Energy Infrastructure Planning: Forecasting

Carlos Abad

November 7, 2014

Who am I?

Who am I?

- PhD student in the IEOR department

Who am I?

- PhD student in the IEOR department
- Advisors: Prof. Vijay Modi and Prof. Garud Iyengar

Who am I?

- PhD student in the IEOR department
- Advisors: Prof. Vijay Modi and Prof. Garud Iyengar
- Research:
  - Robust control algorithms for solar micro-grids
  - Control, signal detection, and forecasting methods for managing DR programs

References

- Hyndman, R. J. & Athanasopoulos, G.(2013) Forecasting: principles and practice.
- www.otexts.org/fpp/
- R package fpp

Outline

Time series data

- Time series consists of sequences of observations collected over time.

Time series data

- Time series consists of sequences of observations collected over time.
- We will assume the time periods are equally spaced

Time series data

- Time series consists of sequences of observations collected over time.
- We will assume the time periods are equally spaced

Time series data

- Time series consists of sequences of observations collected over time.
- We will assume the time periods are equally spaced

Time series examples

- Hourly electricity demand

Forecasting is estimating how the sequence of observations will continue into the future.

Time series data

- Time series consists of sequences of observations collected over time.
- We will assume the time periods are equally spaced

Time series examples

- Hourly electricity demand
- Daily maximum temperature

Forecasting is estimating how the sequence of observations will continue into the future.

Time series data

- Time series consists of sequences of observations collected over time.
- We will assume the time periods are equally spaced

Time series examples

- Hourly electricity demand
- Daily maximum temperature
- Weekly wind generation

Forecasting is estimating how the sequence of observations will continue into the future.

Time series data

- Time series consists of sequences of observations collected over time.
- We will assume the time periods are equally spaced

Time series examples

- Hourly electricity demand
- Daily maximum temperature
- Weekly wind generation
- Monthly rainfall

Forecasting is estimating how the sequence of observations will continue into the future.

Time series in R

Main package used in this course

```
> library(fpp)
```

Time series in R

Main package used in this course

```
> library(fpp)
```

This loads:

- some data for use in examples and exercises
- **forecast** package (for forecasting functions)
- **tseries** package (for a few time series functions)
- **fma** package (for lots of time series data)
- expsmooth package (for more time series data)
- lmtest package (for some regression functions)

Time series in R

Main package used in this course

```
> library(fpp)
```

This loads:

- some data for use in examples and exercises
- **forecast** package (for forecasting functions)
- **tseries** package (for a few time series functions)
- **fma** package (for lots of time series data)
- **expsmooth** package (for more time series data)
- **lmtest** package (for some regression functions)

Time series in R

Main package used in this course

```
> library(fpp)
```

This loads:

- some data for use in examples and exercises
- **forecast** package (for forecasting functions)
- **tseries** package (for a few time series functions)
- **fma** package (for lots of time series data)
- **expsmooth** package (for more time series data)
- **lmtest** package (for some regression functions)

Time series in R

Main package used in this course

```
> library(fpp)
```

This loads:

- some data for use in examples and exercises
- **forecast** package (for forecasting functions)
- **tseries** package (for a few time series functions)
- **fma** package (for lots of time series data)
- **expsmooth** package (for more time series data)
- **lmtest** package (for some regression functions)

Time series in R

Main package used in this course

```
> library(fpp)
```

This loads:

- some data for use in examples and exercises
- **forecast** package (for forecasting functions)
- **tseries** package (for a few time series functions)
- **fma** package (for lots of time series data)
- **expsmooth** package (for more time series data)
- **lmtest** package (for some regression functions)

Time series in R

Main package used in this course

```
> library(fpp)
```

This loads:

- some data for use in examples and exercises
- **forecast** package (for forecasting functions)
- **tseries** package (for a few time series functions)
- **fma** package (for lots of time series data)
- **expsmooth** package (for more time series data)
- **lmtest** package (for some regression functions)

Time series in R

Other packages

```
> library(xts)
```

- Order time series by timestamp

Time series in R

Other packages

```
> library(xts)
```

- Order time series by timestamp
- Nicer plots

Time series in R

Other packages

```
> library(xts)
```

- Order time series by timestamp
- Nicer plots
- Easier time aggregation

Outline

Notation

- y_t: observed value at time t

Notation

- y_t: observed value at time t
- $\hat{y}_{T+h|T}$: forecast for time $T + h$ made at time T with historical information up to time T

Some simple forecasting methods

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \dots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.

Seasonal naïve method

Some simple forecasting methods

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \ldots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.
- Forecasts: $\hat{y}_{T+h|T} = y_T$.
- Consequence of efficient market hypothesis.

Seasonal naïve method

Some simple forecasting methods

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \dots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.
- Forecasts: $\hat{y}_{T+h|T} = y_T$.
- Consequence of efficient market hypothesis.

Seasonal naïve method

Some simple forecasting methods

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \dots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.
- Forecasts: $\hat{y}_{T+h|T} = y_T$.
- Consequence of efficient market hypothesis.

Seasonal naïve method

Some simple forecasting methods

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \ldots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.
- Forecasts: $\hat{y}_{T+h|T} = y_T$.
- Consequence of efficient market hypothesis.

Seasonal naïve method

- Forecasts equal to last value from same season.

Some simple forecasting methods

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \dots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.
- Forecasts: $\hat{y}_{T+h|T} = y_T$.
- Consequence of efficient market hypothesis.

Seasonal naïve method

- Forecasts equal to last value from same season.

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \dots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.
- Forecasts: $\hat{y}_{T+h|T} = y_T$.
- Consequence of efficient market hypothesis.

Seasonal naïve method

- Forecasts equal to last value from same season.
- Forecasts: $\hat{y}_{T+h|T} = y_{T+h-km}$ where $m =$ seasonal period and $k = \lfloor (h-1)/m \rfloor + 1$.

Some simple forecasting methods

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \dots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.
- Forecasts: $\hat{y}_{T+h|T} = y_T$.
- Consequence of efficient market hypothesis.

Seasonal naïve method

- Forecasts equal to last value from same season.
- Forecasts: $\hat{y}_{T+h|T} = y_{T+h-km}$ where $m =$ seasonal period and $k = \lfloor (h-1)/m \rfloor + 1$.

Some simple forecasting methods

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \dots, y_T\}$.
- Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$

Naïve method (for time series only)

- Forecasts equal to last observed value.
- Forecasts: $\hat{y}_{T+h|T} = y_T$.
- Consequence of efficient market hypothesis.

Seasonal naïve method

- Forecasts equal to last value from same season.
- Forecasts: $\hat{y}_{T+h|T} = y_{T+h-km}$ where $m =$ seasonal period and $k = \lfloor (h-1)/m \rfloor + 1$.

Drift method

- Forecasts equal to last value plus average change.
- Forecasts:

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1})$$
$$= y_T + \frac{h}{T-1}(y_T - y_1).$$

- Equivalent to extrapolating a line drawn between first and last observations.

Drift method

- Forecasts equal to last value plus average change.
- Forecasts:

$$\begin{aligned}\hat{y}_{T+h|T} &= y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1})\\ &= y_T + \frac{h}{T-1}(y_T - y_1).\end{aligned}$$

- Equivalent to extrapolating a line drawn between first and last observations.

Drift method

- Forecasts equal to last value plus average change.
- Forecasts:

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1})$$
$$= y_T + \frac{h}{T-1}(y_T - y_1).$$

- Equivalent to extrapolating a line drawn between first and last observations.

Some simple forecasting methods

- Mean: `meanf(x, h=20)`
- Naive: `naive(x, h=20)` or `rwf(x, h=20)`
- Seasonal naive: `snaive(x, h=20)`
- Drift: `rwf(x, drift=TRUE, h=20)`

Some simple forecasting methods

- Mean: `meanf(x, h=20)`
- Naive: `naive(x, h=20)` or `rwf(x, h=20)`
- Seasonal naive: `snaive(x, h=20)`
- Drift: `rwf(x, drift=TRUE, h=20)`

Some simple forecasting methods

- Mean: `meanf(x, h=20)`
- Naive: `naive(x, h=20)` or `rwf(x, h=20)`
- Seasonal naive: `snaive(x, h=20)`
- Drift: `rwf(x, drift=TRUE, h=20)`

Some simple forecasting methods

- Mean: `meanf(x, h=20)`
- Naive: `naive(x, h=20)` or `rwf(x, h=20)`
- Seasonal naive: `snaive(x, h=20)`
- Drift: `rwf(x, drift=TRUE, h=20)`

Outline

Forecasting residuals

Residuals in forecasting: difference between observed value and its forecast based on all previous observations: $e_t = y_t - \hat{y}_{t|t-1}$.

Assumptions

1. $\{e_t\}$ uncorrelated. If they aren't, then information left in residuals that should be used in computing forecasts.
2. $\{e_t\}$ have mean zero. If they don't, then forecasts are biased.

Useful properties (for prediction intervals)

3. $\{e_t\}$ have constant variance.
4. $\{e_t\}$ are normally distributed.

Forecasting residuals

Residuals in forecasting: difference between observed value and its forecast based on all previous observations: $e_t = y_t - \hat{y}_{t|t-1}$.

Assumptions

1. $\{e_t\}$ uncorrelated. If they aren't, then information left in residuals that should be used in computing forecasts.
2. $\{e_t\}$ have mean zero. If they don't, then forecasts are biased.

Useful properties (for prediction intervals)

3. $\{e_t\}$ have constant variance.
4. $\{e_t\}$ are normally distributed.

Forecasting residuals

Residuals in forecasting: difference between observed value and its forecast based on all previous observations: $e_t = y_t - \hat{y}_{t|t-1}$.

Assumptions

1. $\{e_t\}$ uncorrelated. If they aren't, then information left in residuals that should be used in computing forecasts.
2. $\{e_t\}$ have mean zero. If they don't, then forecasts are biased.

Useful properties (for prediction intervals)

3. $\{e_t\}$ have constant variance.
4. $\{e_t\}$ are normally distributed.

Measures of forecast accuracy

Let y_t denote the tth observation and $\hat{y}_{t|t-1}$ denote its forecast based on all previous data, where $t = 1, \dots, T$. Then the following measures are useful.

$$\text{MAE} = T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|$$

$$\text{MSE} = T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2 \qquad \text{RMSE} = \sqrt{T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2}$$

$$\text{MAPE} = 100T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|/|y_t|$$

- MAE, MSE, RMSE are all scale dependent.
- MAPE is scale independent but is only sensible if $y_t \gg 0$ for all t, and y has a natural zero.

