followed by an additive ETS model is often better than an ETS model without transformation.
- It makes no sense to use a Box-Cox transformation and a *non-additive* ETS model.

ETS and transformations

- A Box-Cox transformation followed by an additive ETS model is often better than an ETS model without transformation.
- It makes no sense to use a Box-Cox transformation and a *non-additive* ETS model.

Outline

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

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

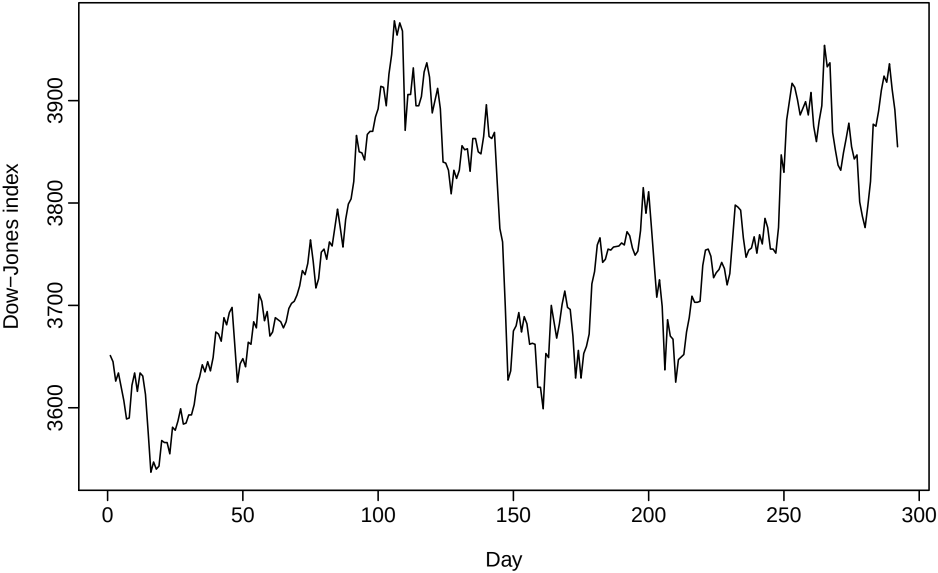
Definition

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:

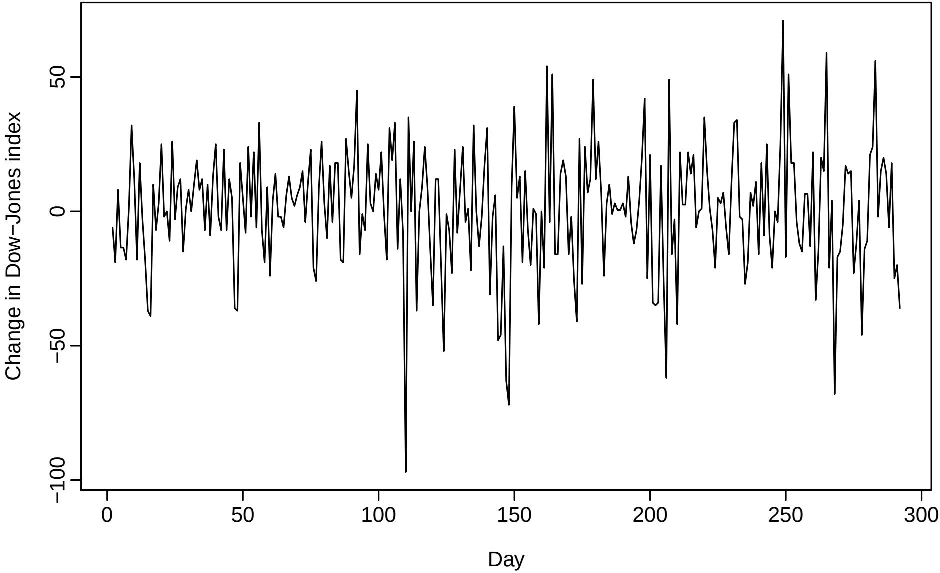
- roughly horizontal
- constant variance
- no patterns predictable in the long-term

Stationary?



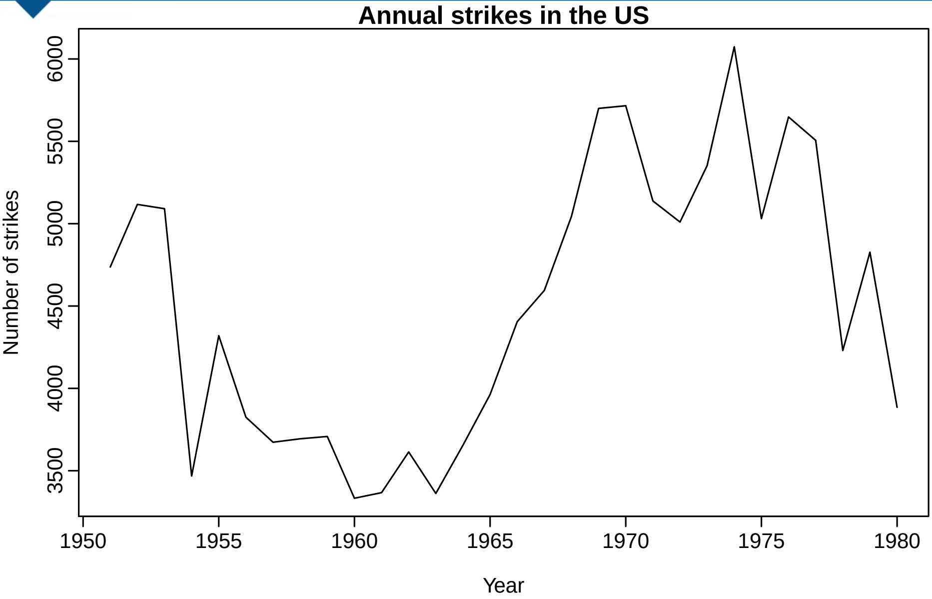
Dow–Jones index
3600
3700
3800
3900
0
50
100
150
200
250
300
Day

Stationary?



