Forecasting

Forecasting

three main components needed to produce a forecast

- information set $\mathcal{I}_h = \{y_0, y_1, \ldots, y_h\}$ at forecast origin h
- forecast horizon ℓ
- loss function $L(y_{h+\ell} - \hat{y}_h(\ell))$ or $L(e_h(\ell))$

where $\hat{y}_h(\ell)$ is the ℓ-step ahead forcast at forecast origin h given information set $\mathcal{I}_h$ and $e_h(\ell) = y_{h+\ell} - \hat{y}_h(\ell)$ is the forecast error

optimal forecast: forecaster wants to construct a forecast $\hat{y}_h^*(\ell)$ that minimizes the expected loss

$$E\left[L(y_{h+\ell} - \hat{y}_h(\ell))|\mathcal{I}_h\right] = \int L(y_{h+\ell} - \hat{y}_h(\ell))f(y_{h+\ell}|\mathcal{I}_h)dy_{h+\ell}$$

thus

$$\hat{y}_h^*(\ell) = \arg\min_{\hat{y}_h(\ell)} E\left[L(y_{h+\ell} - \hat{y}_h(\ell))|\mathcal{I}_h\right]$$

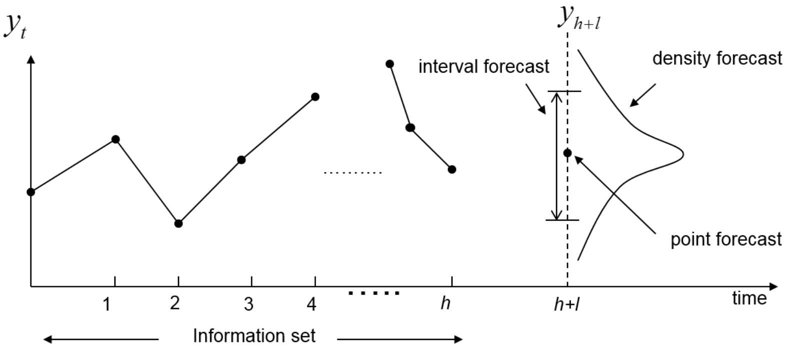
Point, Interval and Density Forecasts

first, we need to obtain conditional distribution for $y_{h+\ell}$ given information set $\mathcal{I}_h$

- conditional probability density function $f(y_{h+\ell}|\mathcal{I}_h)$
- conditional mean $\mu_{h+\ell|h} = E_h(y_{h+\ell}|\mathcal{I}_h)$
- conditional variance $\sigma^2_{h+\ell|h} = var_h(y_{h+\ell}|\mathcal{I}_h)$

these will be used to build the point, interval and density forecasts

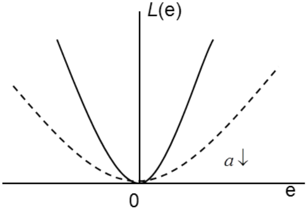
Point, Interval and Density Forecasts



Symmetric Loss Function

Quadratic loss function

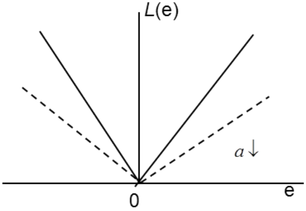
$$L(e) = ae^2, \quad a > 0$$



$$L(e) = L(-e)$$

Absolute value loss function

$$L(e) = a|e|, \quad a > 0$$



$$L(e) = L(-e)$$

Point, Interval and Density Forecasts

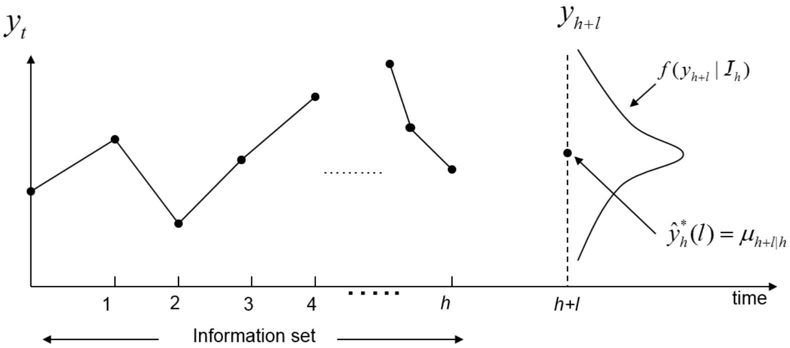
suppose that conditional density $f(y_{h+\ell}|\mathcal{I}_h)$ is $N(\mu_{h+\ell|h}, \sigma^2_{h+\ell|h})$ then density forecast is $N(\mu_{h+\ell|h}, \sigma^2_{h+\ell|h})$ and

