MONASH University

MONASH BUSINESS SCHOOL

Time Series in R: Forecasting and Visualisation

Forecast evaluation

29 May 2017

Outline

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”.
- Often not true forecasts since parameters are estimated on all data.

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

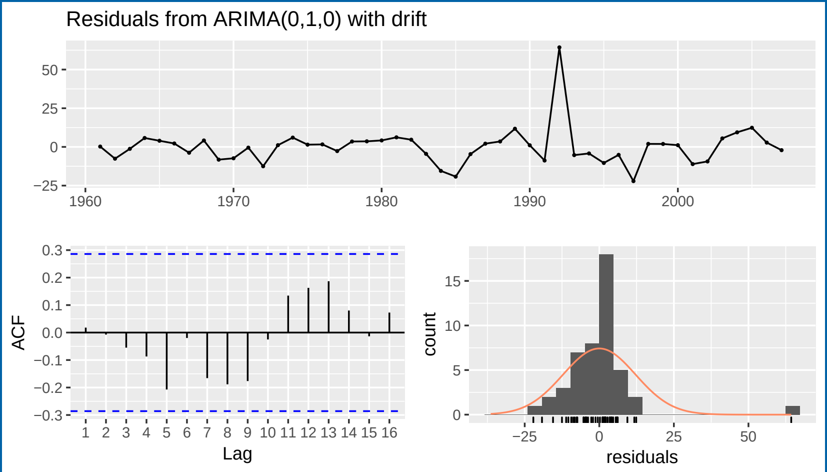
1. $\{e_t\}$ uncorrelated. If they aren't, then information left in residuals that should be used in computing forecasts.
2. $\{e_t\}$ have mean zero. If they don't, then forecasts are biased.

Useful properties (for prediction intervals)

3. $\{e_t\}$ have constant variance.
4. $\{e_t\}$ are normally distributed.

`checkresiduals()` function

```
livestock %>% auto.arima %>% checkresiduals(test=FALSE)
```



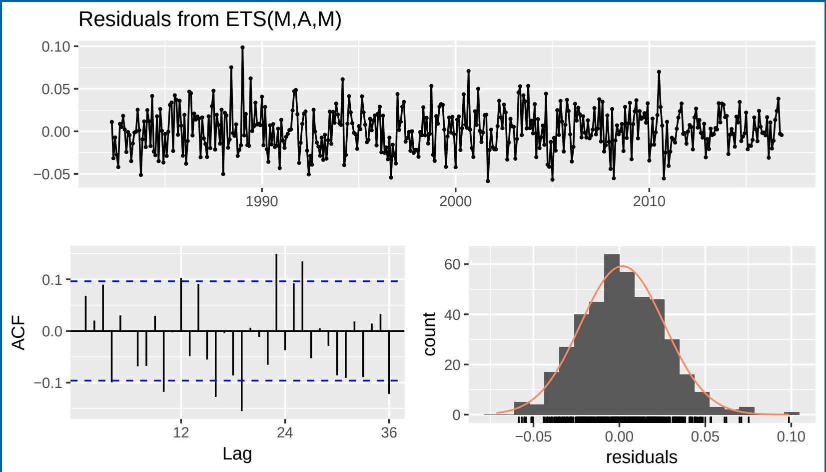
checkresiduals() function

```r
livestock %>% auto.arima %>% checkresiduals(plot=FALSE)
```

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,0) with drift
## Q* = 8.6, df = 9, p-value = 0.5
##
## Model df: 1.   Total lags used: 10
```

`checkresiduals()` function

```
auscafe %>% ets %>% checkresiduals(test=FALSE)
```



checkresiduals() function

