

# Forecasting: principles and practice

Rob J Hyndman

1.2 The forecaster's toolbox

# Outline

**1** The statistical forecasting perspective

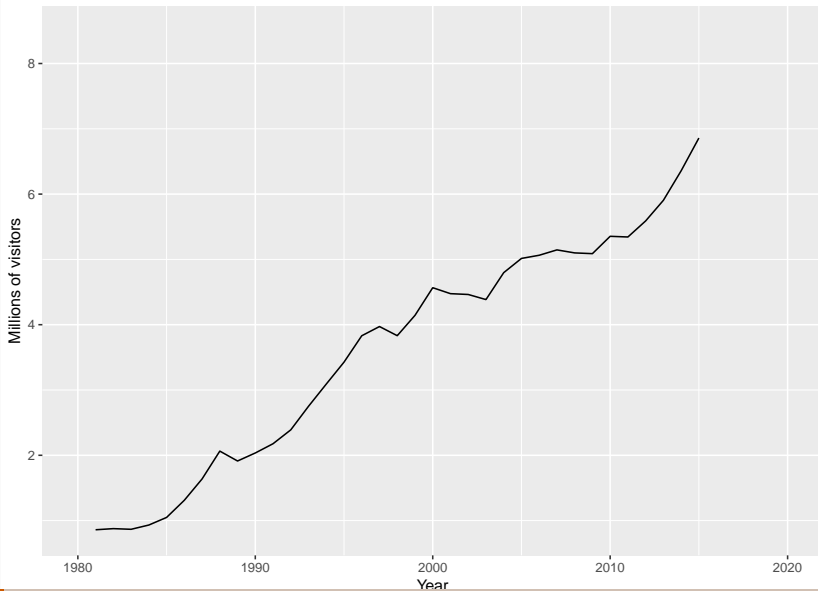
**2** Some simple forecasting methods

**3** Forecasting residuals

**4** Evaluating forecast accuracy

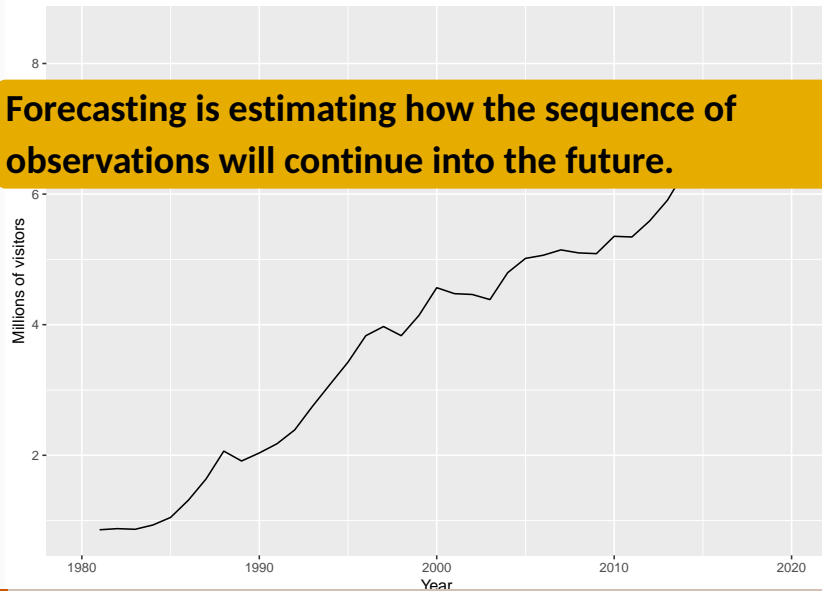
# The statistical forecasting perspective

Total international visitors to Australia



# The statistical forecasting perspective

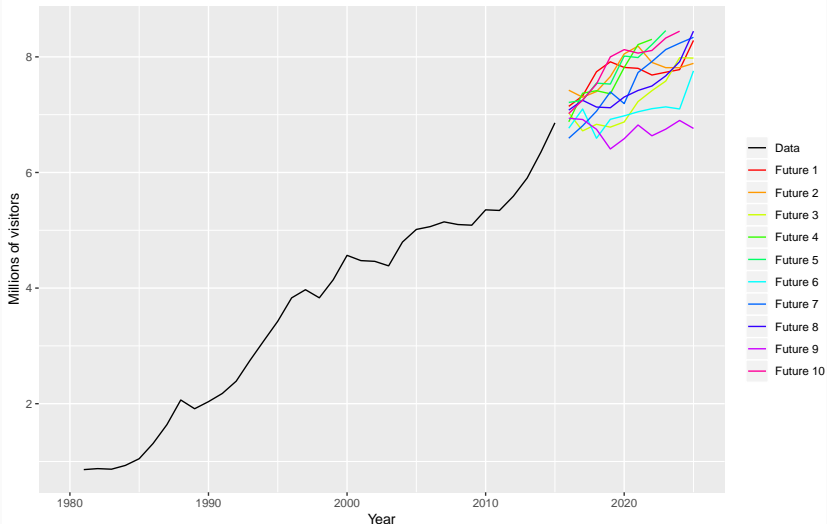
Total international visitors to Australia