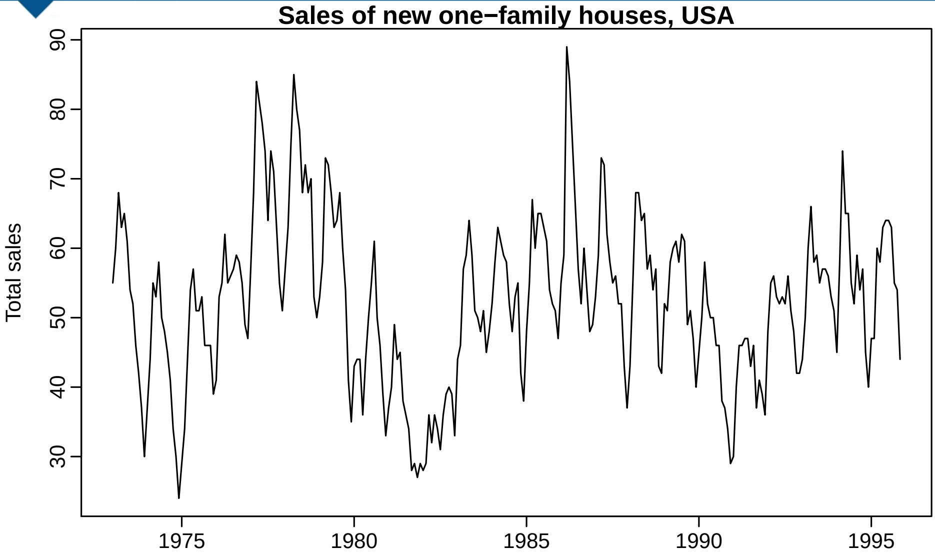
Change in Dow−Jones index
Day
50
0
−50
−100
0
50
100
150
200
250
300

Stationary?



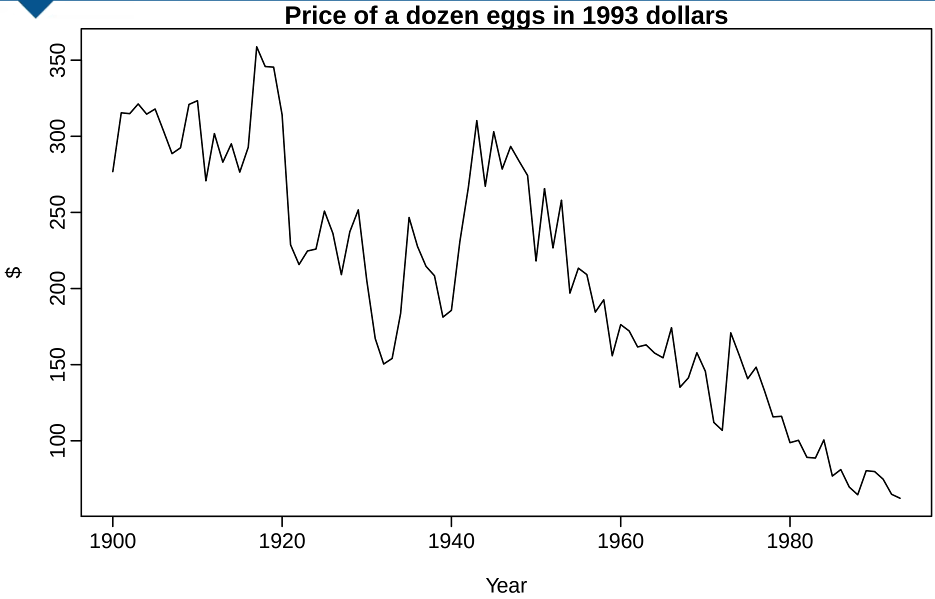
Annual strikes in the US
Number of strikes
6000
5500
5000
4500
4000
3500
1950
1955
1960
1965
1970
1975
1980
Year

Stationary?