```
auscafe %>% ets %>% checkresiduals(plot=FALSE)
```

```
##
##  Ljung-Box test
##
## data:  Residuals from ETS(M,A,M)
## Q* = 64, df = 8, p-value = 7e-11
##
## Model df: 16.   Total lags used: 24
```

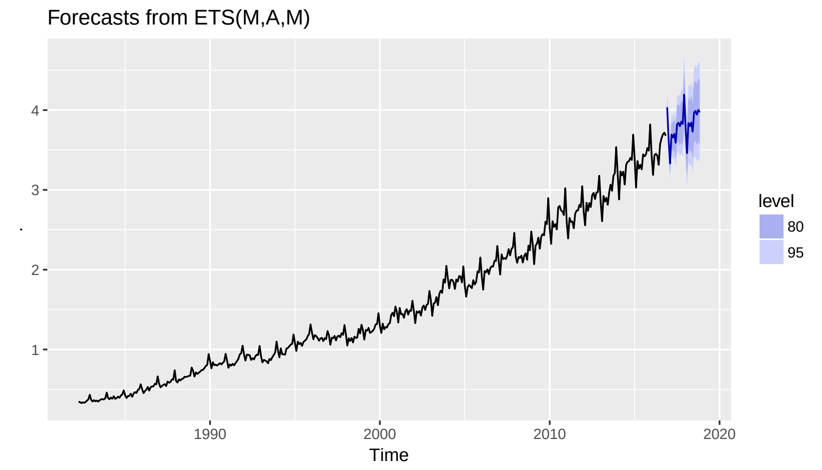
Residuals and forecasting

- Autocorrelations left in residuals suggest the forecast method can be improved (in theory).
- Small autocorrelations have little effect, even if significant.
- Non-Gaussian residuals can be handled using bootstrapped forecast intervals:

```
forecast(..., bootstrap=TRUE)
```

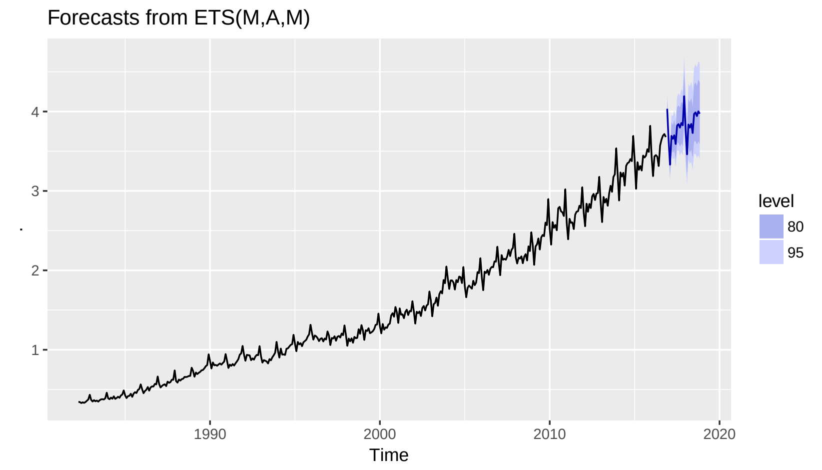
Bootstrapped forecast intervals

```
auscafe %>% ets %>% forecast %>% autoplot
```



Bootstrapped forecast intervals

```
auscafe %>% ets %>% forecast(bootstrap=TRUE) %>% autoplot
```



Outline

Training and test sets

- A model which fits the training data well will not necessarily forecast well.
- A perfect fit can always be obtained by using a model with enough parameters.
- Over-fitting a model to data is just as bad as failing to identify a systematic pattern in the data.
- The test set must not be used for *any* aspect of model development or calculation of forecasts.

Forecast errors

Forecast "error": the difference between an observed value and its forecast.

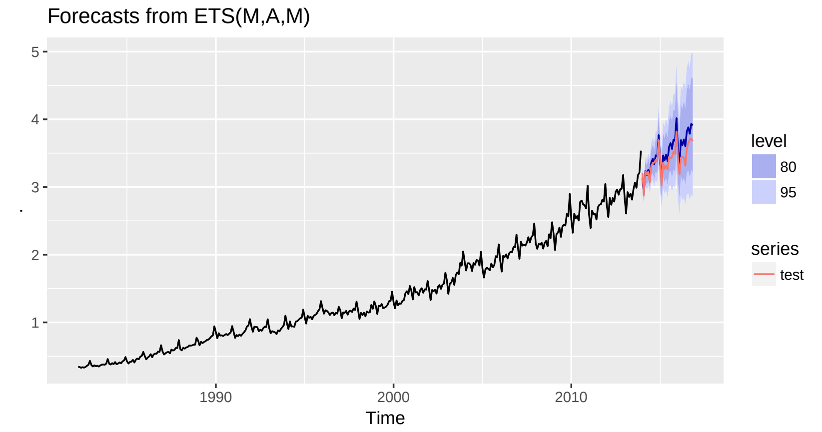
$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T},$$

where the training data is given by $\{y_1, \dots, y_T\}$

- Unlike residuals, forecast errors on the test set involve multi-step forecasts.
- These are *true* forecast errors as the test data is not used in computing $\hat{y}_{T+h|T}$.

Measures of forecast accuracy

