Chapter 4- Curve Fitting and Exponential Smoothing

🔗 Forecasting Book Club

Dr. Bahman Rostami-Tabar

Outline

- Introduction (section 4.1)
- Curve Fitting (section 4.2)
- Exponential Smoothing Methods (section 4.3)
  - Simple Exponential Smoothing
  - Holt's Method
  - Damped Holt's Method
  - Holt Winters Method
- Forecasting Intermittent demand (section 4.4)
- Wrap-up (section 4.5)
- Summary of key terms (section 4.6)

Introduction

- When using the forecasting software, you don't need to deal with the math and formula. The software will do the hard work and produce your forecast,
- However, it is important to understand how these forecasts are obtained so you are aware of the rationales that underpins them,
- That way you will have a better understanding of the strength and limitations of different methods.

What is this chapter about?

- In this chapter, we look at methods that simply use your historical data to identify and project patterns into the future,
- These methods called `time series` methods -**they don't take any account of external factors**,
- Because they only process data on single variable, `e.g. past sales`, they are referred to as univariate,
- Despite ignoring exogenious/external factors, these method can be robust and accurate,
- They also avoid the need to collect and record data on potential drivers, which could be expensive and time consuming,
- They can be run automatically, a huge advantage with many time series.

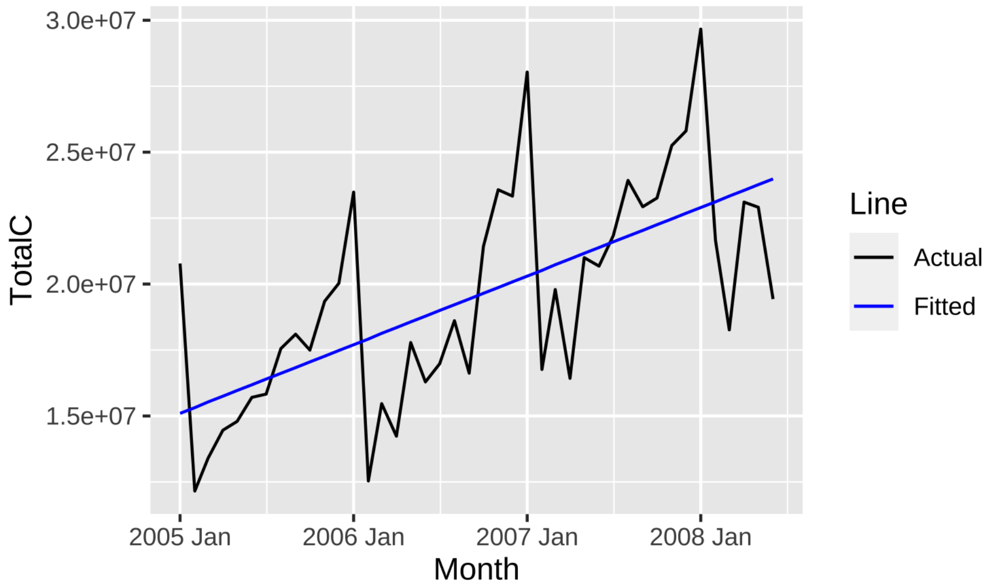
Curve Fitting

- What are the common types of curve?
- What is a good fit?
- What are strengths and Limitations of Curve Fitting?

Curve fitting

- The idea is to find a line that best fits the historical data,
- We want to find the underlying trend in the data and to extrapolate this into the future to produce forecasts,
- We have fitted a curve to the data and, by extending the curve beyond month 42, we can get the forecasts.

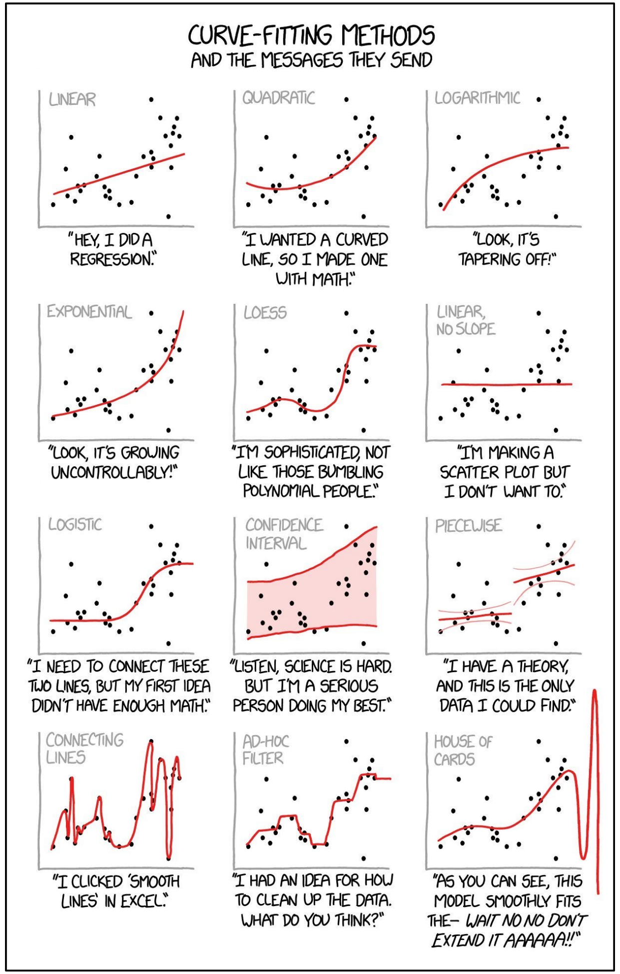
Sales of a product over 42 months



Common types of curves

- Linear, quadratic, exponential, logarithmic, etc
- Computers identify the curve that has the best fit by using a procedure called the least squares method.
- On the graph, the vertical distance between actual and fitted lines is called an residual/error.
- Best fitting curve is the one that minimizes the sum of squared errors. this is called Ordinary Least Squared (OLS) method.

