

ETF3231/5231: Business forecasting

Week 4: The forecasters' toolbox

<https://bf.numbat.space/>



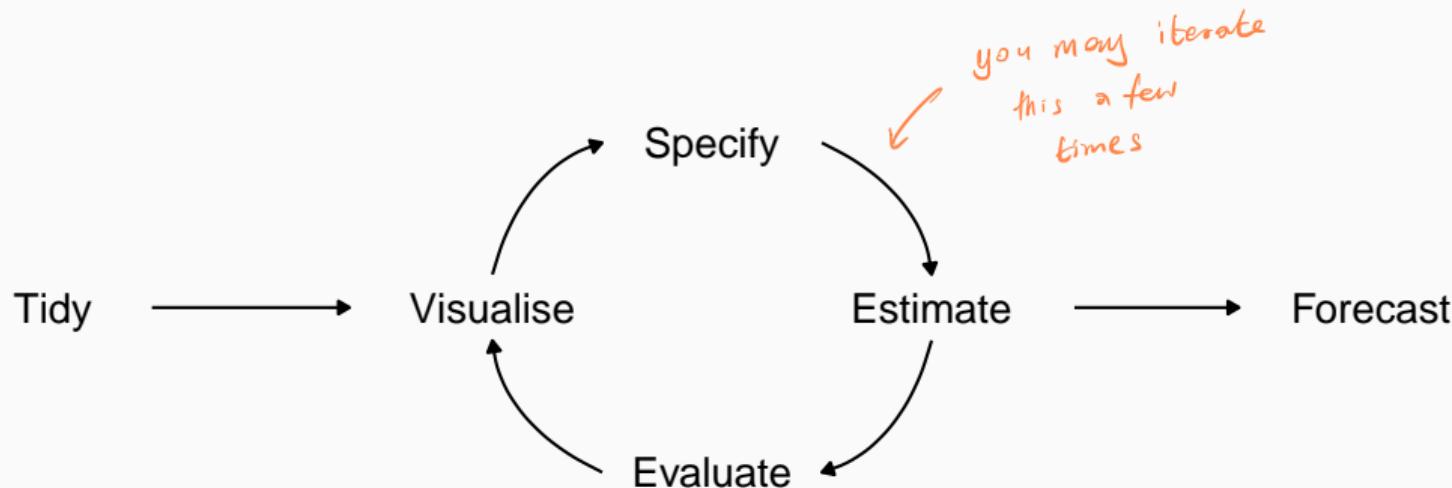
Outline

- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation

Outline

- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation

A tidy forecasting workflow



R: Demo GDP example

Outline

- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation

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

Model fitting

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

SNAIVE($y \sim \text{lag}(m)$): 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($y \sim \text{lag}(m)$): 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 ($m=4$)

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

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



SNAIVE($y \sim \text{lag}(m)$): 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 $(h - 1)/m$.

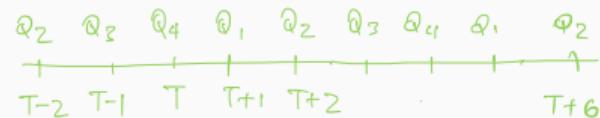
Example : quarterly data ($m=4$)