```r
training <- window(auscafe, end=c(2013,12))
test <- window(auscafe, start=c(2014,1))
training %>% ets %>% forecast(h=length(test)) -> fc
autoplot(fc) + autolayer(test)
```



Measures of forecast accuracy

Let $\hat{y}_{t+h|t}$ denote the forecast of y_{t+h} using data up to time t.

Training set measures:

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

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

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

Measures of forecast accuracy

Let $\hat{y}_{t+h|t}$ denote the forecast of y_{t+h} using data up to time t.

Training set measures:

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

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

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

- 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 $\hat{y}_{t+h|t}$ denote the forecast of y_{t+h} using data up to time t.

Test set measures:

$$\text{MAE} = \frac{1}{H}\sum_{h=1}^{H} |y_{T+h} - \hat{y}_{T+h|T}|$$

$$\text{MSE} = \frac{1}{H}\sum_{h=1}^{H} (y_{T+h} - \hat{y}_{T+h|T})^2 \qquad \text{RMSE} = \sqrt{\frac{1}{H}\sum_{h=1}^{H} (y_{T+h} - \hat{y}_{T+h|T})^2}$$

$$\text{MAPE} = \frac{100}{H}\sum_{h=1}^{H} |y_{T+h} - \hat{y}_{T+h|T}|/|y_t|$$

- 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

$$\text{MASE} = \frac{1}{H}\sum_{h=1}^{H} \left|y_{T+h} - \hat{y}_{T+h|T}\right| / Q$$

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,

$$Q = \frac{1}{T-1}\sum_{t=2}^{T} \left|y_t - y_{t-1}\right|$$

works well. Then MASE is equivalent to MAE relative to a naïve method.

Measures of forecast accuracy

Mean Absolute Scaled Error

$$\text{MASE} = \frac{1}{H}\sum_{h=1}^{H} |y_{T+h} - \hat{y}_{T+h|T}|/Q$$

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,

$$Q = \frac{1}{T-m}\sum_{t=m+1}^{T} |y_t - y_{t-m}|$$

works well. Then MASE is equivalent to MAE relative to a seasonal naïve method.

Measures of forecast accuracy

```r
training <- window(auscafe, end=c(2013,12))
test <- window(auscafe, start=c(2014,1))
training %>% ets %>% forecast(h=length(test)) -> fc
accuracy(fc, test)
```