Examples of curves



Graph from xkcd

Fitted curves are described by their equations

- Linear curve: $Sales = \alpha + \beta_0 \times Time$, e.g. $Sales = 15 + 5 \times 43$
- Quadratic curve: $Sales = \alpha + \beta_0 \times Time + \beta_1 \times Time^2$, e.g. $Sales = 8 + 20 \times 43 - 0.5 \times 43^2$
- Exponential curve: $Sales = e^{\alpha + \beta_0 \times Time}$, e.g. $Sales = e^{2 + .17 \times 43}$
- α and β are parameters that can be estimated using OLS. Time is the period number, e.g. Time =43 is the period number 43.

Assessing how well the curve fists the historical data

- Using the error measure discussed in `Chapter 3`, e.g RMSE, MAE, MAPE,

- R-Squared = Explained variation / Total variation: tells us what fraction of the variation in the actual data can be explained by the fitted curve,

- R-squared is always between 0 and 100. If we had a complex curve that exactly replicated the historical data, then R-squared would be 100. if, instead, we fitted a horizontal line, this would not reflect any of the variation, so R-squared would be ?

Assessing how well the curve fists the historical data

- But there is a problem with R-squared. If we use more complex curve with more parameters, the R-squared will increase. Curves with more parameters tend to have higher R squared,

- This phenomenon is known as overfitting in forecasting. In forecasting we try to identify systematic patterns in data and use them to forecast the future. If we fail this is bad.

- Overfitting happens because we try to model the noise as well. This is bad too. Random variations almost certainly won't be repeated in the future.

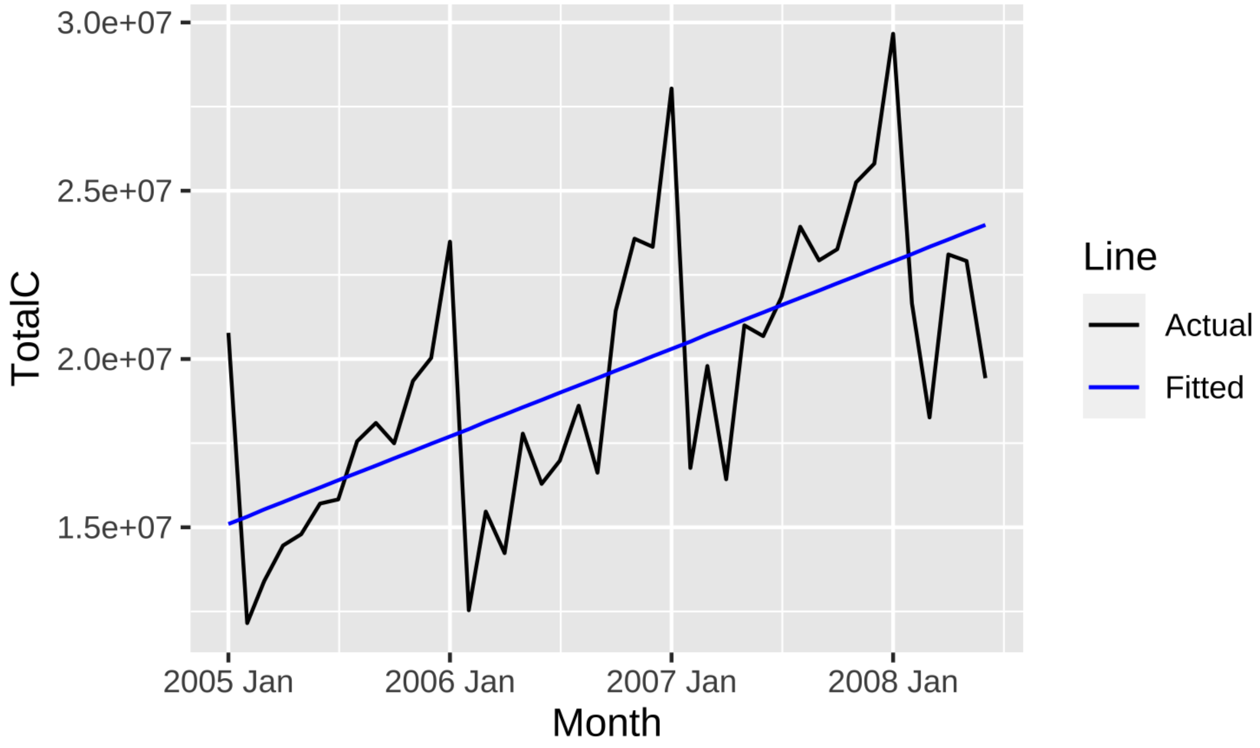
Assessing how well the curve fists the historical data

- To counter this, the Adjusted R-squared is introduced which measures the variation explained by fitted curve while penalizing curves with more parameters.

- `Information Criteria (IC)` such as BIC, AIC introduced in `Chapter 3`, can also be useful in comparing curves with different level of complexity as they penalizes those that have more parameters. It balances a good fit with a resonable number of parameters.

Assessing how well the curve fists the historical data

$R^2 = 0.413$ and $AdjR^2 = 0.398$



Strength and limitations of forecast based on curve fitting

- When we fit a curve to data and extrapolate it to obtain the forecasts, we are said to be using a global model. This could be dangerous!!,
- It assumes the underlying trend has the same form across all time periods, e.g. linear equation assumes that underlying sales will always change by the same amount. however the nature of trend can change over time. Don't be surprised if you get negative values,
- It could be sensitive to the sample of time series you have, A sales history that ends on a high summer peak may lead to the estimation of a different curve than a history that ends in a winter trough,
- While it is simple to understand and implement, it can not take into account changing patterns in trends over time.

To explain or to predict?

There is no guarantee that a good fit leads to an accurate forecast.

Shmueli, Galit. "To explain or to predict?." Statistical science 25.3 (2010): 289-310.

Exponential Smoothing

Exponential Smoothing

In the next section, we discuss some widely used methods that can handle changes in the pattern over time which also account for both trend and seasonality. They update their estimates of what is happening as each new information arrives.