Measures of forecast accuracy

Let y_t denote the tth observation and $\hat{y}_{t|t-1}$ denote its forecast based on all previous data, where $t = 1, \dots, T$. Then the following measures are useful.

$$\text{MAE} = T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|$$

$$\text{MSE} = T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2 \qquad \text{RMSE} = \sqrt{T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2}$$

$$\text{MAPE} = 100T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|/|y_t|$$

- MAE, MSE, RMSE are all scale dependent.
- MAPE is scale independent but is only sensible if $y_t \gg 0$ for all t, and y has a natural zero.

Measures of forecast accuracy

Let y_t denote the tth observation and $\hat{y}_{t|t-1}$ denote its forecast based on all previous data, where $t = 1, \dots, T$. Then the following measures are useful.

$$\text{MAE} = T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|$$

$$\text{MSE} = T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2 \qquad \text{RMSE} = \sqrt{T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2}$$

$$\text{MAPE} = 100T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|/|y_t|$$

- MAE, MSE, RMSE are all scale dependent.
- MAPE is scale independent but is only sensible if $y_t \gg 0$ for all t, and y has a natural zero.

Measures of forecast accuracy

Mean Absolute Scaled Error

$$\text{MASE} = T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}| / Q$$

where Q is a stable measure of the scale of the time series $\{y_t\}$.

Measures of forecast accuracy

Mean Absolute Scaled Error

$$\text{MASE} = T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|/Q$$

where Q is a stable measure of the scale of the time series $\{y_t\}$.

For non-seasonal time series,

$$Q = (T-1)^{-1} \sum_{t=2}^{T} |y_t - y_{t-1}|$$

works well. Then MASE is equivalent to MAE relative to a naive method.

Measures of forecast accuracy

Mean Absolute Scaled Error

$$\text{MASE} = T^{-1} \sum_{t=1}^{T} |y_t - \hat{y}_{t|t-1}|/Q$$

where Q is a stable measure of the scale of the time series $\{y_t\}$.

For seasonal time series,

$$Q = (T - m)^{-1} \sum_{t=m+1}^{T} |y_t - y_{t-m}|$$

works well. Then MASE is equivalent to MAE relative to a seasonal naive method.

Training and test sets

Available data

Training set (e.g., 80%)	Test set (e.g., 20%)

- The test set must not be used for *any* aspect of model development or calculation of forecasts.
- Forecast accuracy is based only on the test set.

Training and test sets

Available data

Training set (e.g., 80%)	Test set (e.g., 20%)

- The test set must not be used for *any* aspect of model development or calculation of forecasts.
- Forecast accuracy is based only on the test set.

Beware of over-fitting

- A model which fits the data well does not necessarily forecast well.
- A perfect fit can always be obtained by using a model with enough parameters. (Compare R^2)
- Over-fitting a model to data is as bad as failing to identify the systematic pattern in the data.
- Problems can be overcome by measuring true *out-of-sample* forecast accuracy. That is, total data divided into “training” set and “test” set. Training set used to estimate parameters. Forecasts are made for test set.
- Accuracy measures computed for errors in test set only.

Beware of over-fitting

- A model which fits the data well does not necessarily forecast well.
- A perfect fit can always be obtained by using a model with enough parameters. (Compare R^2)
- Over-fitting a model to data is as bad as failing to identify the systematic pattern in the data.
- Problems can be overcome by measuring true *out-of-sample* forecast accuracy. That is, total data divided into "training" set and "test" set. Training set used to estimate parameters. Forecasts are made for test set.
- Accuracy measures computed for errors in test set only.

Beware of over-fitting

- A model which fits the data well does not necessarily forecast well.
- A perfect fit can always be obtained by using a model with enough parameters. (Compare R^2)
- Over-fitting a model to data is as bad as failing to identify the systematic pattern in the data.
- Problems can be overcome by measuring true *out-of-sample* forecast accuracy. That is, total data divided into "training" set and "test" set. Training set used to estimate parameters. Forecasts are made for test set.
- Accuracy measures computed for errors in test set only.

Beware of over-fitting

- A model which fits the data well does not necessarily forecast well.
- A perfect fit can always be obtained by using a model with enough parameters. (Compare R^2)
- Over-fitting a model to data is as bad as failing to identify the systematic pattern in the data.
- Problems can be overcome by measuring true *out-of-sample* forecast accuracy. That is, total data divided into “training” set and “test” set. Training set used to estimate parameters. Forecasts are made for test set.
- Accuracy measures computed for errors in test set only.

Beware of over-fitting

- A model which fits the data well does not necessarily forecast well.
- A perfect fit can always be obtained by using a model with enough parameters. (Compare R^2)
- Over-fitting a model to data is as bad as failing to identify the systematic pattern in the data.
- Problems can be overcome by measuring true *out-of-sample* forecast accuracy. That is, total data divided into "training" set and "test" set. Training set used to estimate parameters. Forecasts are made for test set.
- Accuracy measures computed for errors in test set only.

Outline

Time series graphics

- **Time plots**
  R command: `plot` or `plot.ts`
- **Seasonal plots**
  R command: `seasonplot`
- **Seasonal subseries plots**
  R command: `monthplot`
- **Lag plots**
  R command: `lag.plot`
- **ACF plots**
  R command: `Acf`

Seasonal plots

- Data plotted against the individual "seasons" in which the data were observed. (In this case a "season" is a month.)
- Something like a time plot except that the data from each season are overlapped.
- Enables the underlying seasonal pattern to be seen more clearly, and also allows any substantial departures from the seasonal pattern to be easily identified.
- In R: `seasonplot`

Seasonal plots

- Data plotted against the individual “seasons” in which the data were observed. (In this case a “season” is a month.)
- Something like a time plot except that the data from each season are overlapped.
- Enables the underlying seasonal pattern to be seen more clearly, and also allows any substantial departures from the seasonal pattern to be easily identified.
- In R: `seasonplot`

Seasonal plots

- Data plotted against the individual "seasons" in which the data were observed. (In this case a "season" is a month.)
- Something like a time plot except that the data from each season are overlapped.
- Enables the underlying seasonal pattern to be seen more clearly, and also allows any substantial departures from the seasonal pattern to be easily identified.
- In R: `seasonplot`

Seasonal plots

- Data plotted against the individual "seasons" in which the data were observed. (In this case a "season" is a month.)
- Something like a time plot except that the data from each season are overlapped.
- Enables the underlying seasonal pattern to be seen more clearly, and also allows any substantial departures from the seasonal pattern to be easily identified.
- In R: `seasonplot`

Seasonal subseries plots

- Data for each season collected together in time plot as separate time series.
- Enables the underlying seasonal pattern to be seen clearly, and changes in seasonality over time to be visualized.
- In R: `monthplot`

Seasonal subseries plots

- Data for each season collected together in time plot as separate time series.
- Enables the underlying seasonal pattern to be seen clearly, and changes in seasonality over time to be visualized.
- In R: `monthplot`

Seasonal subseries plots

- Data for each season collected together in time plot as separate time series.
- Enables the underlying seasonal pattern to be seen clearly, and changes in seasonality over time to be visualized.
- In R: `monthplot`

Time series patterns

Trend pattern exists when there is a long-term increase or decrease in the data.

Seasonal pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).

Cyclic pattern exists when data exhibit rises and falls that are *not of fixed period* (duration usually of at least 2 years).

Seasonal or cyclic?

Differences between seasonal and cyclic patterns:

- seasonal pattern constant length; cyclic pattern variable length
- average length of cycle longer than length of seasonal pattern
- magnitude of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

Seasonal or cyclic?

Differences between seasonal and cyclic patterns:

- seasonal pattern constant length; cyclic pattern variable length
- average length of cycle longer than length of seasonal pattern
- magnitude of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

Seasonal or cyclic?

Differences between seasonal and cyclic patterns:

- seasonal pattern constant length; cyclic pattern variable length
- average length of cycle longer than length of seasonal pattern
- magnitude of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

Seasonal or cyclic?

Differences between seasonal and cyclic patterns:

- seasonal pattern constant length; cyclic pattern variable length
- average length of cycle longer than length of seasonal pattern
- magnitude of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

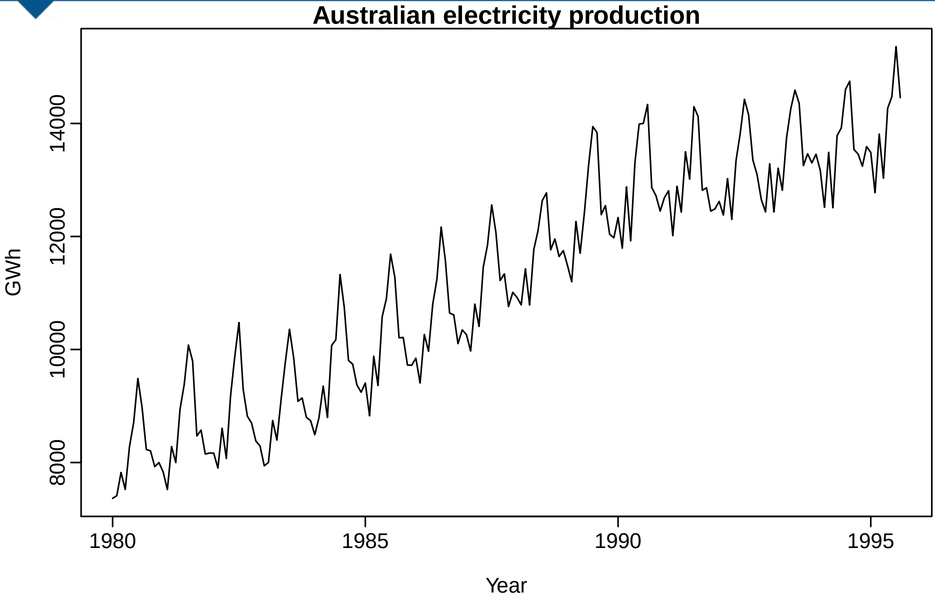
Seasonal or cyclic?

