MONASH University

MONASH BUSINESS SCHOOL

# ETF3231/5231: Business forecasting

## Week 4: The forecasters’ toolbox

https://bf.numbat.space/

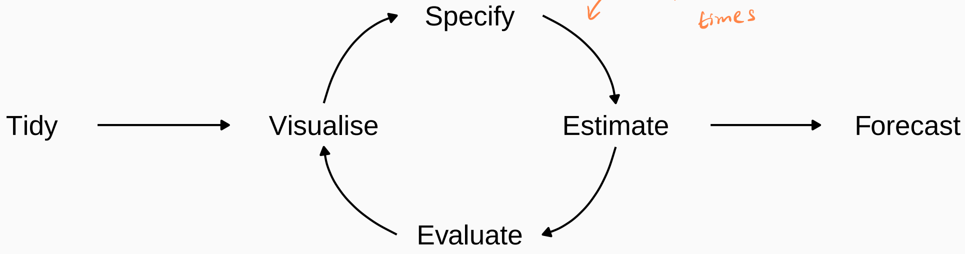
AACSB ACCREDITED EQUIS ACCREDITED ASSOCIATION OF MBAS AMBA ACCREDITED

# Outline

# Outline

# A tidy forecasting workflow

you may iterate this a few times




R: Demo GDP example

# Outline

# Some simple forecasting methods - Benchmarks

- `MEAN(y)`: Average method
- `NAIVE(y)`: Naïve method
- `SNAIVE(y ~ lag(m))`: Seasonal naïve method
- `RW(y ~ drift())`: Drift method

Note: distinguish between a method and a model

| Method | Forecasts | Implicit model $\varepsilon_t \overset{iid}{\sim} N(0,\sigma^2)$ |
|---|---|---|
| MEAN( ) | $\hat{y}_{T+h\|T} = \bar{y} = (y_1 + \dots + y_T)/T$ | |
| NAIVE( ) | $\hat{y}_{T+h\|T} = y_T$ | |
| SNAIVE( ) | $\hat{y}_{T+h\|T} = y_{T+h-m(k+1)}$ | |
| DRIFT( ) | $\hat{y}_{T+h\|T} = y_T + \frac{h}{(T-1)} \sum_{t=2}^{T} (y_t - y_{t-1})$ $= y_T + \frac{h}{T-1}(y_T - y_1)$ | |

| Method | Forecasts | Implicit model $\varepsilon_t \overset{iid}{\sim} N(0,\sigma^2)$ |
|---|---|---|
| MEAN( ) | $\hat{y}_{T+h\mid T} = \bar{y} = (y_1 + \cdots + y_T)/T$ | $y_t = c + \varepsilon_t$ |
| NAIVE( ) | $\hat{y}_{T+h\mid T} = y_T$ | |
| SNAIVE( ) | $\hat{y}_{T+h\mid T} = y_{T+h-m(k+1)}$ | |
| DRIFT( ) | $\hat{y}_{T+h\mid T} = y_T + \frac{h}{(T-1)} \sum_{t=2}^{T} (y_t - y_{t-1})$ $= y_T + \frac{h}{T-1}(y_T - y_1)$ | |

| Method | Forecasts | Implicit model $\varepsilon_t \overset{iid}{\sim} N(0,\sigma^2)$ |
| --- | --- | --- |
| MEAN( ) | $\hat{y}_{T+h\mid T} = \bar{y} = (y_1 + \cdots + y_T)/T$ | $y_t = c + \varepsilon_t$ |
| NAIVE( ) | $\hat{y}_{T+h\mid T} = y_T$ | $y_t = y_{t-1} + \varepsilon_t$ |
| SNAIVE( ) | $\hat{y}_{T+h\mid T} = y_{T+h-m(k+1)}$ | |
| DRIFT( ) | $\hat{y}_{T+h\mid T} = y_T + \frac{h}{(T-1)} \sum_{t=2}^{T} (y_t - y_{t-1})$ | |

