

Forecasting in R

Exponential smoothing in ETS form



Outline




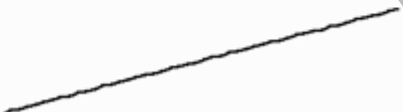
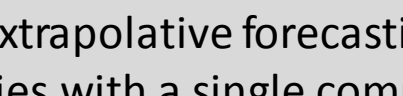

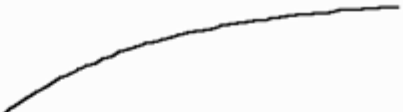



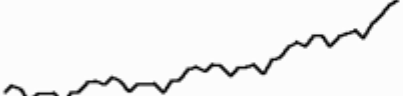


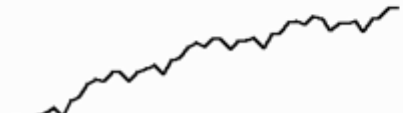

1. Forecasting level series;
2. Simple Exponential Smoothing;
3. Introduction to ETS;
4. Local level model;
5. Trend and seasonal models;
6. Model estimation and selection.

Outline

1. Forecasting level series;
2. Simple Exponential Smoothing;
3. Introduction to ETS;
4. Local level model;
5. Trend and seasonal models;
6. Model estimation and selection.

Introduction to ETS

- Different types of time series:

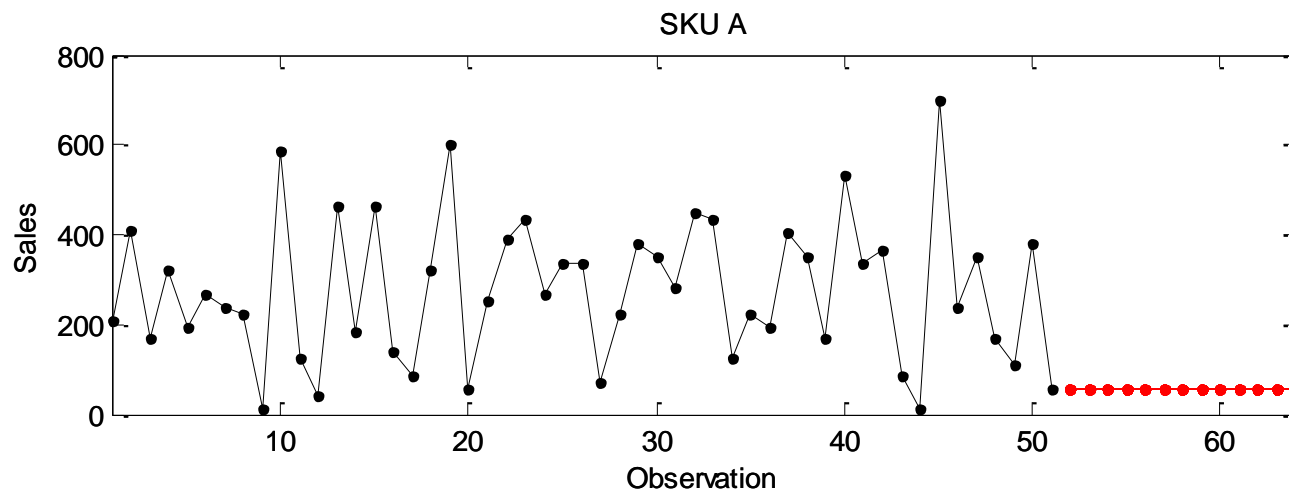
Trend	Seasonality		
	None	Additive	Multiplicative
None			
Additive			
Additive Damped			
Multiplicative			
Multiplicative Damped			

Let us understand the principles of extrapolative forecasting with series with a single component

Naïve forecast

What is the simplest forecast you can think of for a time series?
For example: what will the temperature be like in your room after 5 minutes?

$$\hat{y}_{t+1} = y_t$$



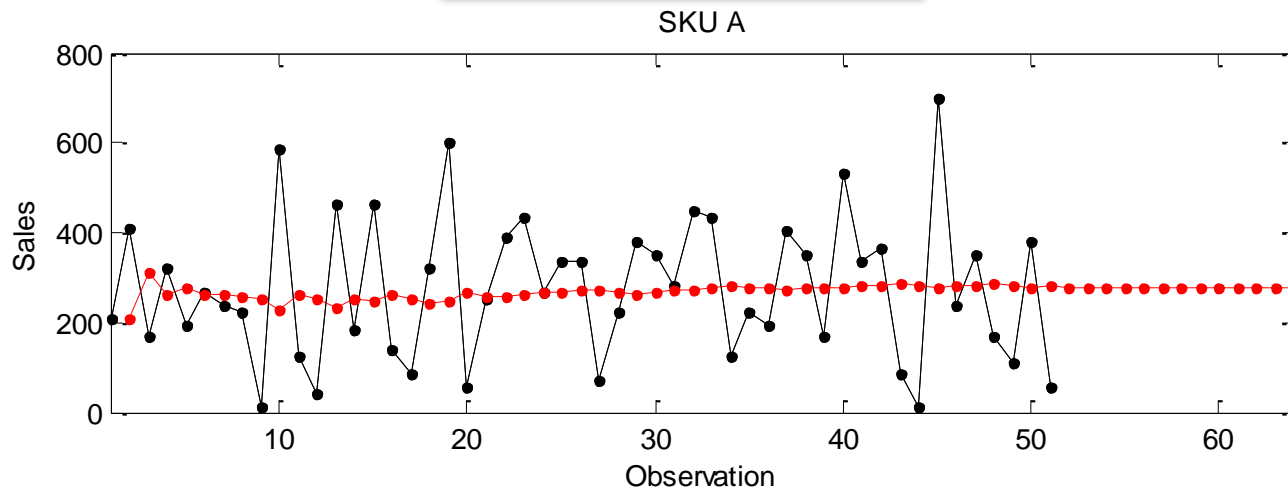
- The forecast is a straight line → always equal to the last observation.
- Is this a good forecast?

Arithmetic mean

Another approach would be to calculate the average and use this as a forecast.

For example: calculate the average temperature in your room over all the years you live there...

$$\hat{y}_{t+1} = \frac{1}{t} \sum_{i=1}^t y_i$$



- The average has long memory and the random movements of the noise will be cancelled out.
- Is this a good forecast?

Simple Moving Average

Simple Moving Average allows us to select the appropriate memory (length of the average).

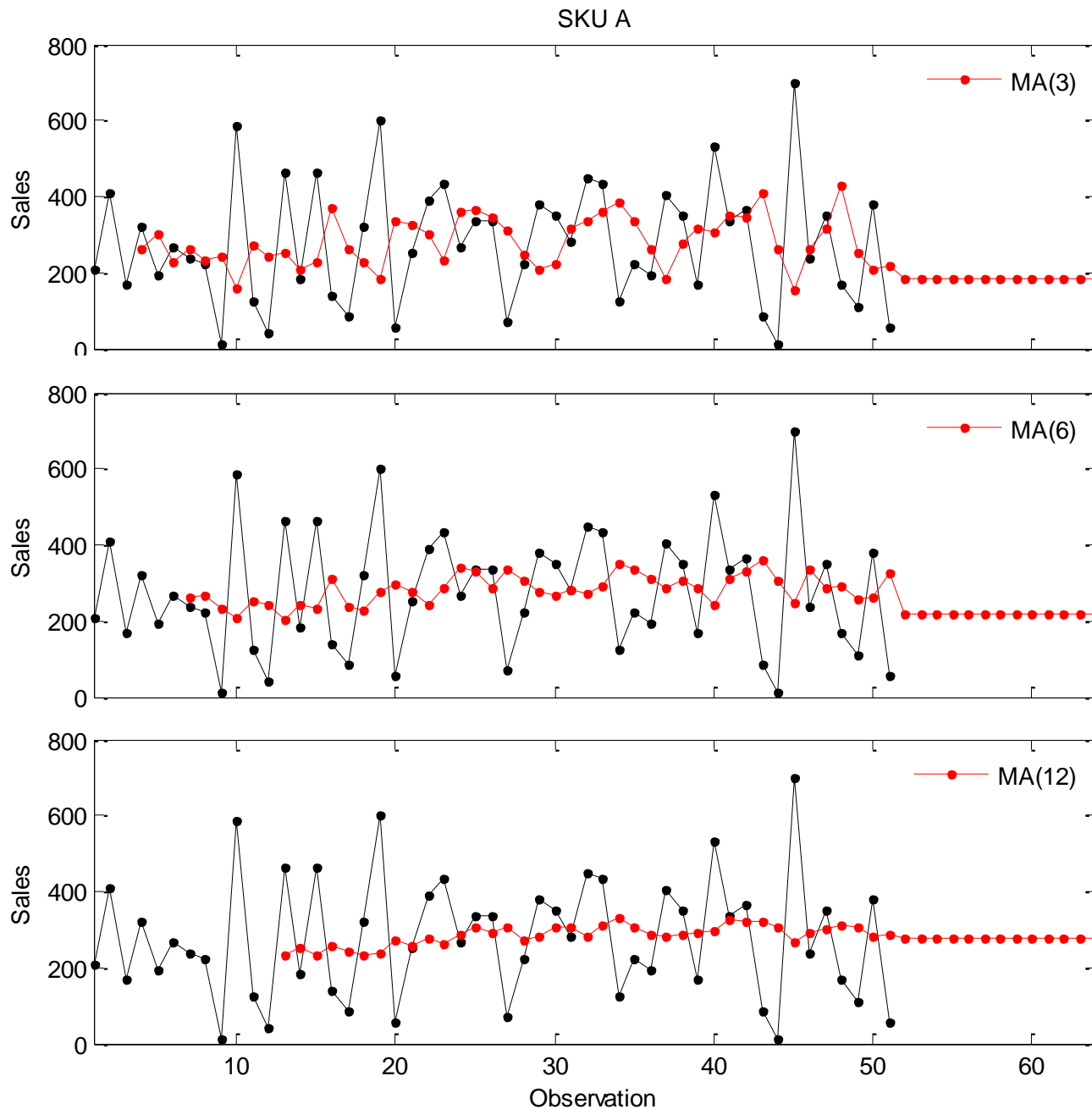
e.g. only consider the temperature over the last week

$$\hat{y}_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^t y_i$$

The **simple moving average**:

- Has a single parameter **k**. This controls the length of the moving average and it is also known as its order.
- Its variable length allows us to control how reactive we are to new information and how robust we are against noise.
- Gives equal importance to all **k** observations.

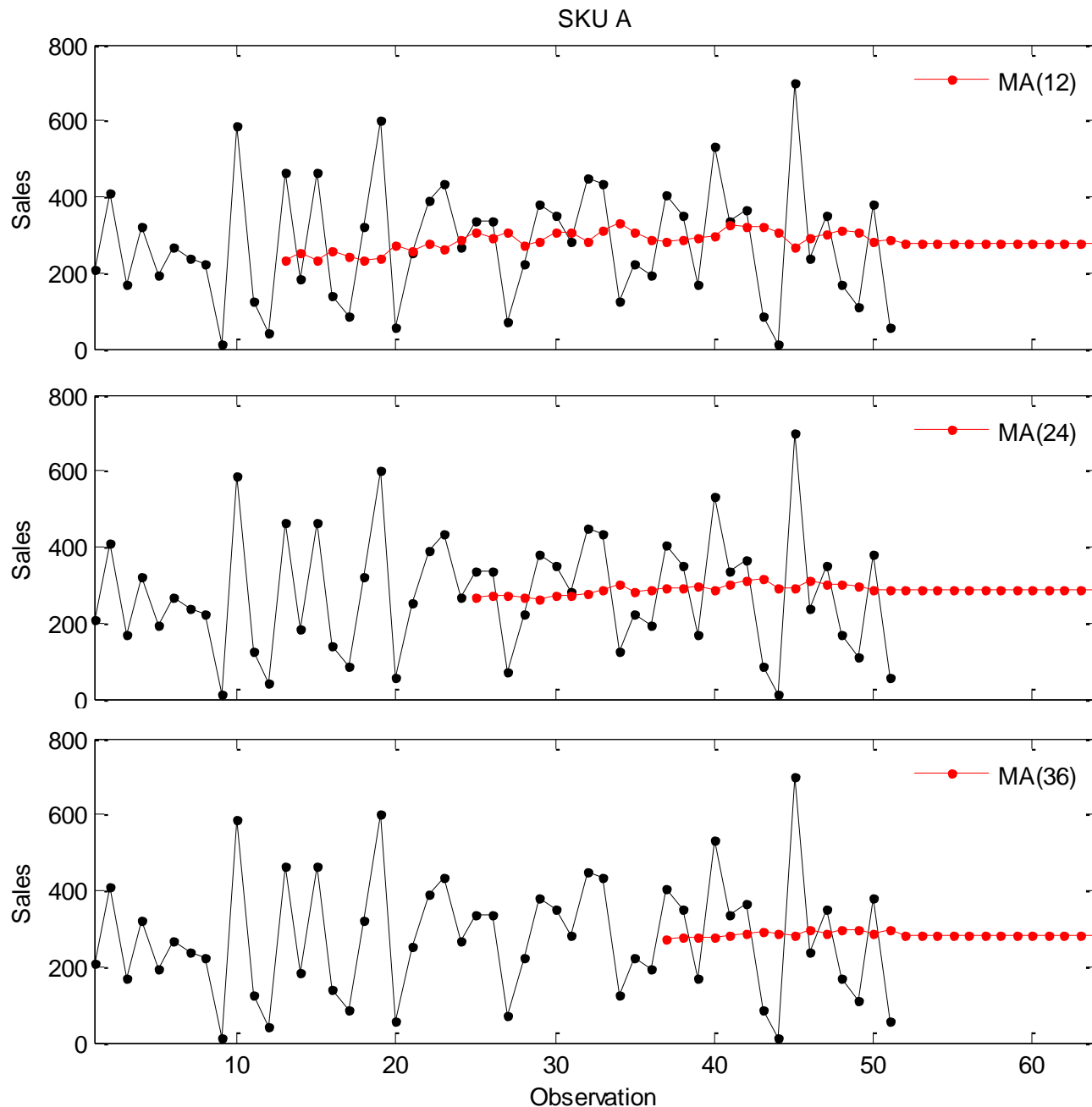
Simple Moving Average



Which of the different length moving averages is the most appropriate for this SKU?

We choose the one that gives us a smooth estimate of the level, here MA(12)

Simple Moving Average



Which of the different length moving averages is the most appropriate for this SKU?

We do not need excessive moving average lengths. These will be far too insensitive to new information.

Weighted Moving Average

Should the weights be the same for all k observations?

We can overcome this limitation by allowing different weights for each observation in the average:

$$\hat{y}_{t+1} = \sum_{i=t-k+1}^t w_i y_i, \quad \text{w. r. t. } \sum_{i=1}^k w_i = 1$$

With the **weighted moving average**:

- We can control the length of the average and the importance of each observation
- All weights must add up to 100% or 1. Normally the older the observation the smaller the weight.
- Has $k+1$ parameters, the length of the average and k weights.
- The number of weights makes it very challenging to use in practice.

Outline

1. Forecasting level series;
2. Simple Exponential Smoothing;
3. Introduction to ETS;
4. Local level model;
5. Trend and seasonal models;
6. Model estimation and selection.

The Exponential Smoothing Concept

Starting from the **weighted moving average** we can construct a heuristic to select the weights easily and consequently its order (k).

Data	y_t	y_{t-1}	y_{t-2}	y_{t-3}	...
Weights	w_t	w_{t-1}	w_{t-2}	w_{t-3}	...

1. Make the more recent information more relevant, bigger weights
2. Remember! Weights must add up to 100% (or 1)

→ Take 50% for the first and then always take 50% of the remaining weight. (Sum of all weights \approx 100%)

Weights	w_t	w_{t-1}	w_{t-2}	w_{t-3}	w_{t-4}	w_{t-5}	w_{t-6}
Weights	50%	25%	12.5%	6.25%	3.12%	1.56%	$\approx 0\%$

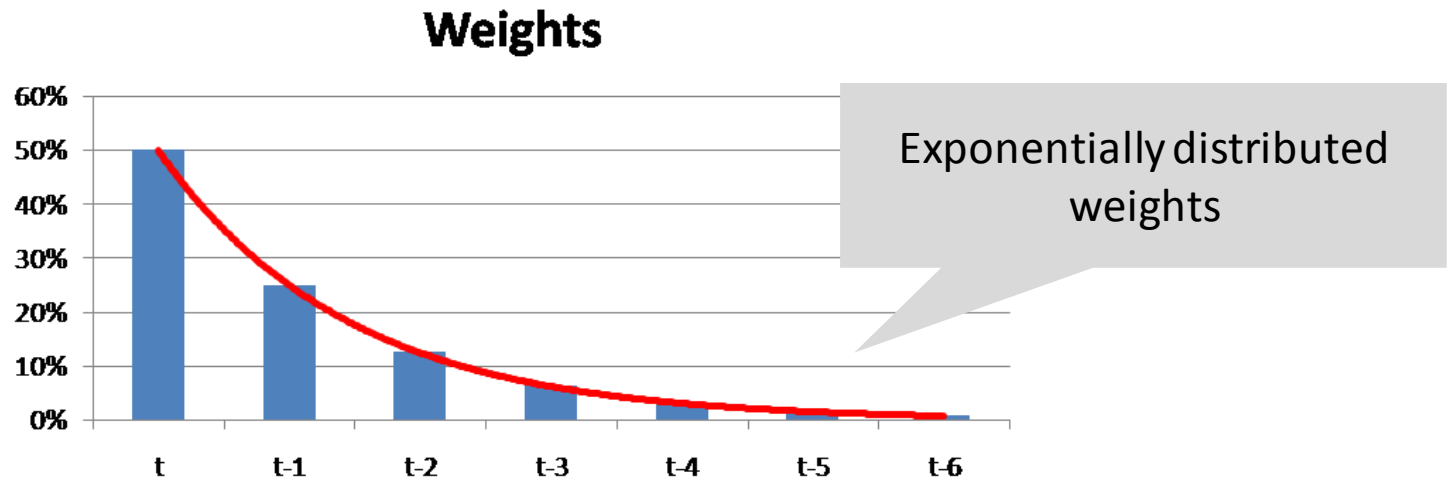
→ The length of the average is set automatically!

