Time Series (TS) Forecast Models
Validation & Evaluation

Nagiza F. Samatova, samatova@csc.ncsu.edu
Professor, Department of Computer Science
North Carolina State University

Senior Scientist, Computer Science & Mathematics Division
Oak Ridge National Laboratory

Basic Notation

Symbol	Definition
$\boldsymbol{t = 1, 2, 3, \ldots,}$	An index for the time period of interest; e.g., for a *daily* time period, $t = 1$ means day 1, $t = 2$ means day 2, etc.
$\boldsymbol{y_1, y_2, \ldots, y_T}$	A series of T values measure over T time periods; e.g., for the annual average stock price, y_1 denotes the price for year 1, y_2 denotes the price for year 2, etc.
$\boldsymbol{F_t}$ or $\boldsymbol{\hat{y}_t}$	The forecast value for time period t
$\boldsymbol{F_{t+k}}$ or $\boldsymbol{\widehat{y_{t+k}}}$	The k-step-ahead forecast when forecasting time is t; e.g., F_{t+1} is the forecast for time period $(t + 1)$ made during the time period t
$\boldsymbol{e_t = y_t - F_t}$ or $\boldsymbol{e_t = y_t - \hat{y}_t}$	The forecast error for time period t

Scale-dependent Metrics

The forecast error ($e_t = y_t - \hat{y}_t$) is on the same scale as the data. Accuracy measures that are based on e_t are scale-dependent and can NOT be used to compare series that are on different scales.

Error Measures

- **Error**: $\boldsymbol{e_t = y_t - \hat{y}_t}$
  - Cons: Dependent of the scale of data

- **Percentage Error**: $\boldsymbol{p_t = 100 e_t / y_t}$
  - Cons of Percentage Errors:
    - Infinite or undefined if y_t = 0
    - Have extreme values when any y_t is close to zero
    - Put a heavier penalty on negative errors than on positive errors
  - Pros of Percentage Errors:
    - Independent of the scale of data

- **Scaled Error**: $\boldsymbol{q_t = \frac{e_t}{\frac{1}{T-1}\sum_{j=2}^{T} |y_j - y_{j-1}|}}$
  - Scaling the errors based on the training MAE from a naive forecast
    - $q_t < 1$: a better forecast than the average naive forecast computed on the training data.
    - $q_t > 1$: the forecast is worse than the average naive forecast computed on the training data.
  - Pros of Scaled Errors:
    - Independent of the scale of data

Scaled Error for Seasonal & Cross-Sectional Data

- **Scaled Error for Seasonal Naive Forecast:**

$$q_t = \frac{e_t}{\frac{1}{T-m}\sum_{j=m+1}^{T} |y_j - y_{j-m}|}$$

- **Scaled Error for Cross-Sectional Data:**

$$q_j = \frac{e_j}{\frac{1}{N}\sum_{i=1}^{N} |y_i - \bar{y}|}$$

Forecast Accuracy Measures

Measure	Formula	Scale-dependent?
Mean Absolute Error	$MAE = mean(\|e_t\|)$	YES
Root Mean Squared Error	$RMSE = sqrt(mean(e_t^2))$	YES
Mean Absolute Percentage Error	$MAPE = mean(\|p_t\|)$	NO
Mean Absolute Scaled Error	$MASE = mean(\|q_t\|)$	NO
Mean Squared Scaled Error	$MSSE = sqrt(mean(q_t^2))$	NO

Baseline: Simple Forecasting Methods

- **Average**: meanf (ts.data, h=20)
  - Forecast of all future values is the mean of historical data $\{y_1, \ldots, y_T\}$
  - $F_{T+h} = \hat{y}_{T+h} = \bar{y} = (y_1 + \cdots + y_T)/T$

- **Naive:** naive (ts.data, h=20) or rwf (ts.data, h=20)
  - Forecast is equal to the last observed value
  - $F_{T+h|T} = \hat{y}_{T+h|T} = y_T$

- **Seasonal naive:** snaive (ts.data, h=20)
  - Forecast is equal to the last value from the same season
  - $\hat{y}_{T+h|T} = y_{T+h-km}$, where m is the seasonal period and $k = round\left(\frac{h-1}{m}\right) + 1$

- **Drift:** rwf (ts.data, drift=TRUE, h=20)
  - Forecast is equal to the last value plus the average change
  - Equivalent to extrapolating a line between the first and last observation
  - $F_{T+h|T} = \hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + \frac{h}{T-1}(y_T - y_1)$

Training and Test Data 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.