$$= y_T + \frac{h}{T-1}(y_T - y_1)$$

| Method | Forecasts | Implicit model $\varepsilon_t \overset{iid}{\sim} N(0,\sigma^2)$ |
|---|---|---|
| MEAN( ) | $\hat{y}_{T+h\vert T} = \bar{y} = (y_1 + \dots + y_T)/T$ | $y_t = c + \varepsilon_t$ |
| NAIVE( ) | $\hat{y}_{T+h\vert T} = y_T$ | $y_t = y_{t-1} + \varepsilon_t$ |
| SNAIVE( ) | $\hat{y}_{T+h\vert T} = y_{T+h-m(k+1)}$ | $y_t = y_{t-m} + \varepsilon_t$ |
| DRIFT ( ) | $\hat{y}_{T+h\vert T} = y_T + \frac{h}{(T-1)} \sum_{t=2}^{T} (y_t - y_{t-1})$ $= y_T + \frac{h}{T-1}(y_T - y_1)$ | |

| Method | Forecasts | Implicit model $\varepsilon_t \overset{iid}{\sim} N(0,\sigma^2)$ |
|---|---|---|
| MEAN( ) | $\hat{y}_{T+h\|T} = \bar{y} = (y_1 + \dots + y_T)/T$ | $y_t = c + \varepsilon_t$ |
| NAIVE( ) | $\hat{y}_{T+h\|T} = y_T$ | $y_t = y_{t-1} + \varepsilon_t$ |
| SNAIVE( ) | $\hat{y}_{T+h\|T} = y_{T+h-m(k+1)}$ | $y_t = y_{t-m} + \varepsilon_t$ |
| DRIFT( ) | $\hat{y}_{T+h\|T} = y_T + \frac{h}{(T-1)} \sum_{t=2}^{T} (y_t - y_{t-1})$ $= y_T + \frac{h}{T-1}(y_T - y_1)$ | $y_t = c + y_{t-1} + \varepsilon_t$ |

# Model fitting

- The `model()` function trains models to data.
- The `forecast()` function generates forecasts.

# `SNAIVE(y ~ lag(m))`: Seasonal naïve method

- Forecasts equal to last value from same season.
- Forecasts: $\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$, where $m$ = seasonal period and $k$ is the integer part of $(h-1)/m$.

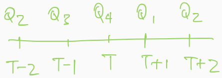
# `SNAIVE(y ~ lag(m))`: Seasonal naïve method

- Forecasts equal to last value from same season.
- Forecasts: $\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$, where $m$ = seasonal period and $k$ is the integer part of $(h-1)/m$.

Example : quarterly data $(m=4)$

$h=2$ : $(h-1)/m = (2-1)/4 = 0.25 \Rightarrow k=0$ (integer part)

$\hat{y}_{T+2|T} = \hat{y}_{T+2-m(0+1)} = y_{T+2-4} = y_{T-2}$



# `SNAIVE(y ~ lag(m))`: Seasonal naïve method

- Forecasts equal to last value from same season.
- Forecasts: $\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$, where $m$ = seasonal period and $k$ is the integer part of $(h-1)/m$.

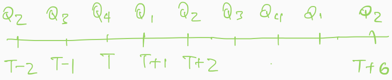
Example : quarterly data $(m=4)$

$h=2$ : $(h-1)/m = (2-1)/4 = 0.25 \Rightarrow k=0$ (integer part)