The Exponential Smoothing Concept

Weights	W_t	W_{t-1}	W_{t-2}	W_{t-3}	W_{t-4}	W_{t-5}	W_{t-6}
	50%	25%	12.5%	6.25%	3.12%	1.56%	$\approx 0\%$

Only one parameter, the initial weight! Let this weight be **Alpha (α)**...

$$\alpha(1-\alpha)^0 \quad \alpha(1-\alpha)^1 \quad \alpha(1-\alpha)^2 \quad \alpha(1-\alpha)^3 \quad \alpha(1-\alpha)^4 \quad \alpha(1-\alpha)^5 \quad \alpha(1-\alpha)^6$$



The exponential weighting scheme allows us to select reasonable weights and the length of the weighted moving average with a single parameter, the α .

The Exponential Smoothing Concept

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \alpha(1 - \alpha)^3 y_{t-3} + \dots$$

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \underbrace{(\alpha y_{t-1} + \alpha(1 - \alpha)y_{t-2} + \alpha(1 - \alpha)^2 y_{t-3} + \dots)}$$

What is this?

$$\hat{y}_t = \alpha y_{t-1} + \alpha(1 - \alpha)y_{t-2} + \alpha(1 - \alpha)^2 y_{t-3} + \dots$$

A simpler form of the model:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$$

Simple Exponential Smoothing

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t$$

The parameter α , is called **smoothing parameter** and is bounded between 0 and 1.

The exponential smoothing formula can be read as: the forecast is α times the most recent observation and $(1-\alpha)$ times all the previous information.

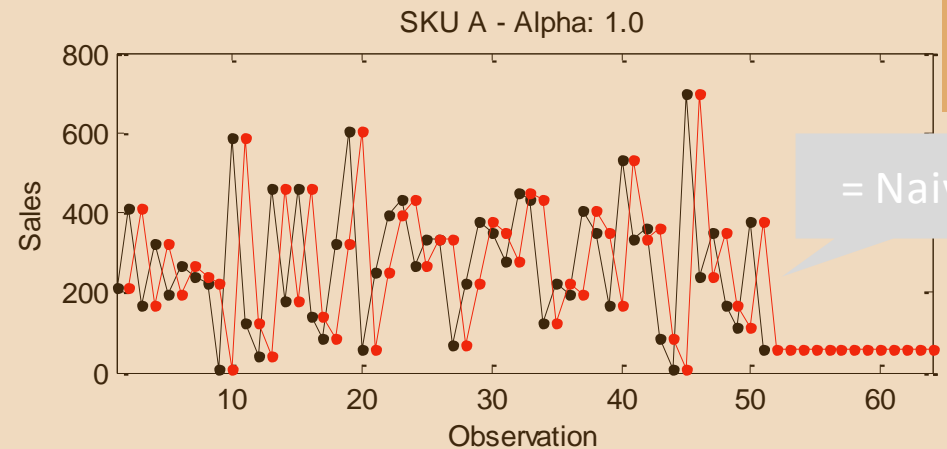
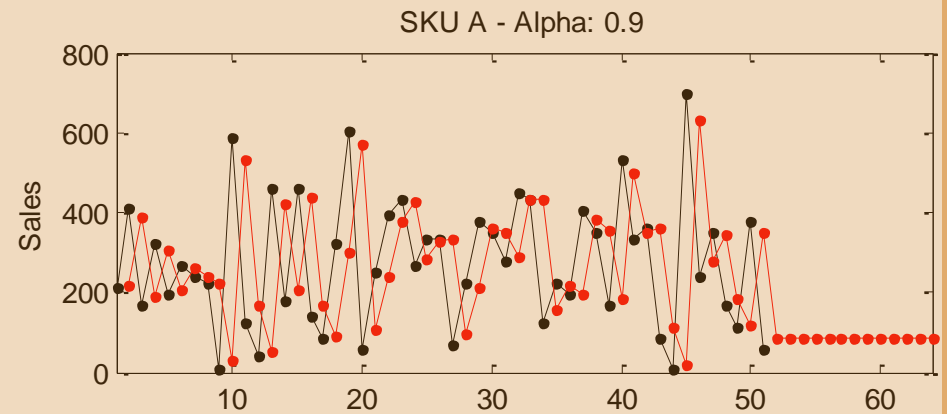
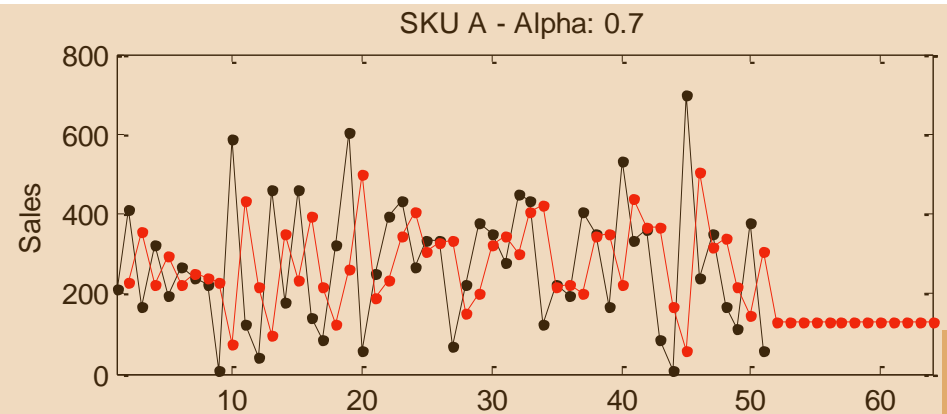
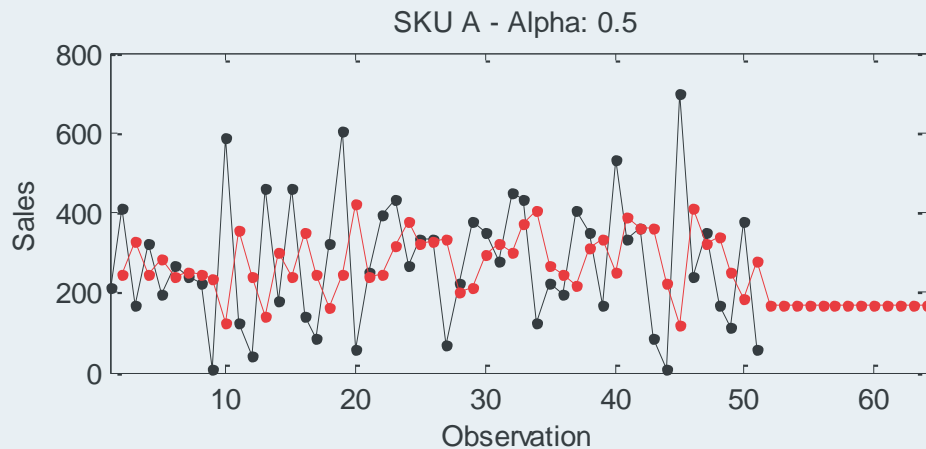
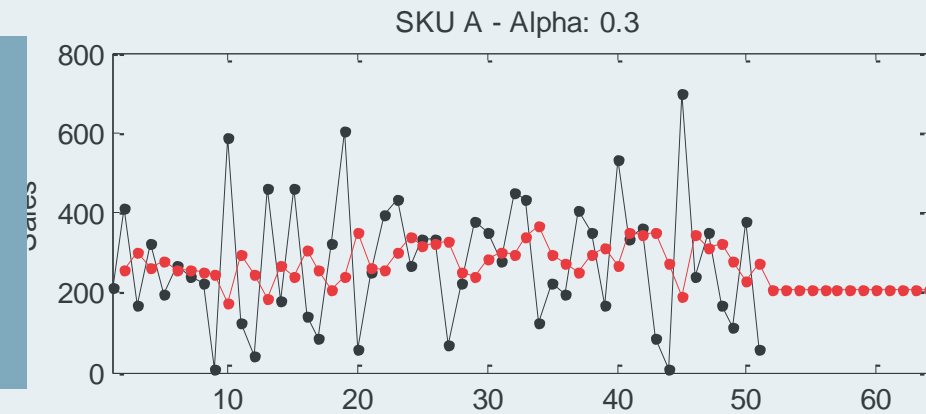
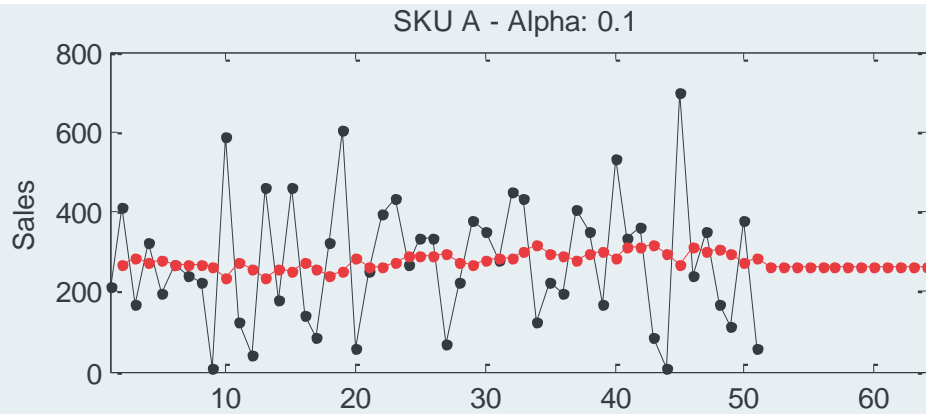
- A low α implies that the forecast is mostly based on the previous information
- A high α implies that the forecast is mostly based on the last information

Therefore the smoothing parameter α controls how reactive is the forecast to new information.

This form was proposed by Brown (1956).

Much has changed since then...

Simple Exponential Smoothing

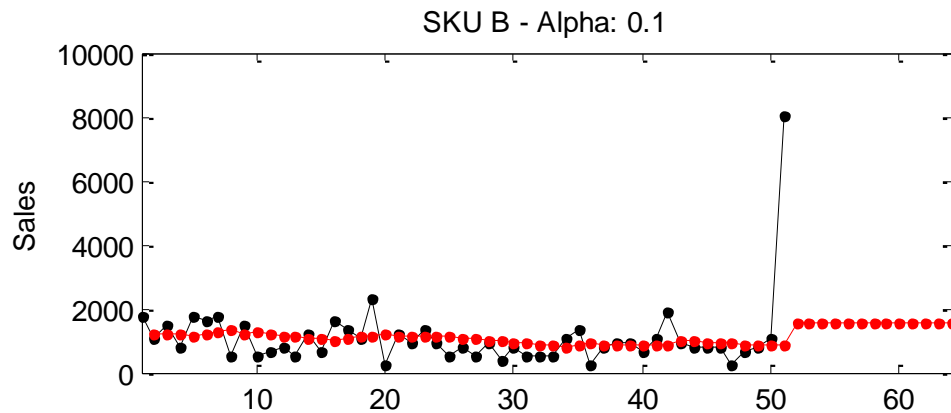


Noise is filtered

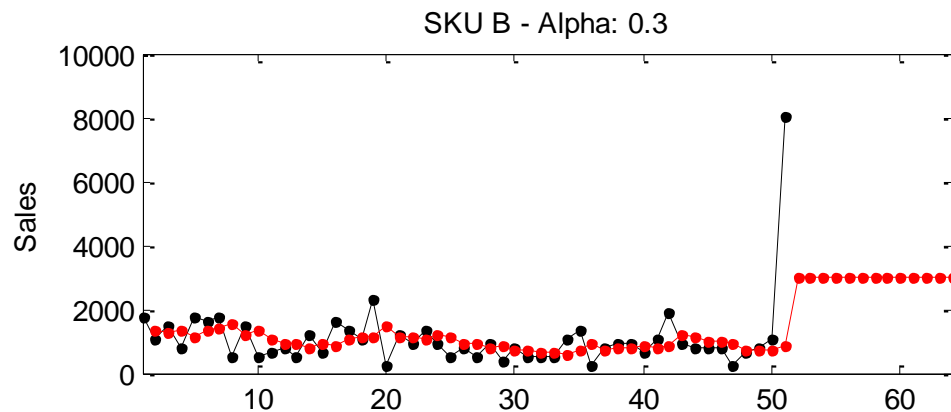
Noise is not filtered → Avoid

= Naive

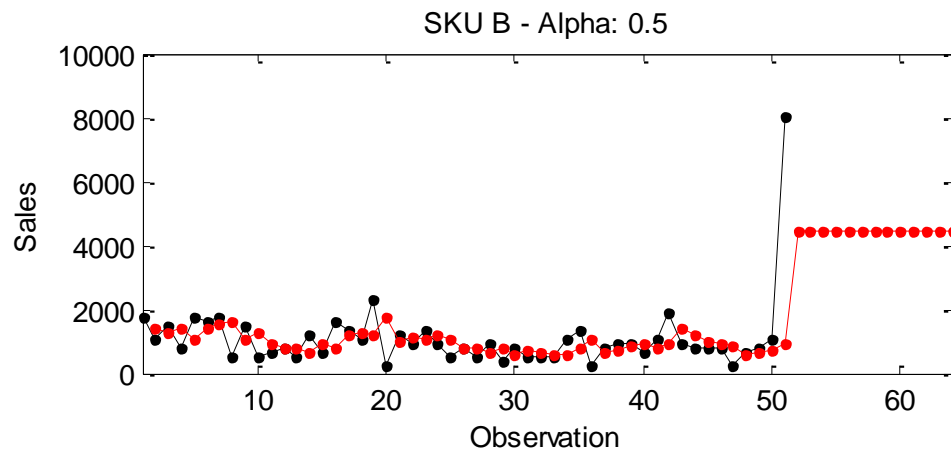
Simple Exponential Smoothing



In the presence of high noise or outliers we need to use low values of alpha to make our forecasts more robust.

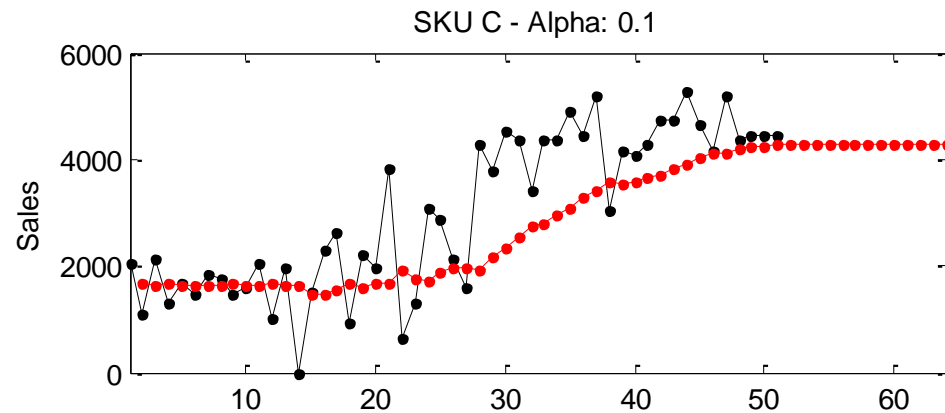


Here the outlier affects strongly our forecast.

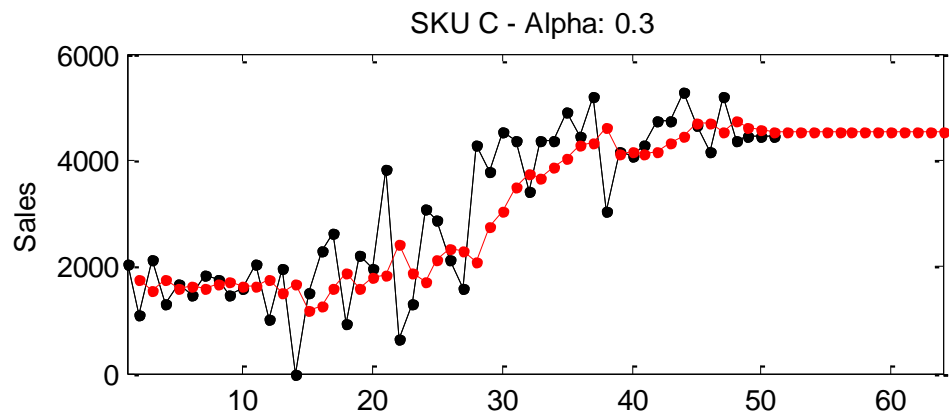


Here the effect of outlier is even stronger.