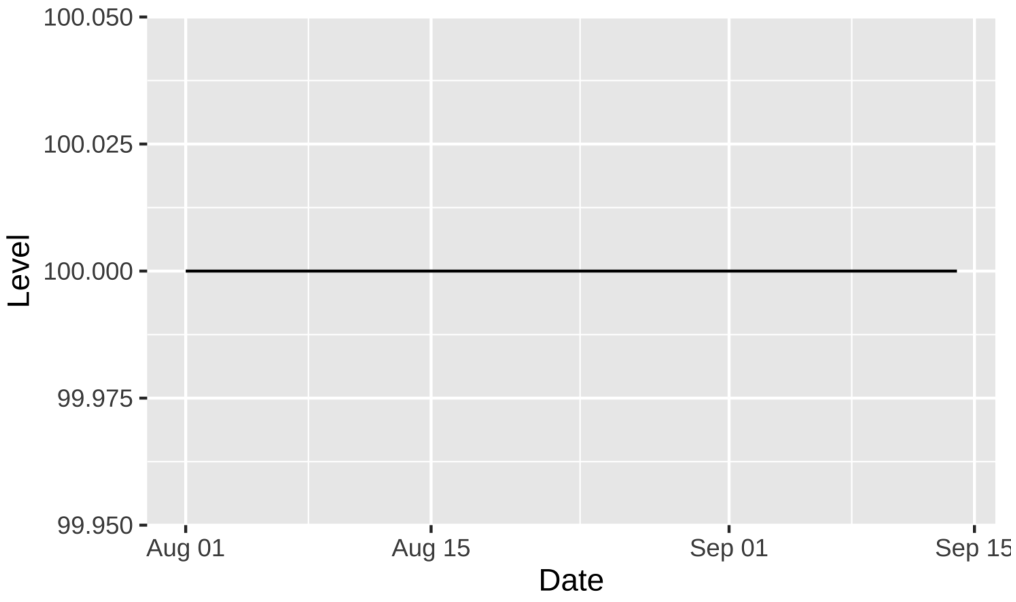
- Combine a `level`, `trend` (slope) and `seasonal` component to describe a time series.
- The rate of change of the components are controlled by `smoothing parameters`: α, β, ϕ and γ

Simple (Single) Exponential Smoothing (SES)

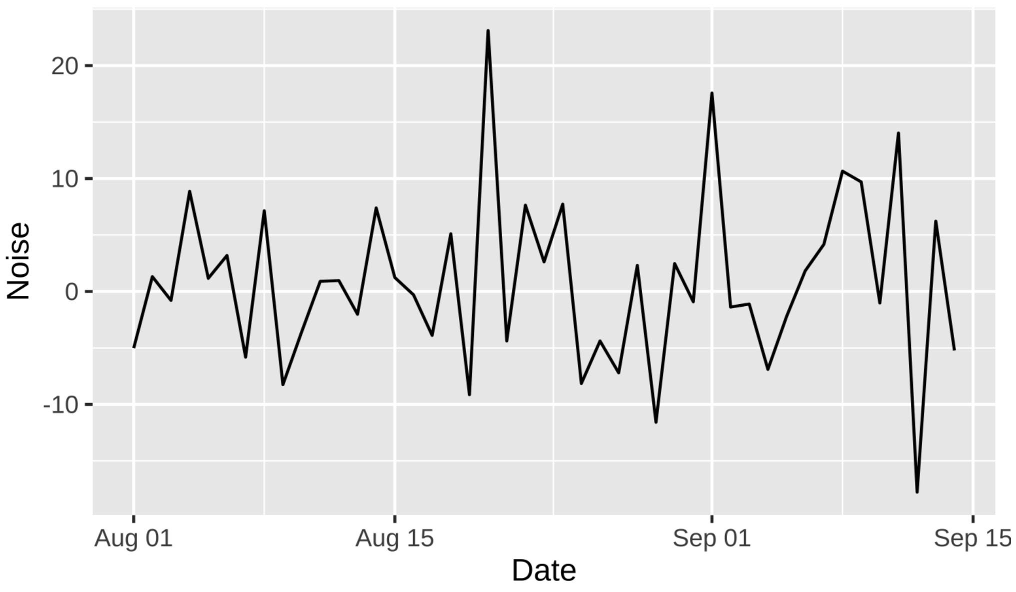
- SES is widely used in practice and appropriate when the historical series does not have a trend or seasonal pattern, but the underlying level of sales can change,

- Actual time series will vary randomly around the current level, so the task of SES is to estimate what that level is