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

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

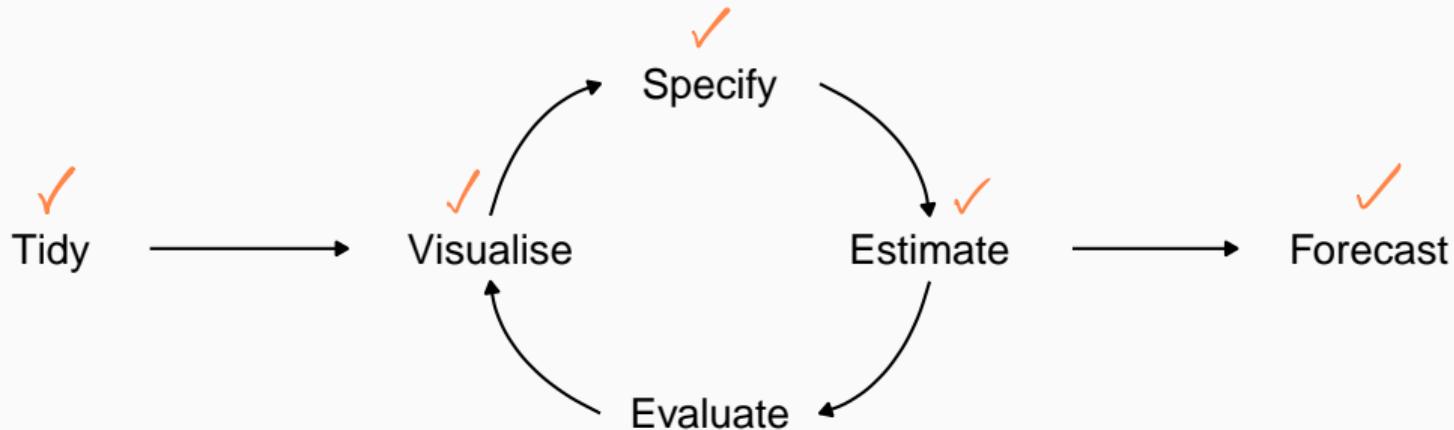
$$h=6 : (h-1)/m = (6-1)/4 = 1.25 \Rightarrow k=1$$

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



A tidy forecasting workflow - Recap

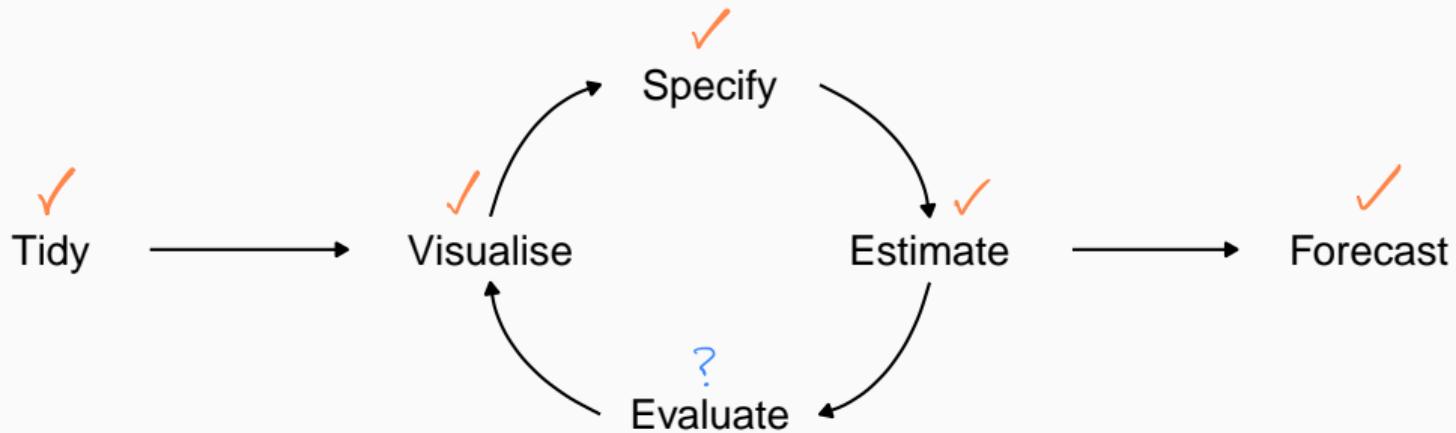
Complete Seminar Activities 1-3 before going through the slide