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

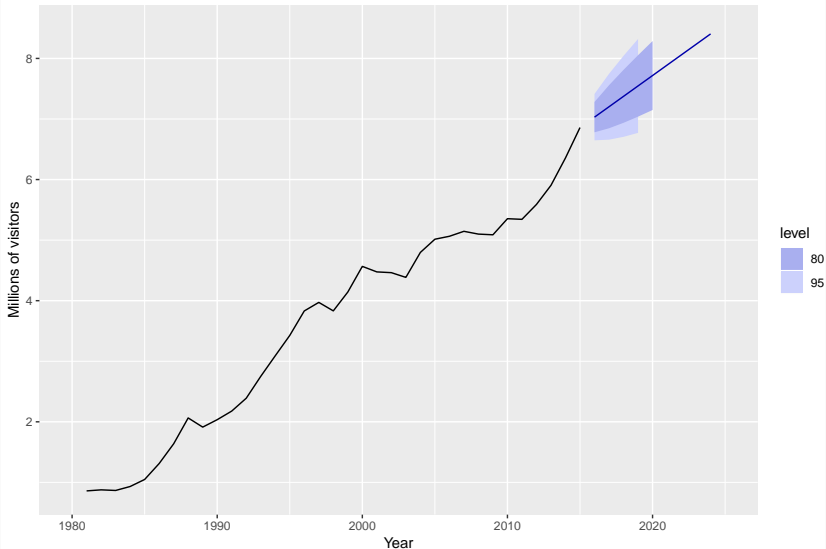
# Sample futures

Total international visitors to Australia



# Forecast intervals

Total international visitors to Australia



# Outline

**1** The statistical forecasting perspective

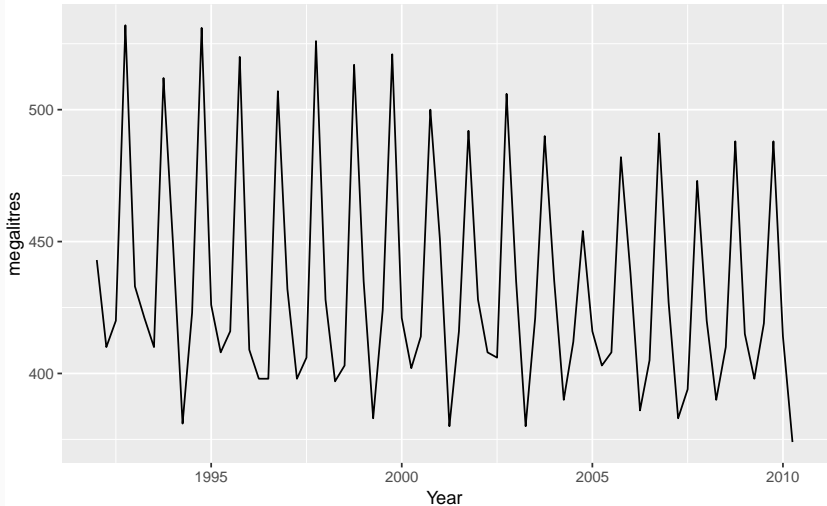
**2** Some simple forecasting methods

**3** Forecasting residuals

**4** Evaluating forecast accuracy

# Some simple forecasting methods

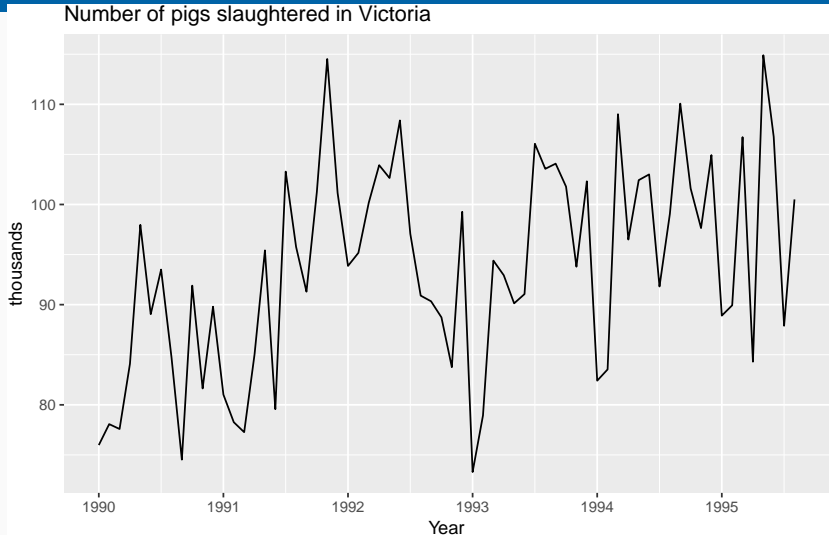
Australian quarterly beer production



How would you forecast these data?

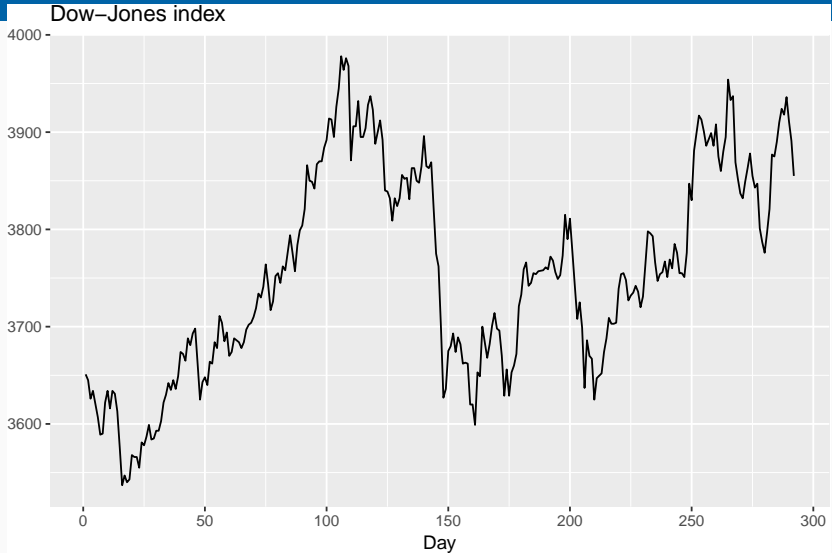


# Some simple forecasting methods



How would you forecast these data?

