Hochschule Esslingen
University of Applied Sciences

04_Forecasting and Prediction

Lecture: Intelligent Data Analytics

(these slides are based on the online book from Prof. Rob Hyndman)

Overview

1. Motivation
2. Fundamentals
3. Simple forecasting methods
4. Evaluating forecast accuracy

Hochschule Esslingen
University of Applied Sciences

Famous predictions:

"I think there is a world market for maybe five computers."

(Thomas J. Watson, Chairman of IBM, 1943)

"There is no reason anyone would want a computer in their home."

(Ken Olsen, founder of Digital Equipment Corporation, 1977)

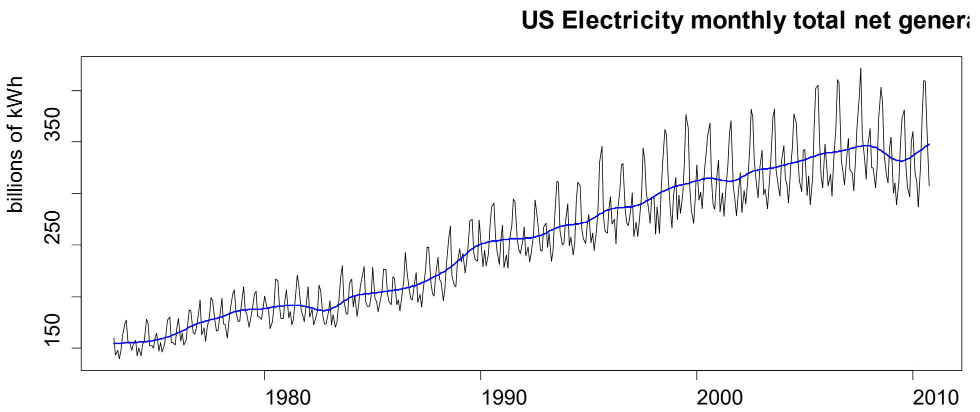
[http://www.techhive.com/article/155984/worst_tech_predictions.html?page=1, accessed on Feb 26nd]

The quality of forecasting depends on:

1. how well we understand the factors that contribute to it
2. availability of data
3. if the forecasts affect what were are trying to forecast
(e.g. a forecast of changes on the stock market will affect the behavior of investors at the stock market)

Example: Forecasting electricity demand

1. Main contributing factors are known:
   - Temperatures
   - Calendar variations, e.g. holidays
   - Economic conditions
2. data is available:
   - history of electricity demand
   - weather conditions



3. Forecasting the electricity demand does not affect the electricity demand.

Fundamentals

Hochschule Esslingen
University of Applied Sciences

this course will focus on quantitative methods

- **Forecasting Techniques**
  - Qualitative Methods
    - Delphi Method
    - Analogies
    - Sales Force Composite
    - Consumer Market Survey
    - …
  - Quantitative Methods
    - Time Series Methods
      - Naïve methods
      - Exponential smoothing
      - Moving Averages
      - Neural Networks — Nicht klausurrelevant
      - Autoregression ARIMA
      - …
    - Causal Methods
      - Linear Regression
      - Multiple Regression
      - …

[cf. Tutorial Slides of S. F. Crone: Slide 12 of "Forecasting with Neural Networks"]

Process of forecasting

1. models are fitted using training data
2. forecast accuracy is determined using test/data