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

A tidy forecasting workflow - Recap

Complete Seminar Activities 1-2 before going through the slide



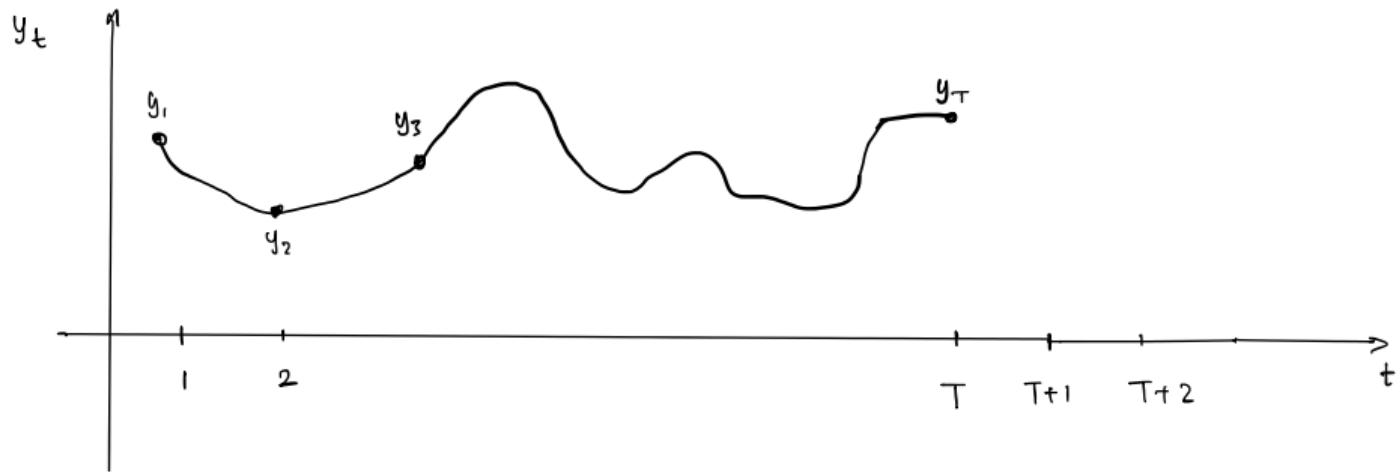
- Introduced four simple methods/benchmarks (mean, naive, seas naive, drift)
- Evaluate :
 1. How well do these fit the data
 2. How well do they actually forecast