$\hat{y}_{T+2|T} = \hat{y}_{T+2-m(0+1)} = y_{T+2-4} = y_{T-2}$

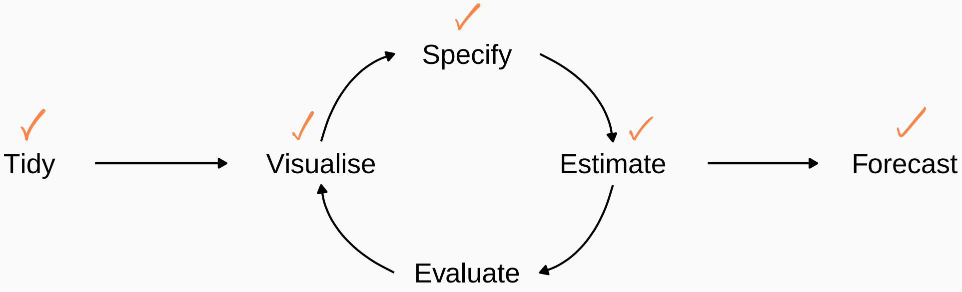
$h=6$ : $(h-1)/m = (6-1)/4 = 1.25 \Rightarrow k=1$

$\hat{y}_{T+6|T} = y_{T+6-4(1+1)} = y_{T-2}$



# A tidy forecasting workflow - Recap

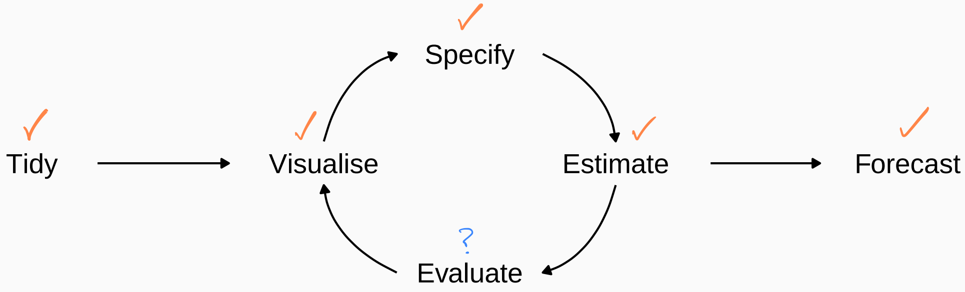
Complete Seminar Activities 1-3 before going through the slide




- Introduced four simple methods / benchmarks (mean, naïve, seas naïve, drift)

# A tidy forecasting workflow - Recap

Complete Seminar Activities 1-2 before going through the slide

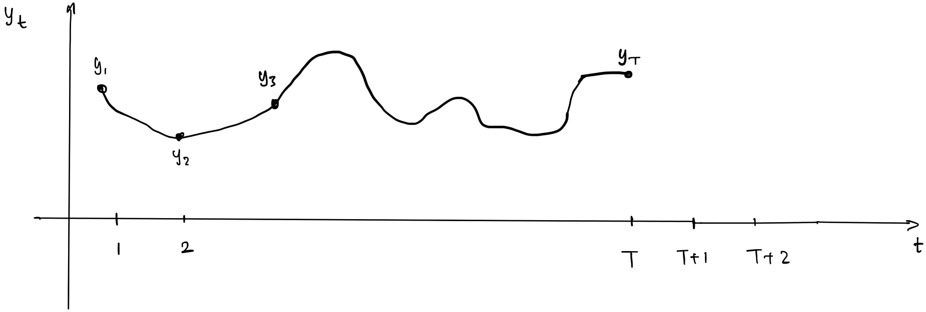


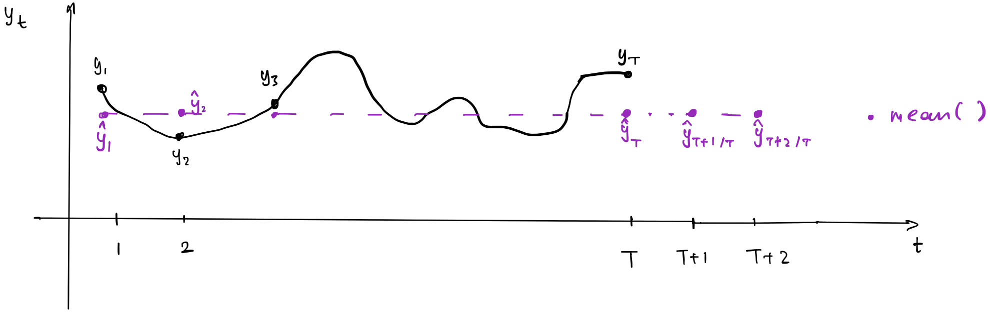

- Introduced four simple methods/benchmarks (mean, naïve, seas naïve, drift)
- Evaluate: 1. How well do these fit the data
  2. How well do they actually forecast