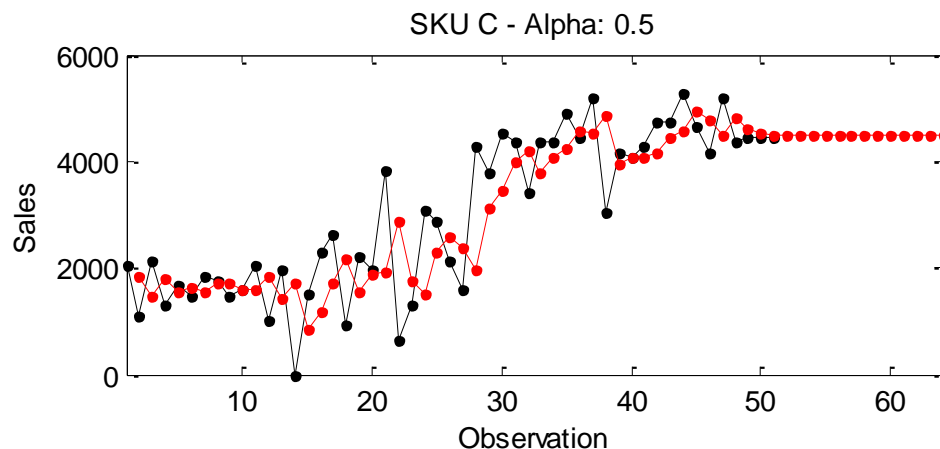
Simple Exponential Smoothing



Very low alpha parameter makes our forecast too slow to adjust to the new level of sales.



Here the alpha achieves a good compromise between reactivity and robustness to noise.



Very high alpha parameter makes our forecast to react very fast, but now it does not filter out noise adequately.

Simple Exponential Smoothing

We can formulate exponential smoothing in a different way:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$$

$$\hat{y}_{t+1} = \alpha y_t + \hat{y}_t - \alpha \hat{y}_t$$

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t)$$

The difference between the Actuals and the Forecast is the forecast error.

$$\hat{y}_{t+1} = \hat{y}_t + \alpha e_t$$

This is known as the **error correction form** of exponential smoothing.

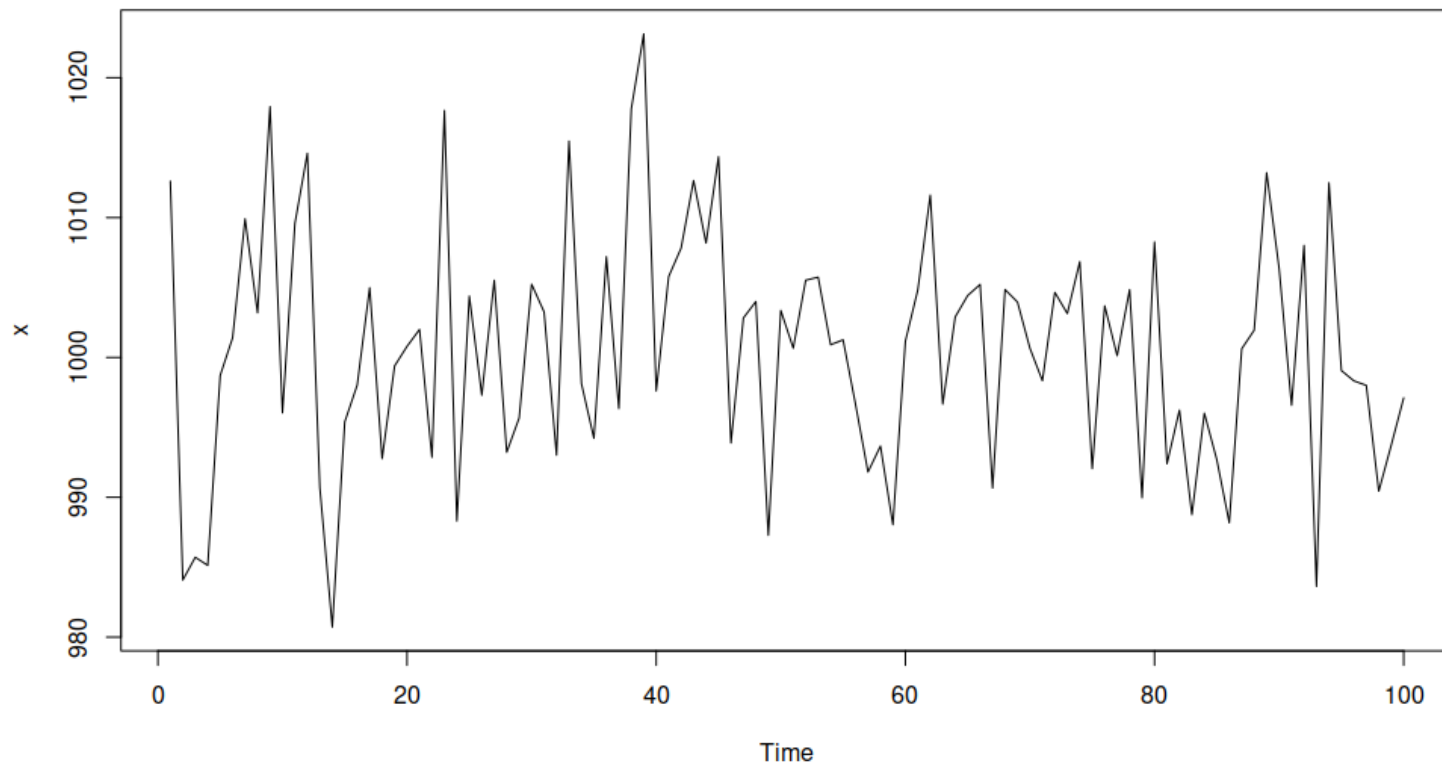
Why is this useful?

Let's find out after a short quiz...

Forecasting Level Series, Quiz

Please, follow the link: <http://etc.ch/V7Ss>

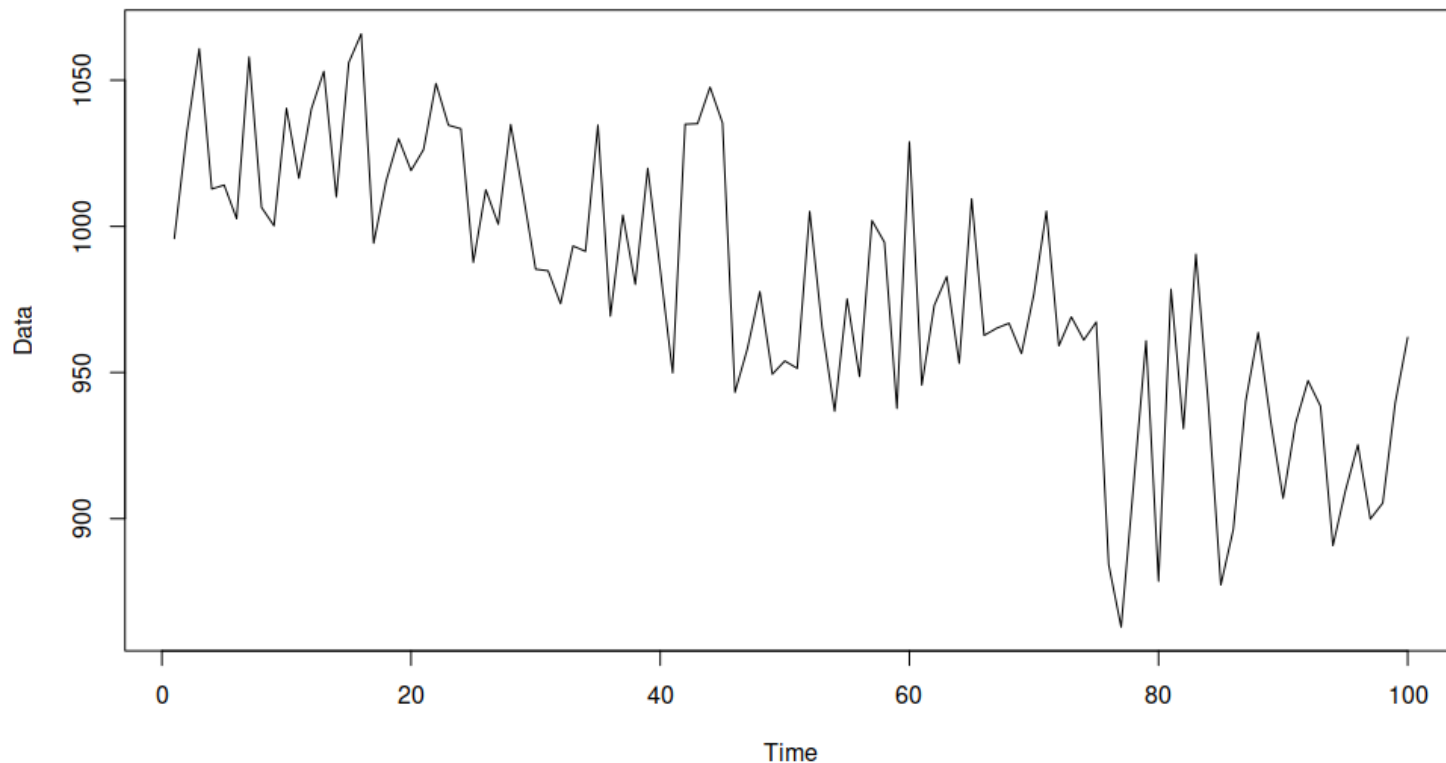
1. Which of the methods is more appropriate for the following data?



Forecasting Level Series, Quiz

Please, follow the link: <http://etc.ch/V7Ss>

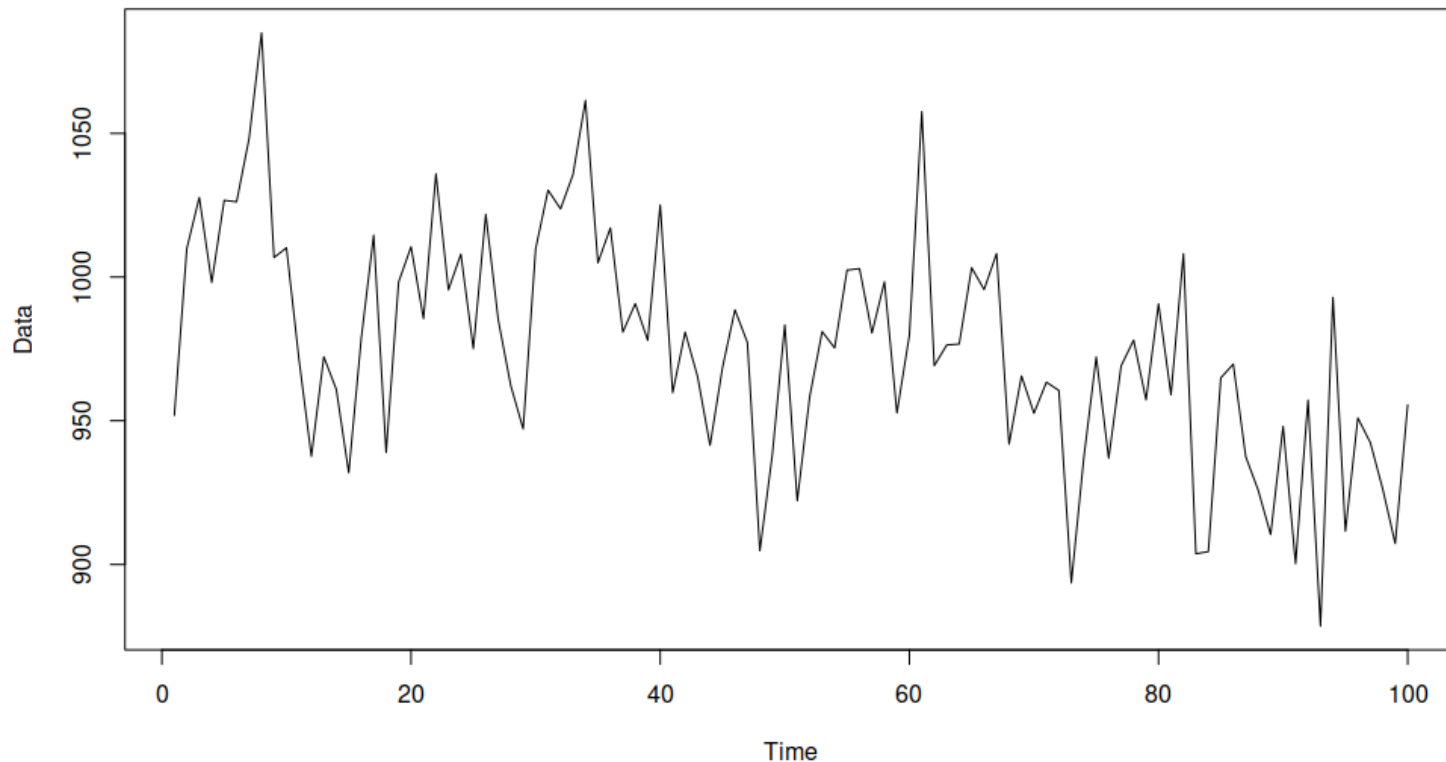
2. Which of the methods is more appropriate for the following data (2nd example)?



Forecasting Level Series, Quiz

Please, follow the link: <http://etc.ch/V7Ss>

3. Which of the smoothing parameters is more appropriate for this data if we use SES?



Outline

1. Forecasting level series;
2. Simple Exponential Smoothing;
3. Introduction to ETS;
4. Local level model;
5. Trend and seasonal models;
6. Model estimation and selection.

Introduction to ETS

SES models the level of a time series

So, we can write $\hat{y}_{t+1} = l_t$

By shifting the indices by 1 period we can now write:

$$y_t = l_{t-1} + e_t \quad (1)$$




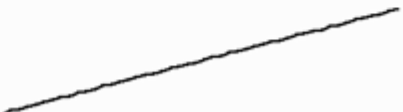


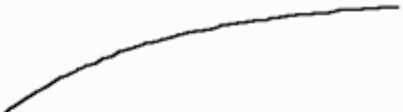



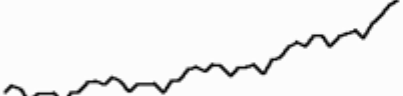


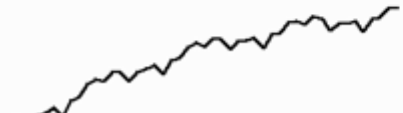

$$l_t = l_{t-1} + ae_t \quad (2)$$

This will lead us to the so called **State Space Models**:

- Eq. (1) – the **measurement equation**: says that the observed actuals are the result of some structure (l_t) and noise (e_t).
- Eq. (2) – the **transition equation**: says that there is an unobserved process describing how the level of the time series evolves. For our case this is all the structure of the series.
- We can have other components as well...

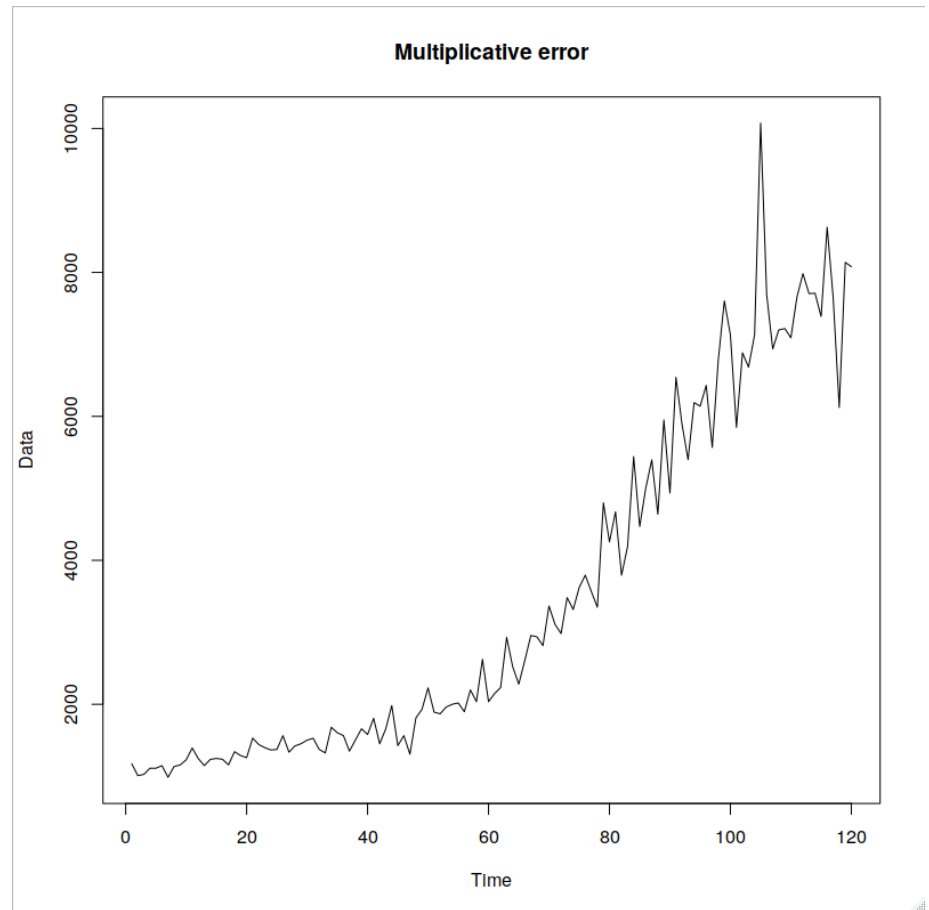
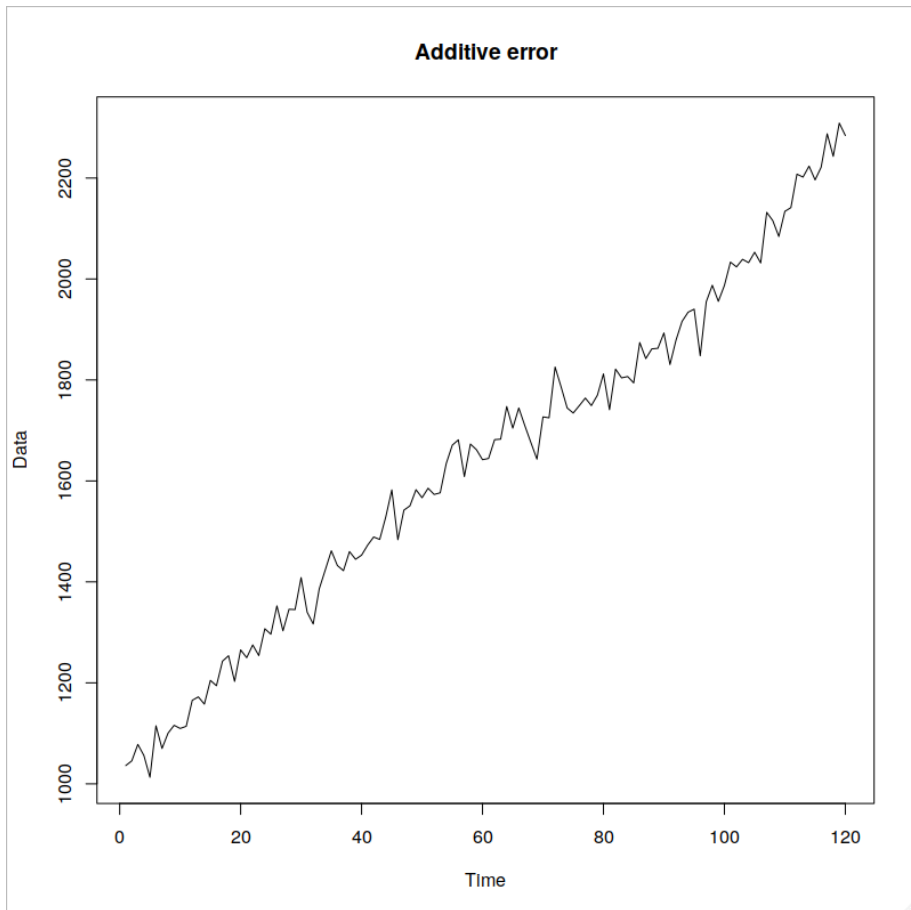
Introduction to ETS

- Different types of components:

Trend	Seasonality		
	None "N"	Additive "A"	Multiplicative "M"
None "N"			
Additive "A"			
Additive Damped "Ad"			
Multiplicative "M"			
Multiplicative Damped "Md"			

Introduction to ETS

- And two types of errors:



Introduction to ETS

- ETS taxonomy includes:
 - 2 types of errors,
 - 5 types of trends,
 - 3 types of seasonality.
- Which gives us 30 models:
 - 6 pure additive models,
 - 6 pure multiplicative models,
 - 18 mixed models.