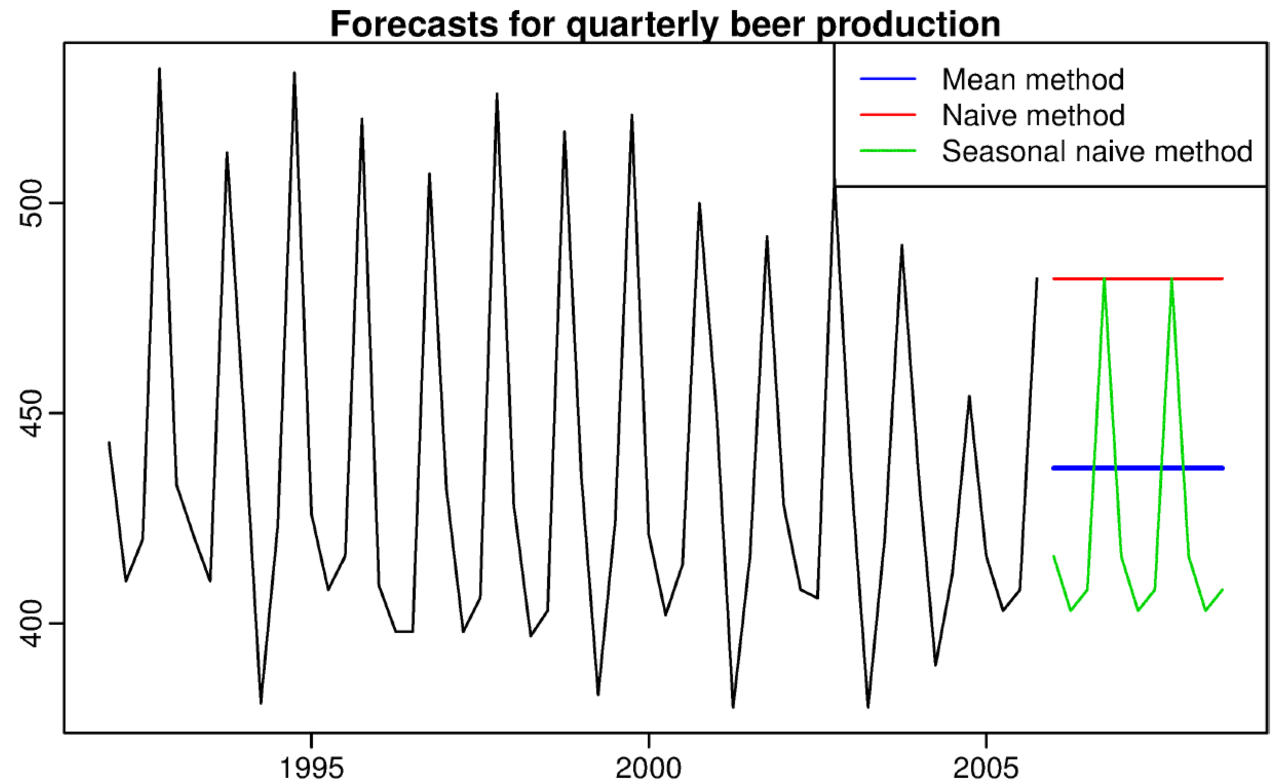
Example: Beer Production Forecast

```
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)

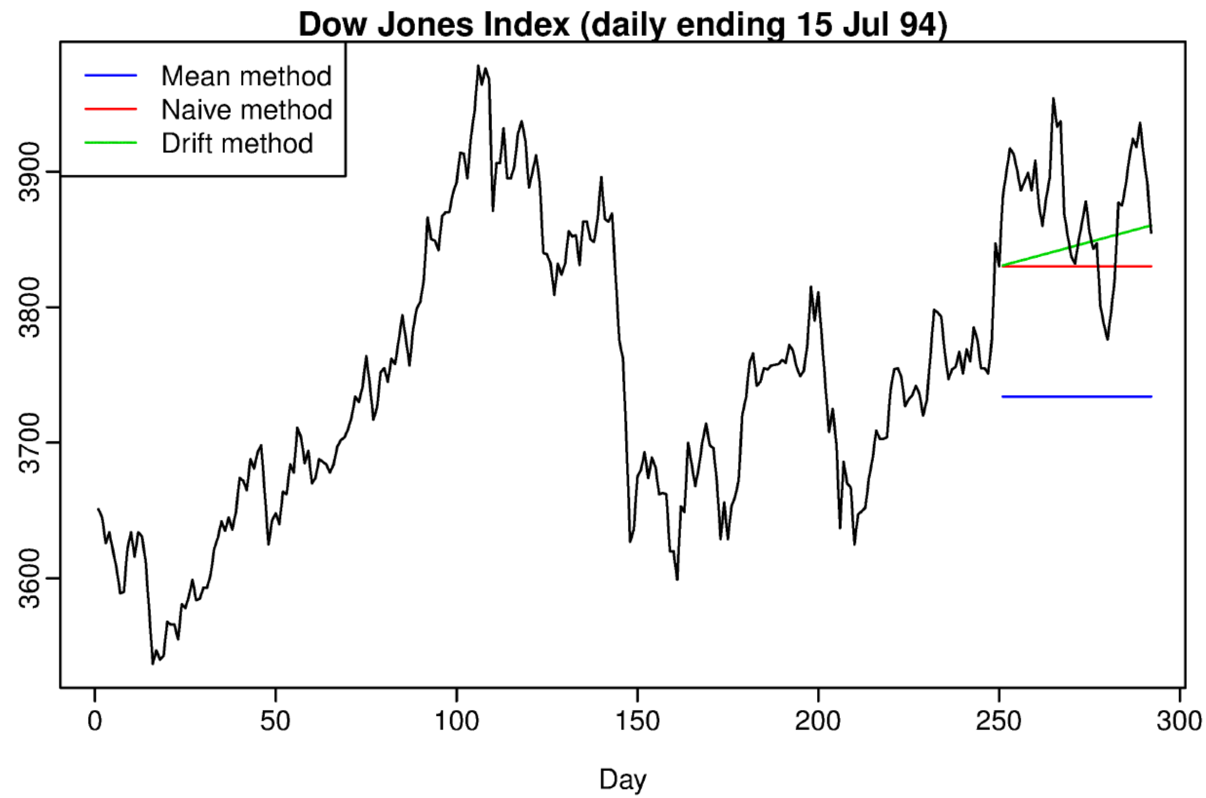
```
accuracy(fit1)
accuracy(fit2)
```

Example: Beer Production Forecast 2006-08



Method	RMSE	MAE	MAPE	MASE
Mean method	38.01	33.78	8.17	2.30
Naïve method	70.91	63.91	15.88	4.35
Seasonal naïve method	12.97	11.27	2.73	0.77

Example: Dow Jones Index Forecast



Method	RMSE	MAE	MAPE	MASE
Mean method	148.24	142.42	3.66	8.70
Naïve method	62.03	54.44	1.40	3.32
Drift method	53.70	45.73	1.18	2.79

Cross-Validation for Cross-Sectional Data

- **Leave-one-out Cross-Validation (LOOCV) Steps:**

1. Select observation i for the test set, and use the remaining observations in the training set. Compute the error on the test observation.
2. Repeat the above step for $i = 1, 2, \ldots, N$ where N is the total number of observations.
3. Compute the forecast accuracy measures based on the errors obtained.

Cross-Validation for Time-Series Data

- For time series data, the procedure is similar to LOOCV for cross-sectional data but the **training set consists only of observations that occurred prior to the observation that forms the test set**.
  - No future observations can be used in constructing the forecast.
  - It is not possible to get a reliable forecast based on a very small training set, so the earliest observations are not considered as test sets.
- Suppose k observations are required to produce a reliable forecast. Then the cross-validation works as follows:

  1. Select the observation at time $k + i$ for the test set, and use the observations at times $1, 2, \ldots, k + i - 1$ to estimate the forecasting model. Compute the error on the forecast for time $k + i$.
  2. Repeat the above step for $i = 1, 2, \ldots, T - k$ where T is the total number of observations.
  3. Compute the forecast accuracy measures based on the errors obtained.

- This procedure is sometimes known as a "rolling forecasting origin" because the "origin" (k + i -1) at which the forecast is based rolls forward in time.

Cross-Validation for Multi-Step-Ahead Forecast

- The cross-validation procedure based on a rolling forecasting origin can be modified to allow for multi-step errors to be used:

1. Select the observation at time $k+h+i-1$ for the test set, and use the observations at times $1, 2, \ldots, k+i-1$ to estimate the forecasting model. Compute the h-step error on the forecast for time $k+h+i-1$.
2. Repeat the above step for $i = 1, 2, \ldots, T-k-h+1$ where T is the total number of observations.
3. Compute the forecast accuracy measures based on the errors obtained.