# Outline

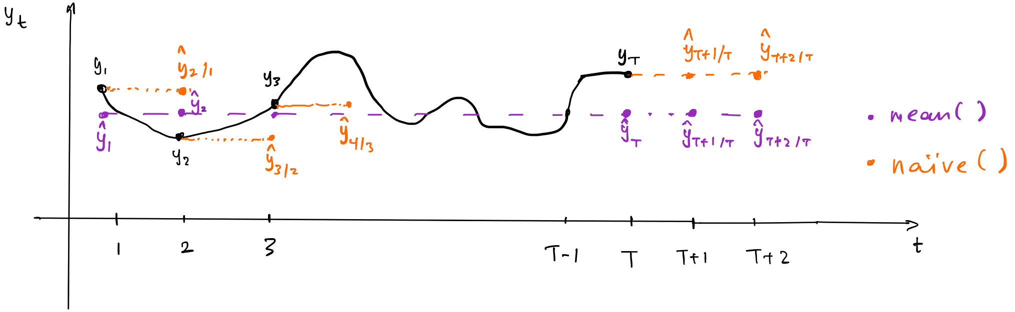
# Forecasting residuals

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


$y_t$
$y_1$
$y_2$
$y_3$
$y_T$
1
2
T
T+1
T+2
t



$y_t$
$y_1$
$y_2$
$y_3$
$y_T$
$\hat{y}_1$
$\hat{y}_2$
$\hat{y}_T$
$\hat{y}_{T+1/T}$
$\hat{y}_{T+2/T}$
• mean( )
1
2
T
T+1
T+2
t



$y_t$
$y_1$
$\hat{y}_{2|1}$
$\hat{y}_2$
$y_3$
$\hat{y}_1$
$y_2$
$\hat{y}_{3|2}$
$\hat{y}_{4|3}$
$y_T$
$\hat{y}_{T+1|T}$
$\hat{y}_{T+2|T}$
$\hat{y}_T$
$\hat{y}_{T+1|T}$
$\hat{y}_{T+2|T}$
mean( )
naïve( )
1
2
3
T-1
T
T+1
T+2
t

# 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 distributions & prediction intervals)

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

# 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.

$H_0 : \rho_\ell = 0$

- 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.

## Ljung-Box test

$H_0: \rho_1 = \rho_2 = \dots = \rho_\ell = 0 \sim WN$

$$Q^* = T(T+2)\sum_{k=1}^{\ell}(T-k)^{-1}r_k^2 \sim \chi^2_{\ell-k}$$

$k =$ no. of parameters in the model

$\ell > k$

where $\ell$ is max lag being considered and $T$ is number of observations.

- My preferences: $\ell = 10$ for non-seasonal data, $\ell = 2m$ for seasonal data.
- Better performance, especially in small samples.