Sales of new one−family houses, USA
Total sales
90
80
70
60
50
40
30
1975
1980
1985
1990
1995

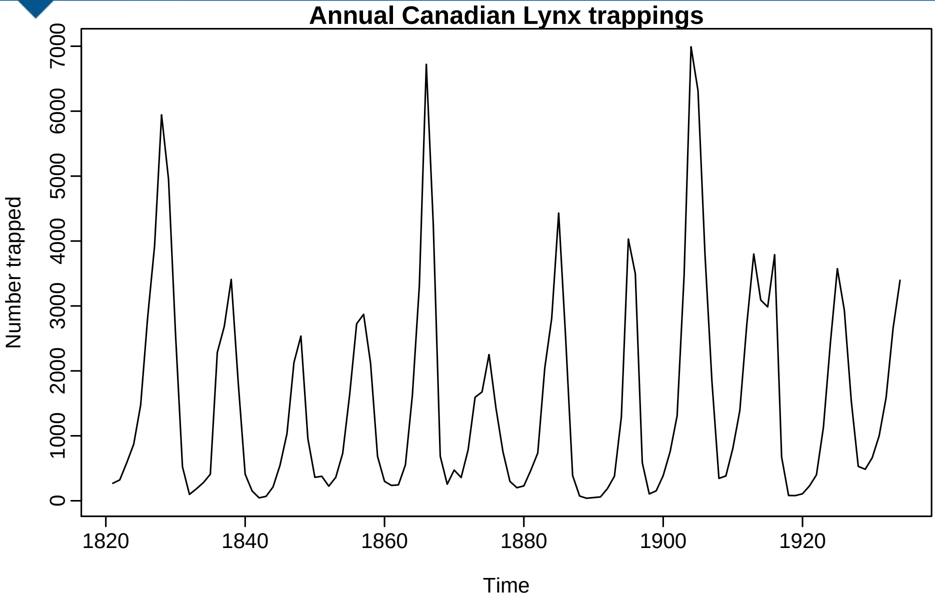
Stationary?


Price of a dozen eggs in 1993 dollars
$
350
300
250
200
150
100
1900
1920
1940
1960
1980
Year

Stationary?

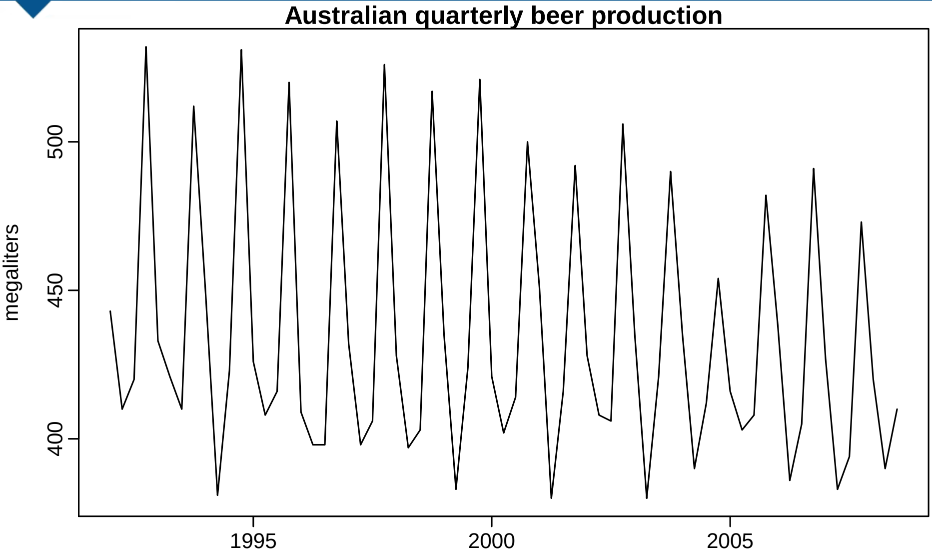
Number of pigs slaughtered in Victoria
thousands
80
90
100
110
1990
1991
1992
1993
1994
1995

Stationary?



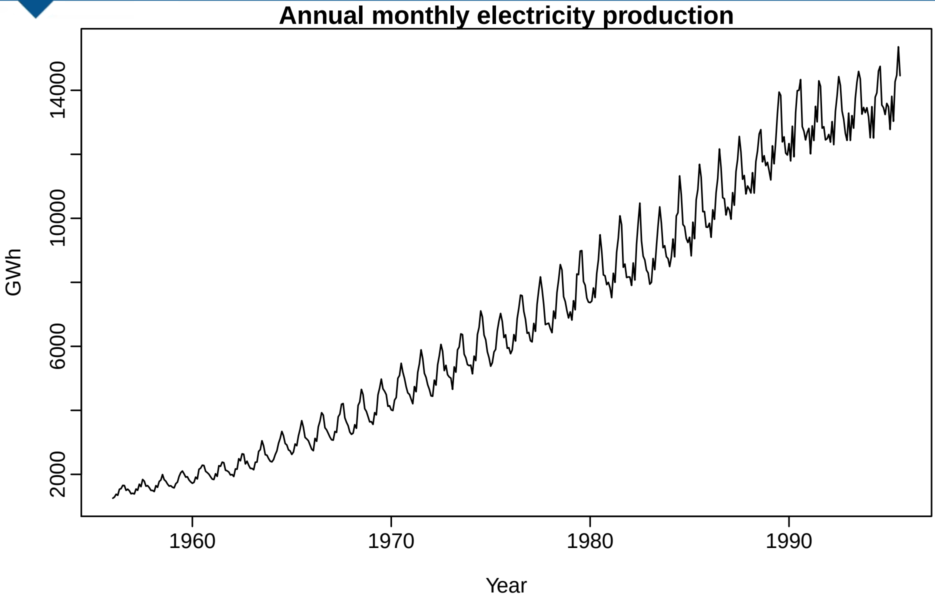
Annual Canadian Lynx trappings
Number trapped
0
1000
2000
3000
4000
5000
6000
7000
1820
1840
1860
1880
1900
1920
Time

Stationary?