validation

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

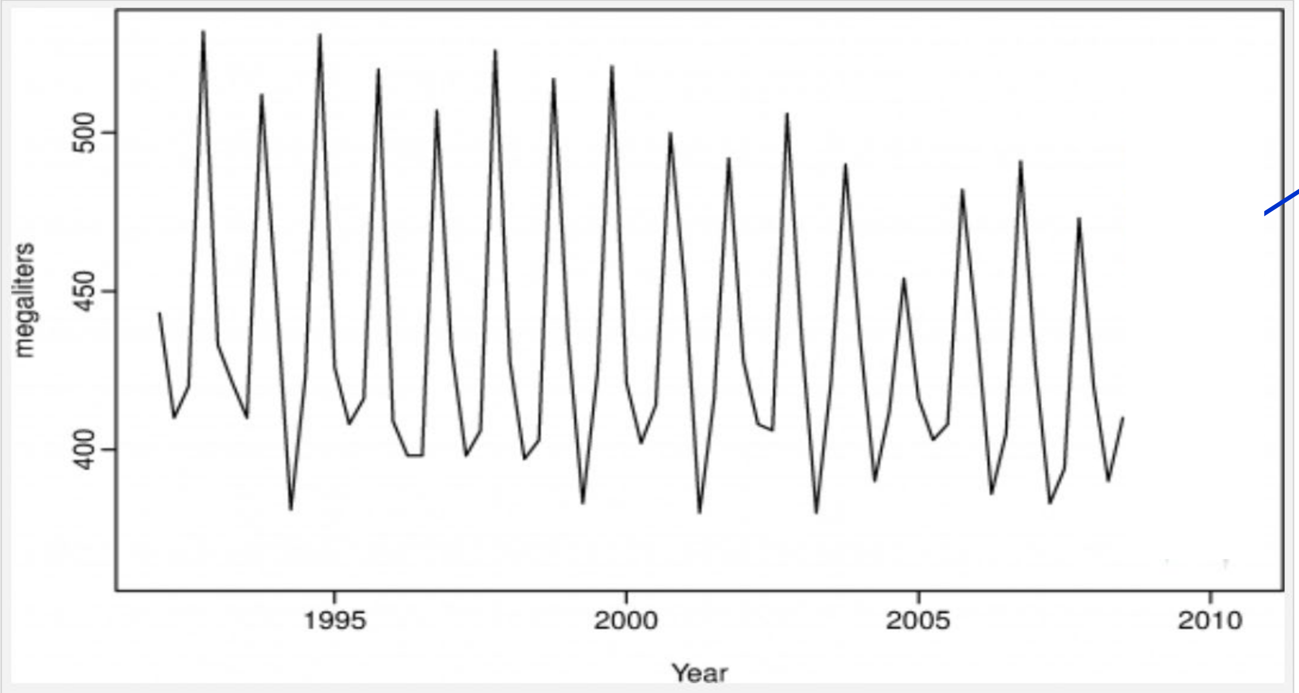
Note:

- do not use any data from the test set to fit your model
- forecast accuracy is computed *only* on the test set.
- a model that fits the data well does not necessarily forecast well

Time series forecasting

The term "forecasting" usually refers to time series forecasting. The goal is to estimate how the sequence of observations will continue in the future by using the information of their past and present.

Example: Forecasting the Australian beer production



Forecast with „prediction intervals"

Predictor variables and forecasting

As an alternative or in addition to historical values of the time series, we can use “predictor variables”.

With “predictor variables” you work as in a classical regression model.

Example: Forecasting the hourly electricity demand (=ED)

- **Time series model:**

$$ED_{t+1} = f(\overbrace{ED_t, ED_{t-1}, ED_{t-2}, \dots}^{\text{vergangene Daten \&}}, \overbrace{error}^{\text{Fehler}})$$

- **Model with predictor variables:**

$$ED = f\begin{pmatrix} current\ temperature, strength\ of\ economy, \\ population, time\ of\ day, day\ of\ week, error \end{pmatrix}$$

- **Mixed model:**

$$ED_{t+1} = f(ED_t, current\ temperature, time\ of\ day, day\ of\ week, error)$$

Notation for forecasting:

Note, that the notation is not consistent throughout literature, this is the notation used in the free online book from Prof. Hyndman:

- Historical data, i.e. observed values, until time T:
  $y_1, y_2, \dots, y_T$.
- forecast values are denoted with a "^":
  $\hat{y}_t$: forecast of the value of y at time t.
- It is useful to specify on what information the forecasts are based on:
  $\hat{y}_{t|T}$: forecast of the value of y at time t taking account of all observations up to time T, i.e. $y_1, y_2, \dots, y_T$.
- We can also specify the forecast horizon h (= time span of forecast)
  $\hat{y}_{T+h|T}$: an h-step forecast taking account of $y_1, y_2, \dots, y_T$.