if loss function is quadratic $L(e_h(\ell)) = ae_h(\ell)^2$
- optimal point forecast is $\hat{y}^*_h(\ell) = \mu_{h+\ell|h}$
- 95% interval forecast is $\mu_{h+\ell|h} \pm 1.96\sigma_{h+\ell|h}$

if loss function is absolute value $L(e_h(\ell)) = a|e_h(\ell)|$
- optimal point forecast is the conditional median $\hat{y}^*_h(\ell) = median(y_{h+\ell}|\mathcal{I}_h)$

note: if $f(y_{h+\ell}|\mathcal{I}_h)$ is symmetric then mean and median coincide

Quadratic Loss Function



Example: AR(1) model

suppose that y_t follows an AR(1) model $y_t = \phi_0 + \phi_1 y_{t-1} + a_t$ with $a_t \sim N(0, \sigma_a^2)$ and that $L(e_h(\ell)) = a e_h(\ell)^2$ then:

- for conditional mean we have

$$\mu_{h+1|h} = E_h(y_{h+1}|\mathcal{I}_h) = \phi_0 + \phi_1 y_h$$

- for conditional variance

$$\sigma^2_{h+1|h} = var_h(y_{h+1}|\mathcal{I}_h) = var(a_{h+1}) = \sigma_a^2$$

- thus the 1 step ahead point forecast of y_{h+1} is

$$\hat{y}_h(1) = \mu_{h+1|h} = \phi_0 + \phi_1 y_h$$

- the conditional density forecast for y_{h+1} is $N(\phi_0 + \phi_1 y_h, \sigma_a^2)$
- the 95% interval forecast is $\mu_{h+1|h} \pm 1.96\sigma_{h+1|h}$ that is $\phi_0 + \phi_1 y_h \pm 1.96\sigma_a$

Example: AR(1) model

for forecast step $\ell \in \{1, 2, 3, \ldots\}$

- for conditional mean we have

$$\mu_{h+1|h} = E_h(y_{h+1}|\mathcal{I}_h) = \phi_0 + \phi_1 y_h$$
$$\mu_{h+2|h} = E_h(y_{h+2}|\mathcal{I}_h) = \phi_0 + \phi_1 E_h(y_{h+1}|\mathcal{I}_h) = (1+\phi_1)\phi_0 + \phi_1^2 y_h$$
$$\mu_{h+3|h} = E_h(y_{h+3}|\mathcal{I}_h) = \phi_0 + \phi_1 E_h(y_{h+2}|\mathcal{I}_h) = (1+\phi_1+\phi_1^2)\phi_0 + \phi_1^3 y_h$$
$$\vdots$$

  and so $\mu_{h+\ell|h} \rightarrow \frac{\phi_0}{1-\phi_1}$ as $\ell \rightarrow \infty$

- for conditional variance

$$\sigma^2_{h+1|h} = var_h(y_{h+1}|\mathcal{I}_h) = var(a_{h+1}) = \sigma^2_a$$
$$\sigma^2_{h+2|h} = var_h(y_{h+2}|\mathcal{I}_h) = var(\phi_1 y_{h+1} + a_{h+2}|\mathcal{I}_h) = (1+\phi_1^2)\sigma^2_a$$
$$\sigma^2_{h+3|h} = var_h(y_{h+3}|\mathcal{I}_h) = var(\phi_1 y_{h+2} + a_{h+3}|\mathcal{I}_h) = (1+\phi_1^2+\phi_1^4)\sigma^2_a$$
$$\vdots$$

  and so $\sigma^2_{h+\ell|h} \rightarrow \frac{\sigma^2_a}{1-\phi_1^2}$ as $\ell \rightarrow \infty$