Differences between seasonal and cyclic patterns:

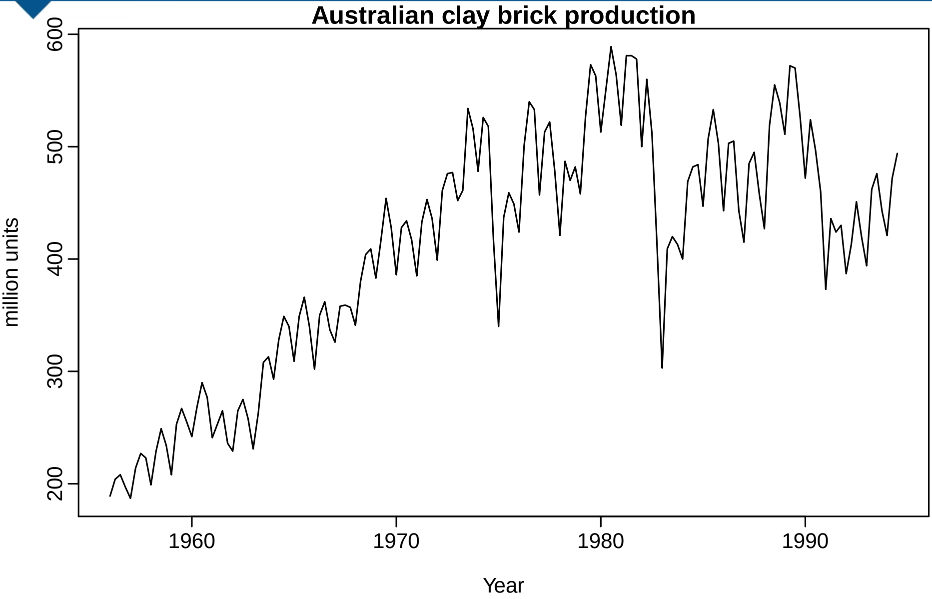
- seasonal pattern constant length; cyclic pattern variable length
- average length of cycle longer than length of seasonal pattern
- magnitude of cycle more variable than magnitude of seasonal pattern

The timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data.

Time series patterns



Time series patterns



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

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

A **stationary series** is:

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

Stationarity

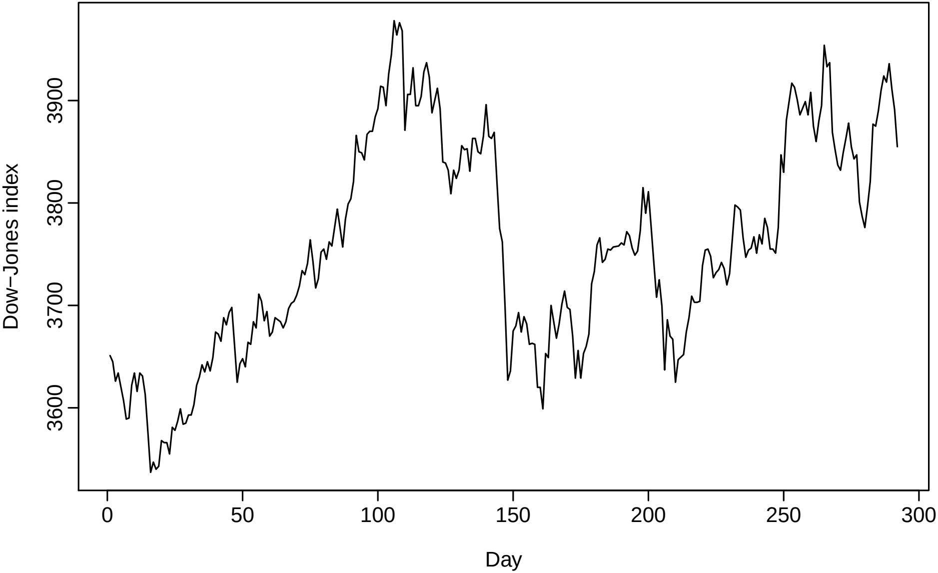
Definition

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.

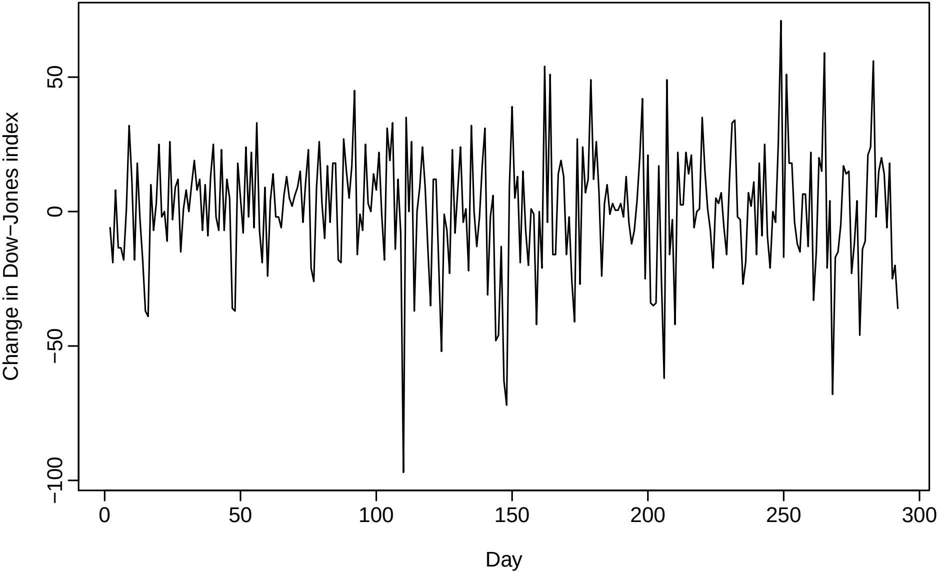
A **stationary series** is:

- 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
50
0
−50
−100
0
50
100
150
200
250
300
Day

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.

Autocorrelation

Covariance and **correlation**: measure extent of **linear relationship** between two variables (y and X).

Autocovariance and **autocorrelation**: measure linear relationship between **lagged values** of a time series y.

We measure the relationship between:
y_t and y_{t-1}
y_t and y_{t-2}
y_t and y_{t-3}
etc.

Autocorrelation

Covariance and **correlation**: measure extent of **linear relationship** between two variables (y and X).

Autocovariance and **autocorrelation**: measure linear relationship between **lagged values** of a time series y.

We measure the relationship between: y_t and y_{t-1}
y_t and y_{t-2}
y_t and y_{t-3}
etc.

Autocorrelation

Covariance and **correlation**: measure extent of **linear relationship** between two variables (y and X).

Autocovariance and **autocorrelation**: measure linear relationship between **lagged values** of a time series y.

We measure the relationship between:

y_t and y_{t-1}
y_t and y_{t-2}
y_t and y_{t-3}
etc.

Autocorrelation

We denote the sample autocovariance at lag k by c_k and the sample autocorrelation at lag k by r_k. Then define

$$c_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

and $$r_k = c_k / c_0$$

- r_1 indicates how successive values of y relate to each other
- r_2 indicates how y values two periods apart relate to each other
- r_k is *almost* the same as the sample correlation between y_t and y_{t-k}.

Autocorrelation

We denote the sample autocovariance at lag k by c_k and the sample autocorrelation at lag k by r_k. Then define

$$c_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

$$\text{and} \qquad r_k = c_k / c_0$$

- r_1 indicates how successive values of y relate to each other
- r_2 indicates how y values two periods apart relate to each other
- r_k is *almost* the same as the sample correlation between y_t and y_{t-k}.

Autocorrelation

We denote the sample autocovariance at lag k by c_k and the sample autocorrelation at lag k by r_k. Then define

$$c_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

and $$r_k = c_k/c_0$$

- r_1 indicates how successive values of y relate to each other
- r_2 indicates how y values two periods apart relate to each other
- r_k is *almost* the same as the sample correlation between y_t and y_{t-k}.

Autocorrelation

We denote the sample autocovariance at lag k by c_k and the sample autocorrelation at lag k by r_k. Then define

$$c_k = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

and $$r_k = c_k/c_0$$

- r_1 indicates how successive values of y relate to each other
- r_2 indicates how y values two periods apart relate to each other
- r_k is *almost* the same as the sample correlation between y_t and y_{t-k}.

Recognizing seasonality in a time series

If there is seasonality, the ACF at the seasonal lag (e.g., 12 for monthly data) will be **large and positive**.

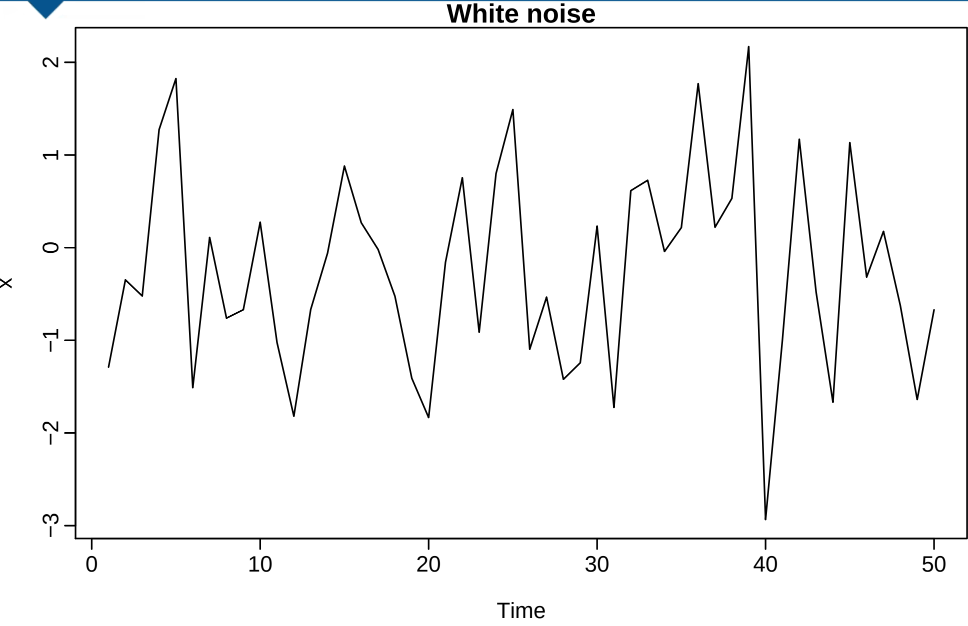
- For seasonal monthly data, a large ACF value will be seen at lag 12 and possibly also at lags 24, 36, . . .
- For seasonal quarterly data, a large ACF value will be seen at lag 4 and possibly also at lags 8, 12, . . .