Average method: *"mean value of historical data"*

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

Time series model (h=1): $y_t = \mu + z_t$, with z_t normal, iid with variance σ^2

```
meanf(y, h=10, level=c(80,95))
#y: the time series; h: forecasting horizon
#level: confidence levels for prediction intervals
```

Naïve method: *"last observed value"*

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

Time series model (h=1): $y_t = y_{t-h} + z_t$, with z_t normal, iid with variance σ^2

```
naive(y, h=10, level=c(80,95))
#y: the time series; h: forecasting horizon
#level: confidence levels for prediction intervals
```

Seasonal naïve method: *"last value from same season"*

$$\hat{y}_{T+h|T} = y_{T+h-km}$$

m = seasonal period, e.g. months, quarters,…

$$k = \lfloor (h-1)/m \rfloor + 1$$

Time series model (h=1): $y_t = y_{t-m} + z_t$, with z_t normal, iid with variance σ^2

```
snaive(y, h=2*frequency(y), level=c(80,95))
#y: the time series; h: forecasting horizon
#level: confidence levels for prediction intervals
```

vielfaches d. Frequenz

Drift method: *"last observed value plus average change"*

Durchschnittsbildung

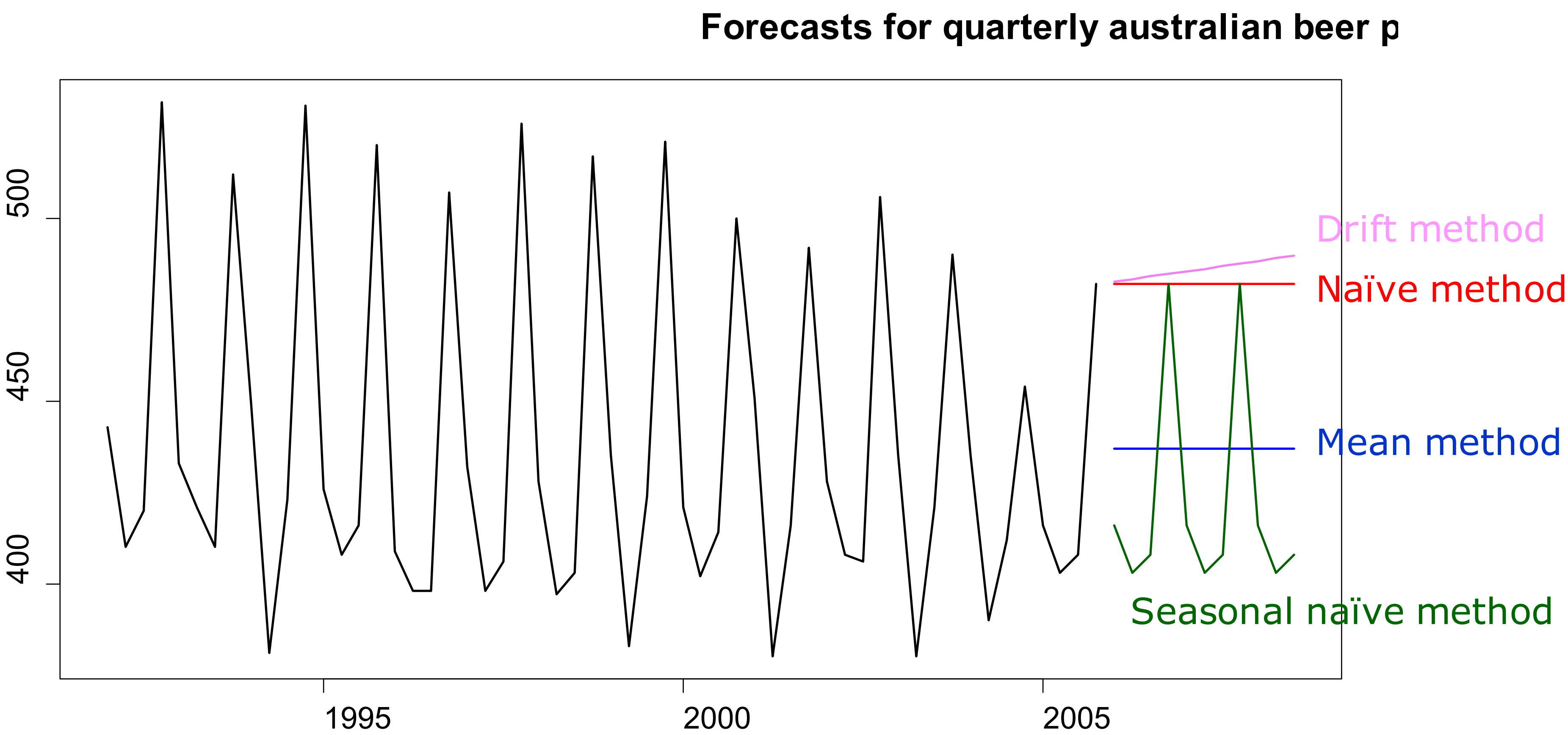
$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + \frac{h}{T-1}(y_T - y_1)$$