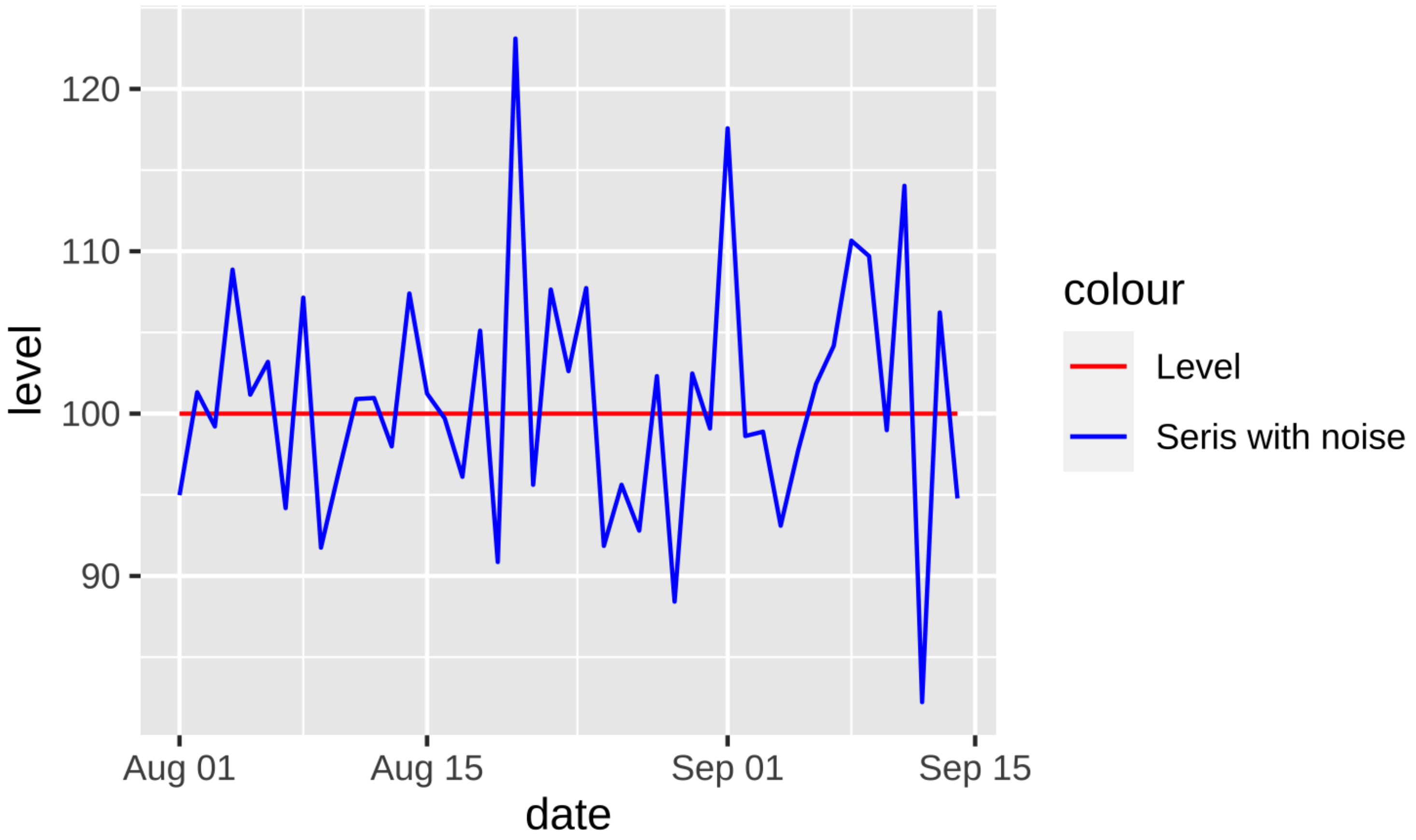
level



Random series/noise



Combining level and noise



date	level	noise	series
2020-08-01	100	-5.0219235	94.97808
2020-08-02	100	1.3153117	101.31531
2020-08-03	100	-0.7891709	99.21083
2020-08-04	100	8.8678481	108.86785
2020-08-05	100	1.1697127	101.16971
2020-08-06	100	3.1863009	103.18630
2020-08-07	100	-5.8179068	94.18209
2020-08-08	100	7.1453271	107.14533
2020-08-09	100	-8.2525943	91.74741
2020-08-10	100	-3.5986213	96.40138

How

SES updates the estimate of level?

How SES produces a forecast

- Forecast for next period = Current forecast+ $\alpha\times$ Forecast Error = Current forecast+ $\alpha\times$ (Latest observation - Current forecast)

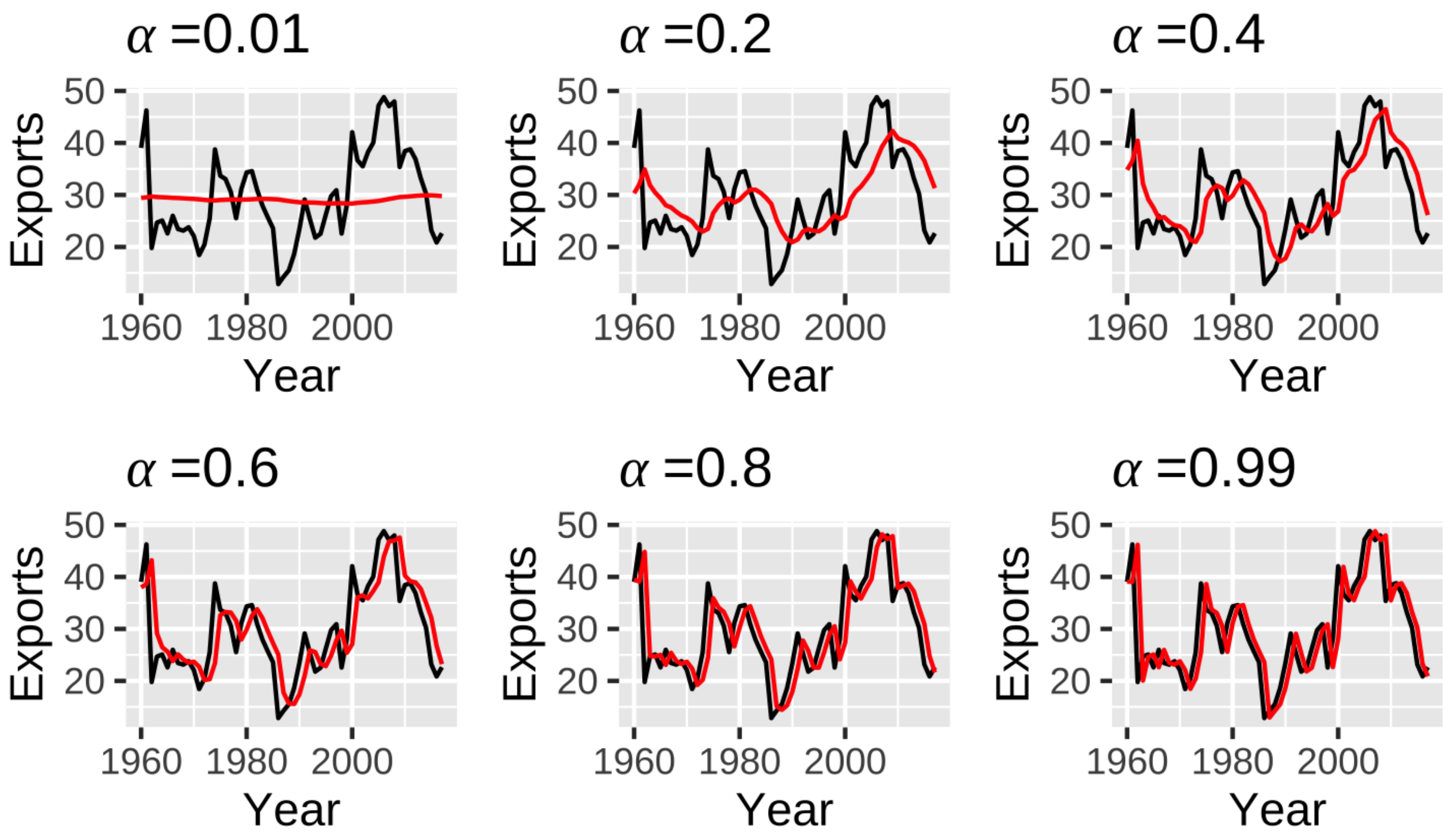
- We can rewrite this as: Forecast for next period = $\alpha\times$ (Latest observation) + $(1-\alpha)\times$ Current forecast