Australian quarterly beer production
megaliters
500
450
400
1995
2000
2005

Stationary?


Annual monthly electricity production
GWh
2000
6000
10000
14000
1960
1970
1980
1990
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.

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**.

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.

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.

Non-stationarity in the mean

Identifying non-stationary series

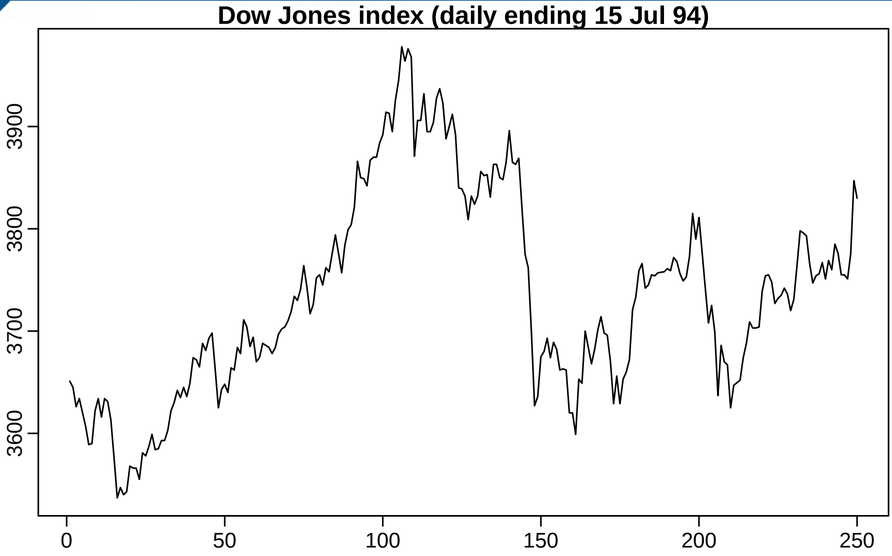
- time plot.
- The ACF of stationary data drops to zero relatively quickly
- The ACF of non-stationary data decreases slowly.
- For non-stationary data, the value of r_1 is often large and positive.

Non-stationarity in the mean

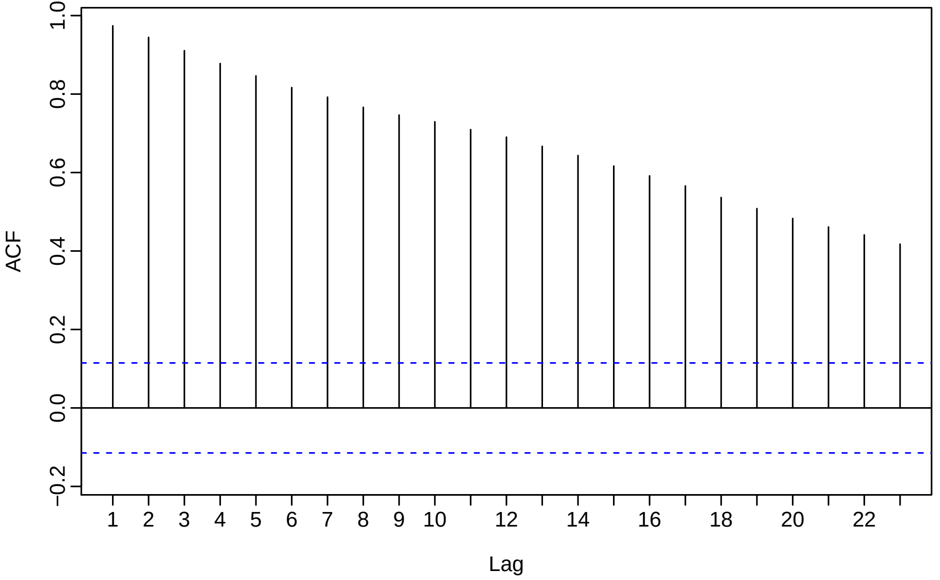
Identifying non-stationary series

- 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.