# Some simple forecasting methods



How would you forecast these data?

# Some simple forecasting methods

## Average method

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

# Some simple forecasting methods

## Average method

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

## Naïve method

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

# Some simple forecasting methods

## Average method

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

## Naïve method

- 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

## Drift method

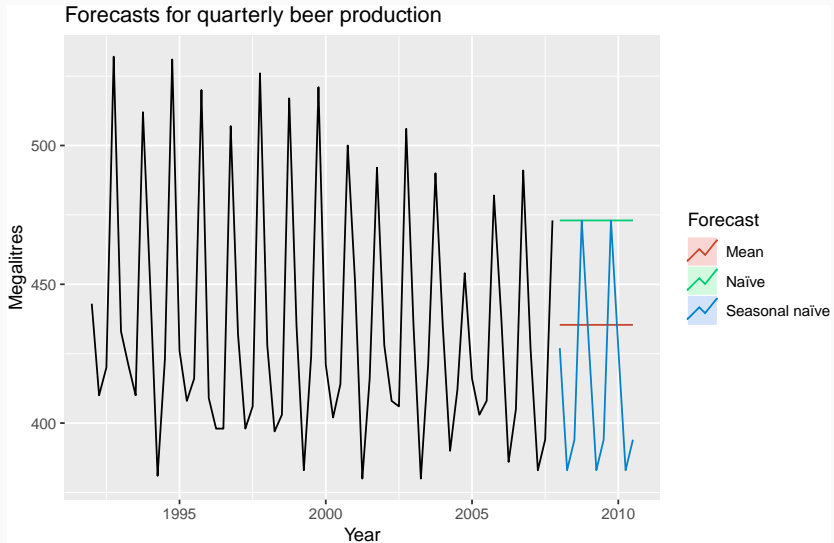
- Forecasts equal to last value plus average change.