Recognizing seasonality in a time series

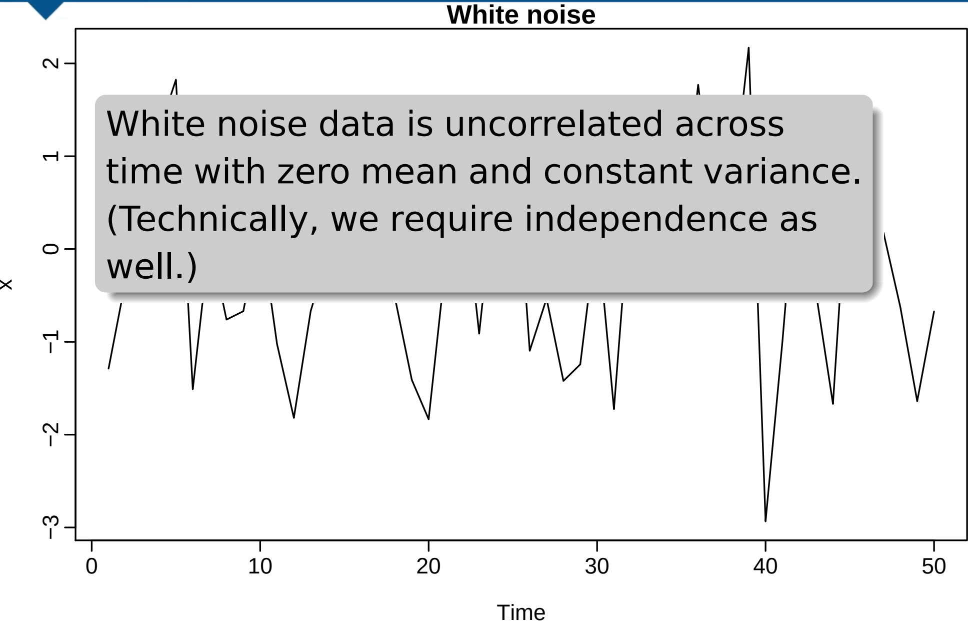
If there is seasonality, the ACF at the seasonal lag (e.g., 12 for monthly data) will be **large and positive**.

- For seasonal monthly data, a large ACF value will be seen at lag 12 and possibly also at lags 24, 36, . . .
- For seasonal quarterly data, a large ACF value will be seen at lag 4 and possibly also at lags 8, 12, . . .

Example: White noise

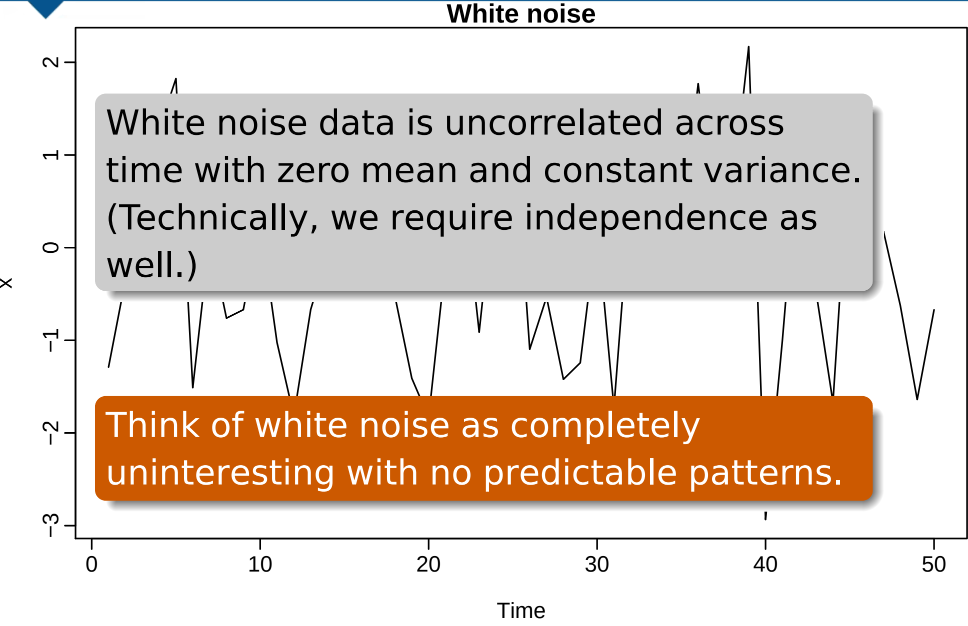


Example: White noise



White noise data is uncorrelated across time with zero mean and constant variance.
(Technically, we require independence as well.)

Example: White noise

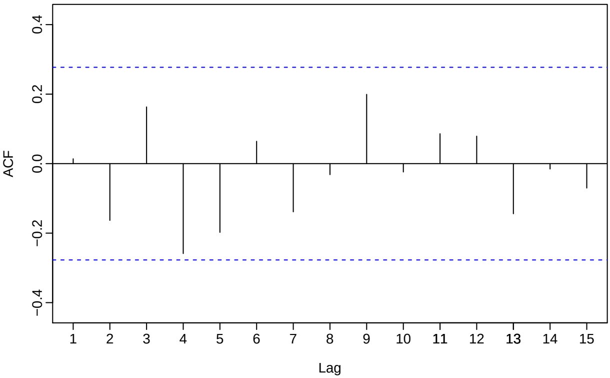


White noise data is uncorrelated across time with zero mean and constant variance. (Technically, we require independence as well.)

Think of white noise as completely uninteresting with no predictable patterns.

Example: White noise

$$
\begin{aligned}
r_1 &= 0.013 \\
r_2 &= -0.163 \\
r_3 &= 0.163 \\
r_4 &= -0.259 \\
r_5 &= -0.198 \\
r_6 &= 0.064 \\
r_7 &= -0.139 \\
r_8 &= -0.032 \\
r_9 &= 0.199 \\
r_{10} &= -0.240
\end{aligned}
$$



Sample autocorrelations for white noise series.
For uncorrelated data, we would expect each autocorrelation to be close to zero.

Sampling distribution of autocorrelations

Sampling distribution of r_k for white noise data is asymptotically $\text{N}(0,1/T)$.

- 95% of all r_k for white noise must lie within $\pm 1.96/\sqrt{T}$.
- If this is not the case, the series is probably not WN.
- Common to plot lines at $\pm 1.96/\sqrt{T}$ when plotting ACF. These are the *critical values*.

Sampling distribution of autocorrelations

Sampling distribution of r_k for white noise data is asymptotically N$(0,1/T)$.

- 95% of all r_k for white noise must lie within $\pm 1.96/\sqrt{T}$.
- If this is not the case, the series is probably not WN.
- Common to plot lines at $\pm 1.96/\sqrt{T}$ when plotting ACF. These are the ***critical values***.

Sampling distribution of autocorrelations

Sampling distribution of r_k for white noise data is asymptotically N(0,1$/T$).

- 95% of all r_k for white noise must lie within $\pm 1.96/\sqrt{T}$.
- If this is not the case, the series is probably not WN.
- Common to plot lines at $\pm 1.96/\sqrt{T}$ when plotting ACF. These are the ***critical values***.

Sampling distribution of autocorrelations

Sampling distribution of r_k for white noise data is asymptotically $\text{N}(0,1/T)$.

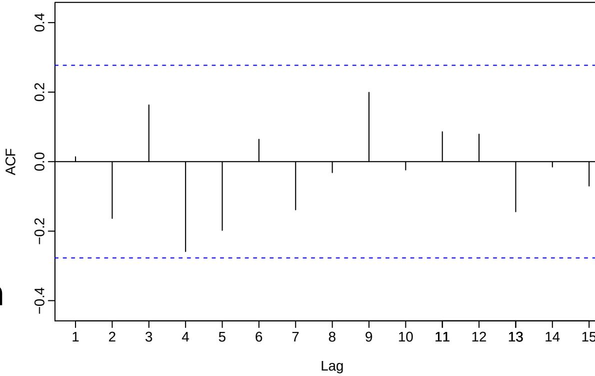
- 95% of all r_k for white noise must lie within $\pm 1.96/\sqrt{T}$.
- If this is not the case, the series is probably not WN.
- Common to plot lines at $\pm 1.96/\sqrt{T}$ when plotting ACF. These are the ***critical values***.

Autocorrelation

Example:

$T = 50$ and so critical values at $\pm 1.96/\sqrt{50} = \pm 0.28$.

All autocorrelation coefficients lie within these limits, confirming that the data are white noise.



(More precisely, the data cannot be distinguished from white noise.)

ACF of residuals

- We assume that the residuals are white noise (uncorrelated, mean zero, constant variance). If they aren't, then there is information left in the residuals that should be used in computing forecasts.
- So a standard residual diagnostic is to check the ACF of the residuals of a forecasting method.
- We *expect* these to look like white noise.

Dow-Jones naive forecasts revisited

$$\hat{y}_{t|t-1} = y_{t-1}$$
$$e_t = y_t - y_{t-1}$$

ACF of residuals

- We assume that the residuals are white noise (uncorrelated, mean zero, constant variance). If they aren't, then there is information left in the residuals that should be used in computing forecasts.
- So a standard residual diagnostic is to check the ACF of the residuals of a forecasting method.
- We *expect* these to look like white noise.

Dow-Jones naive forecasts revisited

$$\hat{y}_{t|t-1} = y_{t-1}$$
$$e_t = y_t - y_{t-1}$$

ACF of residuals

- We assume that the residuals are white noise (uncorrelated, mean zero, constant variance). If they aren't, then there is information left in the residuals that should be used in computing forecasts.
- So a standard residual diagnostic is to check the ACF of the residuals of a forecasting method.
- We *expect* these to look like white noise.

Dow-Jones naive forecasts revisited

$$\hat{y}_{t|t-1} = y_{t-1}$$
$$e_t = y_t - y_{t-1}$$

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.

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.

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.