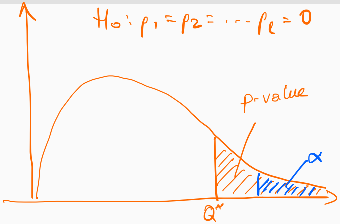
# Portmanteau tests

- If data are WN, $Q^*$ has $\chi^2$ distribution with $(\ell - K)$ degrees of freedom where $K$ = no. parameters in model.
- When applied to raw data, set $K = 0$.
- `lag` = $\ell$, `dof` = $K$

```r
augment(fit) %>%
  features(.resid, ljung_box, lag=10, dof=0)
```

$Q^*$ p-value

```
## # A tibble: 1 x 4
##   Symbol .model       lb_stat lb_pvalue
##   <chr>  <chr>          <dbl>     <dbl>
## 1 FB     NAIVE(Close)    12.1     0.276
```




Seminar Activities 3-4

Cannot reject $H_0$ if $p > \alpha$, i.e. WN

# Outline

# Forecast distributions

Assuming residuals: have zero mean, are uncorrelated, normal, with variance = $\hat{\sigma}^2$:

| | |
|---|---|
| **Mean:** | $y_{T+h\|T} \sim N(\bar{y}, (1 + 1/T)\hat{\sigma}^2)$ |
| **Naïve:** | $y_{T+h\|T} \sim N(y_T, h\hat{\sigma}^2)$ |
| **Seasonal naïve:** | $y_{T+h\|T} \sim N(y_{T+h-m(k+1)}, (k+1)\hat{\sigma}^2)$ |
| **Drift:** | $y_{T+h\|T} \sim N(y_T + \frac{h}{T-1}(y_T - y_1), h\frac{T+h}{T}\hat{\sigma}^2)$ |

where $k$ is the integer part of $(h-1)/m$.

Note that when $h = 1$ and $T$ is large, these all give the same approximate forecast variance: $\hat{\sigma}^2$.

So for h=1 in-sample variance can be used (no need to do anything more)

NAÏVE METHOD : $\hat{y}_{T+h/T} = y_T$

NAÏVE METHOD : $\hat{y}_{T+h/T} = y_T$

RANDOM WALK MODEL : $y_t = y_{t-1} + \varepsilon_t \quad , \varepsilon_t \sim iid\, N(0, \sigma^2)$

$t = 1, \ldots, T$

NAÏVE METHOD : $\hat{y}_{T+h/T} = y_T$

RANDOM WALK MODEL : $y_t = y_{t-1} + \varepsilon_t$ , $\varepsilon_t \sim iid\ N(0, \sigma^2)$

$t = 1, \dots, T$

Distinguish between residuals $e_t = y_t - \hat{y}_{t/t-1}$ & model errors $\varepsilon_t$

we estimate $\sigma^2$ by $\hat{\sigma}^2$ var of residuals

NAÏVE METHOD : $\hat{y}_{T+h/T} = y_T$

RANDOM WALK MODEL : $y_t = y_{t-1} + \varepsilon_t$ , $\varepsilon_t \sim iid\, N(0, \sigma^2)$

$t = 1, \ldots, T$

Distinguish between residuals $e_t = y_t - \hat{y}_{t|t-1}$ & model errors $\varepsilon_t$

we estimate $\sigma^2$ by $\hat{\sigma}^2$ var of residuals

Question: What would the RW model look like at Tth? Let's do it sequentially.

$y_t = y_{t-1} + \varepsilon_t \qquad t = 1, \ldots, T. \qquad \varepsilon_t \sim iid\, N(0, \sigma^2)$

Let's go out-of-sample/future $t = T + \underline{1}$.

$y_t = y_{t-1} + \varepsilon_t \qquad t = 1, \ldots, T. \qquad \varepsilon_t \sim \text{iid } N(0, \sigma^2)$

Let's go out-of-sample/future $t = T+1$.

$$y_{T+1} = y_T + \varepsilon_{T+1}$$

$$y_{T+2} = y_{T+1} + \varepsilon_{T+2} = y_T + \varepsilon_{T+1} + \varepsilon_{T+2}$$

$$y_{T+3} = y_{T+2} + \varepsilon_{T+3} = y_T + \varepsilon_{T+1} + \varepsilon_{T+2} + \varepsilon_{T+3}$$

$$y_{T+h} = y_T + \sum_{i=0}^{h} \varepsilon_{T+h-i}$$

So $y_T$ + all errors between $y_T$ + $y_{T+h}$

<u>mean</u>

$$E(y_{T+h}) = E(y_T) + \sum_{i=0}^{h-1} E(\varepsilon_{T+h-i}) = y_T$$

$$\Rightarrow \quad \hat{y}_{T+h|T} = y_T$$ a sensible model for the naive method.