- Forecasts:

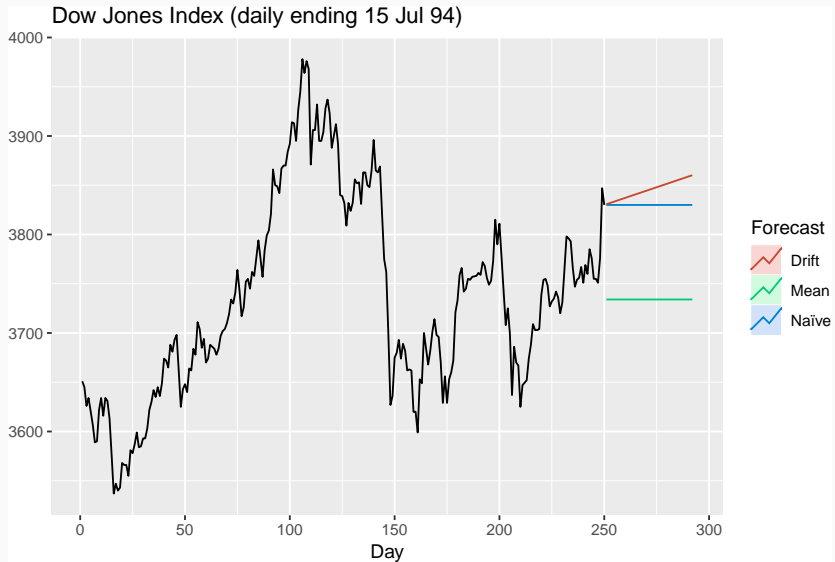
$$\begin{aligned}\hat{y}_{T+h|T} &= y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) \\ &= y_T + \frac{h}{T-1} (y_T - y_1).\end{aligned}$$

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

# Some simple forecasting methods



# Some simple forecasting methods





# Some simple forecasting methods

- Mean: `meanf(y, h=20)`
- Naïve: `naive(y, h=20)`
- Seasonal naïve: `snaive(y, h=20)`
- Drift: `rwf(y, drift=TRUE, h=20)`

# Outline

**1** The statistical forecasting perspective

**2** Some simple forecasting methods

**3** Forecasting residuals

**4** Evaluating forecast accuracy

# Fitted values

- $\hat{y}_{t|t-1}$  is the forecast of  $y_t$  based on observations  $y_1, \dots, y_t$ .
- We call these “fitted values”.
- Sometimes drop the subscript:  $\hat{y}_t \equiv \hat{y}_{t|t-1}$ .
- Often not true forecasts since parameters are estimated on all data.

###For example:

- $\hat{y}_t = \bar{y}$  for average method.
- $\hat{y}_t = y_{t-1} + (y_T - y_1)/(T - 1)$  for drift method.

# Forecasting residuals

**Residuals in forecasting:** difference between observed value and its fitted value:  $e_t = y_t - \hat{y}_{t|t-1}$ .

# Forecasting residuals

**Residuals in forecasting:** difference between observed value and its fitted value:  $e_t = y_t - \hat{y}_{t|t-1}$ .

## 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.

# Forecasting residuals

**Residuals in forecasting:** difference between observed value and its fitted value:  $e_t = y_t - \hat{y}_{t|t-1}$ .

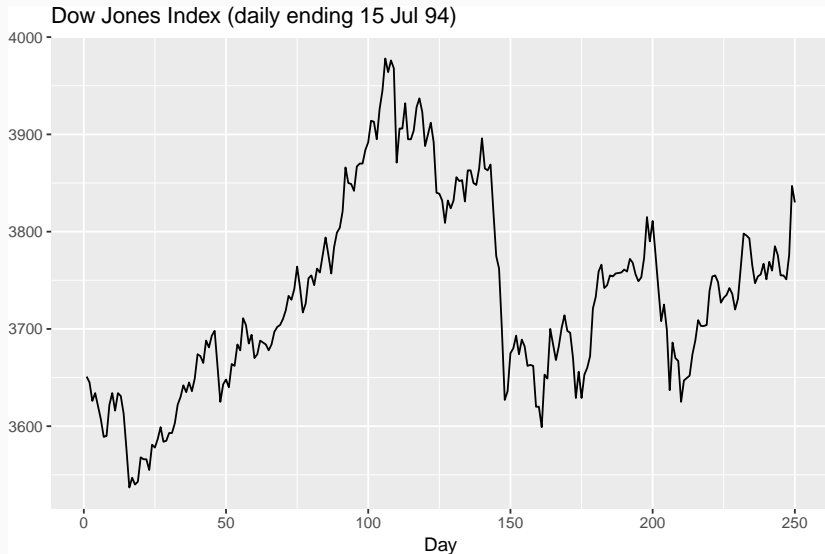
## Assumptions

- 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

# Example: Dow-Jones index



# Example: Dow-Jones index

Naïve forecast:

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



# Example: Dow-Jones index

Naïve forecast:

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

$$e_t = y_t - y_{t-1}$$

# Example: Dow-Jones index

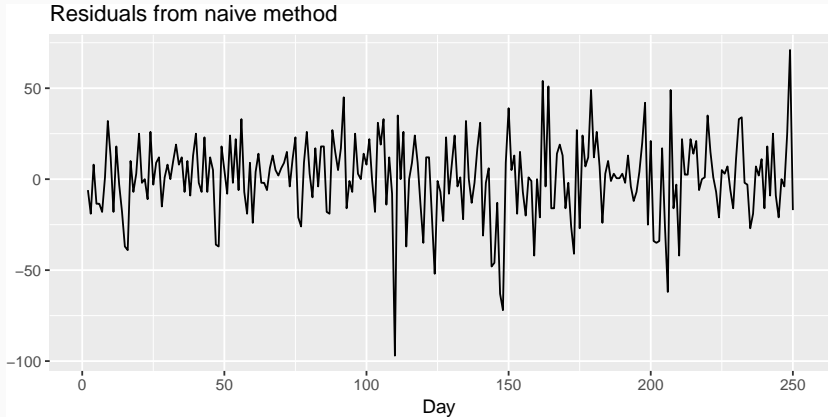
Naïve forecast:

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

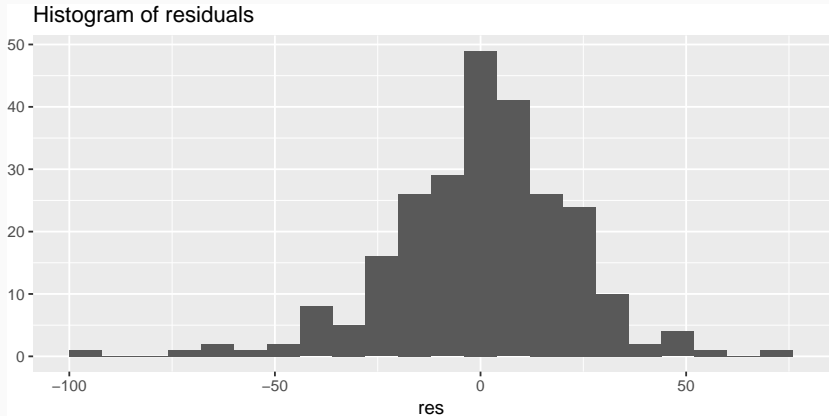
$$e_t = y_t - y_{t-1}$$

Note:  $e_t$  are one-step-forecast residuals

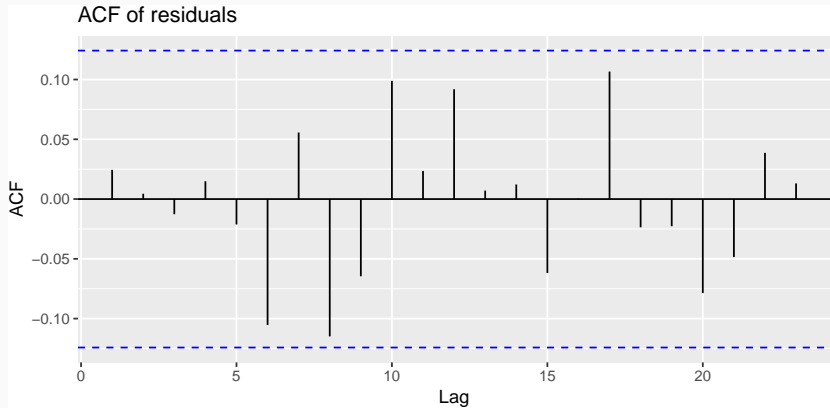
# Example: Dow-Jones index



# Example: Dow-Jones index



# Example: Dow-Jones index



# 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.

## 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**.



# 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  $K$  = no. parameters in model.
- When applied to raw data, set  $K = 0$ .
- For the Dow-Jones example,

```
# lag=h and fitdf=K  
Box.test(res, lag=10, fitdf=0)
```

```
##  
## Box-Pierce test  
##  
## data: res  
## X-squared = 10.655, df = 10, p-value = 0.385
```