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 $m =$

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

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

Autoregressive models

Autoregressive (AR) models:

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

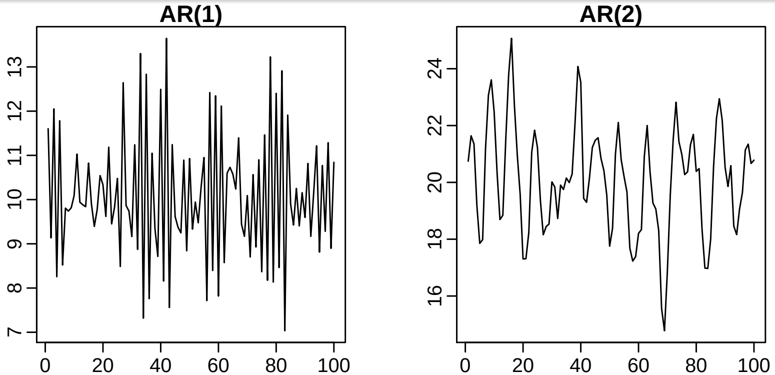
where e_t is white noise. This is a multiple regression with **lagged values** of y_t as predictors.

Autoregressive models

Autoregressive (AR) models:

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

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



AR(1) model

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

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

AR(1) model

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

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

AR(1) model

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

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

AR(1) model

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

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

Moving Average (MA) models

Moving Average (MA) models:

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

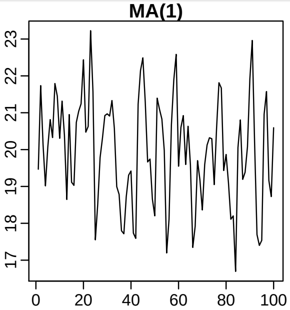
where e_t is white noise. This is a multiple regression with **past *errors*** as predictors. *Don't confuse this with moving average smoothing!*

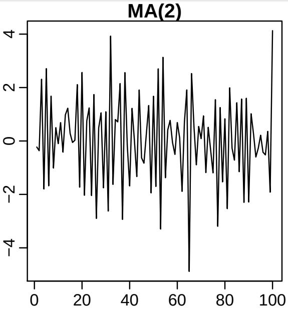
Moving Average (MA) models

Moving Average (MA) models:

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

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





ARIMA models

Autoregressive Moving Average models:

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

- Predictors include both **lagged values of y_t and lagged errors.**
- ARMA models can be used for a huge range of stationary time series.
- They model the short-term dynamics.
- An ARMA model applied to *differenced data* is an **ARIMA model**.

ARIMA models

Autoregressive Moving Average models:

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

- Predictors include both **lagged values of y_t and lagged errors.**
- ARMA models can be used for a huge range of stationary time series.
- They model the short-term dynamics.
- An ARMA model applied to *differenced* data is an **ARIMA model**.

ARIMA models

Autoregressive Moving Average models:

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

- Predictors include both **lagged values of y_t and lagged errors.**
- ARMA models can be used for a huge range of stationary time series.
- They model the short-term dynamics.
- An ARMA model applied to *differenced* data is an **ARIMA model**.

ARIMA models

Autoregressive Moving Average models:

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

- Predictors include both **lagged values of y_t and lagged errors.**
- ARMA models can be used for a huge range of stationary time series.
- They model the short-term dynamics.
- An ARMA model applied to *differenced* data is an **ARIMA model**.

ARIMA models

Autoregressive Moving Average models:

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

- Predictors include both **lagged values of y_t and lagged errors.**
- ARMA models can be used for a huge range of stationary time series.
- They model the short-term dynamics.
- An ARMA model applied to *differenced* data is an **ARIMA 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)

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)

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)

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)

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)

Understanding ARIMA models

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

ACF and PACF plots

- Recall that k-th autocorrelation r_k measures the linear relationship between y_t and y_{t-k}

ACF and PACF plots

- Recall that k-th autocorrelation r_k measures the linear relationship between y_t and y_{t-k}
- Now, if y_t and y_{t-1} are correlated, then y_t and y_{t-2} must be correlated

ACF and PACF plots

- Recall that k-th autocorrelation r_k measures the linear relationship between y_t and y_{t-k}
- Now, if y_t and y_{t-1} are correlated, then y_t and y_{t-2} must be correlated
- What is the correlation between y_t and y_{t-2} after removing the correlation between y_t and y_{t-1}?

ACF and PACF plots

- Recall that k-th autocorrelation r_k measures the linear relationship between y_t and y_{t-k}
- Now, if y_t and y_{t-1} are correlated, then y_t and y_{t-2} must be correlated
- What is the correlation between y_t and y_{t-2} after removing the correlation between y_t and y_{t-1}?
- α_k: k-th partial autocorrelation

ACF and PACF plots

- Recall that k-th autocorrelation r_k measures the linear relationship between y_t and y_{t-k}
- Now, if y_t and y_{t-1} are correlated, then y_t and y_{t-2} must be correlated
- What is the correlation between y_t and y_{t-2} after removing the correlation between y_t and y_{t-1}?
- α_k: k-th partial autocorrelation
- α_k: linear relationship between y_t and y_{t-k} after removing the effects of time lags $\kappa = 1, 2, \ldots, k-1$

ACF and PACF plots

- Recall that k-th autocorrelation r_k measures the linear relationship between y_t and y_{t-k}
- Now, if y_t and y_{t-1} are correlated, then y_t and y_{t-2} must be correlated
- What is the correlation between y_t and y_{t-2} after removing the correlation between y_t and y_{t-1}?
- α_k: k-th partial autocorrelation
- α_k: linear relationship between y_t and y_{t-k} after removing the effects of time lags $\kappa = 1, 2, \ldots, k-1$
- $\alpha_k =$ the estimate of ϕ_k in the autoregression model

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

ACF and PACF plots

- If data follows an ARIMA($p, d, 0$) or amd ARIMA($0, d, q$) model, ACF and PACF plots can help to determine the value of p or q

ACF and PACF plots

- If data follows an ARIMA$(p, d, 0)$ or amd ARIMA$(0, d, q)$ model, ACF and PACF plots can help to determine the value of p or q
- If both p and q are positive, ACF and PACF plots are not helpful

ACF and PACF plots

- If data follows an ARIMA($p, d, 0$) or amd ARIMA($0, d, q$) model, ACF and PACF plots can help to determine the value of p or q
- If both p and q are positive, ACF and PACF plots are not helpful
- Data may follow ARIMA($p, d, 0$) model if

ACF and PACF plots

- If data follows an ARIMA$(p, d, 0)$ or amd ARIMA$(0, d, q)$ model, ACF and PACF plots can help to determine the value of p or q
- If both p and q are positive, ACF and PACF plots are not helpful
- Data may follow ARIMA$(p, d, 0)$ model if
  - ACF is exponentially decaying or sinusoidal

ACF and PACF plots

- If data follows an ARIMA($p, d, 0$) or amd ARIMA($0, d, q$) model, ACF and PACF plots can help to determine the value of p or q
- If both p and q are positive, ACF and PACF plots are not helpful
- Data may follow ARIMA($p, d, 0$) model if
  - ACF is exponentially decaying or sinusoidal
  - Significant spike at lag p in PACF, but none beyond lag p

ACF and PACF plots

- If data follows an ARIMA($p, d, 0$) or amd ARIMA($0, d, q$) model, ACF and PACF plots can help to determine the value of p or q
- If both p and q are positive, ACF and PACF plots are not helpful
- Data may follow ARIMA($p, d, 0$) model if
  - ACF is exponentially decaying or sinusoidal
  - Significant spike at lag p in PACF, but none beyond lag p
- Data may follow ARIMA($0, d, q$) model if

ACF and PACF plots

- If data follows an ARIMA($p, d, 0$) or amd ARIMA($0, d, q$) model, ACF and PACF plots can help to determine the value of p or q
- If both p and q are positive, ACF and PACF plots are not helpful
- Data may follow ARIMA($p, d, 0$) model if
  - ACF is exponentially decaying or sinusoidal
  - Significant spike at lag p in PACF, but none beyond lag p
- Data may follow ARIMA($0, d, q$) model if
  - PACF is exponentially decaying or sinusoidal

ACF and PACF plots

- If data follows an ARIMA($p, d, 0$) or amd ARIMA($0, d, q$) model, ACF and PACF plots can help to determine the value of p or q
- If both p and q are positive, ACF and PACF plots are not helpful
- Data may follow ARIMA($p, d, 0$) model if
  - ACF is exponentially decaying or sinusoidal
  - Significant spike at lag p in PACF, but none beyond lag p
- Data may follow ARIMA($0, d, q$) model if
  - PACF is exponentially decaying or sinusoidal
  - Significant spike at lag p in ACF, but none beyond lag p

Akaike's Information Criterion

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

where p is the number of estimated parameters in the model.

- *Minimizing* the AIC gives the best model for prediction.

AIC corrected (for small sample bias)

$$\text{AIC}_\text{C} = \text{AIC} + \frac{2(p+1)(p+2)}{n-p}$$

Schwartz' Bayesian IC

$$\text{BIC} = \text{AIC} + p(\log(n) - 2)$$

Akaike's Information Criterion

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

where p is the number of estimated parameters in the model.

- *Minimizing* the AIC gives the best model for prediction.

AIC corrected (for small sample bias)

$$\text{AIC}_{\text{C}} = \text{AIC} + \frac{2(p+1)(p+2)}{n-p}$$

Schwartz' Bayesian IC

$$\text{BIC} = \text{AIC} + p(\log(n) - 2)$$

Akaike's Information Criterion

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

where p is the number of estimated parameters in the model.

- *Minimizing* the AIC gives the best model for prediction.

AIC corrected (for small sample bias)

$$\text{AIC}_\text{C} = \text{AIC} + \frac{2(p+1)(p+2)}{n-p}$$

Schwartz' Bayesian IC

$$\text{BIC} = \text{AIC} + p(\log(n) - 2)$$

Akaike's Information Criterion

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

where p is the number of estimated parameters in the model.

- *Minimizing* the AIC gives the best model for prediction.

AIC corrected (for small sample bias)

$$\text{AIC}_\text{C} = \text{AIC} + \frac{2(p+1)(p+2)}{n-p}$$

Schwartz' Bayesian IC

$$\text{BIC} = \text{AIC} + p(\log(n) - 2)$$

Akaike's Information Criterion

$$\text{AIC} = -2\log(\text{Likelihood}) + 2p$$

where p is the number of estimated parameters in the model.

- *Minimizing* the AIC gives the best model for prediction.

