2. The forecaster’s toolbox

OTexts.com/fpp/2/
1 Time series graphics
2 Seasonal or cyclic?
3 Autocorrelation
4 Forecast residuals
5 White noise
6 Evaluating forecast accuracy
Economy class passengers: Melbourne–Sydney

plot(melsyd[, "Economy.Class"])
Seasonal plot: antidiabetic drug sales

Year
$ million
● ● ● ● ●
●
●
● ● ● ● ●
● ● ● ●
●
●●
●... Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

2008
2007
2006
2005
2004
2003
2002
2001
2000
1999
1998
1997
1996
1995
1994
1993
1992
1991
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

Seasonal subseries plot: antidiabetic drug sales

> monthplot(a10)
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`
Quarterly Australian Beer Production

```r
beer <- window(ausbeer,start=1992)
plot(beer)
seasonplot(beer,year.labels=TRUE)
monthplot(beer)
```
Australian quarterly beer production (megaliters)

Time series graphics
Seasonal plot: quarterly beer production
Seasonal subseries plot: quarterly beer production

Megalitres

Quarter

Jan Apr Jul Oct

Quarter
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`
1 Time series graphics
2 Seasonal or cyclic?
3 Autocorrelation
4 Forecast residuals
5 White noise
6 Evaluating forecast accuracy
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).
Time series patterns

Australian electricity production

<table>
<thead>
<tr>
<th>Year</th>
<th>GWh</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>8000</td>
</tr>
<tr>
<td>1985</td>
<td>10000</td>
</tr>
<tr>
<td>1990</td>
<td>12000</td>
</tr>
<tr>
<td>1995</td>
<td>14000</td>
</tr>
</tbody>
</table>
Time series patterns

Australian clay brick production

Year

million units


200 300 400 500 600

Seasonal or cyclic?
Sales of new one-family houses, USA

Total sales


30 40 50 60 70 80 90
### US Treasury bill contracts

<table>
<thead>
<tr>
<th>Day</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>85</td>
</tr>
<tr>
<td>20</td>
<td>86</td>
</tr>
<tr>
<td>40</td>
<td>87</td>
</tr>
<tr>
<td>60</td>
<td>88</td>
</tr>
<tr>
<td>80</td>
<td>89</td>
</tr>
<tr>
<td>100</td>
<td>90</td>
</tr>
</tbody>
</table>
Time series patterns

Annual Canadian Lynx trappings

Number trapped

1820 1840 1860 1880 1900 1920

Time

Forecasting: Principles and Practice

Seasonal or cyclic?
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.
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.
Outline

1. Time series graphics
2. Seasonal or cyclic?
3. Autocorrelation
4. Forecast residuals
5. White noise
6. Evaluating forecast accuracy
**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 \((y \text{ and } X)\).

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

We measure the relationship between: \(y_t \text{ and } y_{t-1}\), \(y_t \text{ and } y_{t-2}\), \(y_t \text{ and } y_{t-3}\) etc.
Example: Beer production

```r
> lag.plot(beer, lags=9)
```
Example: Beer production

> lag.plot(beer, lags=9, do.lines=FALSE)
Lagged scatterplots

- Each graph shows $y_t$ plotted against $y_{t-k}$ for different values of $k$.
- The autocorrelations are the correlations associated with these scatterplots.
Each graph shows $y_t$ plotted against $y_{t-k}$ for different values of $k$.

The autocorrelations are the correlations associated with these scatterplots.
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

$$r_k = \frac{c_k}{c_0}$$

- $r_1$ 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 $y_{t-k}$. 

*Forecasting: Principles and Practice*
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 $$r_k = \frac{c_k}{c_0}$$

- $r_1$ 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 $y_{t-k}$. 
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

$$r_k = \frac{c_k}{c_0}$$

- $r_1$ 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 $y_{t-k}$. 

**Forecasting: Principles and Practice**

*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

$$r_k = \frac{c_k}{c_0}$$

- $r_1$ 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 $y_{t-k}$. 