```
##                     ME    RMSE     MAE      MPE
## Training set  0.001482 0.03816 0.02761  0.09342
## Test set     -0.121597 0.15318 0.13016 -3.51895
##               MAPE   MASE   ACF1 Theil's U
## Training set 2.056 0.2881 0.2006        NA
## Test set     3.780 1.3583 0.6323    0.7647
```

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.

Outline

1. Forecasting residuals
2. Evaluating forecast accuracy
3. **Forecasting benchmark methods**
4. Lab session 7
5. Time series cross-validation
6. Lab session 8

Forecasting benchmark methods

Average method

- Forecasts equal to mean of historical data.

Naïve method

- Forecasts equal to last observed value.
- Consequence of efficient market hypothesis.

Seasonal naïve method

- Forecasts equal to last value from same season.

Drift method

- Forecasts equal to last value plus average change.
- Equivalent to extrapolating a line drawn between first and last observations.

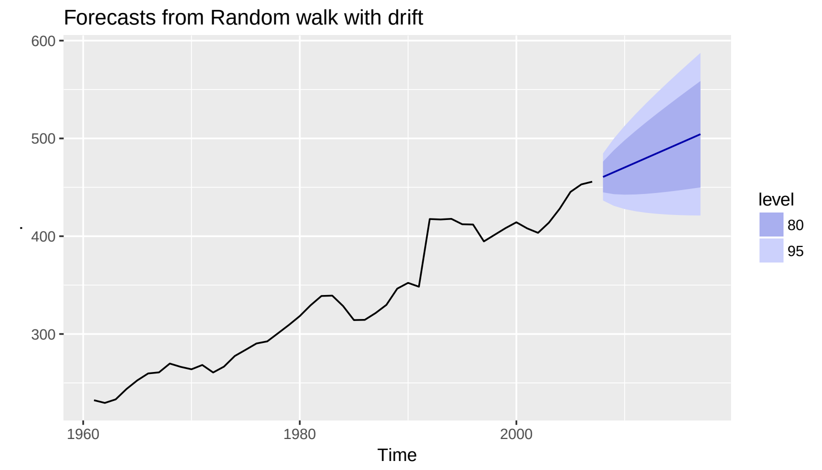
Forecasting benchmark methods

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

Check that your method does better than these standard benchmark methods.

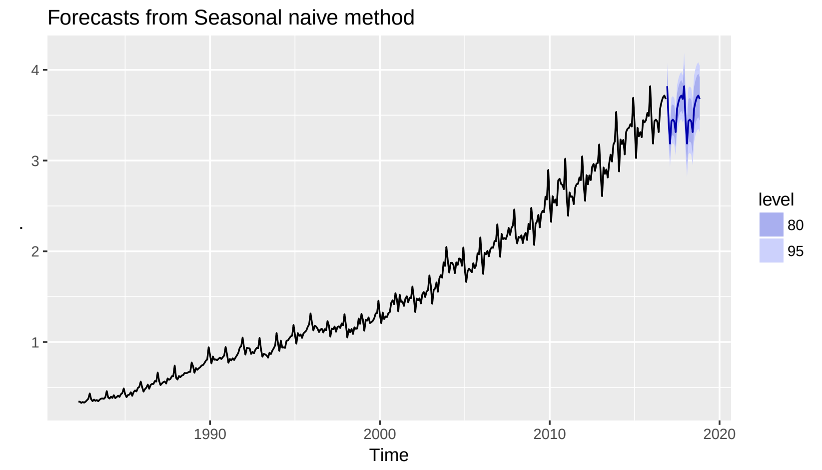
Forecasting benchmark methods

```
livestock %>% rwf(drift=TRUE) %>% autoplot
```



Forecasting benchmark methods

```
auscafe %>% snaive %>% autoplot
```



Forecasting benchmark methods

```
training %>% ets %>% forecast(h=length(test)) -> fc_ets
training %>% snaive(h=length(test)) -> fc_snaive
accuracy(fc_ets, test)
```

```
##                      ME    RMSE     MAE      MPE
## Training set  0.001482 0.03816 0.02761  0.09342
## Test set     -0.121597 0.15318 0.13016 -3.51895
##               MAPE   MASE   ACF1 Theil's U
## Training set 2.056 0.2881 0.2006        NA
## Test set     3.780 1.3583 0.6323    0.7647
```