Time series model (h=1): $y_t = c + y_{t-h} + z_t$, with z_t normal, iid with variance σ^2

```
rwf(y, h=10, drift=TRUE level=c(80,95))
#y: the time series; h: forecasting horizon
#level: confidence levels for prediction intervals
```

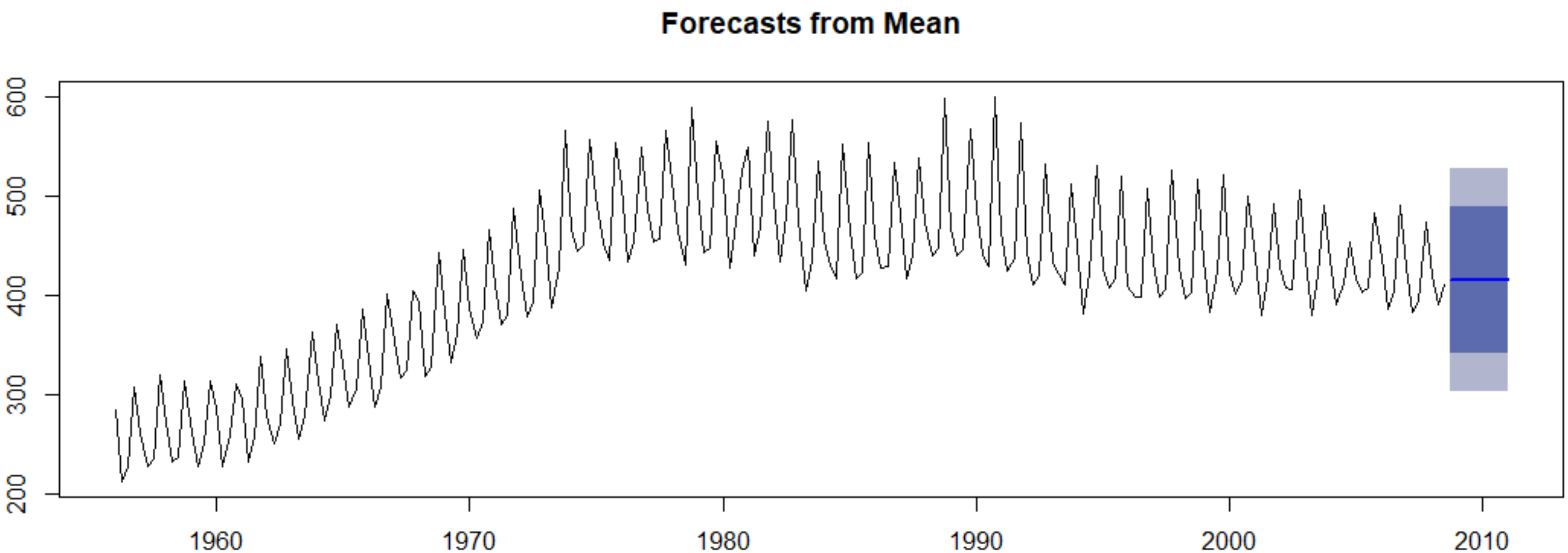
random walk forecast

Simple forecasting methods