- $ 0 \le \alpha \le 1$ is the smoothing constant. α controls the flexibility of the level

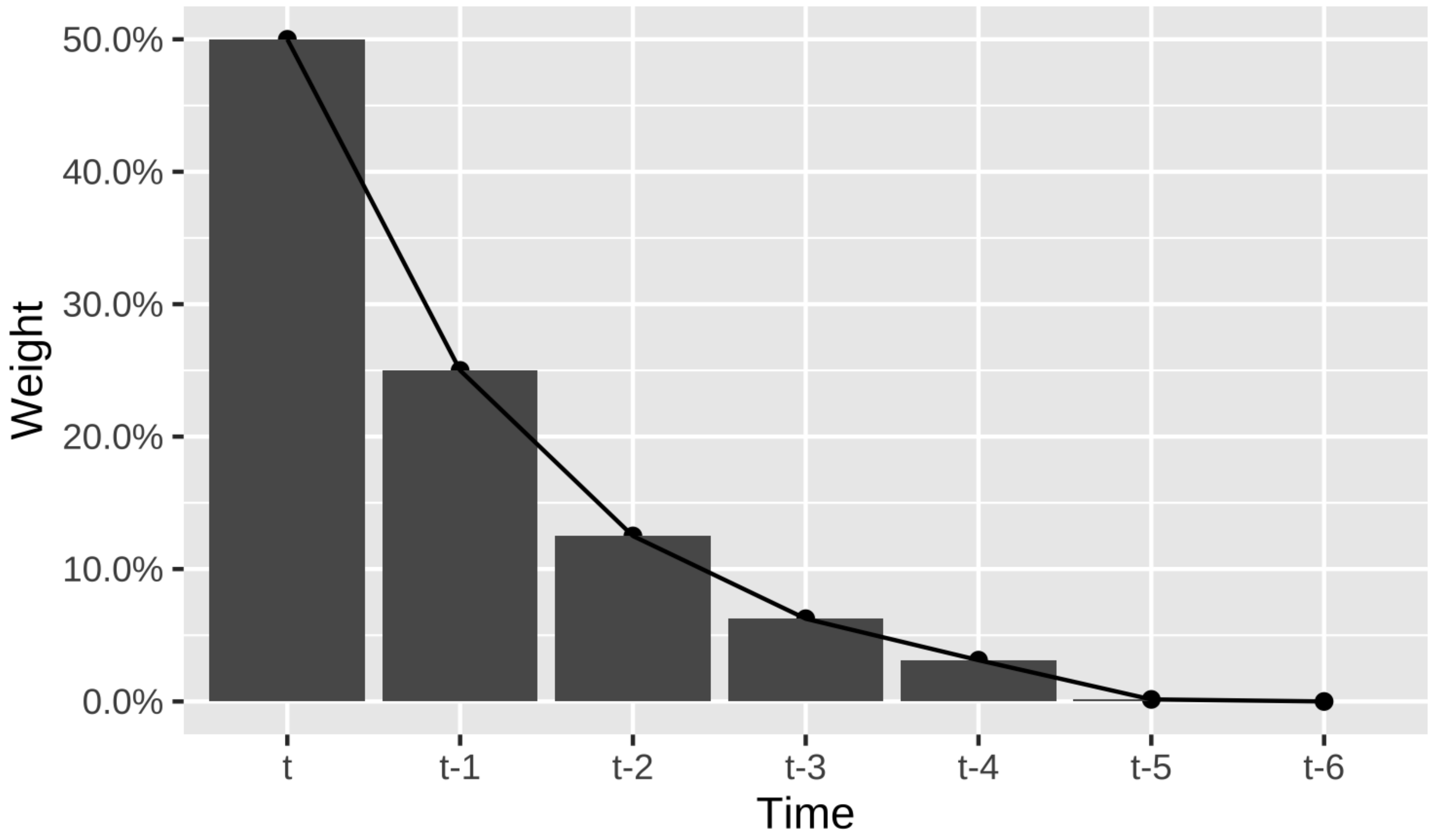
- Larger α, means more attention is given to recent observations

- If $\alpha = 0$, the level never updates (mean)
- If $\alpha = 1$, the level updates completely (naive)

How α controls the change



Exponential smoothing weights



SES: Example

Week	Sales
1	65
2	39
3	57
4	79
5	35
6	46
7	60
8	NA
9	NA
10	NA

SES: produce forecast with $\alpha = .2$

Week	Sales	Forecast
1	65	45
2	39	49
3	57	47
4	79	49
5	35	55
6	46	51
7	60	50
8	NA	52
9	NA	52
10	NA	52

- to start the process, we need an `initial value` that is the forecast for the first period, here we use 45. this could be optimised. -Forecast for next period = .2 $\times$ 65 + $(1 - .2) \times$ 45 = 49
- α could also be optimised, e.g by minimising sum of squared errors or Mean Squared Error
- SES assumes no, trend, no seasonality! so, forecasts for more than one- period-ahead are the same as the one period ahead.