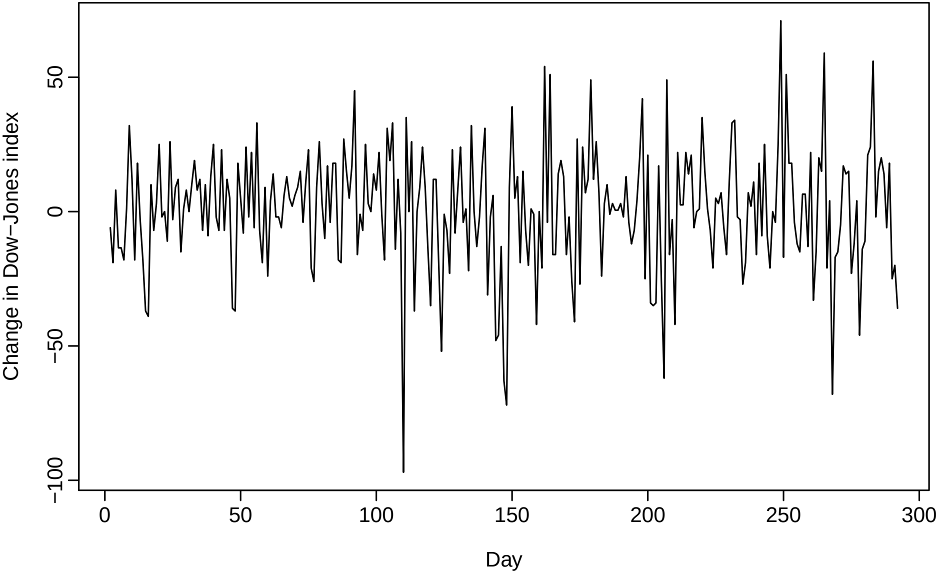
Example: Dow-Jones index



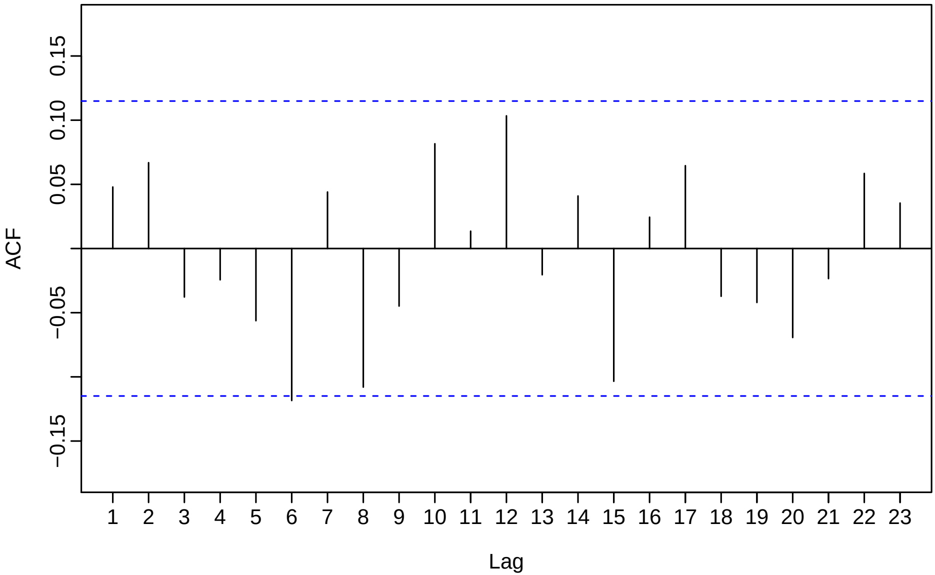
Example: Dow-Jones index



Example: Dow-Jones index



Example: Dow-Jones index



Outline

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.

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.

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.

Example: Dow-Jones index

- The differences of the Dow-Jones index are the **day-today changes**.
- Now the series looks just like **a white noise series**:
  - no autocorrelations outside the 95% limits.
  - Ljung-Box Q^* statistic has a p-value 0.153 for $h = 10$.
- **Conclusion:** The *daily change* in the Dow-Jones index is essentially a random amount uncorrelated with previous days.

Example: Dow-Jones index

- The differences of the Dow-Jones index are the **day-today changes**.
- Now the series looks just like **a white noise series**:
  - no autocorrelations outside the 95% limits.
  - Ljung-Box Q^* statistic has a p-value 0.153 for $h = 10$.
- **Conclusion:** The *daily change* in the Dow-Jones index is essentially a random amount uncorrelated with previous days.

Example: Dow-Jones index

- The differences of the Dow-Jones index are the **day-today changes**.
- Now the series looks just like **a white noise series**:
  - no autocorrelations outside the 95% limits.
  - Ljung-Box Q^* statistic has a p-value 0.153 for $h = 10$.
- **Conclusion:** The *daily change* in the Dow-Jones index is essentially a random amount uncorrelated with previous days.

Example: Dow-Jones index

- The differences of the Dow-Jones index are the **day-today changes**.
- Now the series looks just like **a white noise series**:
  - no autocorrelations outside the 95% limits.
  - Ljung-Box Q^* statistic has a p-value 0.153 for $h = 10$.
- **Conclusion:** The *daily change* in the Dow-Jones index is essentially a random amount uncorrelated with previous days.

Example: Dow-Jones index

- The differences of the Dow-Jones index are the **day-today changes**.
- Now the series looks just like **a white noise series**:
  - no autocorrelations outside the 95% limits.
  - Ljung-Box Q^* statistic has a p-value 0.153 for $h = 10$.
- **Conclusion:** The *daily change* in the Dow-Jones index is essentially a random amount uncorrelated with previous days.

Random walk model

Graph of differenced data suggests model for Dow-Jones index:

$$y_t - y_{t-1} = e_t \quad \text{or} \quad y_t = y_{t-1} + e_t\,.$$

- "Random walk" model very widely used for non-stationary data.
- This is the model behind the naïve method.
- Random walks typically have:

Random walk model

Graph of differenced data suggests model for Dow-Jones index:

$$y_t - y_{t-1} = e_t \quad \text{or} \quad y_t = y_{t-1} + e_t\,.$$

- "Random walk" model very widely used for non-stationary data.
- This is the model behind the naïve method.
- Random walks typically have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction

Random walk model

Graph of differenced data suggests model for Dow-Jones index:

$$y_t - y_{t-1} = e_t \quad \text{or} \quad y_t = y_{t-1} + e_t \,.$$

- "Random walk" model very widely used for non-stationary data.
- This is the model behind the naïve method.
- Random walks typically have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction.