# 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,

```
# lag=h and fitdf=K
```

```
Box.test(res, lag=10, fitdf=0, type="Lj")
```

```
##
```

```
## Box-Ljung test
```

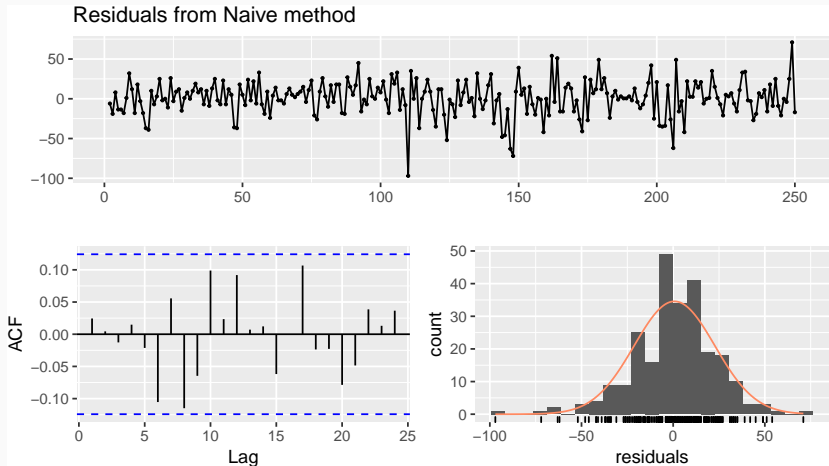
```
##
```

```
## data: res
```

```
## X-squared = 11.088, df = 10, p-value = 0.3507
```

# checkresiduals function

```
checkresiduals(naive(dj2))
```



## checkresiduals function

```
##  
##  Ljung-Box test  
##  
## data:  Residuals from Naive method  
## Q* = 11.088, df = 10, p-value = 0.3507  
## Model df: 0.    Total lags used: 10  
#Lab session 3 ##
```

# Lab Session 3

# Outline

**1** The statistical forecasting perspective

**2** Some simple forecasting methods

**3** Forecasting residuals

**4** 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  $t = 1, \dots, T$ . Then the following measures are useful.

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

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

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

# 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  $t = 1, \dots, T$ . Then the following measures are useful.

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

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

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

- MAE, MSE, RMSE are all scale dependent.
- MAPE is scale independent but is only sensible if



# 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  $\{y_t\}$ .

Proposed by Hyndman and Koehler (IJF, 2006).

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 naïve 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  $\{y_t\}$ .

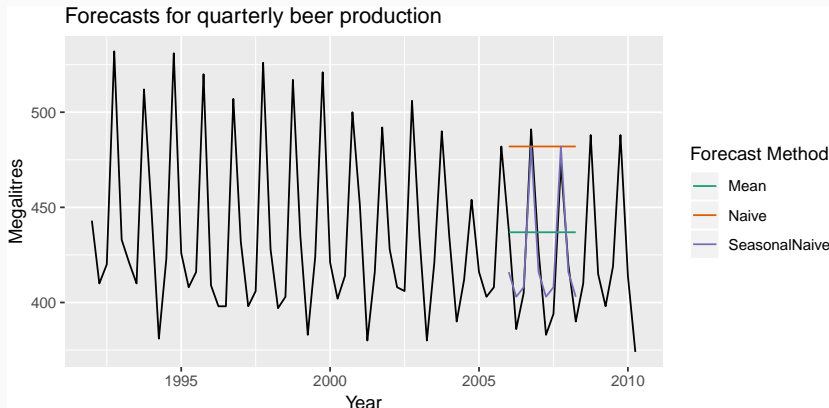
Proposed by Hyndman and Koehler (IJF, 2006).

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 naïve method.

# Measures of forecast accuracy



# Measures of forecast accuracy

	RMSE	MAE	MAPE	MASE
Mean method	38.95	34.46	8.33	2.35
Naïve method	70.80	63.10	15.71	4.29
Seasonal naïve method	13.59	12.20	2.95	0.83

# 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

```
beer2 <- window(ausbeer, start=1992, end=c(2005,4))  
fc <- snaive(beer2, h=10)  
accuracy(fc, ausbeer)
```

# Training and test sets

```
beer2 <- window(ausbeer, start=1992, end=c(2005,4))  
fc <- snaive(beer2, h=10)  
accuracy(fc, ausbeer)
```

In-sample accuracy (one-step forecasts)

```
accuracy(fc)
```

# 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.



## Poll: true or false?

- 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.

#Lab session 4 ##