Holt's method

Trend

Why Holt's method

- If there is a trend in series, SES will always forecast too low or too high depending on trend if is upward or downward,
- We need a method that cope with a trend and also detect changes in the trend as time moves on. When the trend is linear, Holt's method meets these requirements.

How Holt's method works?

It works on the same principles of SES, but when it recieves the lastes actual observation, it updates two things:

1. Its estimate of the underlying level at the current time;

2. its estimate of the current trend, (i.e. the expected change in the level between now and the next period).

Smoothing constant and parameters in Holt's method

- $0 \leq \alpha \leq 1$ to update the estimate of the level
- $0 \leq \beta \leq 1$ to update the estimate of the trend:
  - If $\beta = 0$, the trend is linear (regression trend)
  - If $\beta = 1$, the trend updates every observation
- Initial values for level and trend.

How Holt's method produces a forecast

1. Updates its estimate of the underlying level of sales at the current period using the following formula:

   - Current estimate of level = $\alpha \times$ (Latest observation) + $(1-\alpha)\times$ (previous estimate of level+previous estimate of trend)

2. The next step is to update the estimate of the trend using the following formula.

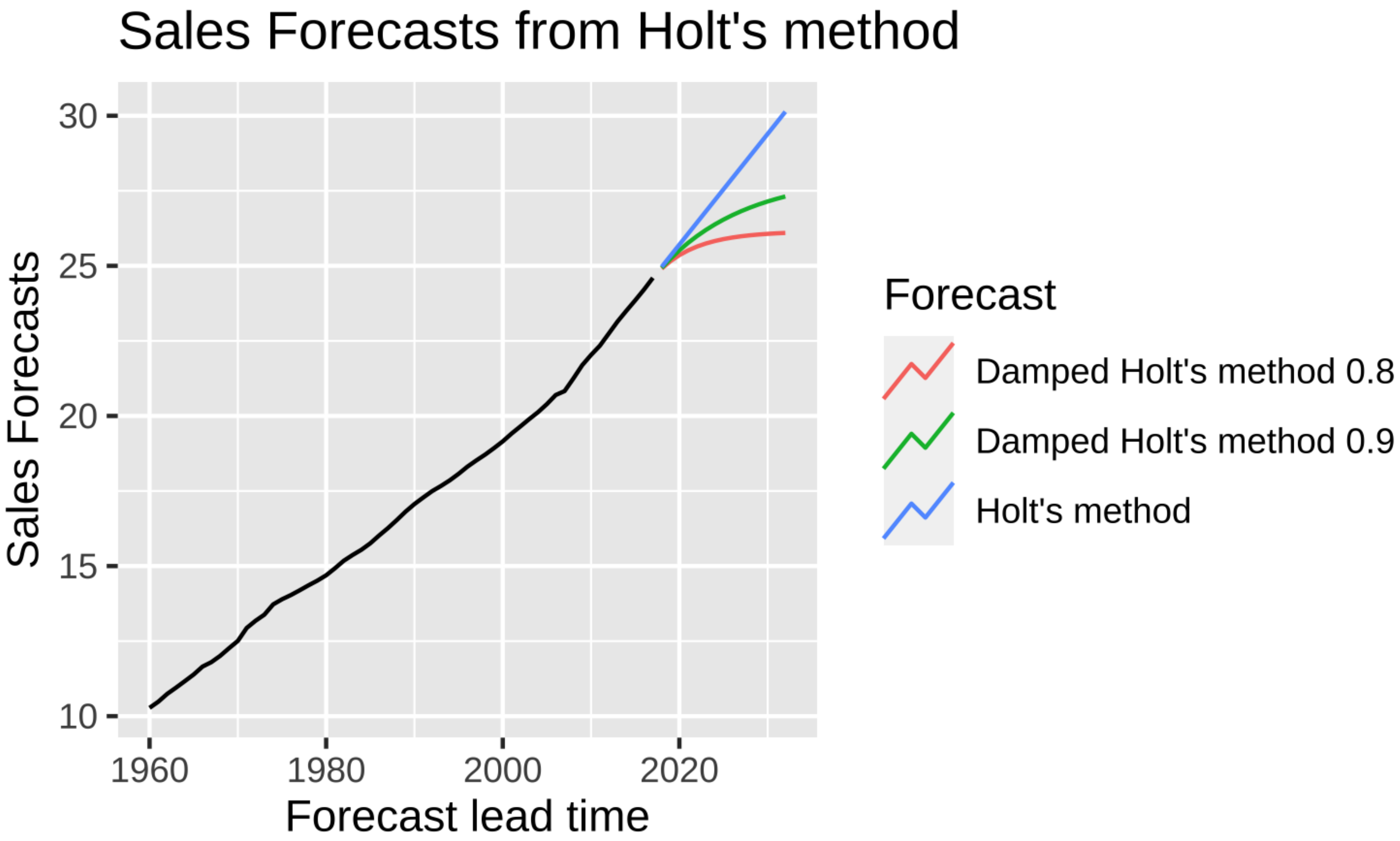
   - Current estimate of trend = $\beta\times$ (Latest observation of trend) + $(1-\beta)\times$ previous estimate of trend

3. Forecast for the next period = Current estimate of level + Current estimate of trend

The Damped Holt's method

- Holt's method assume a linear tend, its forecast imply that sales will go on increasing or decreasing depending on the type of trend forever.
- This might be unrealistic in some circumstance such as sales. This limitation led to the development of the damped Holt's method.
- It projects a trend that gradually slowing down and ulimately approaches a horizontal line.
- Damping is achieved by introducing a new paarmeter, $0 \leq \phi \leq 1$.
  - If $\phi = 0$, the trend is linear (regression trend)
  - If $\phi = 1$, the forecast is not damped at all

Forecasts produced by Holt and Damped Holt



Holt's method with an Exponential Trend

- A linear trend might involve underlying observations increasing by 3 units per period. Our underlying observations might therefore be, 100, 103, 106, etc in successive months,

- When the trend is exponential, Each successive underlying observations level might be 10% greater (i.e. 1.1 times)the previous value. Successive levels might therefore be 100, 110, 121, etc