Question: Which method generated which forecast?

```r
plot(meanf(ausbeer, h=10, level=c(60,80))
```

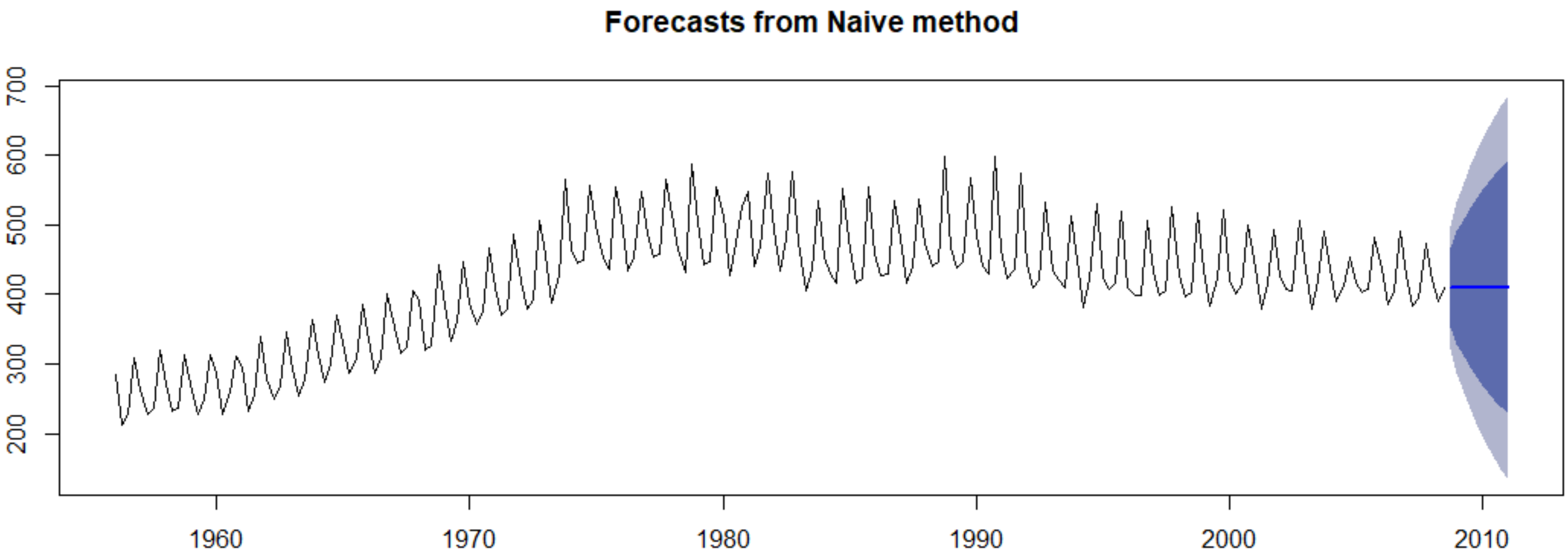



```r
meanf(ausbeer, h=1, level=c(60,80)
```

```
        Point Forecast    Lo 60    Hi 60    Lo 80    Hi 80
2008 Q4            410 352.8062 467.1938 322.9101 497.0899
```