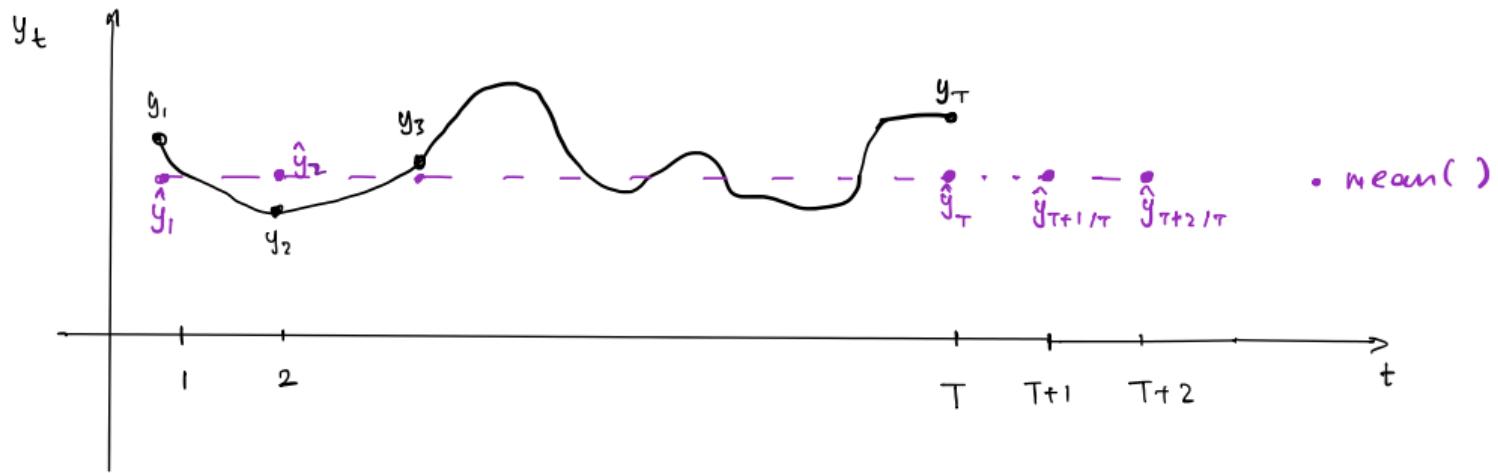
Outline

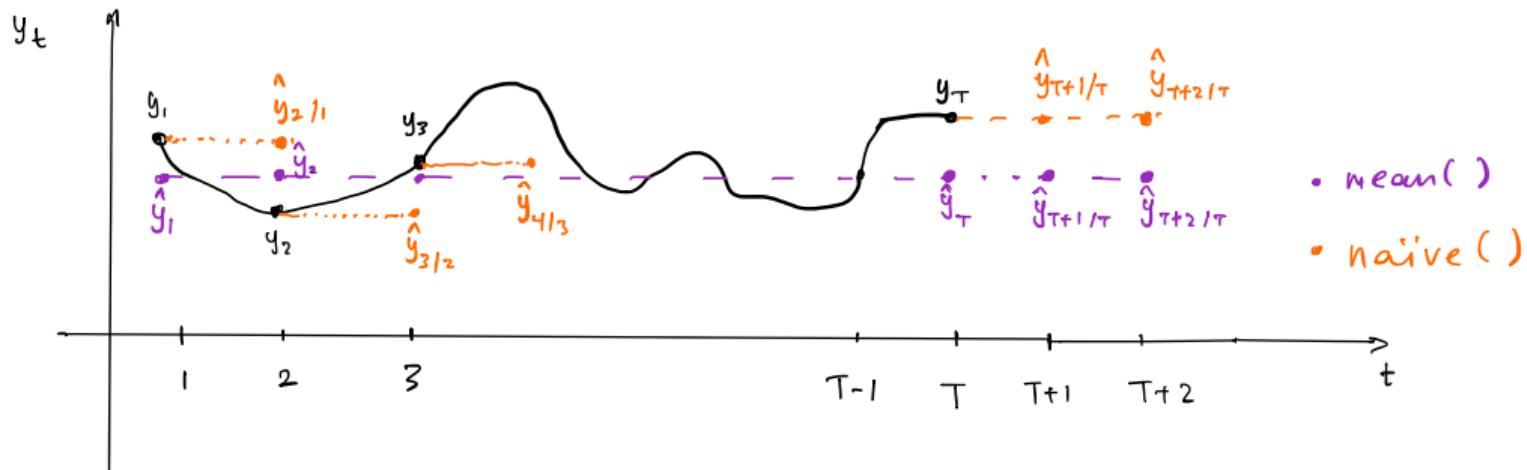
- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation

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.

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.
- We **expect** these to look like white noise.

$$H_0 : \rho_\epsilon = 0$$

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

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

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

$k = \text{no. of parameters in the model}$
 $\ell \leq k$

where ℓ 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.