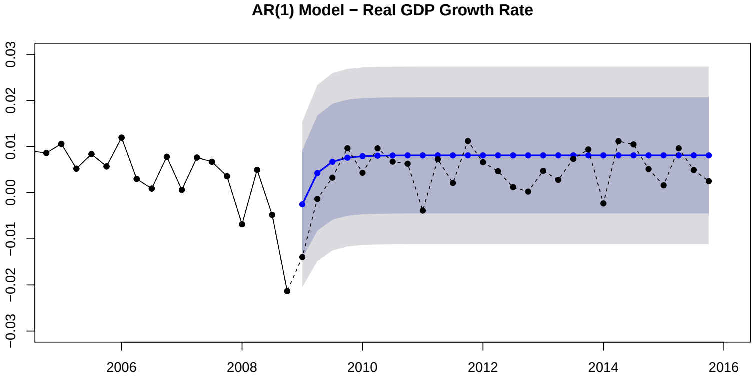
- conditional mean thus converges to the uncontitional mean, conditional variance converges to the uncontitional variance

Example: AR(1) model

```r
rGDP <- Quandl("FRED/GDPC1", type="zoo")
dlrGDP <- diff(log(rGDP))
dlrGDPp1 <- window(dlrGDP, end="2008 Q4")
dlrGDPp2 <- window(dlrGDP, start="2009 Q1")

m1 <- arima(dlrGDPp1, order=c(1,0,0))
library(forecast)
m1.f.1tol <- forecast(m1, length(dlrGDPp2))

plot(m1.f.1tol, type="o", pch=16, xlim=c(2005,2016), ylim=c(-0.03,0.03), main="AR(1) Model - Real GDP Growth Rate")
lines(m1.f.1tol$mean, type="p", pch=16, lty="dashed", col="blue")
lines(dlrGDP, type="o", pch=16, lty="dashed")
```



Example: MA(2) model

suppose that y_t follows an MA(2) model $y_t = \phi_0 + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$ with $a_t \sim N(0, \sigma_a^2)$ and that $L(e_h(\ell)) = ae_h(\ell)^2$ then:

- for conditional mean we have

$$\begin{aligned}
\mu_{h+1|h} &= E_h(y_{h+1}|\mathcal{I}_h) = \phi_0 + \theta_1 a_h + \theta_2 a_{h-1} \\
\mu_{h+2|h} &= E_h(y_{h+2}|\mathcal{I}_h) = \phi_0 + \theta_2 a_h \\
\mu_{h+3|h} &= E_h(y_{h+3}|\mathcal{I}_h) = \phi_0
\end{aligned}$$

- for conditional variance

$$\begin{aligned}
\sigma^2_{h+1|h} &= var_h(y_{h+1}|\mathcal{I}_h) = var(a_{h+1}) = \sigma_a^2 \\
\sigma^2_{h+2|h} &= var_h(y_{h+2}|\mathcal{I}_h) = var(a_{h+2} + \theta_1 a_{h+1}) = (1 + \theta_1^2)\sigma_a^2 \\
\sigma^2_{h+3|h} &= var_h(y_{h+3}|\mathcal{I}_h) = var(a_{h+3} + \theta_1 a_{h+2} + \theta_2 a_{h+1}) = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2
\end{aligned}$$

- the 1, 2, and 3 step ahead point forecasts are thus

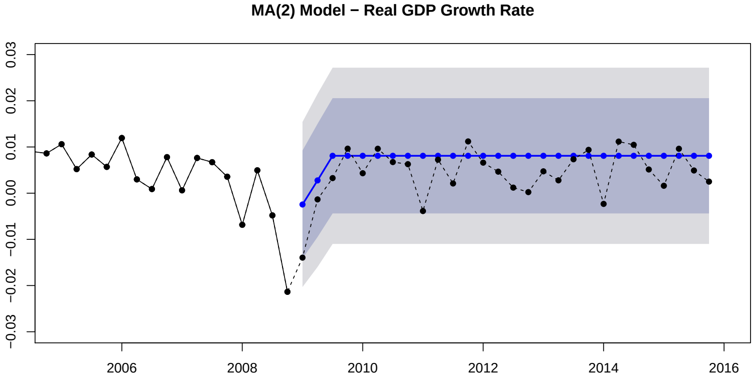
$$\begin{aligned}
\hat{y}_h(1) &= \mu_{h+1|h} = \phi_0 + \theta_1 a_h + \theta_2 a_{h-1} \\
\hat{y}_h(2) &= \mu_{h+2|h} = \phi_0 + \theta_2 a_h \\
\hat{y}_h(3) &= \mu_{h+3|h} = \phi_0
\end{aligned}$$