AIC corrected (for small sample bias)

$$\text{AIC}_{\text{C}} = \text{AIC} + \frac{2(p+1)(p+2)}{n-p}$$

Schwartz' Bayesian IC

$$\text{BIC} = \text{AIC} + p(\log(n) - 2)$$

Akaike's Information Criterion

- Value of AIC/AICc/BIC given in the R output.
- AIC does not have much meaning by itself. Only useful in comparison to AIC value for another model fitted to *same data set.*
- Consider several models with AIC values close to the minimum.
- A difference in AIC values of 2 or less is not regarded as substantial and you may choose the simpler but non-optimal model.
- AIC can be negative.

Akaike's Information Criterion

- Value of AIC/AICc/BIC given in the R output.
- AIC does not have much meaning by itself. Only useful in comparison to AIC value for another model fitted to *same data set.*
- Consider several models with AIC values close to the minimum.
- A difference in AIC values of 2 or less is not regarded as substantial and you may choose the simpler but non-optimal model.
- AIC can be negative.

Akaike's Information Criterion

- Value of AIC/AICc/BIC given in the R output.
- AIC does not have much meaning by itself. Only useful in comparison to AIC value for another model fitted to *same data set.*
- Consider several models with AIC values close to the minimum.
- A difference in AIC values of 2 or less is not regarded as substantial and you may choose the simpler but non-optimal model.
- AIC can be negative.

Akaike's Information Criterion

- Value of AIC/AICc/BIC given in the R output.
- AIC does not have much meaning by itself. Only useful in comparison to AIC value for another model fitted to *same data set.*
- Consider several models with AIC values close to the minimum.
- A difference in AIC values of 2 or less is not regarded as substantial and you may choose the simpler but non-optimal model.
- AIC can be negative.

Akaike's Information Criterion

- Value of AIC/AICc/BIC given in the R output.
- AIC does not have much meaning by itself. Only useful in comparison to AIC value for another model fitted to *same data set.*
- Consider several models with AIC values close to the minimum.
- A difference in AIC values of 2 or less is not regarded as substantial and you may choose the simpler but non-optimal model.
- AIC can be negative.

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

- First difference: $1 - B$.
- Double difference: $(1 - B)^2$.
- dth-order difference: $(1 - B)^d y_t$.
- Seasonal difference: $1 - B^m$.
- Seasonal difference followed by a first difference: $(1 - B)(1 - B^m)$.
- Multiply terms together together to see the combined effect:

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

Backshift notation

- First difference: $1 - B$.
- Double difference: $(1 - B)^2$.
- dth-order difference: $(1 - B)^d y_t$.
- Seasonal difference: $1 - B^m$.
- Seasonal difference followed by a first difference: $(1 - B)(1 - B^m)$.
- Multiply terms together together to see the combined effect:

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

Backshift notation

- First difference: $1 - B$.
- Double difference: $(1 - B)^2$.
- dth-order difference: $(1 - B)^d y_t$.
- Seasonal difference: $1 - B^m$.
- Seasonal difference followed by a first difference: $(1 - B)(1 - B^m)$.
- Multiply terms together 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

- First difference: $1 - B$.
- Double difference: $(1 - B)^2$.
- dth-order difference: $(1 - B)^d y_t$.
- Seasonal difference: $1 - B^m$.
- Seasonal difference followed by a first difference: $(1 - B)(1 - B^m)$.
- Multiply terms together together to see the combined effect:

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

Backshift notation

- First difference: $1 - B$.
- Double difference: $(1 - B)^2$.
- dth-order difference: $(1 - B)^d y_t$.
- Seasonal difference: $1 - B^m$.
- Seasonal difference followed by a first difference: $(1 - B)(1 - B^m)$.
- Multiply terms together together to see the combined effect:

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

Backshift notation

- First difference: $1 - B$.
- Double difference: $(1 - B)^2$.
- dth-order difference: $(1 - B)^d y_t$.
- Seasonal difference: $1 - B^m$.
- Seasonal difference followed by a first difference: $(1 - B)(1 - B^m)$.
- Multiply terms together together to see the combined effect:

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

Backshift notation for ARIMA

- ARMA model:

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

$$\phi(B) y_t = c + \theta(B) e_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

$$\text{and } \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

- ARIMA(1,1,1) model:

$$(1 - \phi_1 B) \quad (1 - B) y_t \quad = \quad c + (1 + \theta_1 B) e_t$$

Backshift notation for ARIMA

- ARMA model:

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

$$\phi(B) y_t = c + \theta(B) e_t$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$

and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$.

- ARIMA(1,1,1) model:

$$(1 - \phi_1 B) \quad (1 - B) y_t \quad = \quad c + (1 + \theta_1 B) e_t$$

Backshift notation for ARIMA

- ARMA model:

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

$$\phi(B) y_t = c + \theta(B) e_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

$$\text{and } \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

- ARIMA(1,1,1) model:

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

Backshift notation for ARIMA

- ARMA model:

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

$$\phi(B) y_t = c + \theta(B) e_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

$$\text{and } \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

- ARIMA(1,1,1) model:

$$\underbrace{(1 - \phi_1 B)}_{\text{AR(1)}} \quad (1 - B) y_t \quad = \quad c + (1 + \theta_1 B) e_t$$

Backshift notation for ARIMA

- ARMA model:

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

$$\phi(B) y_t = c + \theta(B) e_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

$$\text{and } \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

- ARIMA(1,1,1) model:

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

Seasonal ARIMA models

$$\text{ARIMA} \;\; (p, d, q) \;\; (P, D, Q)_m$$

where $m =$ number of periods per season.

Seasonal ARIMA models

$$\text{ARIMA} \quad \underbrace{(p, d, q)} \quad (P, D, Q)_m$$

$$\uparrow$$

$$\left(\begin{array}{l} \text{Non-} \\ \text{seasonal} \\ \text{part of the} \\ \text{model} \end{array} \right)$$

where $m =$ number of periods per season.

Seasonal ARIMA models

$$\text{ARIMA} \quad (p,d,q) \quad \underbrace{(P,D,Q)_m}_{\uparrow \atop \left(\begin{array}{l}\text{Seasonal}\\ \text{part of}\\ \text{the}\\ \text{model}\end{array}\right)}$$

where $m =$ number of periods per season.

Seasonal ARIMA models

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

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

Seasonal ARIMA models

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

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

Seasonal ARIMA models

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

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

Seasonal ARIMA models

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

$$(1-\phi_1 B)(1-\Phi_1 B^4)\underbrace{(1-B)}_{\left(\substack{\text{Non-seasonal}\\ \text{difference}}\right)}(1-B^4)y_t \;=\; (1+\theta_1 B)(1+\Theta_1 B^4)e_t.$$

Seasonal ARIMA models

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

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

$\begin{pmatrix}\text{Seasonal}\\ \text{AR(1)}\end{pmatrix}$

Seasonal ARIMA models

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

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

$\begin{pmatrix}\text{Non-seasonal}\\ \text{AR(1)}\end{pmatrix}$

Seasonal ARIMA models

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

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

$\begin{pmatrix}\text{Seasonal}\\ \text{MA(1)}\end{pmatrix}$

Seasonal ARIMA models

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

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

$$\begin{pmatrix}\text{Non-seasonal}\\ \text{MA(1)}\end{pmatrix}$$

ACF and PACF plots

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF

ACF and PACF plots

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF
- An ARIMA$(0,0,0)(1,0,0)_{12}$ model will show

ACF and PACF plots

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF
- An ARIMA$(0, 0, 0)(1, 0, 0)_{12}$ model will show
  - Exponential decay in the seasonal lags of the ACF: 12, 24, 36, ...

ACF and PACF plots

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF
- An $\text{ARIMA}(0,0,0)(1,0,0)_{12}$ model will show
  - Exponential decay in the seasonal lags of the ACF: 12, 24, 36, ...
  - Single significant spike at lag 12 in the PACF

ACF and PACF plots

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF
- An ARIMA$(0,0,0)(1,0,0)_{12}$ model will show
  - Exponential decay in the seasonal lags of the ACF: 12, 24, 36, ...
  - Single significant spike at lag 12 in the PACF
- An ARIMA$(0,0,0)(0,0,1)_{12}$ model will show

ACF and PACF plots

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF
- An ARIMA$(0,0,0)(1,0,0)_{12}$ model will show
  - Exponential decay in the seasonal lags of the ACF: 12, 24, 36, ...
  - Single significant spike at lag 12 in the PACF
- An ARIMA$(0,0,0)(0,0,1)_{12}$ model will show
  - Exponential decay in the seasonal lags of the PACF: 12, 24, 36, ...

ACF and PACF plots

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF
- An ARIMA$(0,0,0)(1,0,0)_{12}$ model will show
  - Exponential decay in the seasonal lags of the ACF: 12, 24, 36, ...
  - Single significant spike at lag 12 in the PACF
- An ARIMA$(0,0,0)(0,0,1)_{12}$ model will show
  - Exponential decay in the seasonal lags of the PACF: 12, 24, 36, ...
  - Single significant spike at lag 12 in the ACF

Regression with ARIMA errors

Regression models

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

- y_t modeled as function of k explanatory variables $x_{1,t}, \dots, x_{k,t}$.
- Usually, we assume that n_t is WN.
- Now we want to allow n_t to be autocorrelated.

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

where $\quad (1 - \phi_1 B)(1 - B)n_t = (1 - \theta_1 B)e_t$

and e_t is white noise .

Regression with ARIMA errors

Regression models

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

- y_t modeled as function of k explanatory variables $x_{1,t}, \ldots, x_{k,t}$.
- Usually, we assume that n_t is WN.
- Now we want to allow n_t to be autocorrelated.

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

where $\quad (1 - \phi_1 B)(1 - B) n_t = (1 - \theta_1 B) e_t$

and e_t is white noise .

Regression with ARIMA errors

Regression models

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

- y_t modeled as function of k explanatory variables $x_{1,t}, \dots, x_{k,t}$.
- Usually, we assume that n_t is WN.
- Now we want to allow n_t to be autocorrelated.

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

where $\quad (1 - \phi_1 B)(1 - B)n_t = (1 - \theta_1 B)e_t$

and e_t is white noise .

Regression with ARIMA errors

Regression models

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

- y_t modeled as function of k explanatory variables $x_{1,t}, \dots, x_{k,t}$.
- Usually, we assume that n_t is WN.
- Now we want to allow n_t to be autocorrelated.

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

where $\quad (1 - \phi_1 B)(1 - B)n_t = (1 - \theta_1 B)e_t$

and e_t is white noise .