Introduction to ETS

- Based on the time series decomposition we can have the pure additive model:

$$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$$

- And for the pure multiplicative one:

$$y_t = l_{t-1} b_{t-1} s_{t-m} \varepsilon_t$$

- And there are combinations between the two.

- For example, an ETS(M,A,M) model:

$$y_t = (l_{t-1} + b_{t-1}) s_{t-m} \varepsilon_t$$

Introduction to ETS

- All pure models make sense:
 - Additive assume that the variables can be positive, negative or zero;
 - Multiplicative ones assume that the response variable can only be positive.
- Not all mixed models are reasonable
 - For example, ETS(A,M,A) model:
$$y_t = l_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$$
 - Why?
- You can fit them and produce forecasts, but they break easily.

Introduction to ETS

- The list of reasonable ETS models:
 - Additive error ($\varepsilon_t = \epsilon_t$):

		Seasonal		
		"N"	A	M
Trend	"N"	$y_t = l_{t-1} + \epsilon_t$	$y_t = l_{t-1} + s_{t-m} + \epsilon_t$	-
	"A"	$y_t = l_{t-1} + b_{t-1} + \epsilon_t$	$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t$	-
	"Ad"	$y_t = l_{t-1} + \phi b_{t-1} + \epsilon_t$	$y_t = l_{t-1} + \phi b_{t-1} + s_{t-m} + \epsilon_t$	-
	"M"	-	-	-
	"Md"	-	-	-

- It is usually assumed that $\epsilon_t \sim N(0, \sigma^2)$

Introduction to ETS

- The list of reasonable ETS models:
 - Multiplicative error ($\varepsilon_t = 1 + \epsilon_t$):

		Seasonal		
		"N"	A	M
Trend	"N"	$y_t = l_{t-1}(1 + \epsilon_t)$	$y_t = (l_{t-1} + s_{t-m})(1 + \epsilon_t)$	$y_t = l_{t-1}s_{t-m}(1 + \epsilon_t)$
	"A"	$y_t = (l_{t-1} + b_{t-1})(1 + \epsilon_t)$	$y_t = (l_{t-1} + b_{t-1} + s_{t-m})(1 + \epsilon_t)$	$y_t = (l_{t-1} + b_{t-1})s_{t-m}(1 + \epsilon_t)$
	"Ad"	$y_t = (l_{t-1} + \phi b_{t-1})(1 + \epsilon_t)$	$y_t = (l_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \epsilon_t)$	$y_t = (l_{t-1} + \phi b_{t-1})s_{t-m}(1 + \epsilon_t)$
	"M"	$y_t = l_{t-1}b_{t-1}(1 + \epsilon_t)$	-	$y_t = l_{t-1}b_{t-1}s_{t-m}(1 + \epsilon_t)$
	"Md"	$y_t = l_{t-1}b_{t-1}^\phi(1 + \epsilon_t)$	-	$y_t = l_{t-1}b_{t-1}^\phi s_{t-m}(1 + \epsilon_t)$

- Usual assumption is $\epsilon_t \sim N(0, \sigma^2)$, but in smooth it is $1 + \epsilon_t \sim \text{log } N(0, \sigma^2)$

Introduction to ETS

- So far, we've discussed only one part of ETS model.
- It is called “measurement equation” and it shows how the data is formed.
 - For example, with local level model: $y_t = l_{t-1} + \epsilon_t$
- But level, trend and seasonal components might change over time.
- So, there should be a mechanism for update of states.

Introduction to ETS

- **Transition equation** – the equation that shows how the components change over time.
- For example, for ETS(A,N,N):

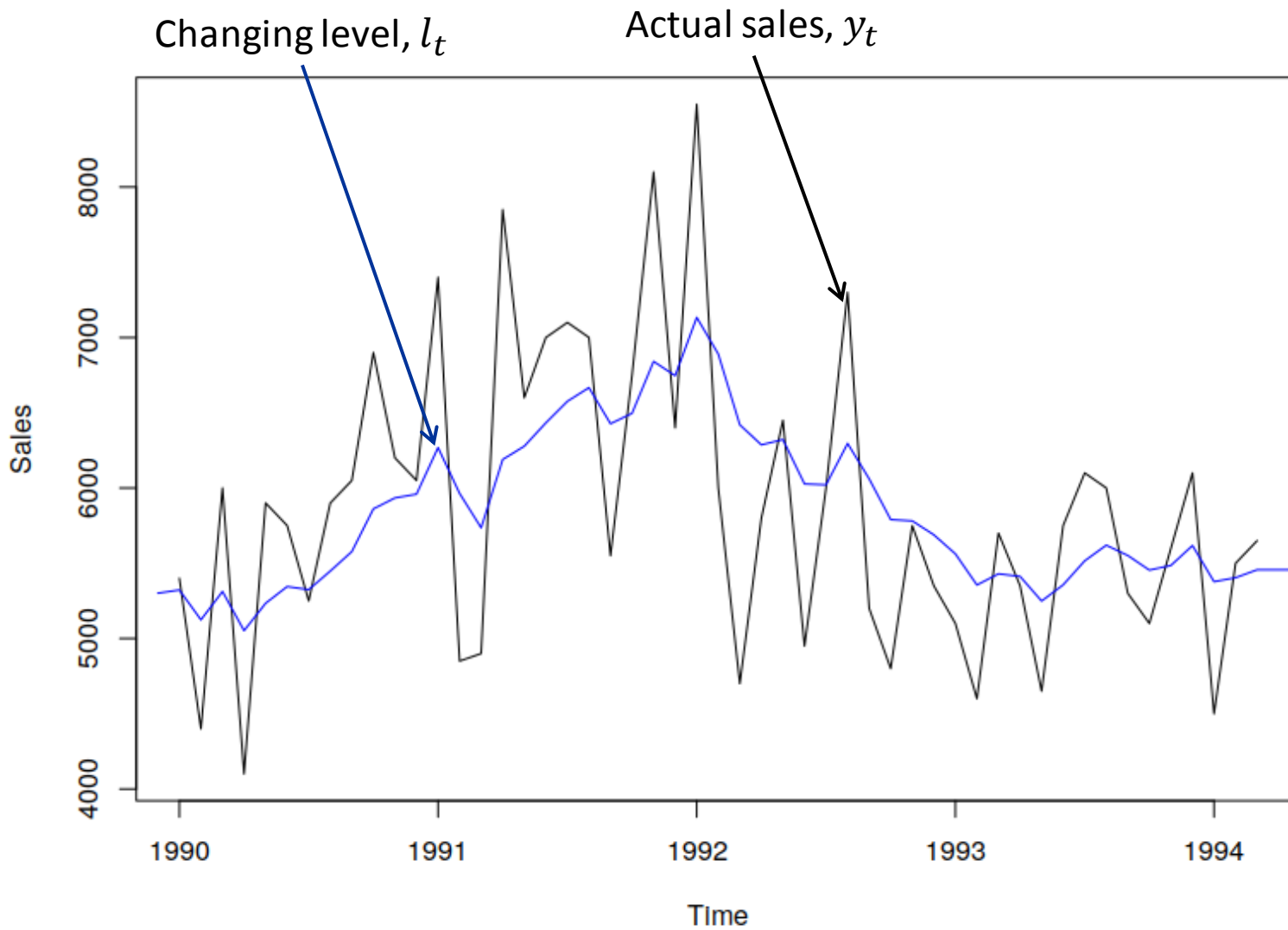
$$l_t = l_{t-1} + \alpha\epsilon_t$$

- Any ETS model consists of these two parts.
- So, ETS(A,N,N) can be represented as:

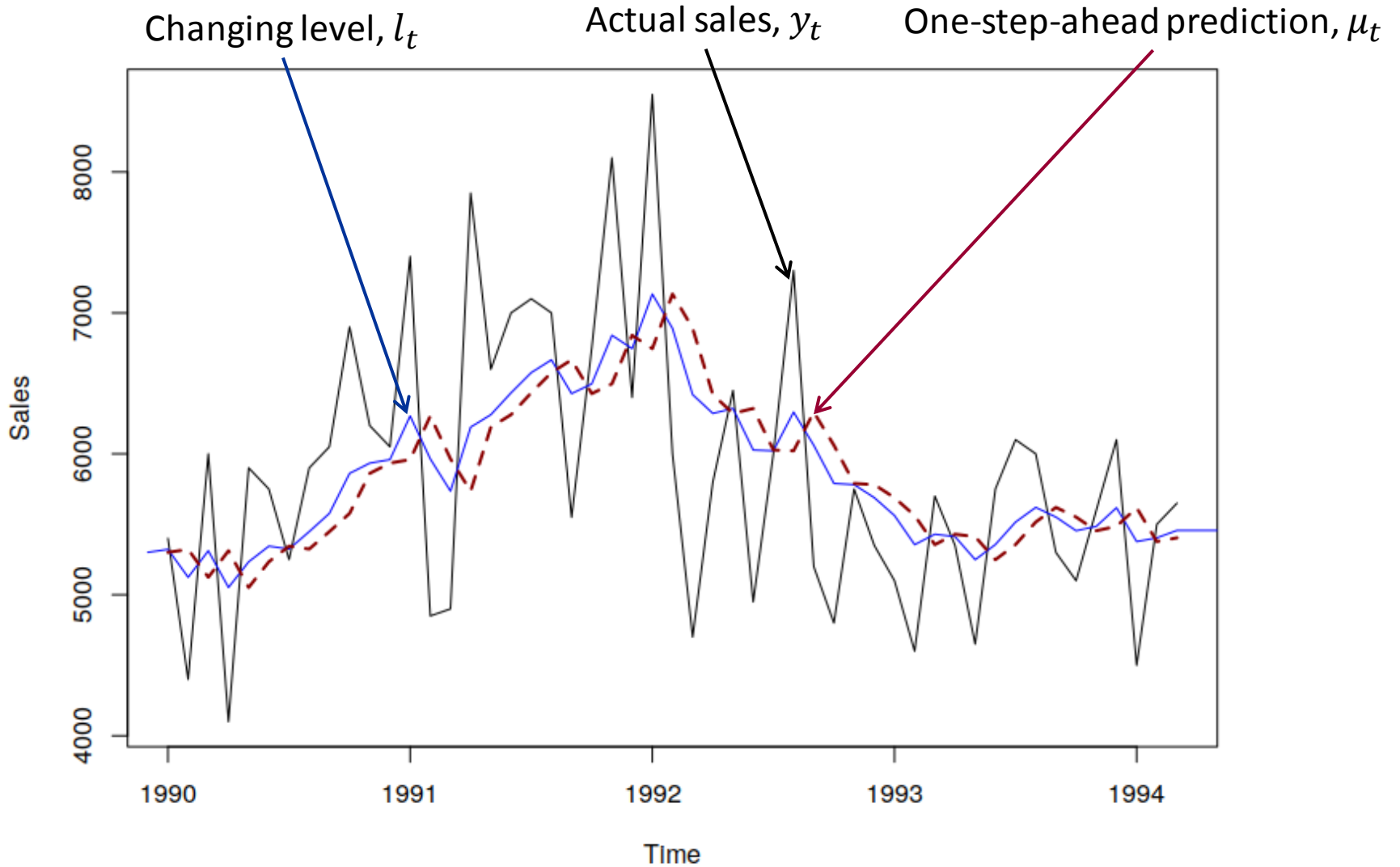
$$y_t = l_{t-1} + \epsilon_t$$

$$l_t = l_{t-1} + \alpha\epsilon_t$$

Introduction to ETS



Introduction to ETS



Introduction to ETS

- In general pure additive model can be summarised as:

$$y_t = \mathbf{w}'\mathbf{v}_{t-1} + \epsilon_t$$
$$\mathbf{v}_t = \mathbf{F}\mathbf{v}_{t-1} + \mathbf{g}\epsilon_t$$

- \mathbf{g} is the persistence vector... The rest is not important.
- See Hyndman et al. (2008) for details.
- Additional resources:
 - For pure additive models: <http://tiny.cc/znxc9y>
 - For pure multiplicative models: <http://tiny.cc/2oxc9y>
 - For the mixed ones: <http://tiny.cc/emxc9y>

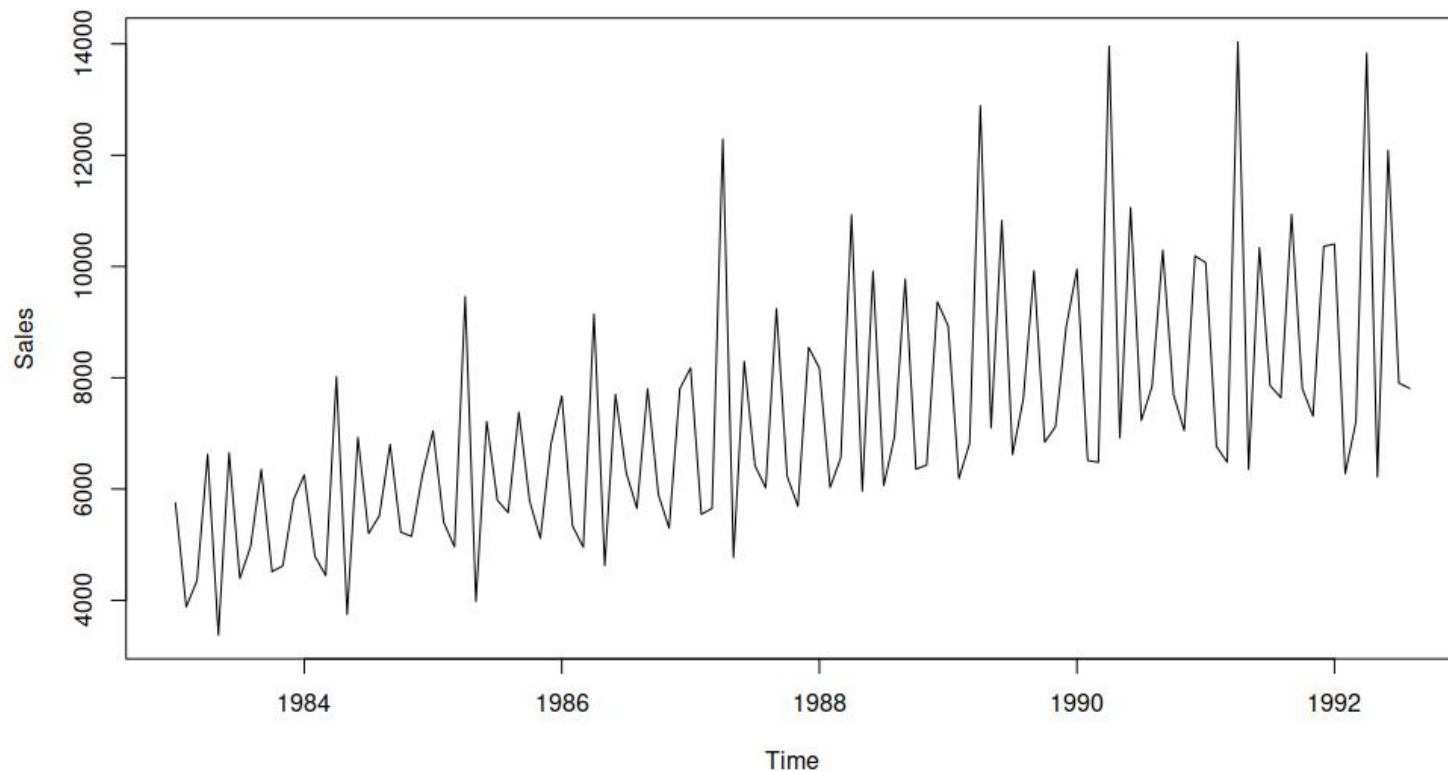
Introduction to ETS

- Why do we bother with ETS **model** and not just stick with **methods**?
- Models allow us:
 - producing point forecasts,
 - producing prediction intervals,
 - selecting the components (error / trend /seasonal),
 - adding explanatory variables (weather, promotions),
- + they can be estimated in a way, guaranteeing that the forecasts will be more stable.