*Forecasting: Principles and Practice*
Autocorrelation

Results for first 9 lags for beer data:

<table>
<thead>
<tr>
<th></th>
<th>$r_1$</th>
<th>$r_2$</th>
<th>$r_3$</th>
<th>$r_4$</th>
<th>$r_5$</th>
<th>$r_6$</th>
<th>$r_7$</th>
<th>$r_8$</th>
<th>$r_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−0.126</td>
<td>−0.650</td>
<td>−0.094</td>
<td>0.863</td>
<td>−0.099</td>
<td>−0.642</td>
<td>−0.098</td>
<td>0.834</td>
<td>−0.116</td>
</tr>
</tbody>
</table>
Autocorrelation

Results for first 9 lags for beer data:

<table>
<thead>
<tr>
<th>Lag</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_1$</td>
<td>$-0.126$</td>
<td>$-0.650$</td>
<td>$-0.094$</td>
<td>$0.863$</td>
<td>$-0.099$</td>
<td>$-0.642$</td>
<td>$-0.098$</td>
<td>$0.834$</td>
<td>$-0.116$</td>
</tr>
</tbody>
</table>
Autocorrelation

- $r_4$ is higher than for the other lags. This is due to the **seasonal pattern in the data**: the peaks tend to be 4 quarters apart and the troughs tend to be 2 quarters apart.

- $r_2$ is more negative than for the other lags because troughs tend to be 2 quarters behind peaks.

- Together, the autocorrelations at lags 1, 2, ... make up the autocorrelation or ACF.

- The plot is known as a correlogram.
$r_4$ is higher than for the other lags. This is due to the seasonal pattern in the data: the peaks tend to be 4 quarters apart and the troughs tend to be 2 quarters apart.

$r_2$ is more negative than for the other lags because troughs tend to be 2 quarters behind peaks.

Together, the autocorrelations at lags 1, 2, \ldots, make up the autocorrelation or ACF.

The plot is known as a correlogram.
\( r_4 \) higher than for the other lags. This is due to the seasonal pattern in the data: the peaks tend to be 4 quarters apart and the troughs tend to be 2 quarters apart.

\( r_2 \) is more negative than for the other lags because troughs tend to be 2 quarters behind peaks.

Together, the autocorrelations at lags 1, 2, \ldots, make up the autocorrelation or ACF.

The plot is known as a correlogram.
Autocorrelation

- $r_4$ higher than for the other lags. This is due to the seasonal pattern in the data: the peaks tend to be 4 quarters apart and the troughs tend to be 2 quarters apart.
- $r_2$ is more negative than for the other lags because troughs tend to be 2 quarters behind peaks.
- Together, the autocorrelations at lags 1, 2, ..., make up the autocorrelation or ACF.
- The plot is known as a correlogram.
Acf(beer)

Lag

ACF

-0.5 0.0 0.5

Lag

ACF

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
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, ...
Australian monthly electricity production forecasting: Principles and Practice

Autocorrelation

Lag

ACF

0 10 20 30 40

Lag

-0.2 0.0 0.2 0.4 0.6 0.8
Time plot shows clear trend and seasonality. The same features are reflected in the ACF.

- The slowly decaying ACF indicates trend.
- The ACF peaks at lags 12, 24, 36, ..., indicate seasonality of length 12.
Time plot shows clear trend and seasonality. The same features are reflected in the ACF.

- The slowly decaying ACF indicates trend.
- The ACF peaks at lags 12, 24, 36, ..., indicate seasonality of length 12.
Which is which?

1. Daily morning temperature of a cow

2. Accidental deaths in USA (monthly)

3. International airline passengers

4. Annual mink trappings (Canada)

A

B

C

D
1. Time series graphics
2. Seasonal or cyclic?
3. Autocorrelation
4. Forecast residuals
5. White noise
6. Evaluating forecast accuracy
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.
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 Dow-Jones index
Naïve forecast:

\[ \hat{y}_{t|t-1} = y_{t-1} \]

\[ e_t = y_t - y_{t-1} \]

Note: \( e_t \) are one-step-forecast residuals
Naïve forecast:

\[ \hat{y}_{t|t-1} = y_{t-1} \]

\[ e_t = y_t - y_{t-1} \]

Note: \( e_t \) are one-step-forecast residuals
Naïve forecast:

\[ \hat{y}_{t|t-1} = y_{t-1} \]

\[ e_t = y_t - y_{t-1} \]

Note: \( e_t \) are one-step-forecast residuals
Forecasting Dow-Jones index
Forecasting Dow-Jones index

Forecast residuals

ACF

Lag

0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15

1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22

Lag
fc <- rwf(dj)

res <- residuals(fc)

plot(res)

hist(res,breaks="FD")

Acf(res,main=""
Example: White noise

![Graph of white noise]

Time

-3 -2 -1 0 1 2

X

0 10 20 30 40 50

Forecasting: Principles and Practice

White noise
White noise data is uncorrelated across time with zero mean and constant variance. (Technically, we require independence as well.)
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.
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 \( N(0, 1/T) \).