```
accuracy(fc_snaive,test)
```

```
##                   ME   RMSE     MAE    MPE   MAPE
## Training set 0.08569 0.1226 0.09583  6.529  7.286
## Test set     0.41363 0.4344 0.41363 12.183 12.183
##               MASE   ACF1 Theil's U
## Training set 1.000 0.8425        NA
## Test set     4.317 0.6438     2.165
```

Outline

Lab Session 7

Outline

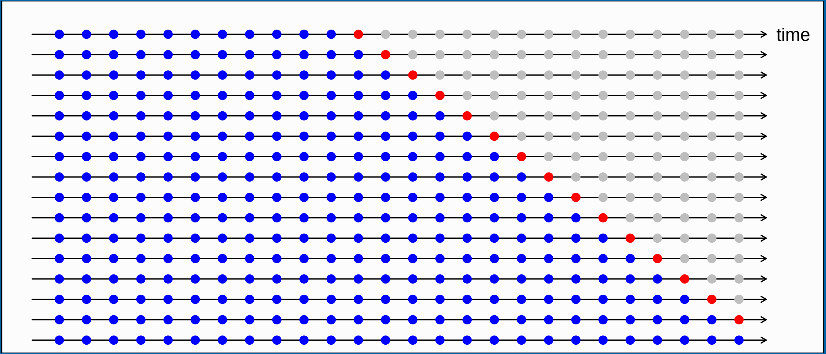
Time series cross-validation

Traditional evaluation

Time series cross-validation

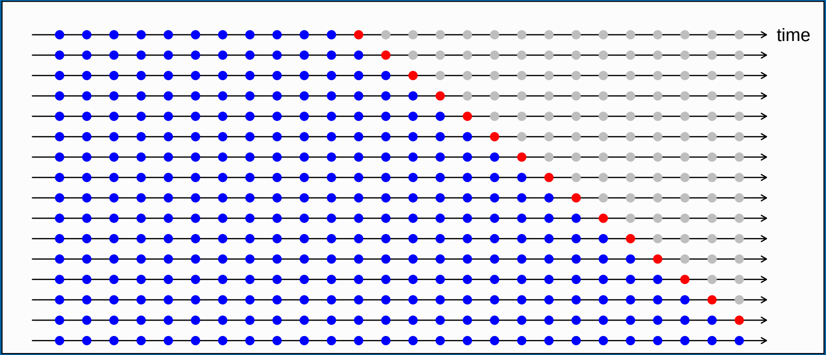
Traditional evaluation

Time series cross-validation



Time series cross-validation

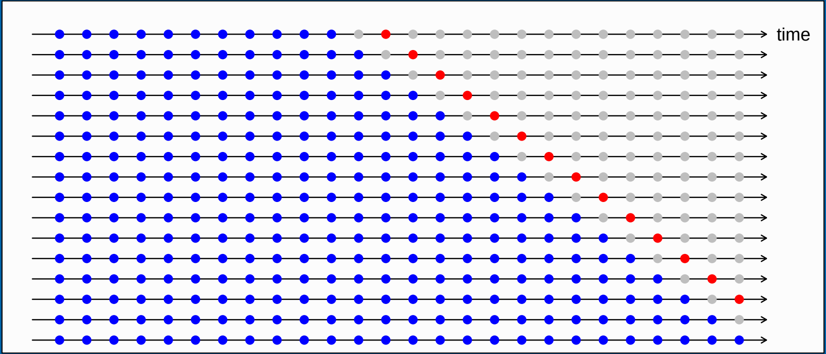
$h = 1$



- Forecast accuracy averaged over test sets.
- Also known as “evaluation on a rolling forecasting origin”

Time series cross-validation

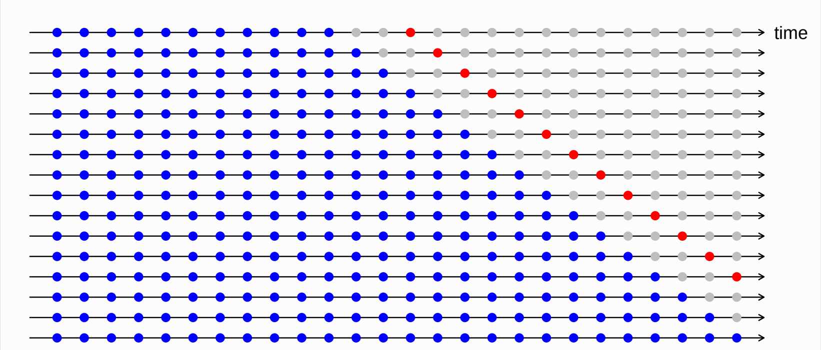
h = 2



- Forecast accuracy averaged over test sets.
- Also known as “evaluation on a rolling forecasting origin”

Time series cross-validation

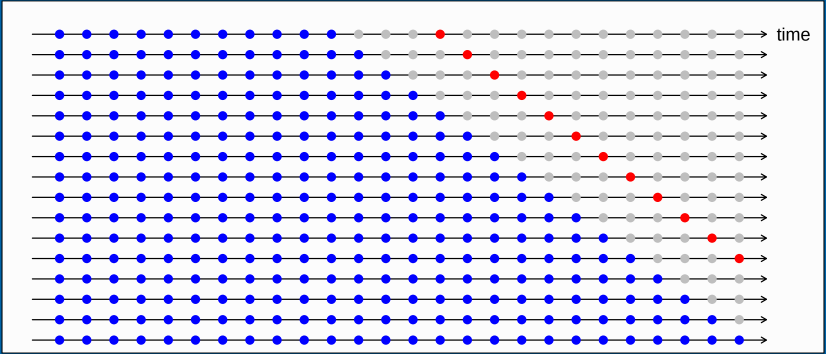
$h = 3$



- Forecast accuracy averaged over test sets.
- Also known as "evaluation on a rolling forecasting origin"

Time series cross-validation

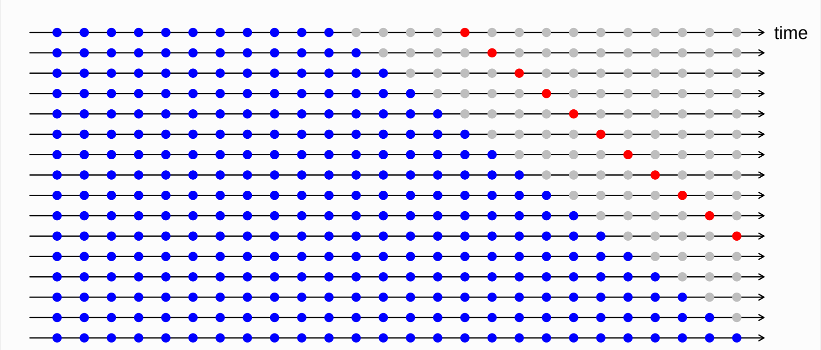
$h = 4$



- Forecast accuracy averaged over test sets.
- Also known as "evaluation on a rolling forecasting origin"

Time series cross-validation

$h = 5$

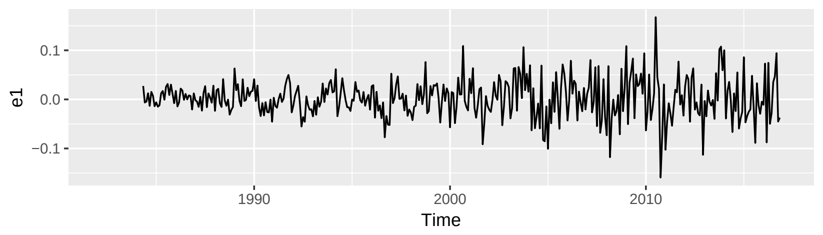


- Forecast accuracy averaged over test sets.
- Also known as “evaluation on a rolling forecasting origin”

tsCV function:

```
e <- tsCV(ts, forecastfunction, h=1, ...)
```

```
e1 <- tsCV(auscafe, stlf,
  etsmodel="AAN", damped=FALSE, lambda=0)
autoplot(e1)
```



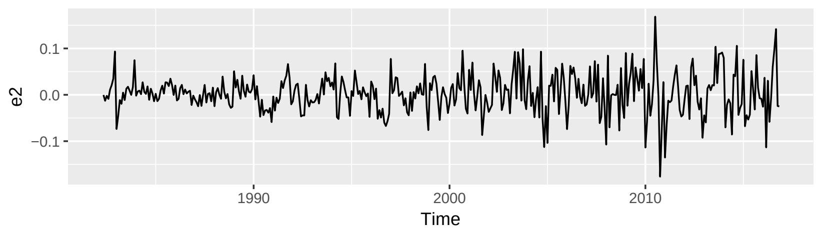

```
sqrt(mean((e1/auscafe)^2, na.rm=TRUE))
```

```
## [1] 0.0259
```

tsCV function:

For `ets` and `auto.arima`, you need to write a single forecasting function:

```
fets <- function(x, h, model="ZZZ", damped=NULL, ...) {
  forecast(ets(x, model=model, damped=damped), h=h)
}
e2 <- tsCV(auscafe, fets, model="MAM", damped=FALSE)
autoplot(e2)
```




```
sqrt(mean((e2/auscafe)^2, na.rm=TRUE))
```

```
## [1] 0.03301
```

tsCV function:

Comparison should be over the same observations:

```r
pe1 <- window(100*e1/auscafe, start=1985)
pe2 <- window(100*e2/auscafe, start=1985)
sqrt(mean(pe1^2, na.rm=TRUE))
```

```
## [1] 2.571
```

```r
sqrt(mean(pe2^2, na.rm=TRUE))
```

```
## [1] 2.733
```

Time series cross-validation

- A good way to choose the best forecasting model is to find the model with the smallest RMSE computed using time series cross-validation.
- Minimizing AICc is asymptotically equivalent to minimizing tscv with $h = 1$.

Outline

Lab Session 8