Introduction to ETS, Quiz

Let's see if you can identify components in time series, please follow the link: <http://etc.ch/V7Ss>

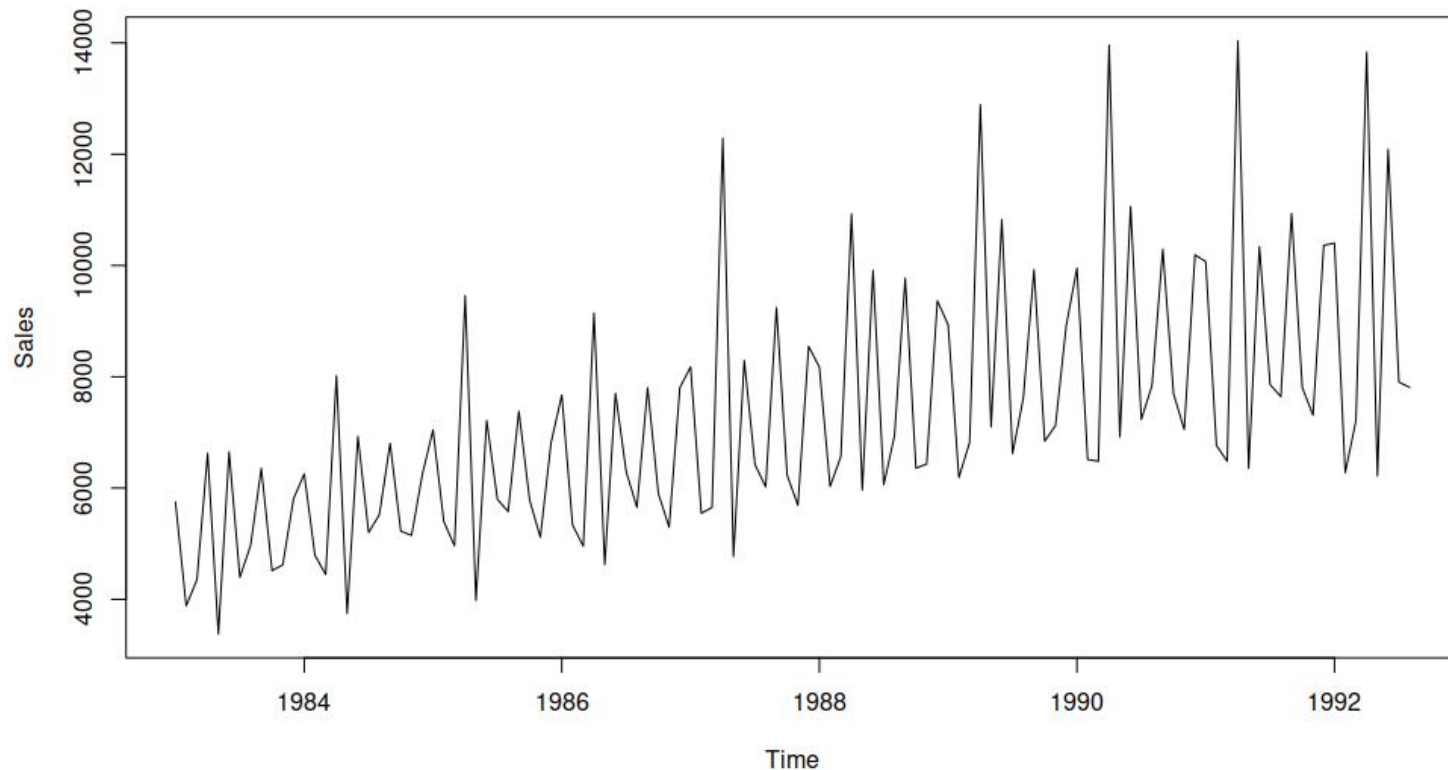
1. What time series components are present here?



Introduction to ETS, Quiz

Please, follow the link: <http://etc.ch/V7Ss>

2. What types of components are present in the same series?



Outline

1. Forecasting level series;
2. Simple Exponential Smoothing;
3. Introduction to ETS;
4. Local level model;
5. Trend and seasonal models;
6. Model estimation and selection.

Local level model

- Local level model underlies SES.
- It can be:
 - either additive – ETS(A,N,N):

$$y_t = l_{t-1} + \epsilon_t$$

$$l_t = l_{t-1} + \alpha\epsilon_t$$

- or multiplicative – ETS(M,N,N):

$$y_t = l_{t-1}(1 + \epsilon_t)$$

$$l_t = l_{t-1}(1 + \alpha\epsilon_t)$$

Local level model

- In the additive case:

$$y_t = l_{t-1} + \epsilon_t$$

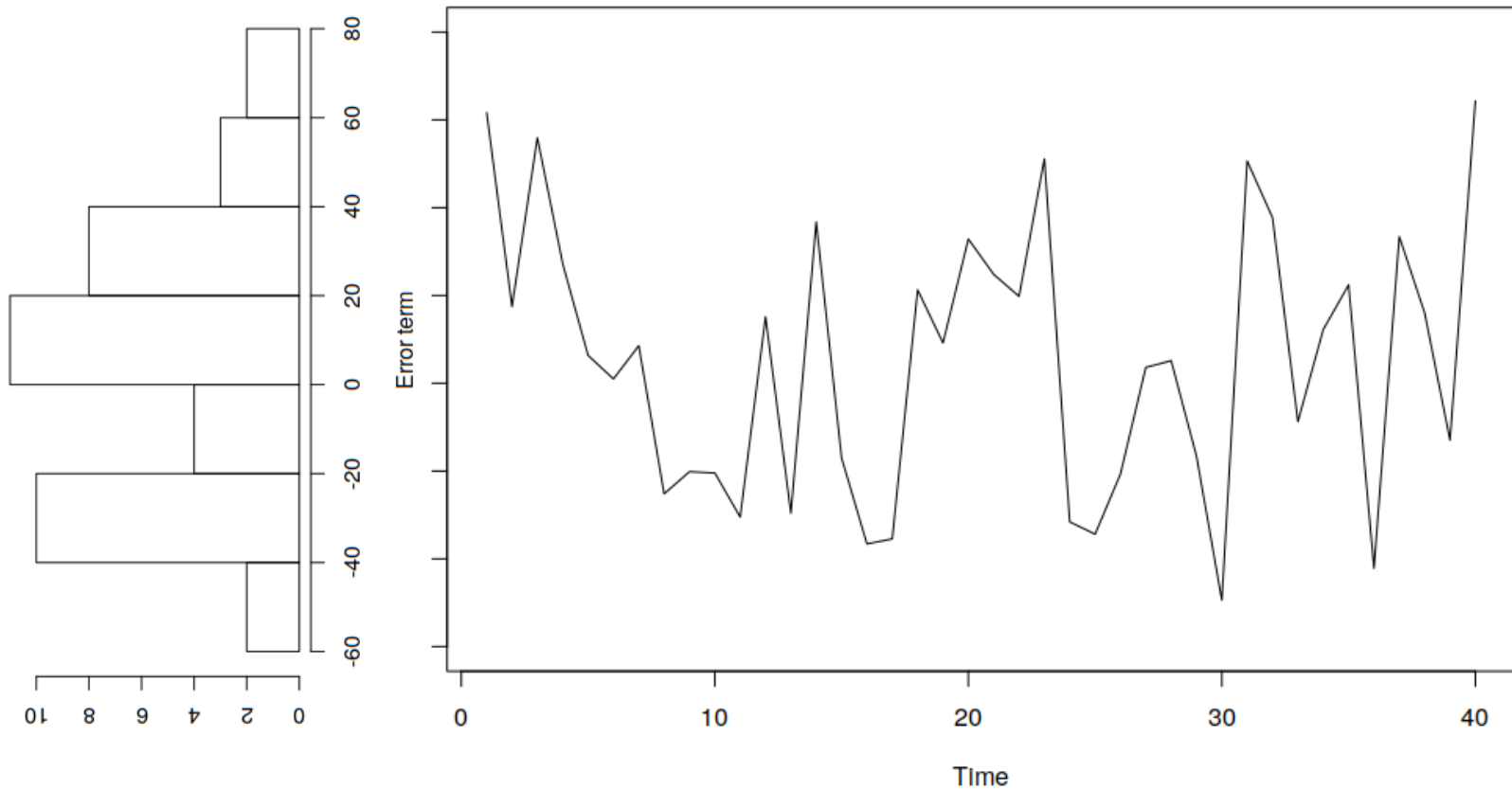
$$l_t = l_{t-1} + \alpha\epsilon_t$$

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

- The l_t represents the anticipated average demand in period t (e.g. average demand on beer in a pub in Cardiff);
- The ϵ_t represents the unexpected demand (e.g. Ivan visits Cardiff);
- σ is the size of the uncertainty about the demand;
- α is the rate of change of the level of demand;
- $\alpha\epsilon_t$ is the persistent effect on the level (e.g. Ivan goes out with his friends);

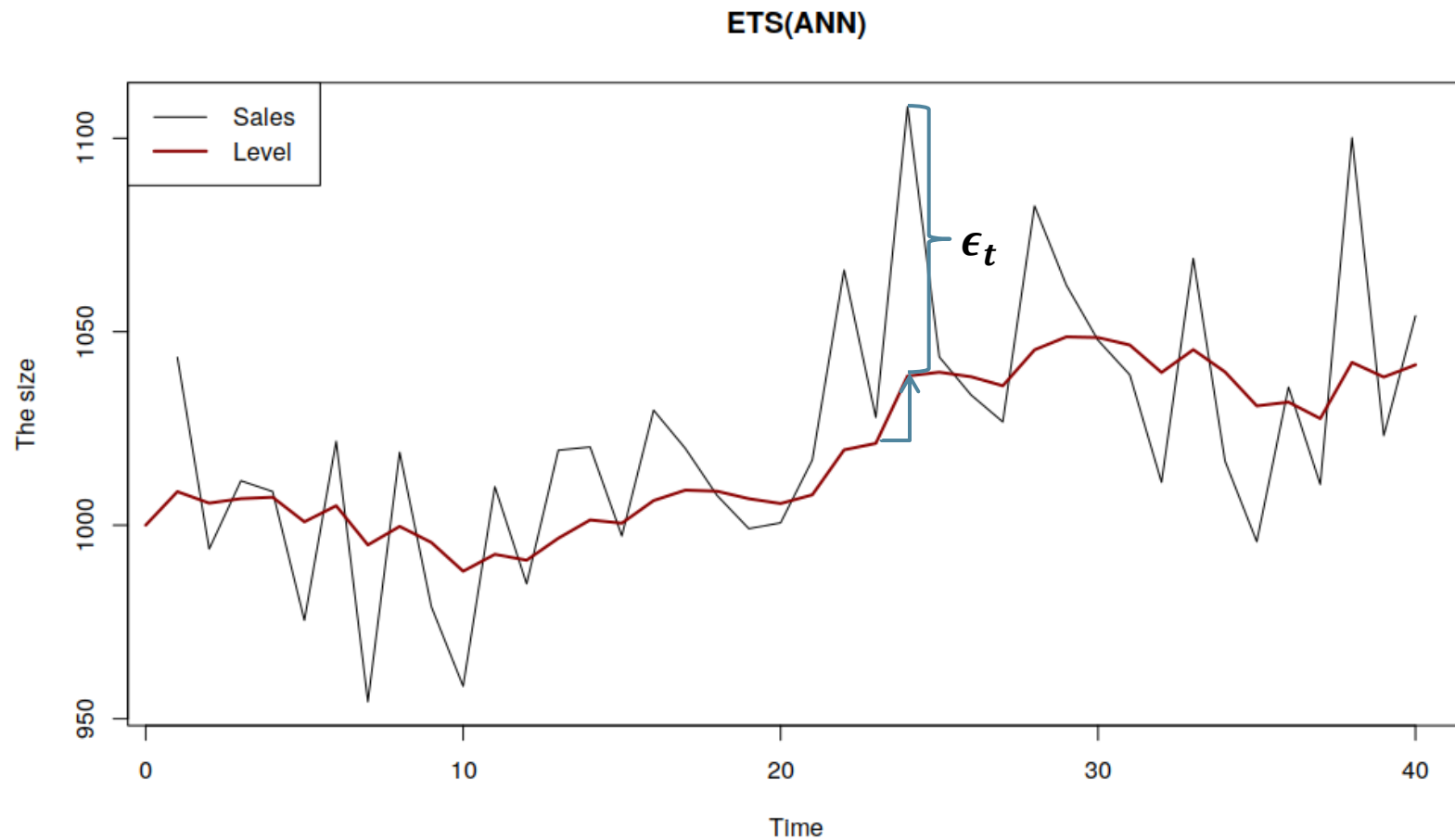
Local level model

- An example. $\sigma = 30$



Local level model

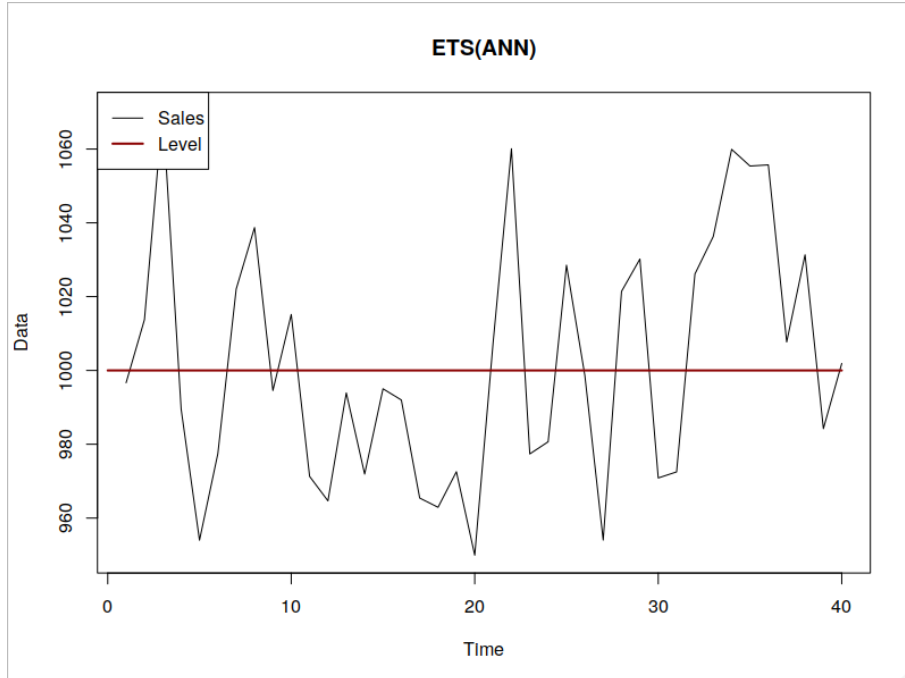
- An example with $\alpha = 0.2$ and $\sigma = 30$