- 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 $r_k$ 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 $r_k$ 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 $r_k$ 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.
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.)
Example: Pigs slaughtered

Number of pigs slaughtered in Victoria

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousands</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>80</td>
</tr>
<tr>
<td>1991</td>
<td>90</td>
</tr>
<tr>
<td>1992</td>
<td>100</td>
</tr>
<tr>
<td>1993</td>
<td>110</td>
</tr>
<tr>
<td>1994</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
</tr>
</tbody>
</table>
Example: Pigs slaughtered

ACF vs. Lag plot showing the autocorrelation function (ACF) for lag values ranging from 0 to 40. The plot displays a series of vertical bars, with the x-axis labeled as 'Lag' and the y-axis labeled as 'ACF'. The bars indicate the correlation of the time series with itself at different lags. The blue dashed lines represent the confidence intervals for the ACF.
Example: Pigs slaughtered

Monthly total number of pigs slaughtered in the state of Victoria, Australia, from January 1990 through August 1995. (Source: Australian Bureau of Statistics.)

- Difficult to detect pattern in time plot.
- ACF shows some significant autocorrelation at lags 1, 2, and 3.
- $r_{12}$ relatively large although not significant. This may indicate some slight seasonality.

These show the series is not a white noise series.
Example: Pigs slaughtered

Monthly total number of pigs slaughtered in the state of Victoria, Australia, from January 1990 through August 1995. (Source: Australian Bureau of Statistics.)

- Difficult to detect pattern in time plot.
- ACF shows some significant autocorrelation at lags 1, 2, and 3.
- \( r_{12} \) relatively large although not significant. This may indicate some slight seasonality.

These show the series is not a white noise series.
Example: Pigs slaughtered

Monthly total number of pigs slaughtered in the state of Victoria, Australia, from January 1990 through August 1995. (Source: Australian Bureau of Statistics.)

- Difficult to detect pattern in time plot.
- ACF shows some significant autocorrelation at lags 1, 2, and 3.
- $r_{12}$ relatively large although not significant. This may indicate some slight seasonality.

These show the series is not a white noise series.
Example: Pigs slaughtered

Monthly total number of pigs slaughtered in the state of Victoria, Australia, from January 1990 through August 1995. (Source: Australian Bureau of Statistics.)

- Difficult to detect pattern in time plot.
- ACF shows some significant autocorrelation at lags 1, 2, and 3.
- $r_{12}$ relatively large although not significant. This may indicate some slight seasonality.

These show the series is not a white noise series.
Example: Pigs slaughtered

Monthly total number of pigs slaughtered in the state of Victoria, Australia, from January 1990 through August 1995. (Source: Australian Bureau of Statistics.)

- Difficult to detect pattern in time plot.
- ACF shows some significant autocorrelation at lags 1, 2, and 3.
- $r_{12}$ relatively large although not significant. This may indicate some slight seasonality.

These show the series is not a white noise series.
Monthly total number of pigs slaughtered in the state of Victoria, Australia, from January 1990 through August 1995. (Source: Australian Bureau of Statistics.)

- Difficult to detect pattern in time plot.
- ACF shows some significant autocorrelation at lags 1, 2, and 3.
- \( r_{12} \) relatively large although not significant. This may indicate some slight seasonality.

These show the series is **not a white noise series**.
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.
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

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

\[ \hat{y}_{t|t-1} = y_{t-1} \]
\[ e_t = y_t - y_{t-1} \]
Forecasting Dow-Jones index

Change in Dow-Jones index

Day

Forecasting: Principles and Practice

White noise
These look like white noise.

But the ACF is a multiple testing problem.
These look like white noise.

But the ACF is a multiple testing problem.
Portmanteau tests

Consider a *whole set* of $r_k$ values, and develop a test to see whether the set is significantly different from a zero set.
Portmanteau tests

Consider a *whole set* of $r_k$ values, and develop a test to see whether the set is significantly different from a zero set.

**Box-Pierce test**

$$Q = T \sum_{k=1}^{h} r_k^2$$

where $h$ is max lag being considered and $T$ is number of observations.

- My preferences: $h = 10$ for non-seasonal data, $h = 2m$ for seasonal data.
- If each $r_k$ close to zero, $Q$ will be small.
- If some $r_k$ values large (positive or negative), $Q$ will be large.
Portmanteau tests

Consider a whole set of $r_k$ values, and develop a test to see whether the set is significantly different from a zero set.

**Box-Pierce test**

$$Q = T \sum_{k=1}^{h} r_k^2$$

where $h$ is max lag being considered and $T$ is number of observations.