Random walk model

Graph of differenced data suggests model for Dow-Jones index:

$$y_t - y_{t-1} = e_t \quad \text{or} \quad y_t = y_{t-1} + e_t\,.$$

- "Random walk" model very widely used for non-stationary data.
- This is the model behind the naïve method.
- Random walks typically have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction.

Random walk model

Graph of differenced data suggests model for Dow-Jones index:

$$y_t - y_{t-1} = e_t \quad \text{or} \quad y_t = y_{t-1} + e_t\,.$$

- "Random walk" model very widely used for non-stationary data.
- This is the model behind the naïve method.
- Random walks typically have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction.

Random walk model

Graph of differenced data suggests model for Dow-Jones index:

$$y_t - y_{t-1} = e_t \quad \text{or} \quad y_t = y_{t-1} + e_t\,.$$

- "Random walk" model very widely used for non-stationary data.
- This is the model behind the naïve method.
- Random walks typically have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction.

Random walk with drift model

$$y_t - y_{t-1} = c + e_t \quad \text{or} \quad y_t = c + y_{t-1} + e_t\,.$$

- c is the average change between consecutive observations.
- This is the model behind the drift method.

Random walk with drift model

$$y_t - y_{t-1} = c + e_t \quad \text{or} \quad y_t = c + y_{t-1} + e_t\,.$$

- c is the average change between consecutive observations.
- This is the model behind the drift method.

Random walk with drift model

$$y_t - y_{t-1} = c + e_t \quad \text{or} \quad y_t = c + y_{t-1} + e_t\,.$$

- c is the average change between consecutive observations.
- This is the model behind the drift method.

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.

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.

Outline

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 $m = 4$.

Seasonal differencing

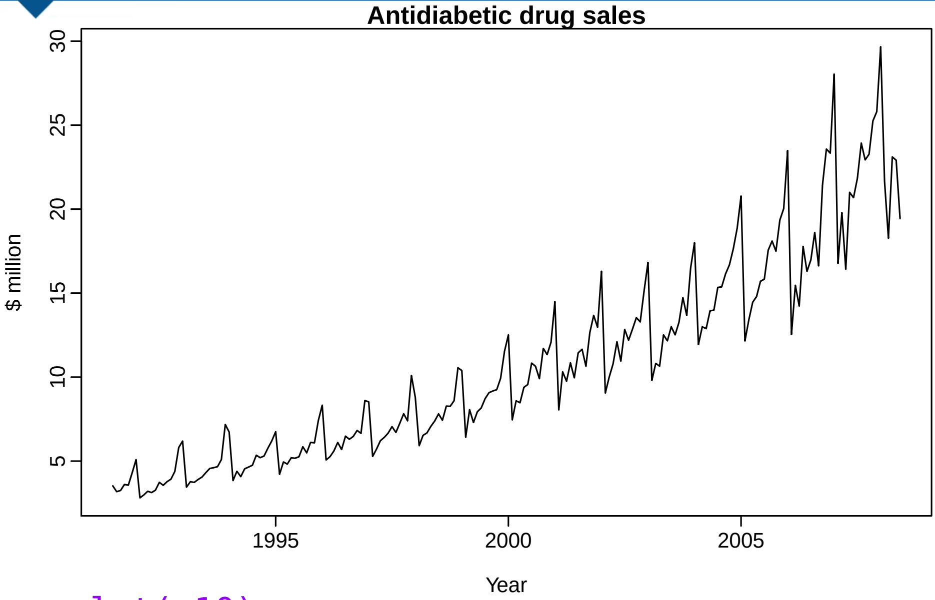
A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

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

where $m =$ number of seasons.

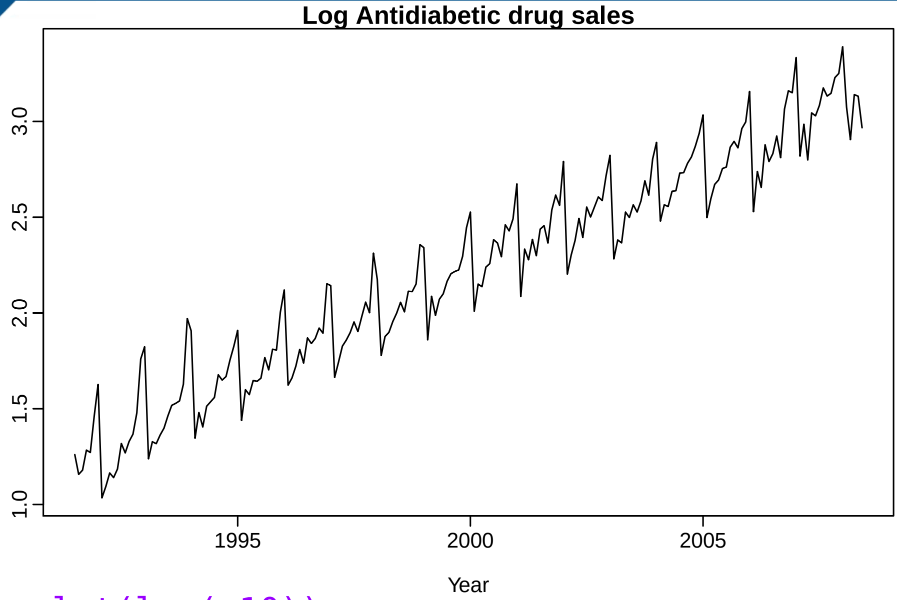
- For monthly data $m = 12$.
- For quarterly data $m = 4$.