Example: MA(2) model

```r
rGDP <- Quandl("FRED/GDPC1", type="zoo")
dlrGDP <- diff(log(rGDP))
dlrGDPp1 <- window(dlrGDP, end="2008 Q4")
dlrGDPp2 <- window(dlrGDP, start="2009 Q1")

m2 <- arima(dlrGDPp1, order=c(0,0,2))
library(forecast)
m2.f.1tol <- forecast(m2, length(dlrGDPp2))

plot(m2.f.1tol, type="o", pch=16, xlim=c(2005,2016), ylim=c(-0.03,0.03), main="MA(2) Model - Real GDP Growth Rate")
lines(m2.f.1tol$mean, type="p", pch=16, lty="dashed", col="blue")
lines(dlrGDP, type="o", pch=16, lty="dashed")
```



Forecasting using ARMA(p, q) models

models mostly suitable for forecasts with a small step, forecasts of distant future not particularly accurate

forecast based on an AR(p) model:

- conditional mean converges to unconditional mean gradually
- conditional variance converges to unconditional variance gradually

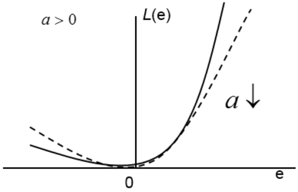
forecast based on an MA(q) model:

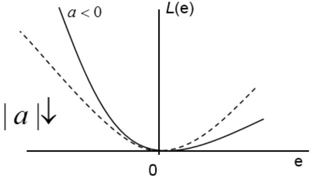
- once $\ell > q$ the conditional mean jumps straight to unconditional mean
- once $\ell > q$ the conditional variance jumps straight to unconditional variance

Asymmetric Loss Function

Linex function

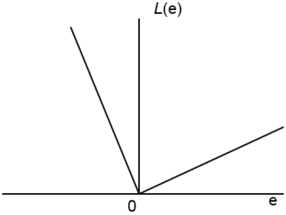
$$L(e) = \exp(ae) - ae - 1, \quad a \neq 0$$





Lin-lin function

$$L(e) = \begin{cases} a\,|e| & e > 0 \\ (1-a)\,|e| & e \leq 0 \end{cases}$$



Point, Interval and Density Forecasts

suppose that conditional density $f(y_{h+\ell}|\mathcal{I}_h)$ is $N(\mu_{h+\ell|h}, \sigma^2_{h+\ell|h})$ so that density forecast is $N(\mu_{h+\ell|h}, \sigma^2_{h+\ell|h})$ and

if loss function is linex $L(e_h(\ell)) = exp(ae_h(\ell)) - ae_h(\ell) - 1$
- optimal point forecast is $\hat{y}_h^*(\ell) = \mu_{h+\ell|h} + \frac{a}{2}\sigma^2_{h+\ell|h}$

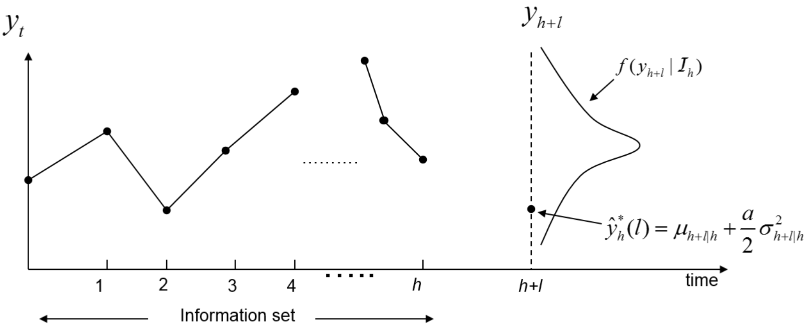
if loss function is linlin

$$L(e_h(\ell)) = \begin{cases} a|e_h(\ell)| & \text{if } e_h(\ell) < 0 \\ (1-a)|e_h(\ell)| & \text{if } e_h(\ell) \geq 0 \end{cases}$$

- optimal point forecast is conditional quintile $\hat{y}_h^*(\ell) = q_a(y_{h+\ell}|\mathcal{I}_h)$

so **for asymmetric loss function optimal forecast is actually biased** - on average forecast error is either positive or negative

Linex Loss Function



Example: AR(1) model

suppose that y_t follows an AR(1) model $y_t = \phi_0 + \phi_1 y_{t-1} + a_t$ with $a_t \sim N(0, \sigma_a^2)$ and that $L(e_h(\ell)) = exp(ae_h(\ell)) - ae_h(\ell) - 1$ then:

- for conditional mean we have

$$\mu_{h+1|h} = E_h(y_{h+1}|\mathcal{I}_h) = \phi_0 + \phi_1 y_h$$

- for conditional variance

$$\sigma^2_{h+1|h} = var_h(y_{h+1}|\mathcal{I}_h) = var(a_{h+1}) = \sigma_a^2$$

- thus the 1 step ahead point forecast of y_{h+1} is

$$\hat{y}_h(1) = \mu_{h+1|h} + \frac{a}{2}\sigma^2_{h+\ell|h} = \phi_0 + \phi_1 y_h + \frac{a}{2}\sigma_a^2$$

- the conditional density forecast for y_{h+1} is $N(\phi_0 + \phi_1 y_h, \sigma_a^2)$

Evaluating Accuracy of Forecasts

general idea:

- split sample into two parts:
  estimation sample $y_1, \ldots, y_t$
  prediction sample $y_{t+1}, \ldots, y_T$
- estimate the model using the first subsample
- evaluate **in-sample accuracy** - compare fitted values $\hat{y}_1, \ldots, \hat{y}_t$ with actual values $y_1, \ldots, y_t$
- use the second subsample to construct set of ℓ step ahead forecasts $\hat{y}_t(\ell), \hat{y}_{t+1}(\ell), \ldots, \hat{y}_{T-l}(\ell)$
- evaluate **out-of-sample accuracy** - compare forecasts $\hat{y}_t(\ell), \hat{y}_{t+1}(\ell), \hat{y}_{T-l}(\ell)$ with actual values $y_{t+\ell}, y_{t+1+\ell}, \ldots, y_T$

In-Sample Evaluation of Accuracy

given the fitted values $\hat{y}_j$ from the model, and in sample residuals $e_j = y_j - \hat{y}_j$

Mean Error - measure of the average bias

$$ME = \frac{1}{t} \sum_{j=0}^{t} e_j$$

Mean Squared Error - sample average loss for quadratic loss function

$$MSE = \frac{1}{t} \sum_{j=0}^{t} e_j^2$$

Mean Absolute Error - sample average loss for absolute value loss function

$$MAE = \frac{1}{t} \sum_{j=0}^{t} |e_j|$$

Mean Absolute Percentage Error

$$MAPE = \frac{1}{t} \sum_{j=0}^{t} \left| \frac{e_j}{y_j} \right|$$

Mean Absolute Scaled Error - compares in sample MAE of the model forecast with in sample MAE for one-step naive forecast method $\hat{y}_{j+1} = y_j$

$$MASE = \frac{\frac{1}{t} \sum_{j=0}^{t} |e_j|}{\frac{1}{t-1} \sum_{j=1}^{t-1} |\hat{y}_{j+1} - y_j|}$$

In-Sample Evaluation of Accuracy

```r
library(Quandl)
rGDP <- Quandl("FRED/GDPC1", type="zoo")
dlrGDP <- diff(log(rGDP))
dlrGDPp1 <- window(dlrGDP, end="2008 Q4")
```

```r
library(forecast)
m1 <- arima(dlrGDPp1, order=c(1,0,0))
accuracy(m1)
```

```
##                        ME        RMSE        MAE       MPE     MAPE      MASE        ACF1
## Training set 1.597568e-05 0.009160185 0.00676911 -46.93084 168.4821 0.8112738 -0.03047387
```

Out-of-Sample Evaluation of Accuracy

given out of sample forecast errors $e_t(\ell), e_{t+1}(\ell), \ldots, e_{T-l}(\ell)$

Mean Error

$$ME = \frac{1}{T-t-\ell+1} \sum_{j=0}^{T-\ell-t} e_{t+j}(\ell)$$

Mean Squared Error

$$MSE = \frac{1}{T-t-\ell+1} \sum_{j=0}^{T-\ell-t} e_{t+j}(\ell)^2$$

Mean Absolute Error

$$MAE = \frac{1}{T-t-\ell+1} \sum_{j=0}^{T-\ell-t} |e_{t+j}(\ell)|$$

Mean Absolute Percentage Error

$$MAPE = \frac{1}{T-t-\ell+1} \sum_{j=0}^{T-\ell-t} \left| \frac{e_{t+j}(\ell)}{y_{t+j+\ell}} \right|$$