Simple forecasting methods

```
plot(naive(ausbeer, h=10, level=c(60,80))
```




```
naive(ausbeer, h=1, level=c(60,80)
```

```
        Point Forecast    Lo 60     Hi 60    Lo 80      Hi 80
2008 Q4 414.9526          341.4463 488.4589 302.8983  527.0069
```

Evaluating forecast accuracy

To be able to evaluate a forecasting model or to compare different methods, we need some evaluation criteria: Test set !

Forecast error or residual:	$e_t = y_t - \hat{y}_{t\|t-1}$

A) Scale-dependent errors:

- The forecast error has the same scale as the data, i.e. it can <u>not</u> be used for comparisons between different scaled time series.
- Most popular scale-dependent accuracy measures based on e_t are:

Mean absolute error:	absolut	$MAE = \frac{1}{T}\sum_{t=1}^{T}\|e_t\|$
Mean squared error:	squared	$MSE = \frac{1}{T}\sum_{t=1}^{T} e_t^2$
Root mean squared error:	root mean squared	$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_t^2}$

Evaluating forecast accuracy

B) scale-independent errors
can be used for comparisons between different scaled time series, e.g.:

Percentage error: $p_t = 100 \cdot {e_t}/{y_t}$

- Disadvantages of percentage errors:
  - Infinite or undefined if $y_t = 0$;
  - Extreme values for y_t values being close to zero.

- Most common accuracy measure based on p_t is:

Mean absolute percentage error: $MAPE = \frac{1}{T}\sum_{t=1}^{T}|p_t|$

- MAPE is only sensible for $y_t \gg 0$ for all t.

Alternative scaled error measure:

Mean absolute scaled error: $MASE = \frac{1}{T}\sum_{t=1}^{T}\frac{|e_t|}{Q}$

- Q is a stable measure of the scale of the time series $\{y_t\}$.
- For *non-seasonal* time series: $Q = \frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|$
- For *seasonal* time series: $Q = \frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|$
- You only need a single R function to evaluate your models:

```
accuracy(f, x)
#f: object of class "forecast", i.e. the forecast model
#x: An optional numerical vector containing actual
    values of the same length as object, or a time
    series overlapping with the times of f.
```

Evaluating forecast accuracy (in-sample)