Local level model

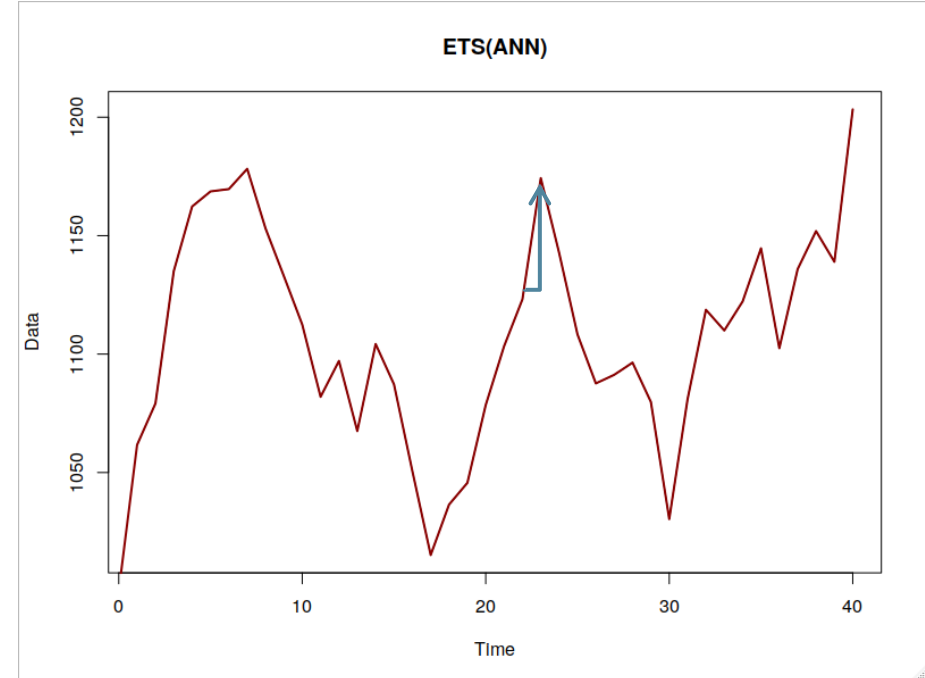
- Two cases of interest:

$$\alpha = 0$$



Global mean (global level)

$$\alpha = 1$$



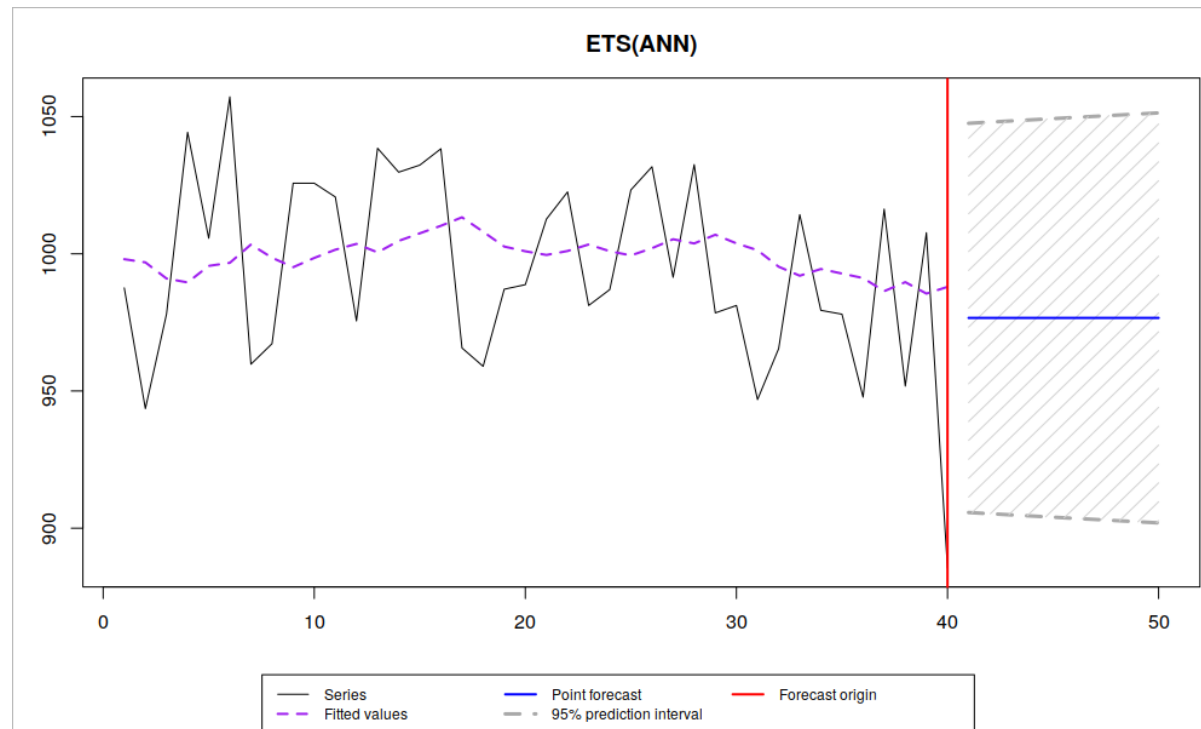
Naïve (random walk)

Local level model

- The forecast is the straight line:

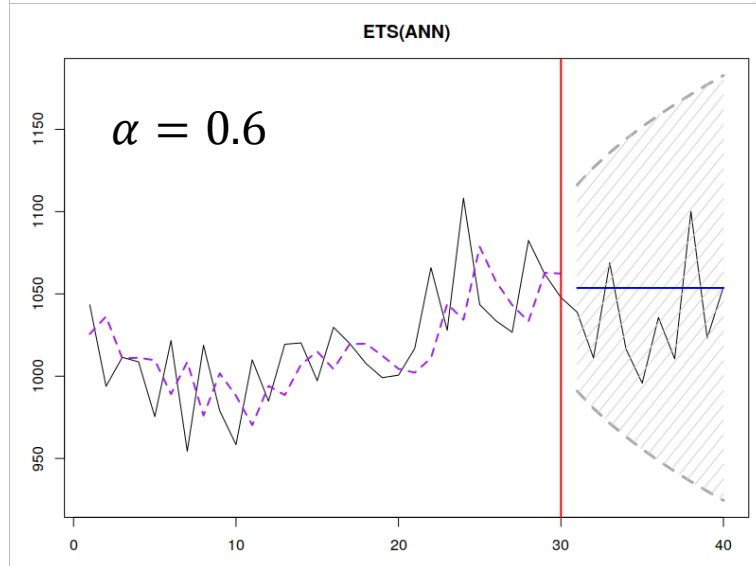
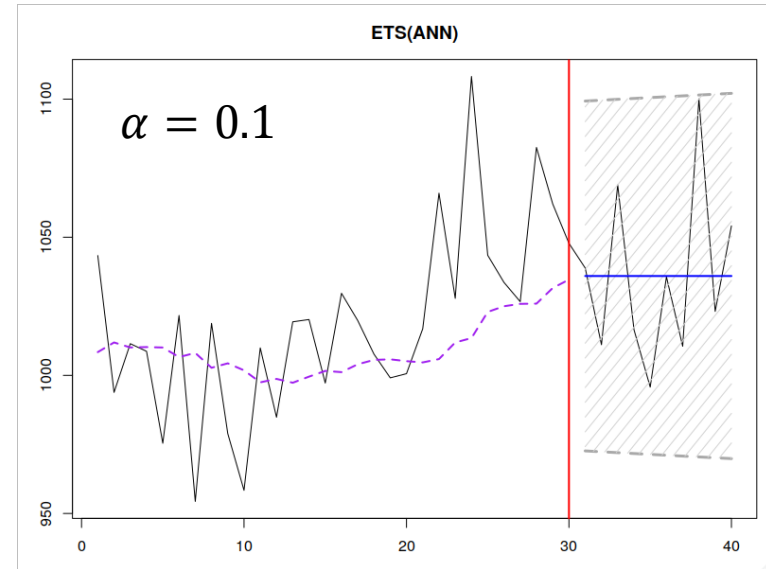
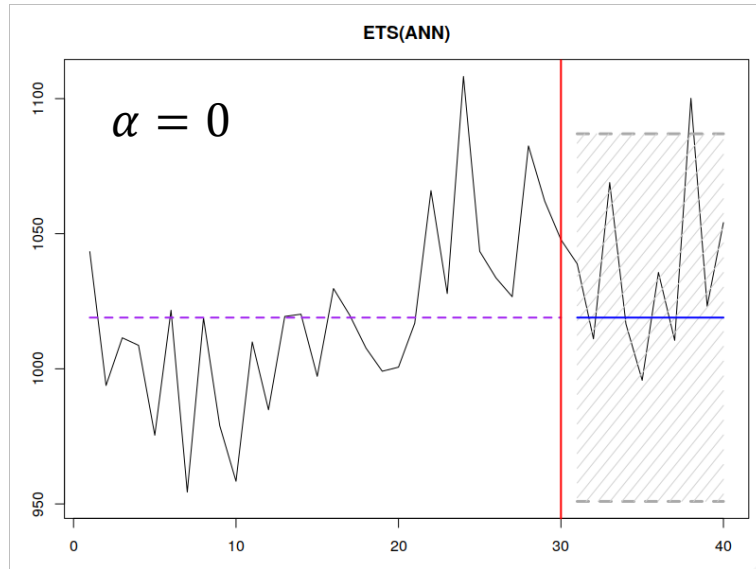
$$\hat{y}_{t+h} = l_t$$

- And we can construct prediction intervals based on $\epsilon_t \sim N(0, \sigma^2)$

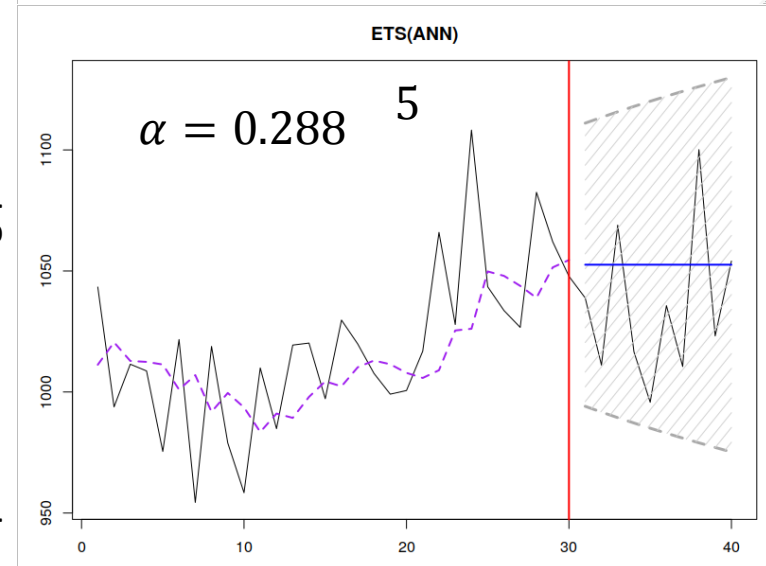


Local level model

- An example with different values of $\hat{\alpha}$:



Optimal smoothing parameter



Local level model

- Summarising:
 1. α regulates the rate of change of the local level;
 2. The higher it is, the higher the responsiveness of the model;
 3. The higher α means higher uncertainty, because of (2);
 4. It also regulates the width of prediction interval;
 5. We can optimise α .

Local level model

- ETS(M,N,N) has properties similar to ETS(A,N,N):

$$y_t = l_{t-1}(1 + \epsilon_t)$$

$$l_t = l_{t-1}(1 + \alpha\epsilon_t)$$

$$1 + \epsilon_t \sim \text{log } N(0, \sigma^2)$$

- The forecast is the straight line again.
- But the prediction interval increases with the increase of level.

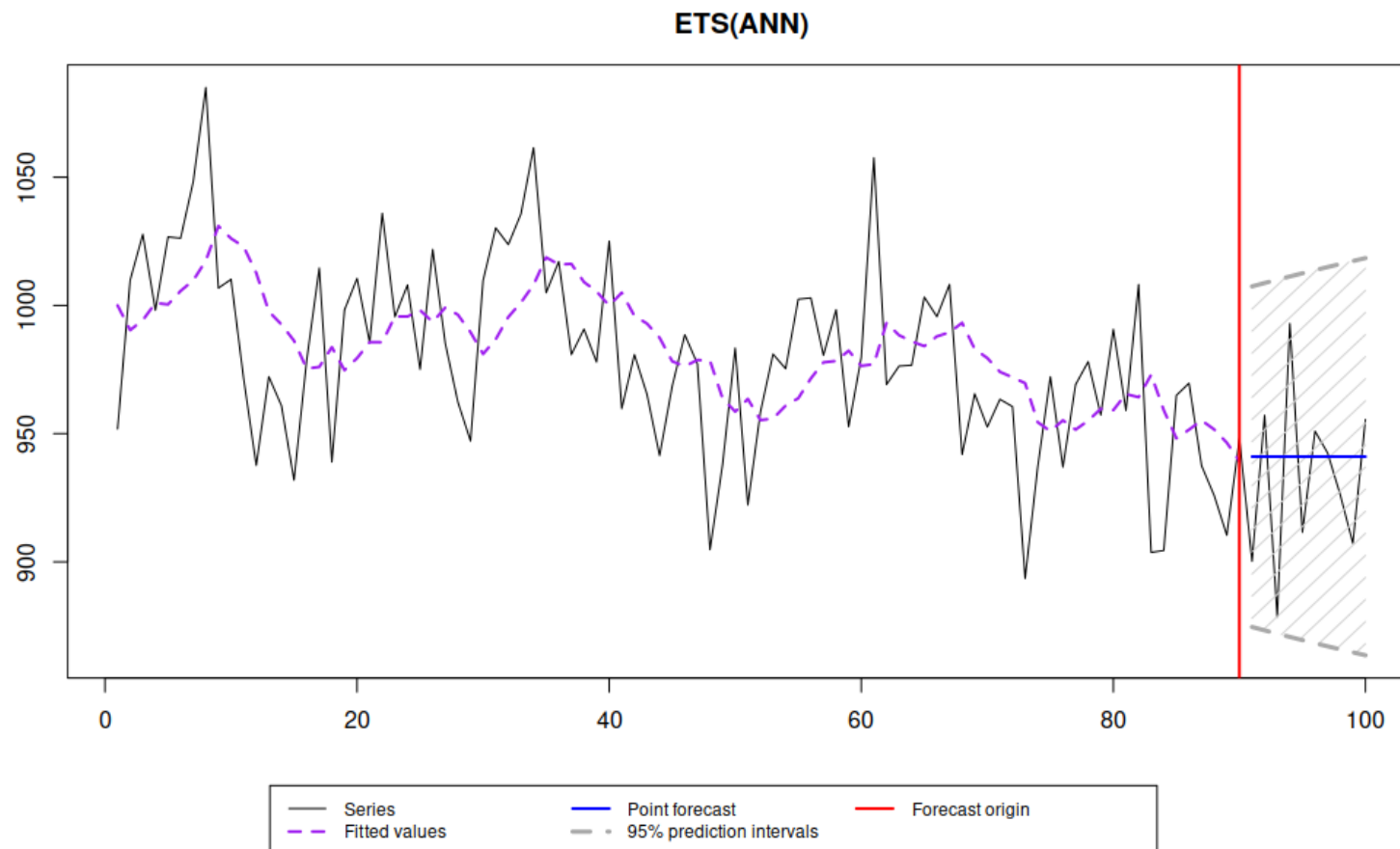
Local level model

- How many parameters do we need to estimate in ETS(A,N,N)?
- Three:
 - \hat{l}_0 , $\hat{\alpha}$ and $\hat{\sigma}^2$.

Local level model, Quiz

Please follow the link: <http://etc.ch/V7Ss>

1. Why does the prediction interval widen with the increase of forecast horizon for ETS(A,N,N) in this case?



Outline

1. Forecasting level series;
2. Simple Exponential Smoothing;
3. Introduction to ETS;
4. Local level model;
5. Trend and seasonal models;
6. Model estimation and selection.

Local trend model

- Are there any other components in time series?
- Why not add a trend component, ETS(A,A,N) :

ETS(A,A,N)

$$\begin{aligned}y_t &= l_{t-1} + b_{t-1} + \epsilon_t \\l_t &= l_{t-1} + b_{t-1} + \alpha \epsilon_t \\b_t &= b_{t-1} + \beta \epsilon_t \\\epsilon_t &\sim N(0, \sigma^2)\end{aligned}$$

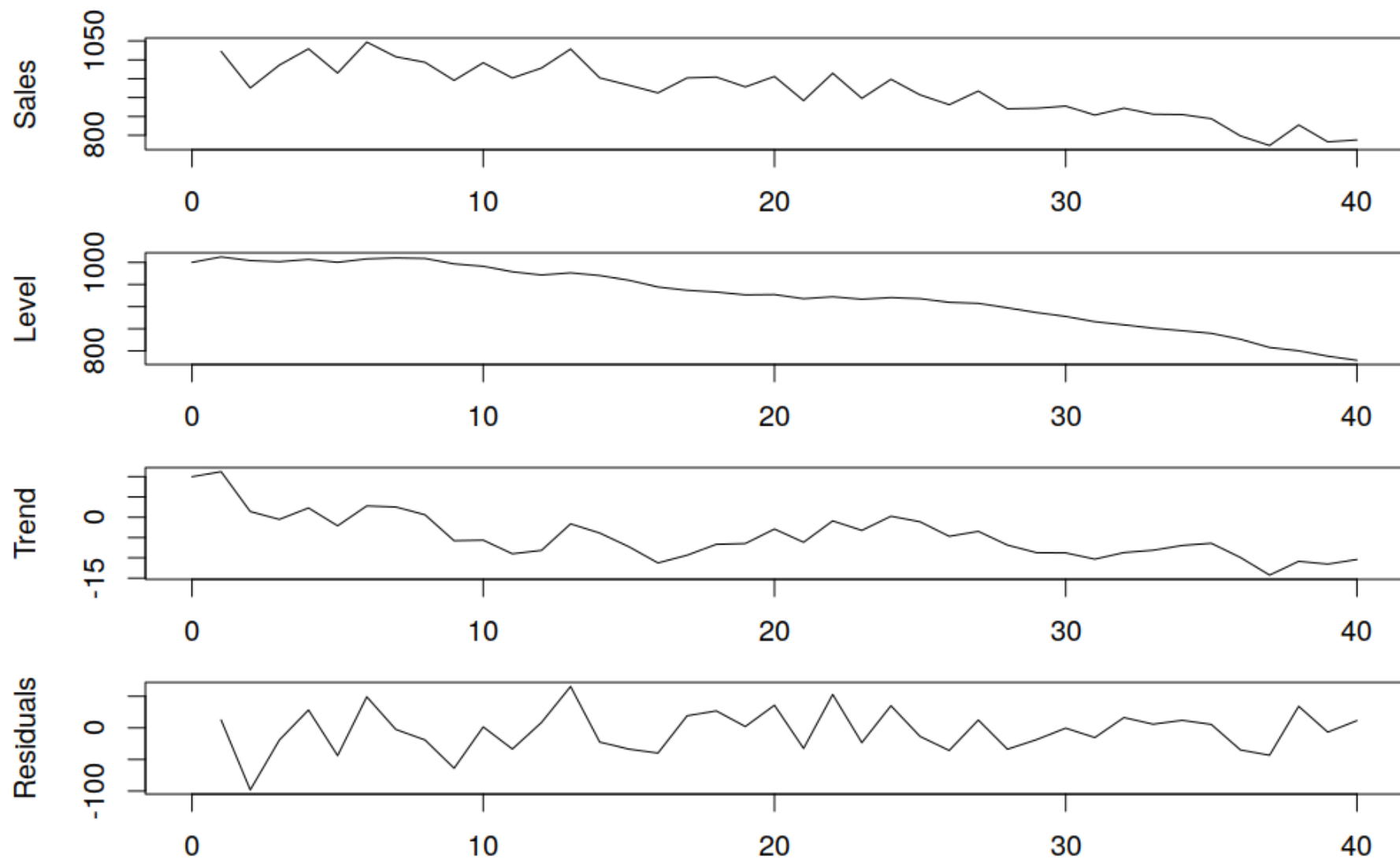
ETS(A,N,N)

$$\begin{aligned}y_t &= l_{t-1} + \epsilon_t \\l_t &= l_{t-1} + \alpha \epsilon_t \\\epsilon_t &\sim N(0, \sigma^2)\end{aligned}$$

- The mechanism is similar to ETS(A,N,N).
- This model underlies “Holt’s method”.
- But now we also update the trend.