Mean Absolute Scaled Error

$$MASE = \frac{\frac{1}{T-l-\ell+1} \sum_{j=0}^{T-\ell-t} |e_{t+j}(l)|}{\frac{1}{t-\ell} \sum_{j=1}^{t-\ell} |y_{j+\ell} - y_j|}$$

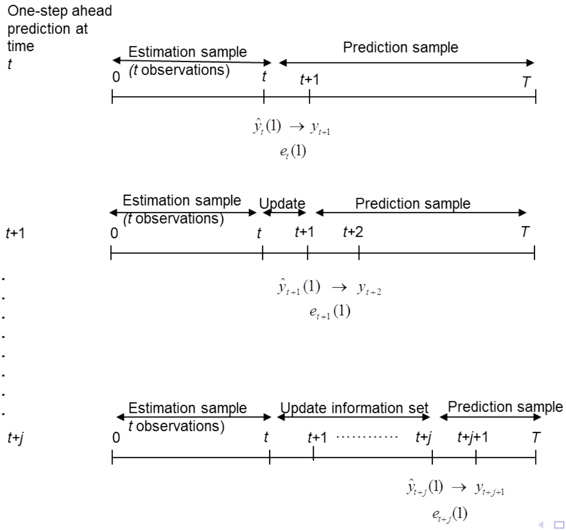
Out-of-Sample Evaluation of Accuracy - Forecasting schemes

- out of sample forecasts and forecast errors used to calculate ME, MSE, MAE, MPE, MAPE, ... can be constructed using one of the three schemes:
  - fixed scheme
  - recursive scheme
  - rolling scheme

Forecasting schemes

Fixed scheme example for one step ahead forecast:

model is estimated only once, each one step ahead forecast is constructed using same parameters



Out-of-Sample Evaluation of Accuracy - Fixed scheme

```r
library(Quandl)
rGDP <- Quandl("FRED/GDPC1", type="zoo")
dlrGDP <- diff(log(rGDP))
fstQ <- 1947.25   # 1947Q2
lstQ <- 2008.75   # 2008Q4
dlrGDPp1 <- window(dlrGDP, end=lstQ)
dlrGDPp2 <- window(dlrGDP, start=lstQ+0.25)
```

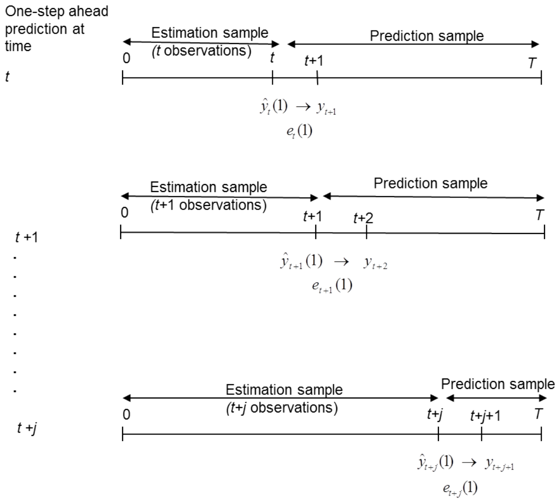
```r
library(forecast)
m1 <- arima(dlrGDPp1, order=c(1,0,0))
m1.fcst <- Arima(x=dlrGDP, model=m1)
m1.fcst.fix <- window(m1.fcst$x-m1.fcst$residuals, start=2009)
accuracy(m1.fcst.fix, dlrGDPp2)
```

```
##                     ME        RMSE         MAE       MPE     MAPE
## Test set -0.002096299 0.005120189 0.004211925 -105.9924 195.7948
```

Forecasting schemes

Recursive scheme example for one step ahead forecast:

estimation sample keeps expanding and model is reestimated again when each new observation is added to the estimation sample



Out-of-Sample Evaluation of Accuracy - Recursive scheme

```r
library(Quandl)
rGDP <- Quandl("FRED/GDPC1", type="zoo")
dlrGDP <- diff(log(rGDP))
fstQ <- 1947.25   # 1947Q2
lstQ <- 2008.75   # 2008Q4
dlrGDPp1 <- window(dlrGDP, end=lstQ)
dlrGDPp2 <- window(dlrGDP, start=lstQ+0.25)
```