- My preferences: $h = 10$ for non-seasonal data, $h = 2m$ for seasonal data.
- If each $r_k$ close to zero, $Q$ will be small.
- If some $r_k$ values large (positive or negative), $Q$ will be large.
Portmanteau tests

Consider a whole set of \( r_k \) values, and develop a test to see whether the set is significantly different from a zero set.

**Box-Pierce test**

\[
Q = T \sum_{k=1}^{h} r_k^2
\]

where \( h \) is max lag being considered and \( T \) is number of observations.

- My preferences: \( h = 10 \) for non-seasonal data, \( h = 2m \) for seasonal data.
- If each \( r_k \) close to zero, \( Q \) will be *small*.
- If some \( r_k \) values large (positive or negative), \( Q \) will be large.
Portmanteau tests

Consider a whole set of $r_k$ values, and develop a test to see whether the set is significantly different from a zero set.

**Box-Pierce test**

$$Q = T \sum_{k=1}^{h} r_k^2$$

where $h$ is max lag being considered and $T$ is number of observations.

- My preferences: $h = 10$ for non-seasonal data, $h = 2m$ for seasonal data.
- If each $r_k$ close to zero, $Q$ will be small.
- If some $r_k$ values large (positive or negative), $Q$ will be large.
Portmanteau tests

Consider a whole set of $r_k$ values, and develop a test to see whether the set is significantly different from a zero set.

**Ljung-Box test**

\[ Q^* = T(T + 2) \sum_{k=1}^{h} (T - k)^{-1} r_k^2 \]

where $h$ is max lag being considered and $T$ is number of observations.

- My preferences: $h = 10$ for non-seasonal data, $h = 2m$ for seasonal data.
- Better performance, especially in small samples.
Portmanteau tests

- If data are WN, $Q^*$ has $\chi^2$ distribution with $(h - K)$ degrees of freedom where $K = \text{no. parameters in model}$.

- When applied to raw data, set $K = 0$.

- For the Dow-Jones example,

  ```r
  res <- residuals(naive(dj))
  # lag=h and fitdf=K
  > Box.test(res, lag=10, fitdf=0)
  Box-Pierce test
  X-squared = 14.0451, df = 10, p-value = 0.1709
  > Box.test(res, lag=10, fitdf=0, type="Lj")
  Box-Ljung test
  X-squared = 14.4615, df = 10, p-value = 0.153
  ```
Portmanteau tests

- If data are WN, $Q^*$ has $\chi^2$ distribution with $(h - K)$ degrees of freedom where $K =$ no. parameters in model.

- When applied to raw data, set $K = 0$.

- For the Dow-Jones example,

```r
res <- residuals(naive(dj))

# lag=h and fitdf=K
> Box.test(res, lag=10, fitdf=0)
  Box-Pierce test
X-squared = 14.0451, df = 10, p-value = 0.1709

> Box.test(res, lag=10, fitdf=0, type="Lj")
  Box-Ljung test
X-squared = 14.4615, df = 10, p-value = 0.153
```
Portmanteau tests

- If data are WN, $Q^*$ has $\chi^2$ distribution with $(h - K)$ degrees of freedom where $K =$ no. parameters in model.

- When applied to raw data, set $K = 0$.

- For the Dow-Jones example,

```
res <- residuals(naive(dj))

# lag=h and fitdf=K
> Box.test(res, lag=10, fitdf=0)
  Box-Pierce test
  X-squared = 14.0451, df = 10, p-value = 0.1709

> Box.test(res, lag=10, fitdf=0, type="Lj")
  Box-Ljung test
  X-squared = 14.4615, df = 10, p-value = 0.153
```
Exercise

1. Calculate the residuals from a seasonal naive forecast applied to the quarterly Australian beer production data from 1992.

2. Test if the residuals are white noise.
Exercise

1 Calculate the residuals from a seasonal naive forecast applied to the quarterly Australian beer production data from 1992.