Antidiabetic drug sales



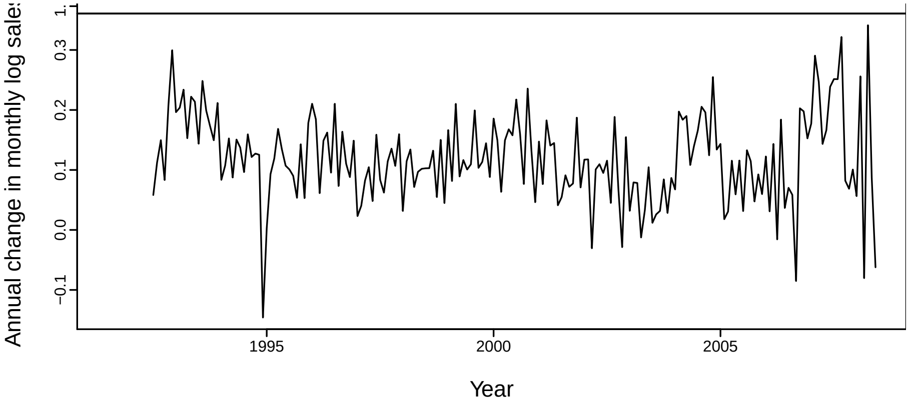

```
> plot(a10)
```

Antidiabetic drug sales



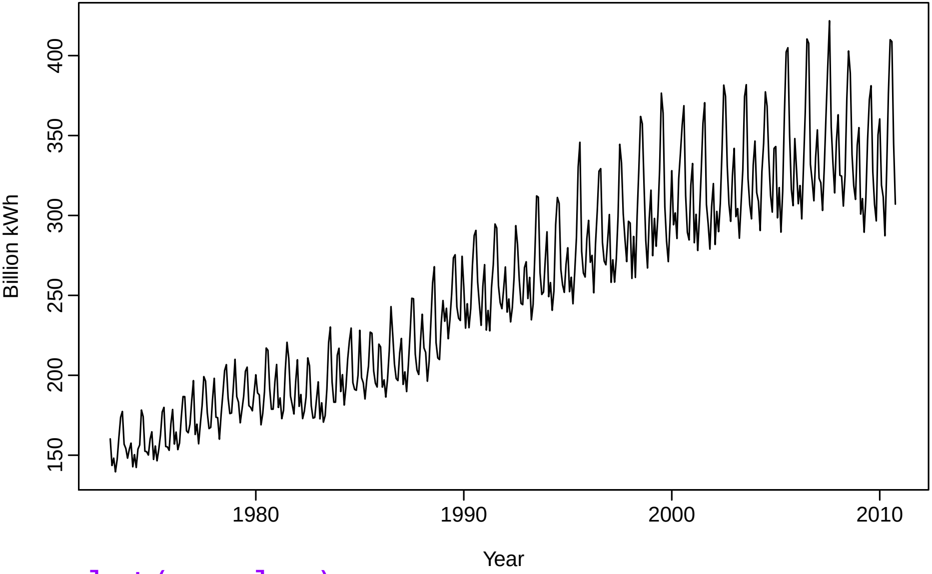

```
> plot(log(a10))
```

Antidiabetic drug sales



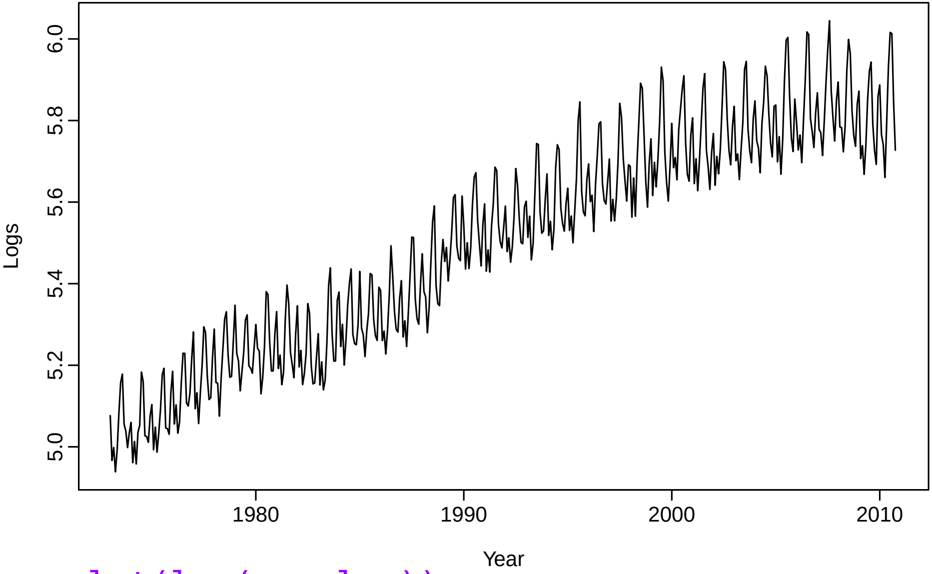

```
> plot(diff(log(a10),12))
```

Electricity production



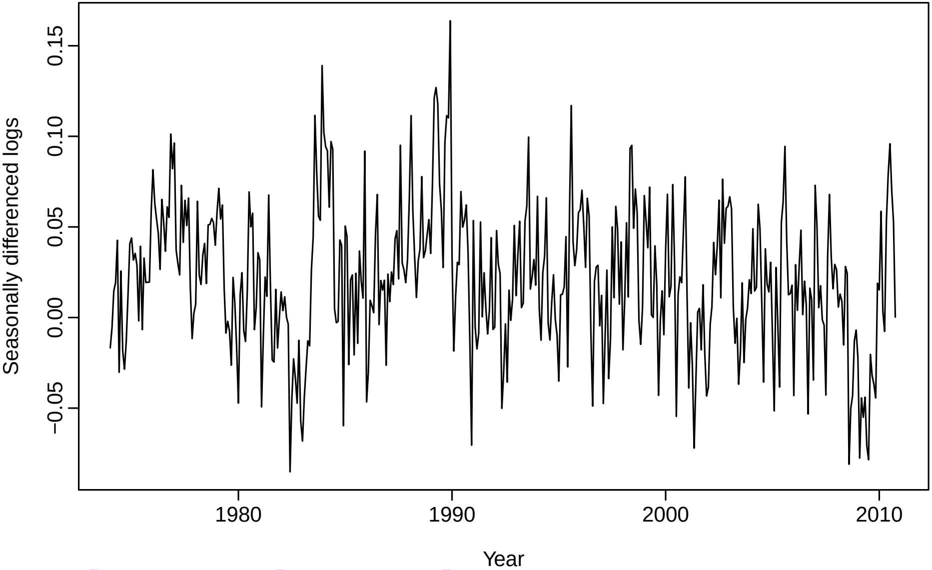

```
> plot(usmelec)
```

Electricity production



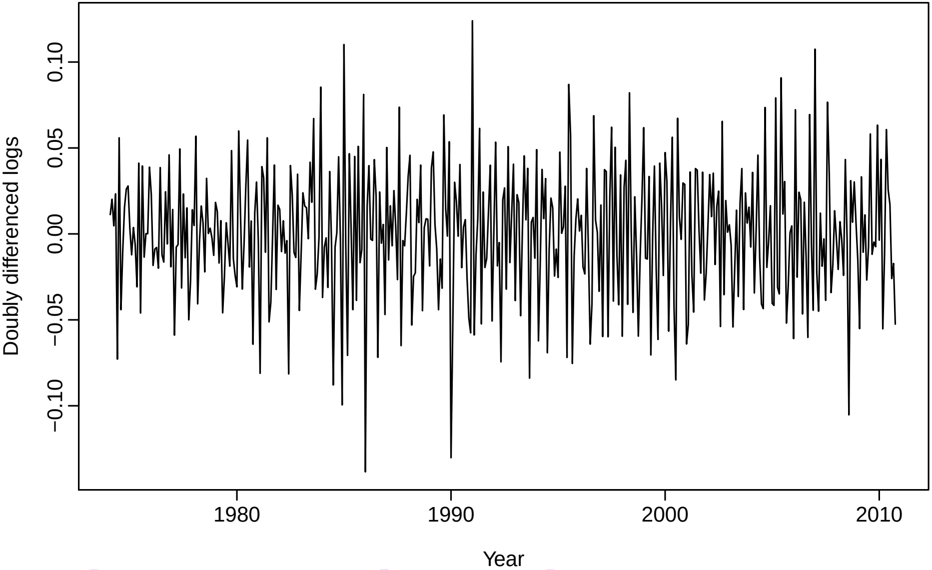

```
> plot(log(usmelec))
```

Electricity production




```
> plot(diff(log(usmelec),12))
```

Electricity production




```
> plot(diff(diff(log(usmelec),12),1))
```

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.

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.

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

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 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.

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 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.

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 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

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

  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 `adf.test()`.

Dickey-Fuller test

Test for "unit root"

- Estimate regression model

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

  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 `adf.test()`.

Dickey-Fuller test

Test for "unit root"

- Estimate regression model

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

  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 `adf.test()`.

Dickey-Fuller test

Test for "unit root"

- Estimate regression model

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

  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 `adf.test()`.

Dickey-Fuller test

Test for "unit root"

- Estimate regression model

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

  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 `adf.test()`.

Dickey-Fuller test in R

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

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

```
> 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

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

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

```
> 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

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

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

```
> 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

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

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

```
> 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

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

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

```
> 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)
```

Automated differencing

```
ns <- nsdiffs(x)
if(ns > 0)
  xstar <- diff(x,lag=frequency(x),
              differences=ns)
else
  xstar <- x
nd <- ndiffs(xstar)
if(nd > 0)
  xstar <- diff(xstar,differences=nd)
```

How many differences?

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

Automated differencing

```
ns <- nsdiffs(x)
if(ns > 0)
  xstar <- diff(x,lag=frequency(x),
            differences=ns)
else
  xstar <- x
nd <- ndiffs(xstar)
if(nd > 0)
  xstar <- diff(xstar,differences=nd)
```

Outline

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

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

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

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*. 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 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

- 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

- 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

- 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.

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

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.