Local trend model

- Decomposition of time series due to ETS(A,A,N):



Local trend model

- ETS(A,A,N) :

$$y_t = l_{t-1} + b_{t-1} + \epsilon_t$$

$$l_t = l_{t-1} + b_{t-1} + \alpha\epsilon_t$$

$$b_t = b_{t-1} + \beta\epsilon_t$$

- α has the same property as in ETS(A,N,N).
- β defines the rate of change of the trend:
 - $\beta = 0$, $b_t = b_{t-1}$, the trend is constant;
 - $\beta = 1$, $b_t = b_{t-1} + \epsilon_t$, the trend is changing rapidly.
- The forecast is a line:
$$\hat{y}_{t+h} = \hat{l}_t + h\hat{b}_t$$
- The width of intervals changes with the change of both smoothing parameters.

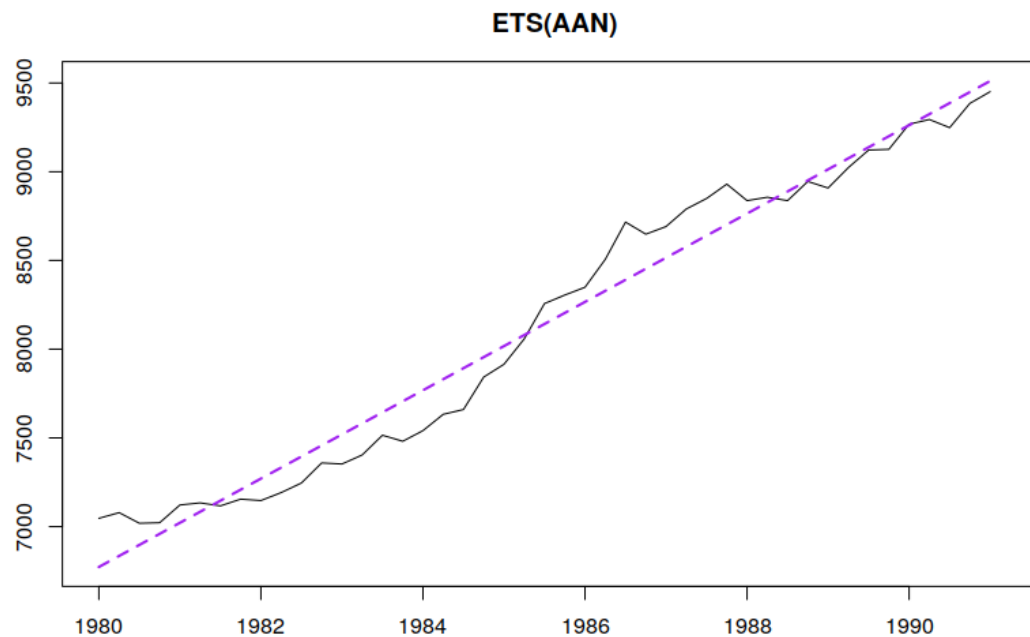
Local trend model

- If both $\alpha = 0$ and $\beta = 0$, then we have a deterministic trend:

$$y_t = l_{t-1} + b_{t-1} + \epsilon_t$$

$$l_t = l_{t-1} + b_{t-1}$$

$$b_t = b_{t-1} = b_0$$

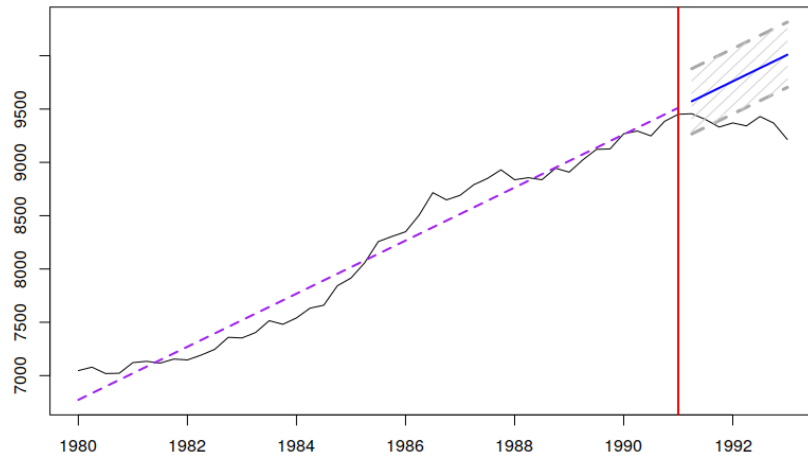


Local trend model

- The influence of parameters on forecasts:

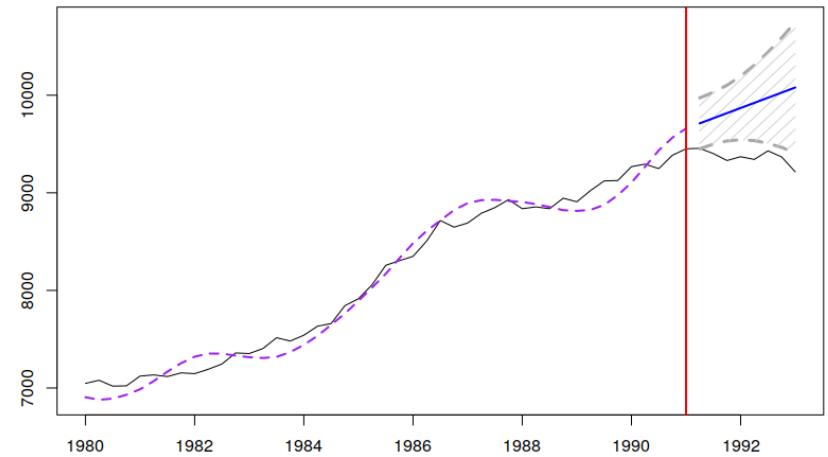
$$\alpha = 0, \beta = 0$$

ETS(AAN)



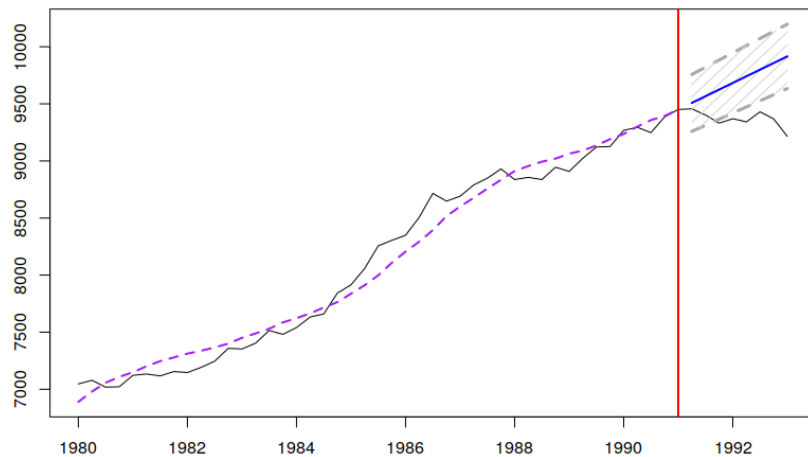
$$\alpha = 0, \beta = 0.2$$

ETS(AAN)



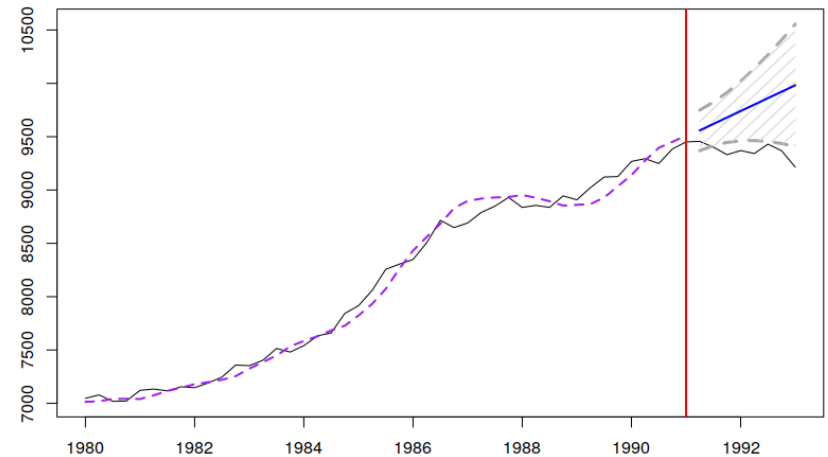
$$\alpha = 0.2, \beta = 0$$

ETS(AAN)



$$\alpha = 0.2, \beta = 0.2$$

ETS(AAN)



Local trend model











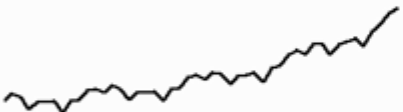




- How many parameters do we need to estimate in ETS(A,A,N)?
- Five:
 - \hat{l}_0 , \hat{b}_0 , $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$.
- ETS(M,A,N) is similar, but assumes a different error term.
 - What does it imply?

Other trend model

- There are other types of trend models:
 - $ETS(A,Ad,N)$ – damped trend model (the trend is not linear, it is slowed down);
 - $ETS(M,M,N)$ – multiplicative trend model (exponential growth / decline);
 - ...
- but we don't have time to discuss all of them.
- The components update is similar to the one for $ETS(A,A,N)$.

ETS taxonomy

- Different types of components:

Trend	Seasonality		
	None "N"	Additive "A"	Multiplicative "M"
None "N"			
Additive "A"			
Additive Damped "Ad"			
Multiplicative "M"			
Multiplicative Damped "Md"			

Trend seasonal model

- Now we can formulate a more complicated model.

- We start with ETS(A,A,A):

$$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \epsilon_t$$

$$b_t = b_{t-1} + \beta \epsilon_t$$

$$s_t = s_{t-m} + \gamma \epsilon_t$$

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

ETS(A,A,N)

$$y_t = l_{t-1} + b_{t-1} + \epsilon_t$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \epsilon_t$$

$$b_t = b_{t-1} + \beta \epsilon_t$$

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

- Almost the same as ETS(A,A,N).
- γ now regulates the rate of change for the seasonal component.
- The forecast is produced as:

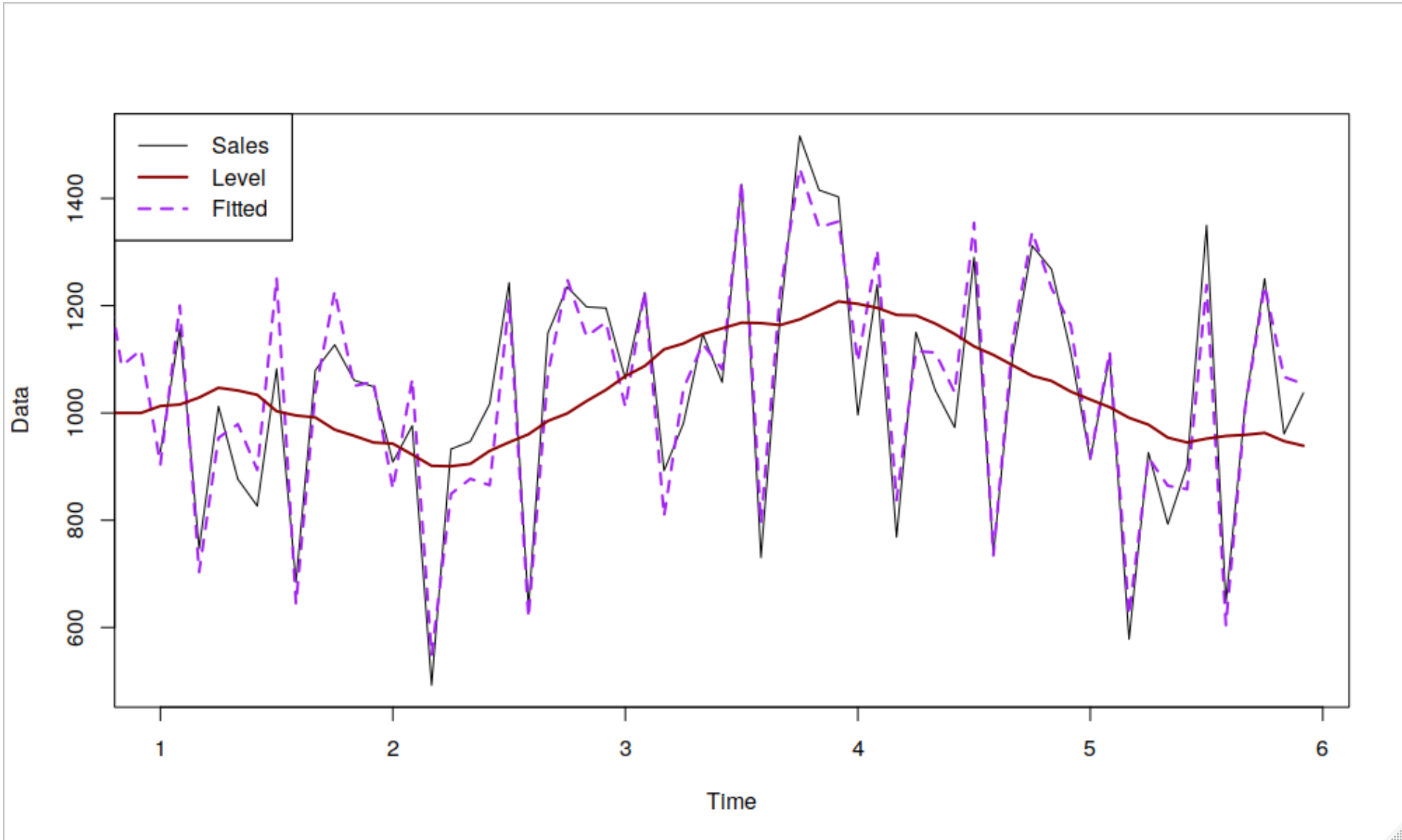
$$\hat{y}_{t+h} = \hat{l}_t + h\hat{b}_t + \hat{s}_{t-m+h}$$

Trend seasonal model

- The model underlies “Holt-Winters method”.
- How many parameters do we have in the trend seasonal model?
- $6 + m$:
 - $\hat{l}_0, \hat{b}_0,$
 - $\hat{\alpha}, \hat{\beta}, \hat{\gamma},$
 - m seasonal indices $s_1, s_2, \dots, s_m,$
 - and $\hat{\sigma}^2$.