<u>variance</u>

$$var(y_{T+h}) = var(y_T) + \sum_{i=0}^{h-1} var(\varepsilon_{T+h-i})$$

$$= 0 + (\underbrace{\sigma^2 + \sigma^2 + \dots + \sigma^2}_{h}) + cov(\quad)$$

$i =$ 0   1   . . . . . .   $h-1$

$T+h-i =$ $T+h$   $T+h-1$   . . .   $T+\cancel{h}-\cancel{h}+1$

$$\Rightarrow \quad var(\hat{y}_{T+h|T}) = h\sigma^2$$ (notice this depends on h)

HOMEWORK: derive mean & variance for the mean model

$y_t = \mu + \varepsilon_t \qquad \varepsilon_t \sim \text{iid } N(0, \sigma^2)$

$$\boxed{\begin{array}{c}\text{RECALL mean method} \\ \hat{y}_{T+h} = \bar{y}\end{array}}$$

At $T+h$ $\quad y_{T+h} = \mu + \varepsilon_{T+h}$

mean $\quad E(y_{T+h}) = \mu + \cancel{E(\varepsilon_{T+h})}^{0} = \mu \quad \Rightarrow \quad \hat{y}_{T+h|T} = \mu$

which we estimate using the sample mean: $\quad \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t = \bar{y}$ $\quad$ a sensible model for the mean method

variance $\quad \mathrm{var}(y_{T+h}) = \mathrm{var}(\mu) + \mathrm{var}(\varepsilon_{T+h}) = \mathrm{var}(\bar{y}) + \mathrm{var}(\varepsilon_{T+h})$

$$\mathrm{var}(\bar{y}) = \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{var}(y_t) = \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{var}(\mu+\varepsilon_t) = \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{var}(\varepsilon_t) = \frac{1}{T^2}T\sigma^2 = \frac{\sigma^2}{T}$$

$$\Rightarrow \quad \mathrm{var}(y_{T+h}) = \frac{\sigma^2}{T} + \sigma^2 = \sigma^2\left(1 + 1/T\right)$$ (this does not depend on $h$)

# Prediction intervals

- Assuming forecast errors are normally distributed, then a 95% PI is

$$\hat{y}_{T+h|T} \pm 1.96\hat{\sigma}_h$$

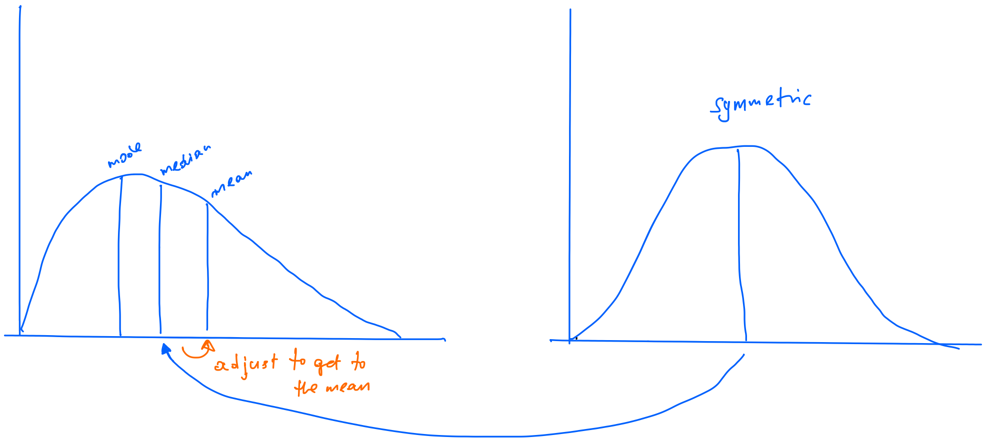
  where $\hat{\sigma}_h$ is the st dev of the $h$-step distribution.
- When $h = 1$, $\hat{\sigma}_h$ can be estimated from the residuals.
- Point forecasts often useless without a measure of uncertainty (such as prediction intervals).
- Prediction intervals require a stochastic model (with random errors, etc).
- Usually too narrow due to unaccounted uncertainty.