Regression with ARIMA errors

Regression models

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

- y_t modeled as function of k explanatory variables $x_{1,t}, \ldots, x_{k,t}$.
- Usually, we assume that n_t is WN.
- Now we want to allow n_t to be autocorrelated.

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

where $\quad (1 - \phi_1 B)(1 - B) n_t = (1 - \theta_1 B) e_t$

and e_t is white noise .

Residuals and errors

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

$$\text{where} \quad (1-\phi_1 B)(1-B)n_t = (1-\theta_1 B)e_t$$

- Be careful in distinguishing n_t from e_t.
- n_t are the "*errors*" and e_t are the "*residuals*".
- In ordinary regression, n_t is assumed to be white noise and so $n_t = e_t$.

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

Now a regression with ARMA(1,1) error

Residuals and errors

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

$$\text{where} \quad (1-\phi_1 B)(1-B)n_t = (1-\theta_1 B)e_t$$

- Be careful in distinguishing n_t from e_t.
- n_t are the "*errors*" and e_t are the "*residuals*".
- In ordinary regression, n_t is assumed to be white noise and so $n_t = e_t$.

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

Now a regression with ARMA(1,1) error

Residuals and errors

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$
$$\text{where} \quad (1-\phi_1 B)(1-B)n_t = (1-\theta_1 B)e_t$$

- Be careful in distinguishing n_t from e_t.
- n_t are the "*errors*" and e_t are the "*residuals*".
- In ordinary regression, n_t is assumed to be white noise and so $n_t = e_t$.

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

Now a regression with ARMA(1,1) error

Residuals and errors

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

$$\text{where} \quad (1-\phi_1 B)(1-B)n_t = (1-\theta_1 B)e_t$$

- Be careful in distinguishing n_t from e_t.
- n_t are the "*errors*" and e_t are the "*residuals*".
- In ordinary regression, n_t is assumed to be white noise and so $n_t = e_t$.

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

Now a regression with ARMA(1,1) error

Residuals and errors

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$
$$\text{where} \quad (1 - \phi_1 B)(1 - B)n_t = (1 - \theta_1 B)e_t$$

- Be careful in distinguishing n_t from e_t.
- n_t are the "*errors*" and e_t are the "*residuals*".
- In ordinary regression, n_t is assumed to be white noise and so $n_t = e_t$.

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

Now a regression with ARMA(1,1) error

Residuals and errors

Example: n_t = ARIMA(1,1,1)

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

where $\quad (1-\phi_1 B)(1-B)n_t = (1-\theta_1 B)e_t$

- Be careful in distinguishing n_t from e_t.
- n_t are the "*errors*" and e_t are the "*residuals*".
- In ordinary regression, n_t is assumed to be white noise and so $n_t = e_t$.

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

Now a regression with ARMA(1,1) error

Regression with ARIMA errors

Any regression with an ARIMA error can be rewritten as a regression with an ARMA error by differencing all variables with the same differencing operator as in the ARIMA model.

Original data

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

$$\text{where} \quad \phi(B)(1-B)^d n_t = \theta(B) e_t$$

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

$$\text{where} \quad \phi(B) n_t = \theta(B) e_t$$

$$\text{and} \quad y'_t = (1-B)^d y_t, \quad \text{etc.}$$

Regression with ARIMA errors

Any regression with an ARIMA error can be rewritten as a regression with an ARMA error by differencing all variables with the same differencing operator as in the ARIMA model.

Original data

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

$$\text{where} \quad \phi(B)(1-B)^d n_t = \theta(B) e_t$$

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

$$\text{where} \quad \phi(B) n_t = \theta(B) e_t$$

$$\text{and} \quad y'_t = (1-B)^d y_t, \quad \text{etc.}$$

Regression with ARIMA errors

Any regression with an ARIMA error can be rewritten as a regression with an ARMA error by differencing all variables with the same differencing operator as in the ARIMA model.

Original data

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_k x_{k,t} + n_t$$

where $\quad \phi(B)(1-B)^d n_t = \theta(B) e_t$

After differencing all variables

$$y'_t = b_1 x'_{1,t} + \cdots + b_k x'_{k,t} + n'_t.$$

where $\quad \phi(B) n_t = \theta(B) e_t$

and $\quad y'_t = (1-B)^d y_t, \quad$ etc.

Modeling procedure

Problems with OLS and autocorrelated errors

1. OLS no longer the best way to compute coefficients as it does not take account of time-relationships in data.
2. Standard errors of coefficients are incorrect — most likely too small. This invalidates tests and prediction intervals.

Second problem more serious because it can lead to misleading results.

Modeling procedure

Problems with OLS and autocorrelated errors

1. OLS no longer the best way to compute coefficients as it does not take account of time-relationships in data.
2. Standard errors of coefficients are incorrect — most likely too small. This invalidates tests and prediction intervals.

- Second problem more serious because it can lead to misleading results.
- If standard errors obtained using OLS too small, some explanatory variables may appear to be significant when, in fact, they are not. This is known as "spurious regression."

Modeling procedure

Problems with OLS and autocorrelated errors

1. OLS no longer the best way to compute coefficients as it does not take account of time-relationships in data.
2. Standard errors of coefficients are incorrect — most likely too small. This invalidates tests and prediction intervals.

- Second problem more serious because it can lead to misleading results.
- If standard errors obtained using OLS too small, some explanatory variables may appear to be significant when, in fact, they are not. This is known as “spurious regression.”

Modeling procedure

Problems with OLS and autocorrelated errors

1. OLS no longer the best way to compute coefficients as it does not take account of time-relationships in data.
2. Standard errors of coefficients are incorrect — most likely too small. This invalidates tests and prediction intervals.

- Second problem more serious because it can lead to misleading results.
- If standard errors obtained using OLS too small, some explanatory variables may appear to be significant when, in fact, they are not. This is known as “spurious regression.”

Modeling procedure

Problems with OLS and autocorrelated errors

1. OLS no longer the best way to compute coefficients as it does not take account of time-relationships in data.
2. Standard errors of coefficients are incorrect — most likely too small. This invalidates tests and prediction intervals.

- Second problem more serious because it can lead to misleading results.
- If standard errors obtained using OLS too small, some explanatory variables may appear to be significant when, in fact, they are not. This is known as “spurious regression.”

Modeling procedure

- Estimation only works when all predictor variables are deterministic or stationary and the errors are stationary.
- So difference stochastic variables as required until all variables appear stationary. Then fit model with ARMA errors.
- `auto.arima()` will handle order selection and differencing (but only checks that errors are stationary).

Modeling procedure

- Estimation only works when all predictor variables are deterministic or stationary and the errors are stationary.
- So difference stochastic variables as required until all variables appear stationary. Then fit model with ARMA errors.
- `auto.arima()` will handle order selection and differencing (but only checks that errors are stationary).

Modeling procedure

- Estimation only works when all predictor variables are deterministic or stationary and the errors are stationary.
- So difference stochastic variables as required until all variables appear stationary. Then fit model with ARMA errors.
- `auto.arima()` will handle order selection and differencing (but only checks that errors are stationary).

Outline

Time series decomposition

$$Y_t = f(S_t, T_t, E_t)$$

where
$$\begin{aligned} Y_t &= \text{data at period } t \\ S_t &= \text{seasonal component at period } t \\ T_t &= \text{trend-cycle component at period } t \\ E_t &= \text{remainder (or irregular or error) component at period } t \end{aligned}$$

Additive decomposition: $Y_t = S_t + T_t + E_t$.

Multiplicative decomposition: $Y_t = S_t \times T_t \times E_t$.

Time series decomposition

$$Y_t = f(S_t, T_t, E_t)$$

where $Y_t =$ data at period t

$S_t =$ seasonal component at period t

$T_t =$ trend-cycle component at period t

$E_t =$ remainder (or irregular or error) component at period t

Additive decomposition: $Y_t = S_t + T_t + E_t$.

Multiplicative decomposition: $Y_t = S_t \times T_t \times E_t$.

Time series decomposition

$$Y_t = f(S_t, T_t, E_t)$$

where
$$\begin{aligned} Y_t &= \text{data at period } t \\ S_t &= \text{seasonal component at period } t \\ T_t &= \text{trend-cycle component at period } t \\ E_t &= \text{remainder (or irregular or error) component at period } t \end{aligned}$$

Additive decomposition: $Y_t = S_t + T_t + E_t$.
Multiplicative decomposition: $Y_t = S_t \times T_t \times E_t$.

Time series decomposition

- Additive model appropriate if magnitude of seasonal fluctuations does not vary with level.
- If seasonal are proportional to level of series, then multiplicative model appropriate.
- Multiplicative decomposition more prevalent with economic series
- Logs turn multiplicative relationship into an additive relationship:

$$Y_t = S_t \times T_t \times E_t \quad \Rightarrow \quad \log Y_t = \log S_t + \log T_t + \log E_t.$$

Time series decomposition

- Additive model appropriate if magnitude of seasonal fluctuations does not vary with level.
- If seasonal are proportional to level of series, then multiplicative model appropriate.
- Multiplicative decomposition more prevalent with economic series
- Logs turn multiplicative relationship into an additive relationship:

$$Y_t = S_t \times T_t \times E_t \quad \Rightarrow \quad \log Y_t = \log S_t + \log T_t + \log E_t.$$

Time series decomposition

- Additive model appropriate if magnitude of seasonal fluctuations does not vary with level.
- If seasonal are proportional to level of series, then multiplicative model appropriate.
- Multiplicative decomposition more prevalent with economic series
- Logs turn multiplicative relationship into an additive relationship:

$$Y_t = S_t \times T_t \times E_t \quad \Rightarrow \quad \log Y_t = \log S_t + \log T_t + \log E_t.$$

Time series decomposition

- Additive model appropriate if magnitude of seasonal fluctuations does not vary with level.
- If seasonal are proportional to level of series, then multiplicative model appropriate.
- Multiplicative decomposition more prevalent with economic series
- Logs turn multiplicative relationship into an additive relationship:

$$Y_t = S_t \times T_t \times E_t \quad \Rightarrow \quad \log Y_t = \log S_t + \log T_t + \log E_t.$$

Seasonal adjustment

- Useful by-product of decomposition: an easy way to calculate seasonally adjusted data.
- Additive decomposition: seasonally adjusted data given by

$$Y_t - S_t = T_t + E_t$$

- Multiplicative decomposition: seasonally adjusted data given by

$$Y_t / S_t = T_t \times E_t$$

Seasonal adjustment

- Useful by-product of decomposition: an easy way to calculate seasonally adjusted data.
- Additive decomposition: seasonally adjusted data given by

$$Y_t - S_t = T_t + E_t$$

- Multiplicative decomposition: seasonally adjusted data given by

$$Y_t / S_t = T_t \times E_t$$

Seasonal adjustment

- Useful by-product of decomposition: an easy way to calculate seasonally adjusted data.
- Additive decomposition: seasonally adjusted data given by

$$Y_t - S_t = T_t + E_t$$

- Multiplicative decomposition: seasonally adjusted data given by

$$Y_t / S_t = T_t \times E_t$$

Forecasting and decomposition

- Forecast seasonal component by repeating the last year
- Forecast seasonally adjusted data using non-seasonal time series method. E.g.,
  - Holt's method — next topic
  - Random walk with drift model
- Combine forecasts of seasonal component with forecasts of seasonally adjusted data to get forecasts of original data.
- Sometimes a decomposition is useful just for understanding the data before building a separate forecasting model.