Trend seasonal model

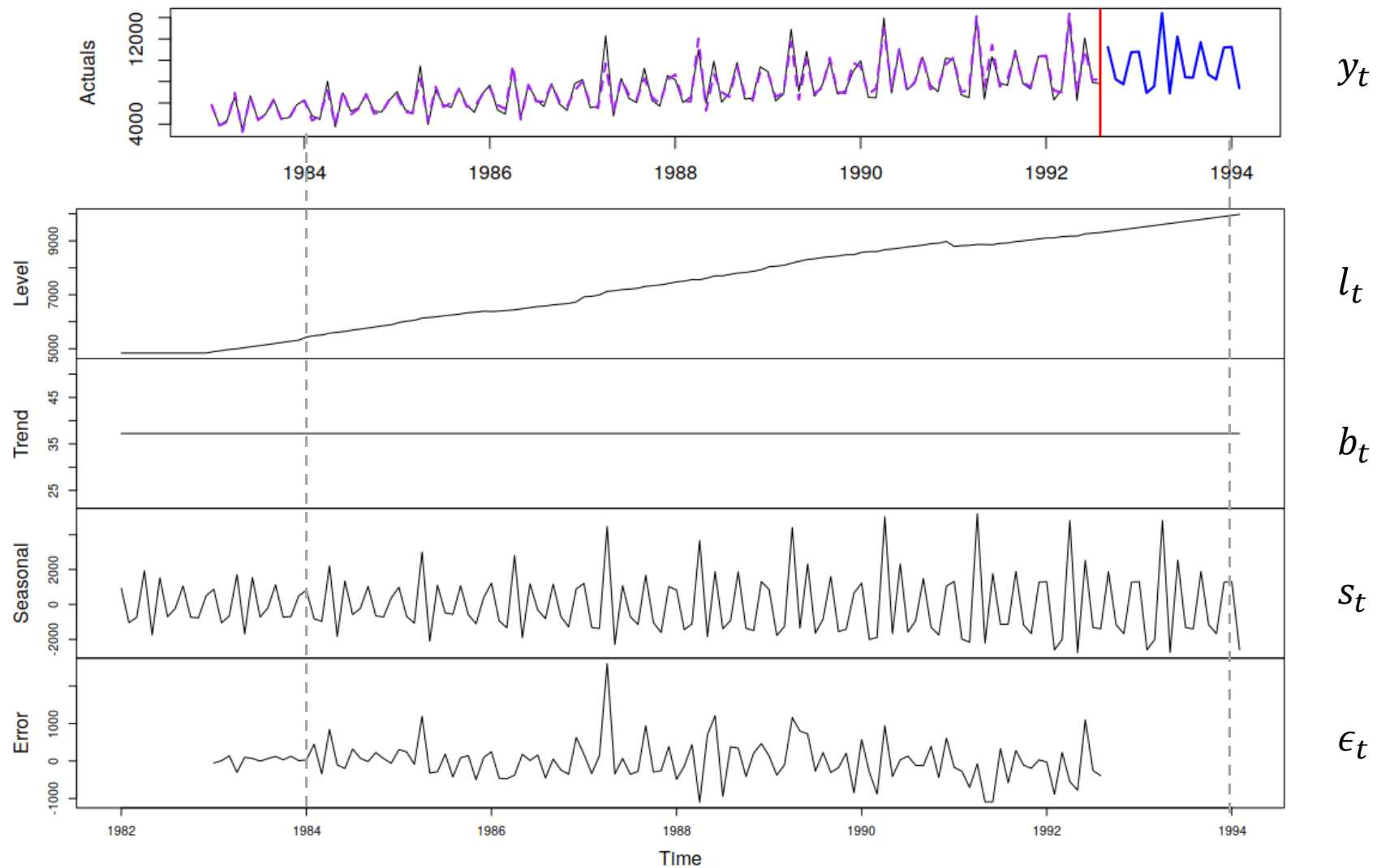
- An example with ETS(A,A,A):



Trend seasonal model

- A series can be decomposed based on ETS(A,A,A):

$$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t$$

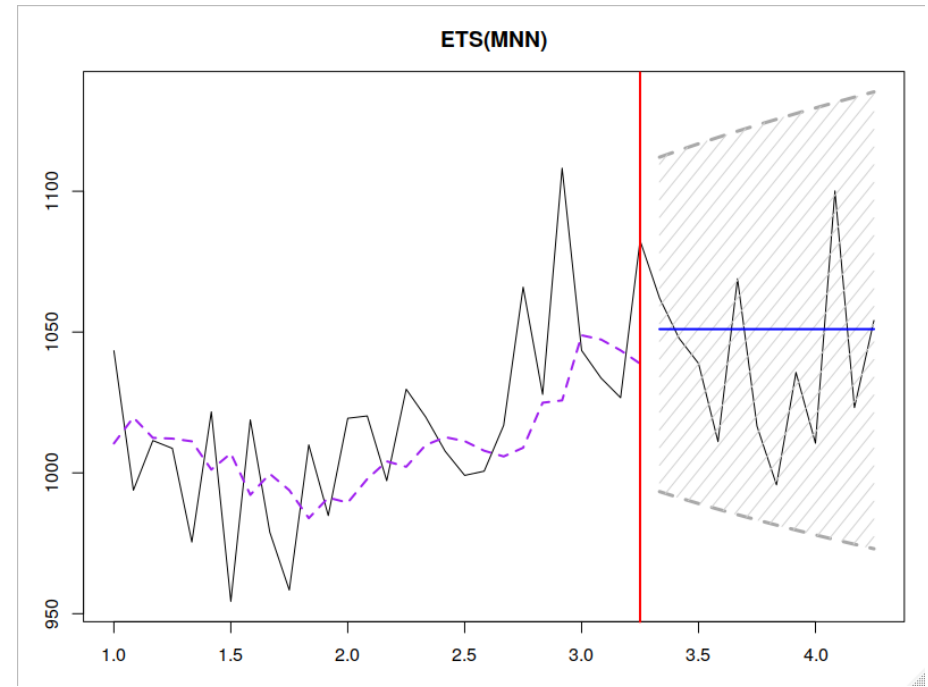
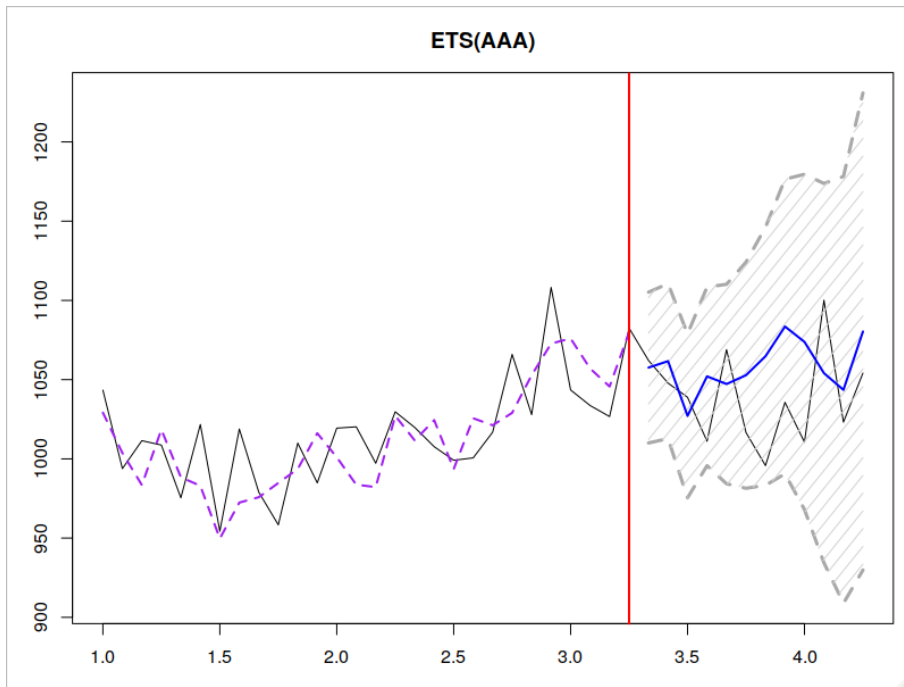


Trend seasonal model

- There are other types of trend-seasonal models
- The update mechanisms are similar.

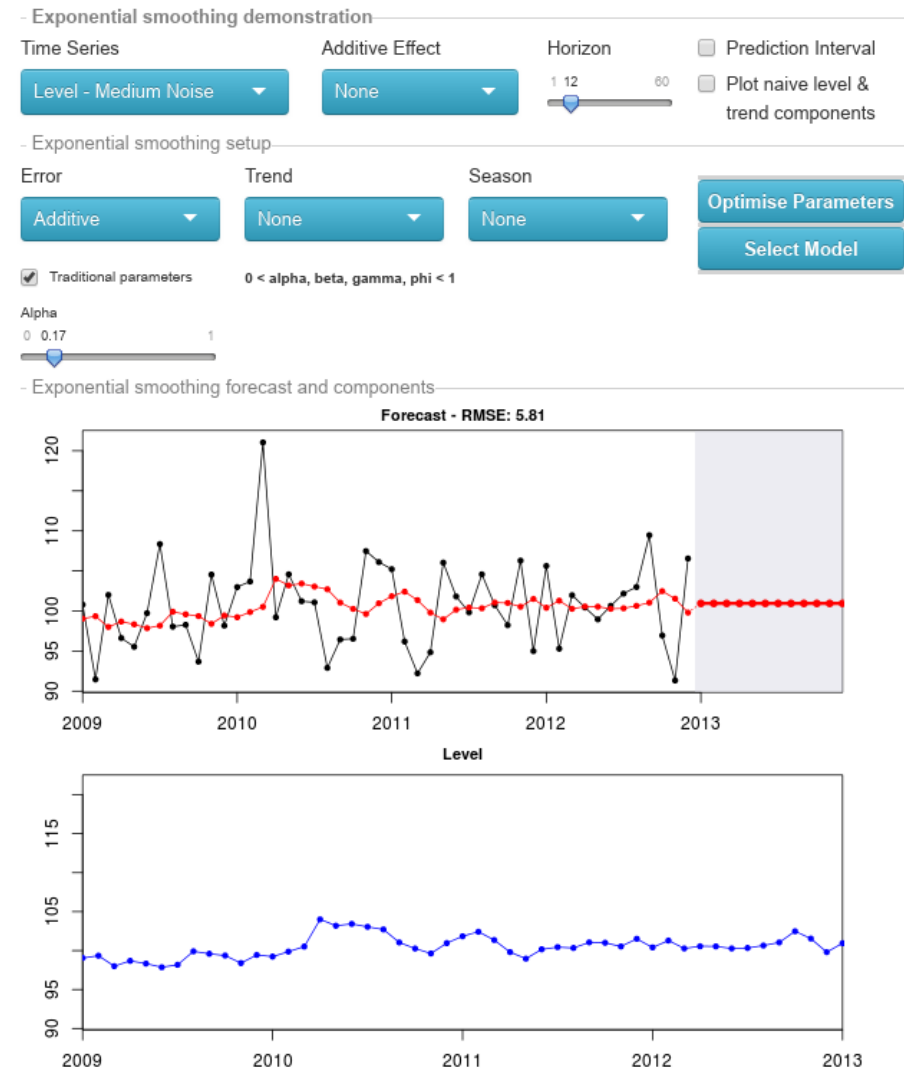
Trend seasonal model, Quiz

- An example, let's go to the quiz:
 1. Which of these two models makes more sense?



Trend seasonal model

- An exercise:
 - <https://kourentzes.com/forecasting/2014/10/30/exponential-smoothing-demo/>



Exponential smoothing demo using the [forecast](#) package for R. Nikolaos Kourentzes, 2014

Outline

1. Forecasting level series;
2. Simple Exponential Smoothing;
3. Introduction to ETS;
4. Local level model;
5. Trend and seasonal models;
6. Model estimation and selection.

Estimation of ETS

- Remember the pure additive ETS model?

$$y_t = \mathbf{w}'\mathbf{v}_{t-1} + \epsilon_t$$
$$\mathbf{v}_t = \mathbf{F}\mathbf{v}_{t-1} + \mathbf{g}\epsilon_t$$

- How can we estimate it?
- We use the assumption that $\epsilon_t \sim N(0, \sigma^2)$
- Based on this assumption we can derive a likelihood function, using pdf of normal distribution:

$$f(y_t|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_t - \hat{y}_t)^2}{2\sigma^2}}$$

- And then maximise it by changing parameters

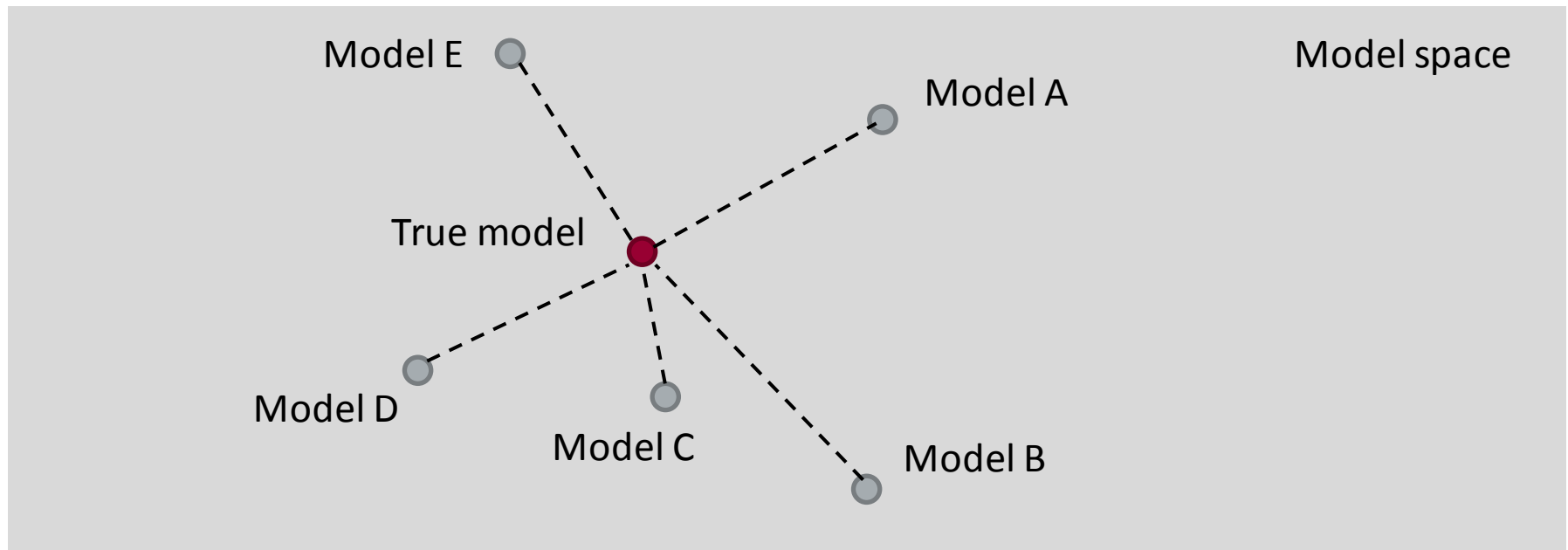
Estimation of ETS

- Why is maximum likelihood estimation (MLE) useful?
- Likelihood has good statistical properties:
 - MLE of parameters are consistent and efficient.
- Likelihood can be used in calculation of information criteria. Thus, model selection is possible.
- What about multiplicative models?
 - The approach is similar, but the likelihood function is different.

Information criteria

Can we measure distance between the true model and our model?

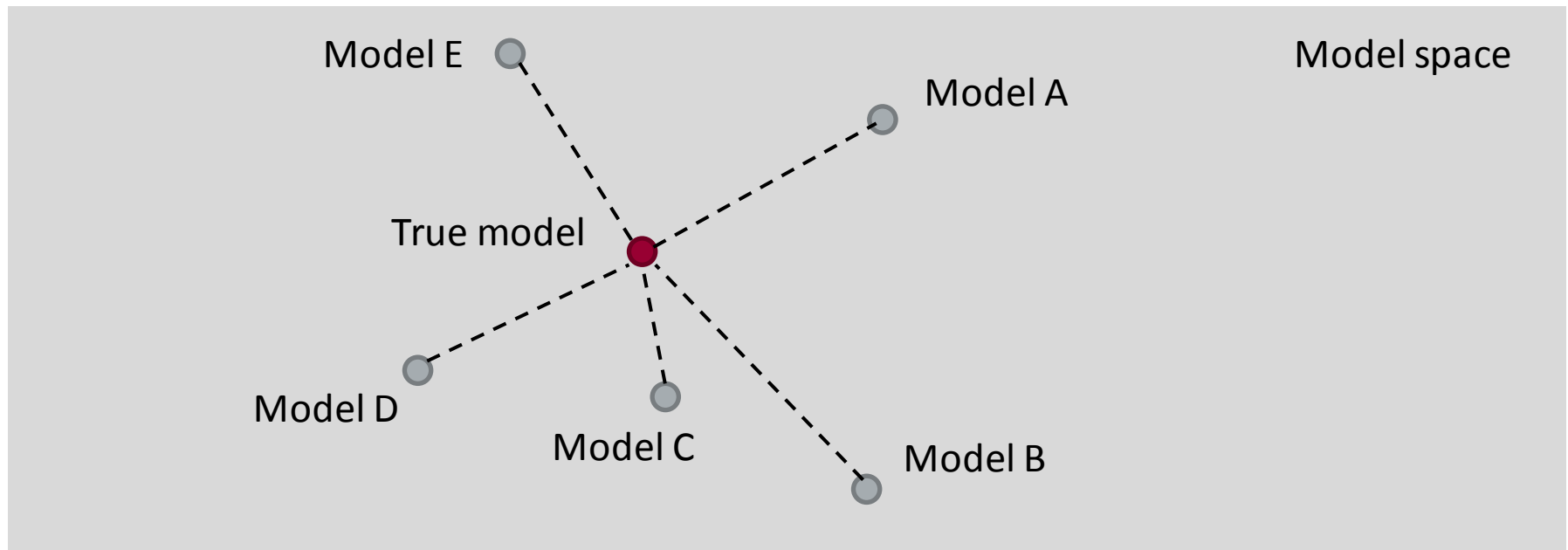
- Yes, if we know the truth:



Information criteria

What makes the model closer to the true one?

- The ETS components,
- The transformation of the variable,
- The estimates of parameters.



Model selection in ETS

- We can compare models using AIC:

$$AIC = 2k - 2\ell(\text{mod})$$

- where ℓ is the log-likelihood value and k is the number of **all** the estimated parameters
- There are other ICs:
 - AICc ← corrected for the sample size AIC;
 - BIC – Bayesian IC (aka Schwartz IC);
 - ...

Assumes normal distribution,
Used by default in R functions.

Model selection in ETS

So, in the ETS framework, we can:

1. fit all the possible models,
 2. calculate their likelihoods,
 3. calculate the number of parameters (including $\hat{\sigma}^2$),
 4. calculate IC values of the models in the pool,
 5. select the model that has the lowest IC.
- This is what all the ETS functions in R do by default.

Summary

1. Forecasting level series;
2. Simple Exponential Smoothing;
3. Introduction to ETS;
4. Local level model;
5. Trend and seasonal models;
6. Model estimation and selection.

Summary

- Packages and functions in R:
 - forecast package:
 - ets() – basic ETS with 19 models;
 - bats(), tbats() – models for multiple frequencies.
 - fable package:
 - ETS() – similar to ets() from forecast.
 - 19 models, only additive trend;
 - smooth package:
 - es() – more flexible ETS:
 - 30 models,
 - different loss functions,
 - allows including explanatory variables.

Thank you!

Thank you for your attention!

Questions?

Ivan Svetunkov

i.svetunkov@lancaster.ac.uk



[@iSvetunkov](https://twitter.com/iSvetunkov)

<https://forecasting.svetunkov.ru>

Full or partial reproduction of the slides is not permitted without author's consent. Please contact i.svetunkov@lancaster.ac.uk for more information.

Marketing Analytics
& Forecasting



Lancaster University
Management School