- Forecast for h periods ahead = Current level estimate $\times$ (Current growth rate estimate) ^{h}

Holt-Winter

Trend and Seasonality

Holt-Winter method

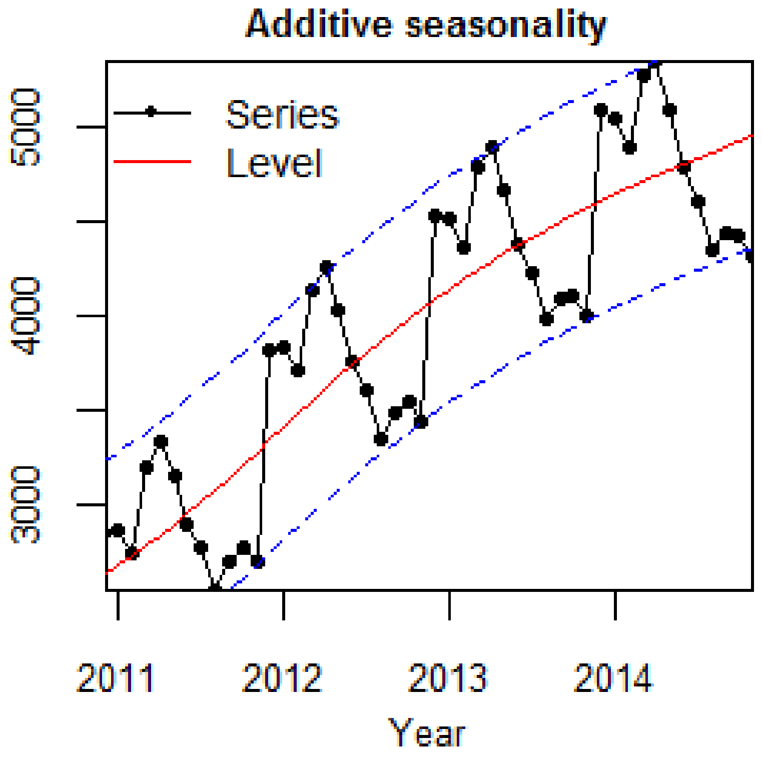
- When there is a seasonal pattern in the series, this also can be subject to changes as time goes on. Examples: a food or drink product may become fashionable at Xmas,

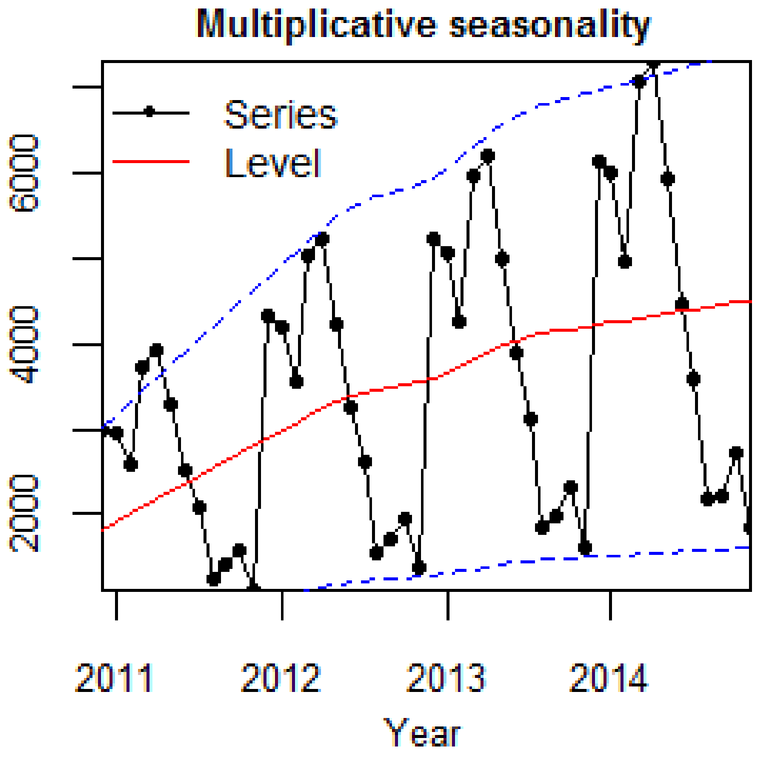
- Holt-Winter method is an extension of Holt. In addition to the updating its estimate of the underlying level and the trend, it also updates its estimate of seasonal pattern as soon as the latest observation is available.

- It does this using a new parameter, $0 \leq \gamma \leq 1$

- There are also initial seasonal indces that need to be estimated such as $S_0, S_1, \ldots$