Forecasting and decomposition

- Forecast seasonal component by repeating the last year
- Forecast seasonally adjusted data using non-seasonal time series method. E.g.,
  - Holt's method — next topic
  - Random walk with drift model
- Combine forecasts of seasonal component with forecasts of seasonally adjusted data to get forecasts of original data.
- Sometimes a decomposition is useful just for understanding the data before building a separate forecasting model.

Forecasting and decomposition

- Forecast seasonal component by repeating the last year
- Forecast seasonally adjusted data using non-seasonal time series method. E.g.,
  - Holt's method — next topic
  - Random walk with drift model
- Combine forecasts of seasonal component with forecasts of seasonally adjusted data to get forecasts of original data.
- Sometimes a decomposition is useful just for understanding the data before building a separate forecasting model.

Forecasting and decomposition

- Forecast seasonal component by repeating the last year
- Forecast seasonally adjusted data using non-seasonal time series method. E.g.,
  - Holt's method — next topic
  - Random walk with drift model
- Combine forecasts of seasonal component with forecasts of seasonally adjusted data to get forecasts of original data.
- Sometimes a decomposition is useful just for understanding the data before building a separate forecasting model.

Forecasting and decomposition

- Forecast seasonal component by repeating the last year
- Forecast seasonally adjusted data using non-seasonal time series method. E.g.,
  - Holt's method — next topic
  - Random walk with drift model
- Combine forecasts of seasonal component with forecasts of seasonally adjusted data to get forecasts of original data.
- Sometimes a decomposition is useful just for understanding the data before building a separate forecasting model.

Forecasting and decomposition

- Forecast seasonal component by repeating the last year
- Forecast seasonally adjusted data using non-seasonal time series method. E.g.,
  - Holt's method — next topic
  - Random walk with drift model
- Combine forecasts of seasonal component with forecasts of seasonally adjusted data to get forecasts of original data.
- Sometimes a decomposition is useful just for understanding the data before building a separate forecasting model.

Simple methods

Random walk forecasts

$$\hat{y}_{T+1|T} = y_T$$

Average forecasts

$$\hat{y}_{T+1|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- Want something in between that weights most recent data more highly.
- Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially.

Simple methods

Random walk forecasts

$$\hat{y}_{T+1|T} = y_T$$

Average forecasts

$$\hat{y}_{T+1|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- Want something in between that weights most recent data more highly.
- Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially.

Simple methods

Random walk forecasts

$$\hat{y}_{T+1|T} = y_T$$

Average forecasts

$$\hat{y}_{T+1|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- Want something in between that weights most recent data more highly.
- Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially.

Simple methods

Random walk forecasts

$$\hat{y}_{T+1|T} = y_T$$

Average forecasts

$$\hat{y}_{T+1|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- Want something in between that weights most recent data more highly.
- Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially.

Simple Exponential Smoothing

Forecast equation

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots ,$$

where $0 \leq \alpha \leq 1$.

	Weights assigned to observations for:			
Observation	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	0.2	0.4	0.6	0.8
y_{T-1}	0.16	0.24	0.24	0.16
y_{T-2}	0.128	0.144	0.096	0.032
y_{T-3}	0.1024	0.0864	0.0384	0.0064
y_{T-4}	$(0.2)(0.8)^4$	$(0.4)(0.6)^4$	$(0.6)(0.4)^4$	$(0.8)(0.2)^4$
y_{T-5}	$(0.2)(0.8)^5$	$(0.4)(0.6)^5$	$(0.6)(0.4)^5$	$(0.8)(0.2)^5$

Simple Exponential Smoothing

Forecast equation

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots,$$

where $0 \le \alpha \le 1$.

	Weights assigned to observations for:			
Observation	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	0.2	0.4	0.6	0.8
y_{T-1}	0.16	0.24	0.24	0.16
y_{T-2}	0.128	0.144	0.096	0.032
y_{T-3}	0.1024	0.0864	0.0384	0.0064
y_{T-4}	$(0.2)(0.8)^4$	$(0.4)(0.6)^4$	$(0.6)(0.4)^4$	$(0.8)(0.2)^4$
y_{T-5}	$(0.2)(0.8)^5$	$(0.4)(0.6)^5$	$(0.6)(0.4)^5$	$(0.8)(0.2)^5$

Simple Exponential Smoothing

Weighted average form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

for $t = 1, \ldots, T$, where $0 \leq \alpha \leq 1$ is the smoothing parameter.

The process has to start somewhere, so we let the first forecast of y_1 be denoted by ℓ_0. Then

$$\begin{aligned}
\hat{y}_{2|1} &= \alpha y_1 + (1-\alpha)\ell_0 \\
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)\hat{y}_{3|2} \\
&\vdots
\end{aligned}$$

Simple Exponential Smoothing

Weighted average form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

for $t = 1, \dots, T$, where $0 \le \alpha \le 1$ is the smoothing parameter.

The process has to start somewhere, so we let the first forecast of y_1 be denoted by ℓ_0. Then

$$\begin{aligned}
\hat{y}_{2|1} &= \alpha y_1 + (1-\alpha)\ell_0 \\
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)\hat{y}_{3|2} \\
&\vdots
\end{aligned}$$

Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Substituting each equation into the following equation:

$$\begin{aligned}
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
&= \alpha y_2 + (1-\alpha)\left[\alpha y_1 + (1-\alpha)\ell_0\right] \\
&= \alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2\ell_0 \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)[\alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2\ell_0] \\
&= \alpha y_3 + \alpha(1-\alpha)y_2 + \alpha(1-\alpha)^2 y_1 + (1-\alpha)^3\ell_0 \\
&\;\vdots \\
\hat{y}_{T+1|T} &= \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots + (1-\alpha)^T\ell_0
\end{aligned}$$

Exponentially weighted average

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T\ell_0$$

Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Substituting each equation into the following equation:

$$\begin{aligned}
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
&= \alpha y_2 + (1-\alpha)\left[\alpha y_1 + (1-\alpha)\ell_0\right] \\
&= \alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2\ell_0 \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)[\alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2\ell_0] \\
&= \alpha y_3 + \alpha(1-\alpha)y_2 + \alpha(1-\alpha)^2 y_1 + (1-\alpha)^3\ell_0 \\
&\ \vdots \\
\hat{y}_{T+1|T} &= \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots + (1-\alpha)^T\ell_0
\end{aligned}$$

Exponentially weighted average

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1}\alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T\ell_0$$

Simple exponential smoothing

Initialization

- Last term in weighted moving average is $(1-\alpha)^T\hat{\ell}_0$.
- So value of ℓ_0 plays a role in *all* subsequent forecasts.
- Weight is small unless α close to zero or T small.
- Common to set $\ell_0 = y_1$. Better to treat it as a parameter, along with α.

Simple exponential smoothing

Initialization

- Last term in weighted moving average is $(1-\alpha)^T\hat{\ell}_0$.
- So value of ℓ_0 plays a role in *all* subsequent forecasts.
- Weight is small unless α close to zero or T small.
- Common to set $\ell_0 = y_1$. Better to treat it as a parameter, along with α.

Simple exponential smoothing

Initialization

- Last term in weighted moving average is $(1-\alpha)^T\hat{\ell}_0$.
- So value of ℓ_0 plays a role in *all* subsequent forecasts.
- Weight is small unless α close to zero or T small.
- Common to set $\ell_0 = y_1$. Better to treat it as a parameter, along with α.

Simple exponential smoothing

Initialization

- Last term in weighted moving average is $(1-\alpha)^T\hat{\ell}_0$.
- So value of ℓ_0 plays a role in *all* subsequent forecasts.
- Weight is small unless α close to zero or T small.
- Common to set $\ell_0 = y_1$. Better to treat it as a parameter, along with α.

Simple exponential smoothing

Optimization

- We can choose α and ℓ_0 by minimizing MSE:

$$\text{MSE} = \frac{1}{T-1}\sum_{t=2}^{T}(y_t - y_{t|t-1})^2$$

- Unlike regression there is no closed form solution — use numerical optimization.

Simple exponential smoothing

Optimization

- We can choose α and ℓ_0 by minimizing MSE:

$$\text{MSE} = \frac{1}{T-1}\sum_{t=2}^{T}(y_t - y_{t|t-1})^2$$

- Unlike regression there is no closed form solution — use numerical optimization.

Simple exponential smoothing

Multi-step forecasts

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \dots$$

- A "flat" forecast function.
- Remember, a forecast is an estimated mean of a future value.
- So with no trend, no seasonality, and no other patterns, the forecasts are constant.

Simple exponential smoothing

Multi-step forecasts

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \ldots$$

- A "flat" forecast function.
- Remember, a forecast is an estimated mean of a future value.
- So with no trend, no seasonality, and no other patterns, the forecasts are constant.

Simple exponential smoothing

Multi-step forecasts

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \ldots$$

- A "flat" forecast function.
- Remember, a forecast is an estimated mean of a future value.
- So with no trend, no seasonality, and no other patterns, the forecasts are constant.

Simple exponential smoothing

Multi-step forecasts

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \ldots$$

- A “flat” forecast function.
- Remember, a forecast is an estimated mean of a future value.
- So with no trend, no seasonality, and no other patterns, the forecasts are constant.