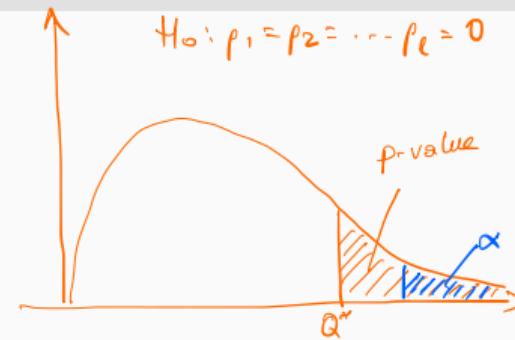
Portmanteau tests

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

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

```
## # 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 $Q^* > \alpha$, i.e. WN

Outline

- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation

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)

NAIVE METHOD : $\hat{y}_{T+h|\tau} = y_\tau$

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

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

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

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

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

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

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

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

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

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

Question: What would the RW model look like at T th? Let's do it sequentially.

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

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

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

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

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

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

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

}

$$y_{T+h} = y_T + \sum_{i=0}^h \varepsilon_{T+h-i} \quad \text{So } y_T + \text{all errors between } y_T + y_{T+h}$$

mean

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

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

Variance

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

$$= 0 + (\sigma^2 + \sigma^2 + \dots + \sigma^2) + \text{cov}(\quad)$$

$\underbrace{\vphantom{\sum_{i=0}^{h-1}}_{\substack{i=0 \\ T+h-i=T+h \\ \vdots \\ T+h-h+1}}_h}_{\substack{1 \\ \dots \\ h-1}}$

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

HOMEWORK: derive mean & variance for the mean model

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

RECALL mean method
 $\hat{y}_{t+h} = \bar{y}$

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

mean $E(y_{T+h}) = \mu + E(\varepsilon_{T+h}) = \mu \Rightarrow \hat{y}_{T+h/T} = \mu$

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

Variance $\text{var}(y_{T+h}) = \text{var}(\mu) + \text{var}(\varepsilon_{T+h}) = \text{var}(\bar{y}) + \text{var}(\varepsilon_{T+h})$

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

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

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. 

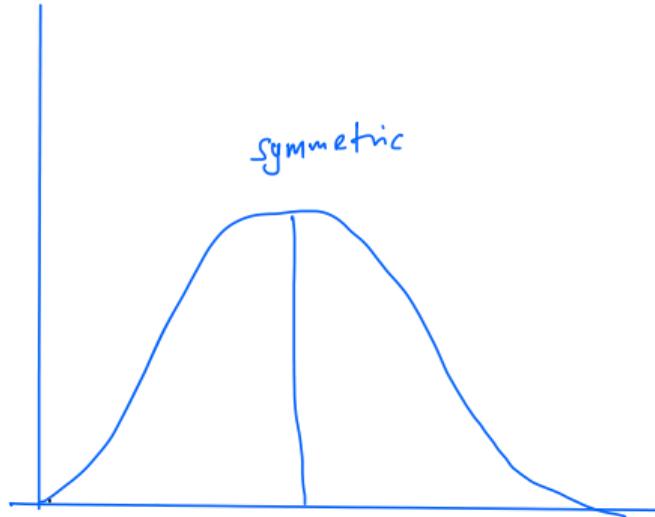
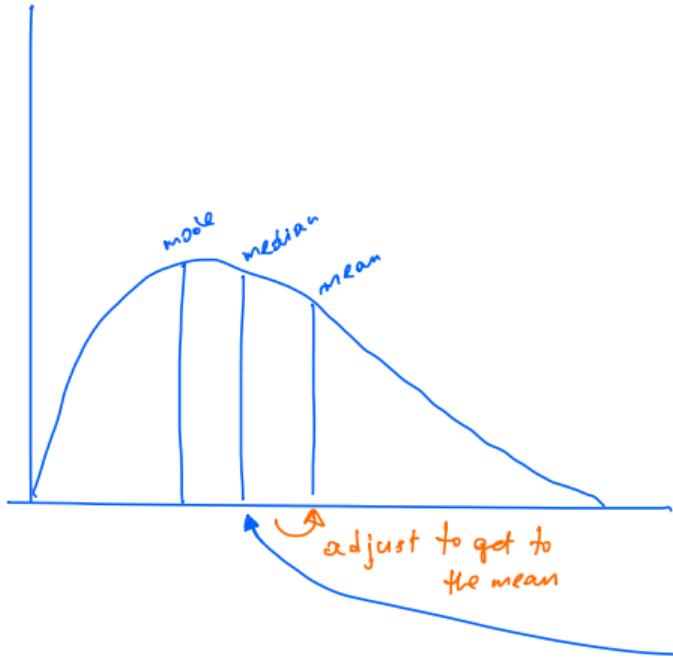
Sources of uncertainty :

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

Seminar Activity 5

Outline

- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation



Outline

- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation

Decomposition models

`decomposition_model()` creates a decomposition model

A
 A_t

- You must provide a method for forecasting the `season_adjust` series.
- A seasonal naive method is used by default for the `seasonal` components.
- The variances from both the seasonally adjusted and seasonal forecasts are combined.

A
 S_t

Outline

- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation

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
 - mean zero
 - constant variance
 - normal
- $\sim \text{iid } N(0, \sigma^2)$

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

Outline

- 1 A tidy forecasting workflow
- 2 Some simple forecasting methods
- 3 Residual diagnostics
- 4 Distributional forecasts and prediction intervals
- 5 Forecasting with transformations
- 6 Forecasting and decomposition
- 7 Evaluating forecast accuracy
- 8 Time series cross-validation

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 commitment

Creating the rolling training sets

Seminar Activity 9