Additive and multiplicative seasonality



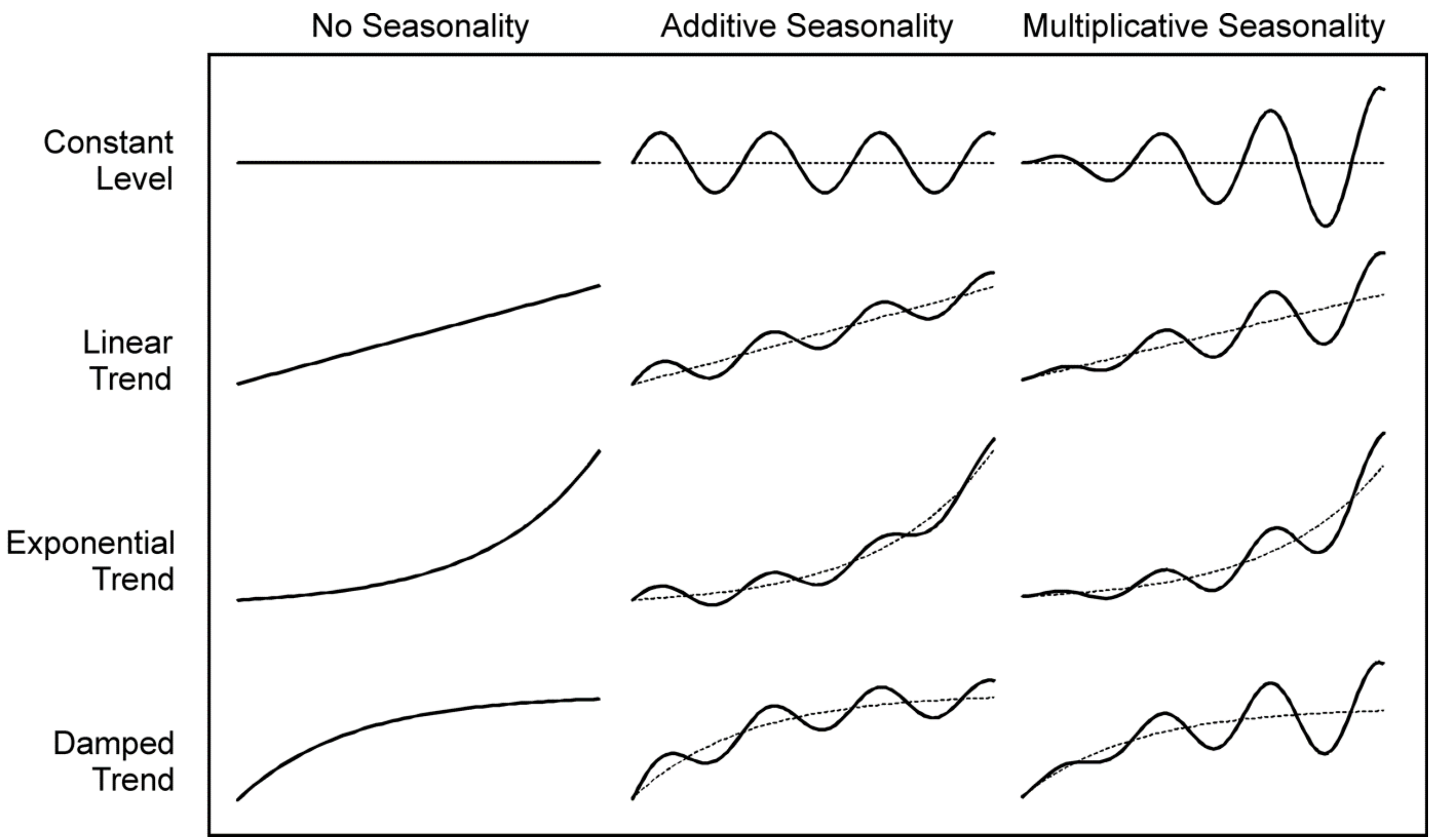


read the blog by Nikolaos Kourentzes

Holt-Winter

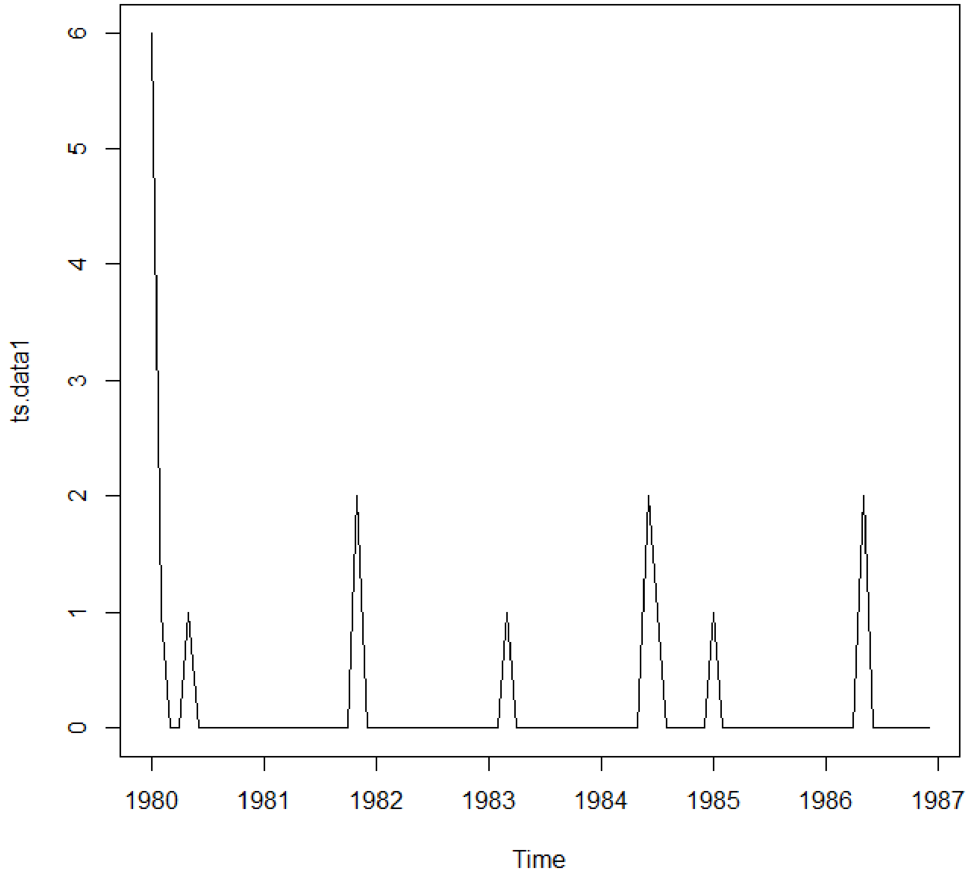
1. Additive:
   - Forecast for each season = Lates level estimate+ Latest trend estimate+ Season's seasonal index
2. Multiplicative:
   - Forecast for each season = (Lates level estimate+ Latest trend estimate) × Season's seasonal index

Summary of Exponential smoothing: Pegel's classification



Forecasting Intermittent demand

- A classic approach called `Croston` is based on SES,
- Forecast of observation = (Forecast of mean demand size / Forecast of mean interval),
- Use SES to update the size of observation,
- Use SES to update the interval between observation periods.



thank you!

- Slides are available at here
- Email rostami-tabarb@cardiff.ac.uk
- Website www.bahmanrt.com
- Twitter @Bahman_R_T