Sources of uncertainty:

1. model errors
2. estimation uncertainty for both model & parameters
3. model choice

Seminar Activity 5

# Outline


mode
median
mean
adjust to get to
the mean
symmetric

# Outline

# Decomposition models

`decomposition_model()` creates a decomposition model

$\hat{A}_t$

- You must provide a method for forecasting the `season_adjust` series.

$\hat{S}_t$

- A seasonal naive method is used by default for the `seasonal` components.
- The variances from both the seasonally adjusted and seasonal forecasts are combined.

Seminar Activity 6

# Outline

# SUMMARY

Train: residual diagnostics

- in-sample
- one-step
- fitted values (or forecasts)

Test: forecast evaluation

- hold-out sample (out-of-sample)
- multi-step
- forecasts

# SUMMARY

Train: residual diagnostics

- in-sample
- one-step
- fitted values (or forecasts)

→

- evaluation versus model assumptions
  - uncorrelated
  - constant variance
  - mean zero
  - normal

  $\sim iid\ N(0, \sigma^2)$

Test: forecast evaluation

- hold-out sample (out-of-sample)
- multi-step
- forecasts

# SUMMARY

Train: residual diagnostics

- in-sample
- one-step
- fitted values (or forecasts)

Test: forecast evaluation

- hold-out sample (out-of-sample)
- multi-step
- forecasts

- ME (bias)
- MAE, MSE → RMSE
- MAPE
- RMSSE, MASE

Seminar Activity 7-8

# Outline

# Creating the rolling training sets

There are three main rolling types which can be used.

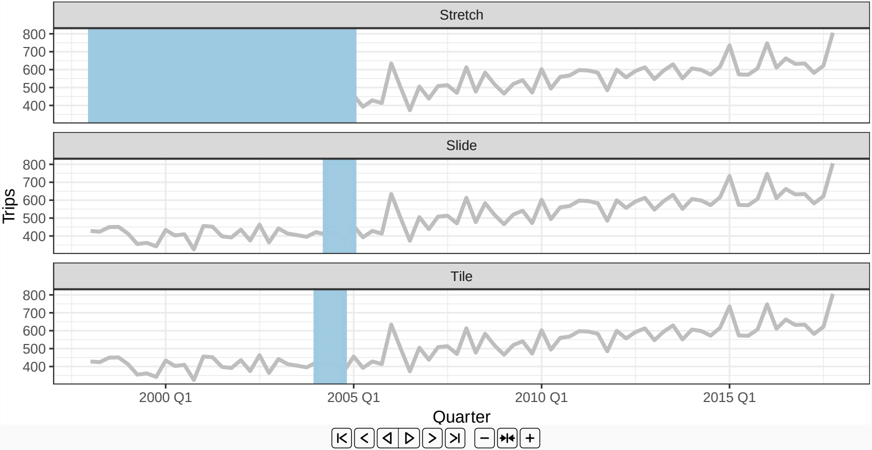
- Stretch: extends a growing length window with new data.
- Slide: shifts a fixed length window through the data.
- Tile: moves a fixed length window without overlap.

Three functions to roll a tsibble: `stretch_tsibble()`, `slide_tsibble()`, and `tile_tsibble()`.

For time series cross-validation, stretching windows are most commonly used.

Disadvantage: computationally intensive / time committment

# Creating the rolling training sets




Seminar Activity 9