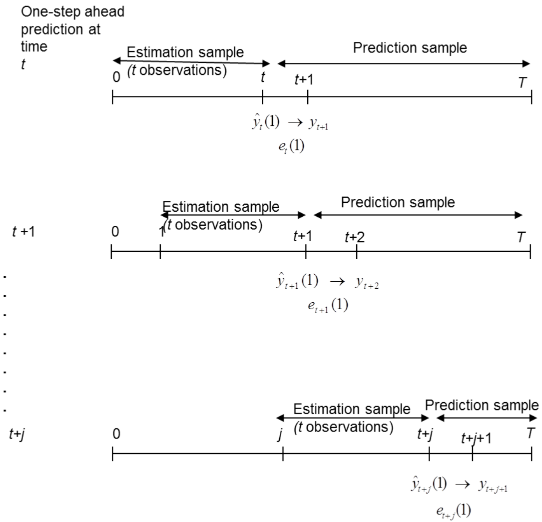
```r
library(forecast)
m1.fcst.rec <- zoo()
for(i in 1:length(dlrGDPp2))
{
    y <- window( dlrGDP, end=lstQ+(i-1)/4 )
    m1new <- arima(y, order=c(1,0,0))
    m1.fcst.rec <- c(m1.fcst.rec, forecast(m1new, 1)$mean)
}
m1.fcst.rec <- as.ts(m1.fcst.rec)
accuracy(m1.fcst.rec, dlrGDPp2)
```

```
##                    ME        RMSE         MAE       MPE     MAPE
## Test set -0.001930672 0.005095709 0.004181166 -100.5574 189.3387
```

Forecasting schemes

Rolling scheme example for one step ahead forecast:

estimation sample always contains the same number of observation and model is reestimated again within each rolling sample



Out-of-Sample Evaluation of Accuracy - Rolling scheme

```r
library(Quandl)
rGDP <- Quandl("FRED/GDPC1", type="zoo")
dlrGDP <- diff(log(rGDP))
fstQ <- 1947.25  # 1947Q2
lstQ <- 2008.75  # 2008Q4
dlrGDPp1 <- window(dlrGDP, end=lstQ)
dlrGDPp2 <- window(dlrGDP, start=lstQ+0.25)
```

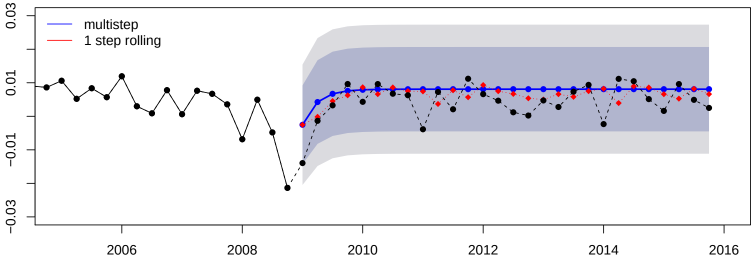
```r
library(forecast)
m1.fcst.rol <- zoo()
for(i in 1:length(dlrGDPp2))
{
    y <- window( dlrGDP, start=fstQ+(i-1)/4, end=lstQ+(i-1)/4 )
    m1new <- arima(y, order=c(1,0,0))
    m1.fcst.rol <- c(m1.fcst.rol, forecast(m1new, 1)$mean)
}
m1.fcst.rol <- as.ts(m1.fcst.rol)
accuracy(m1.fcst.rol, dlrGDPp2)
```

```
##                     ME        RMSE         MAE       MPE     MAPE
## Test set -0.001894988 0.005080691 0.004177216 -99.53401 188.5778
```

Forecasting schemes - Comparison

- fixed scheme is fast and convenient (there is only one estimation), but does not allow for parameter updating
- recursive scheme incorporates in the estimation all information available, advantageous if model is stable over time, but if data has structural breaks model's stability and forecast accuracy are compromised
- rolling scheme is more robust against structural breaks in the data, avoids potential problem with model's stability

Comparison - Multistep Forecast vs 1 step Rolling Scheme Forecast




```r
# multistep forecast
accuracy(m1.fcst.1tol$mean, dlrGDPp2)
```

```
##                    ME        RMSE         MAE       MPE     MAPE
## Test set -0.003166626 0.005246033 0.004239532 -153.0118 252.9683
```

```r
# 1 step rolling scheme forecast
accuracy(m1.fcst.rol, dlrGDPp2)
```

```
##                    ME        RMSE         MAE       MPE     MAPE
## Test set -0.001894988 0.005080691 0.004177216 -99.53401 188.5778
```