2 Test if the residuals are white noise.

```r
beer <- window(ausbeer,start=1992)
fc <- snaive(beer)
res <- residuals(fc)
Acf(res)
Box.test(res, lag=8, fitdf=0, type="Lj")
```
Outline

1 Time series graphics
2 Seasonal or cyclic?
3 Autocorrelation
4 Forecast residuals
5 White noise
6 Evaluating forecast accuracy
Measures of forecast accuracy

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

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

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

\[
\text{MAPE} = 100T^{-1} \sum_{t=1}^{T} \frac{|y_t - \hat{y}_{t|t-1}|}{|y_t|}
\]

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

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

\[
\text{MSE} = T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2
\]

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

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

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

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

\[
\text{MSE} = T^{-1} \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2
\]

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

\[
\text{MAPE} = 100T^{-1} \sum_{t=1}^{T} \frac{|y_t - \hat{y}_{t|t-1}|}{|y_t|}
\]

- 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**

\[
\text{MASE} = T^{-1} \sum_{t=1}^{T} \frac{|y_t - \hat{y}_{t|t-1}|}{Q}
\]

where \( Q \) is a stable measure of the scale of the time series \( \{y_t\} \).
Measures of forecast accuracy

Mean Absolute Scaled Error

\[
\text{MASE} = T^{-1} \sum_{t=1}^{T} \frac{|y_t - \hat{y}_{t|t-1}|}{Q}
\]

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

Proposed by Hyndman and Koehler (IJF, 2006)
Measures of forecast accuracy

**Mean Absolute Scaled Error**

\[
\text{MASE} = T^{-1} \sum_{t=1}^{T} \frac{|y_t - \hat{y}_{t|t-1}|}{Q}
\]

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

\[
\text{MASE} = T^{-1} \sum_{t=1}^{T} \frac{|y_t - \hat{y}_{t|t-1}|}{Q}
\]

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.
Measures of forecast accuracy

Forecasts for quarterly beer production

- Mean method
- Naive method
- Seasonal naive method
Measures of forecast accuracy

Forecasts for quarterly beer production

- Mean method
- Naive method
- Seasonal naive method
## Measures of forecast accuracy

### Mean method

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>38.0145</td>
<td>33.7776</td>
<td>8.1700</td>
<td>2.2990</td>
</tr>
</tbody>
</table>

### Naïve method

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70.9065</td>
<td>63.9091</td>
<td>15.8765</td>
<td>4.3498</td>
</tr>
</tbody>
</table>

### Seasonal naïve method

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12.9685</td>
<td>11.2727</td>
<td>2.7298</td>
<td>0.7673</td>
</tr>
</tbody>
</table>
Measures of forecast accuracy

Dow Jones Index (daily ending 15 Jul 94)

- Mean method
- Naive method
- Drift model

Days: 0, 50, 100, 150, 200, 250, 300
Values: 3600, 3700, 3800, 3900

Day
0 50 100 150 200 250 300
Measures of forecast accuracy

Dow Jones Index (daily ending 15 Jul 94)

- Mean method
- Naive method
- Drift model

Forecasting: Principles and Practice
### Measures of forecast accuracy

#### Mean method

<table>
<thead>
<tr>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>148.2357</td>
<td>142.4185</td>
<td>3.6630</td>
<td>8.6981</td>
</tr>
</tbody>
</table>

#### Naïve method

<table>
<thead>
<tr>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.0285</td>
<td>54.4405</td>
<td>1.3979</td>
<td>3.3249</td>
</tr>
</tbody>
</table>

#### Drift model

<table>
<thead>
<tr>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>53.6977</td>
<td>45.7274</td>
<td>1.1758</td>
<td>2.7928</td>
</tr>
</tbody>
</table>
### Training and test sets

#### Available data

<table>
<thead>
<tr>
<th>Training set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e.g., 80%)</td>
<td>(e.g., 20%)</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Training set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e.g., 80%)</td>
<td>(e.g., 20%)</td>
</tr>
</tbody>
</table>

- 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

beer3 <- window(ausbeer,start=1992,end=2005.99)
beer4 <- window(ausbeer,start=2006)

fit1 <- meanf(beer3,h=20)
fit2 <- rwf(beer3,h=20)

accuracy(fit1,beer4)
accuracy(fit2,beer4)

In-sample accuracy (one-step forecasts)
accuracy(fit1)
accuracy(fit2)
Training and test sets

beer3 <- window(ausbeer,start=1992,end=2005.99)
beer4 <- window(ausbeer,start=2006)

fit1 <- meanf(beer3,h=20)
fit2 <- rwf(beer3,h=20)

accuracy(fit1,beer4)
accuracy(fit2,beer4)

**In-sample accuracy** (one-step forecasts)
accuracy(fit1)
accuracy(fit2)
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.
1. Good forecast methods should have normally distributed residuals.

2. A model with small residuals will give good forecasts.

3. The best measure of forecast accuracy is MAPE.

4. If your model doesn’t forecast well, you should make it more complicated.

5. Always choose the model with the best forecast accuracy as measured on the test set.