Process of forecasting

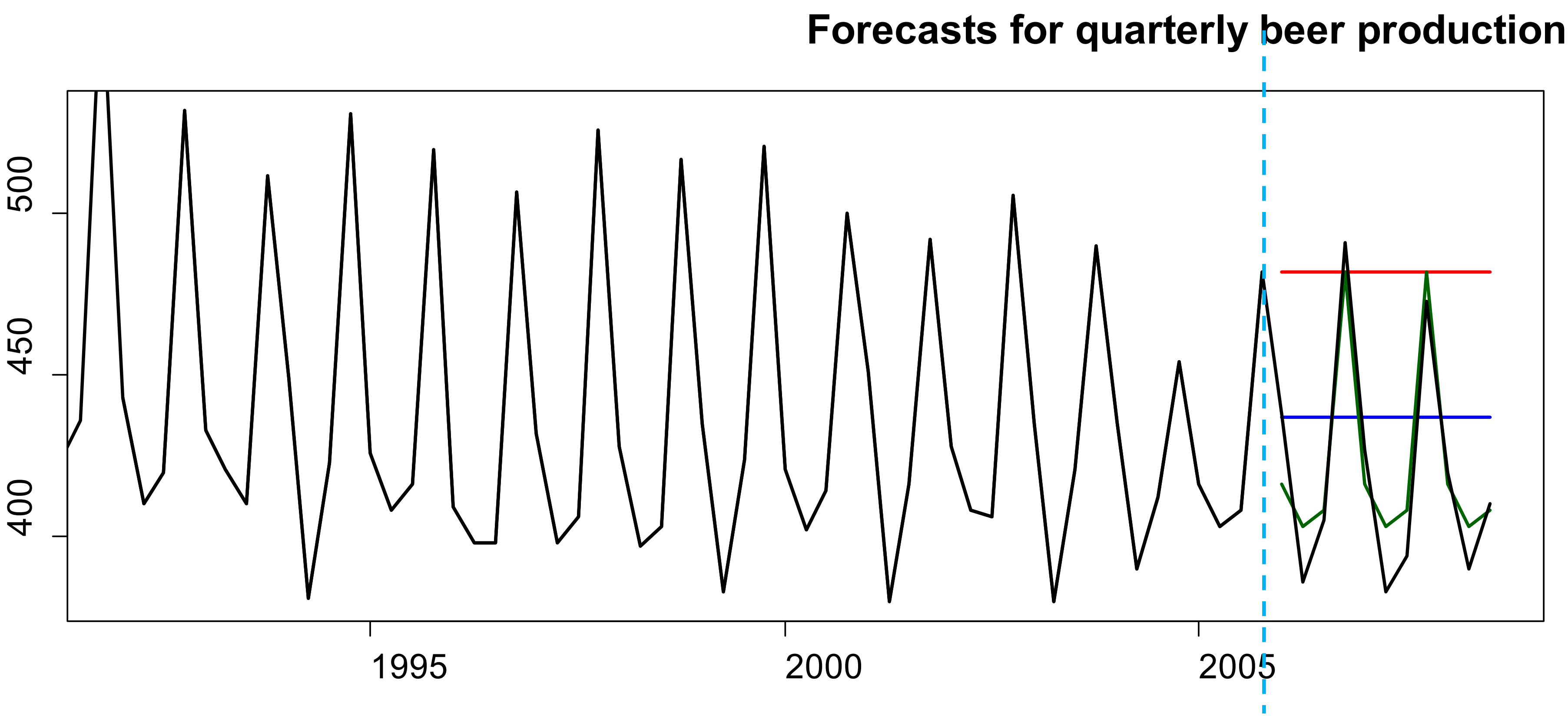
1. models are fitted using training data (in-sample accuracy)
2. forecast accuracy is determined using test data (out-of-sample accuracy)

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

```
#Train data: ts from 1992-2005
beer2 <- window(ausbeer,start=1992,end=2005.99)
beerfit1 <- meanf(beer2,h=11)
beerfit2 <- rwf(beer2,h=11)
beerfit3 <- snaive(beer2,h=11)

#Caluclate accuracies on training set
accuracy(beerfit1)
accuracy(beerfit2)
accuracy(beerfit3)
```

Evaluating forecast accuracy



Mean method

Naïve method

Seasonal naïve method

Evaluating forecast accuracy (out-of-sample)


```r
#Train data: ts from 1992-2005
beer2 <- window(ausbeer,start=1992,end=2006-.1)
beerfit1 <- meanf(beer2,h=11)
beerfit2 <- rwf(beer2,h=11)
beerfit3 <- snaive(beer2,h=11)

#Plot the time series and the forecasts
plot(beerfit1, plot.conf=FALSE, lwd=2,
   main="Forecasts for quarterly beer production")
lines(beerfit2$mean, col=2, lwd=2)
lines(beerfit3$mean, col="darkgreen", lwd=2)
lines(ausbeer, lwd=2)

#Caluclate accuracies of forecast model on test set
beer3 <- window(ausbeer, start=2006)
accuracy(beerfit1, beer3)
accuracy(beerfit2, beer3)
accuracy(beerfit3, beer3)
```

Literature

- R. J. Hyndman, G. Athanasopoulos: *Forecasting: principles and practice.* Available online at https://www.otexts.org/fpp, 2014