Supervised Learning: Regression, Classification & Time Series

Session 2

Bernardo Almada-Lobo, Daniel Pereira, Pedro Amorim
Lisboa, 9 a 12 de outubro 2023

ACCREDITATIONS MEMBERSHIPS RANKINGS

Biography notes – Bernardo Almada-Lobo

Academic

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Porto Business School / University of Porto

Research

INESCTEC ASSOCIATE LABORATORY PORTUGAL

IBM Centers for Advanced Studies Enterprise Engineering and Management U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

inegi driving science & innovation

Consultancy

LTP Advanced Analytics & Business Consultancy

adjustt powered by glintt

Education

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

TU/e Technische Universiteit Eindhoven University of Technology

MIT Massachusetts Institute of Technology

INSEAD The Business School for the World®

Other

FUNDAÇÃO BELMIRO DE AZEVEDO

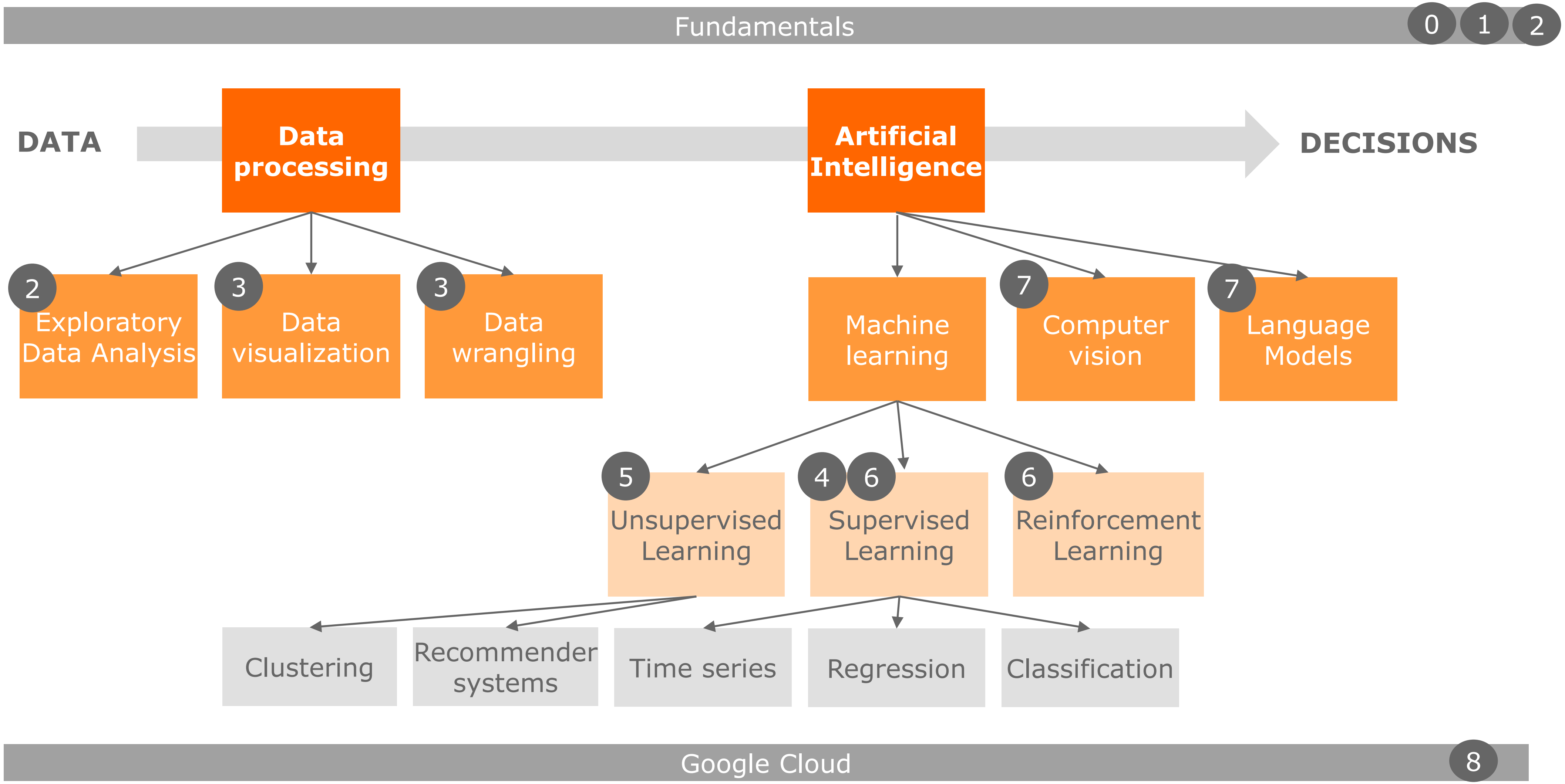
CAP CERTIFIED ANALYTICS PROFESSIONAL

EXPLORER INVESTMENTS

BA

BPI Grupo CaixaBank

Academia DTI



Fundamentals 0 1 2

DATA → DECISIONS

Data processing

- 2 Exploratory Data Analysis
- 3 Data visualization
- 3 Data wrangling

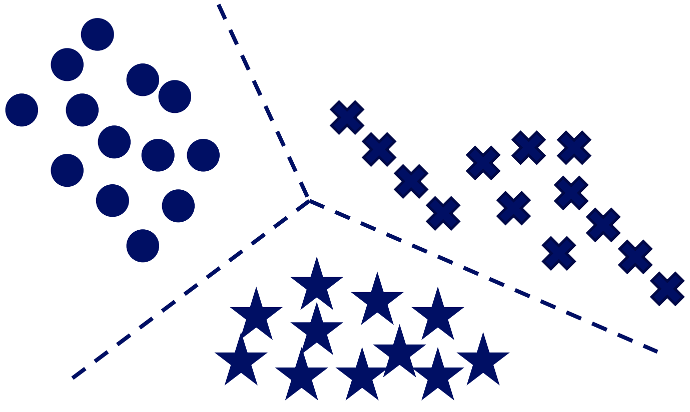
Artificial Intelligence

- Machine learning
  - 5 Unsupervised Learning
    - Clustering
    - Recommender systems
  - 4 6 Supervised Learning
    - Time series
    - Regression
    - Classification
  - 6 Reinforcement Learning
- 7 Computer vision
- 7 Language Models

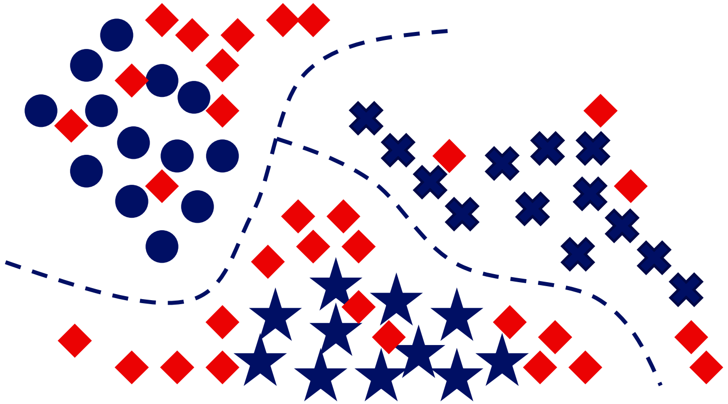
Google Cloud 8

Algorithms can be classified based on the way they "learn" about data to make predictions

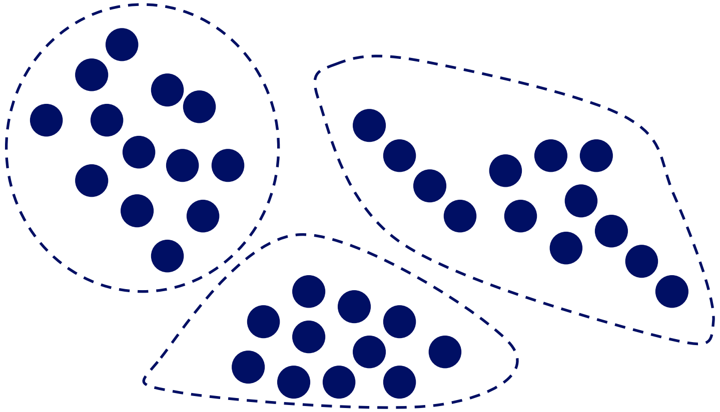
Supervised learning



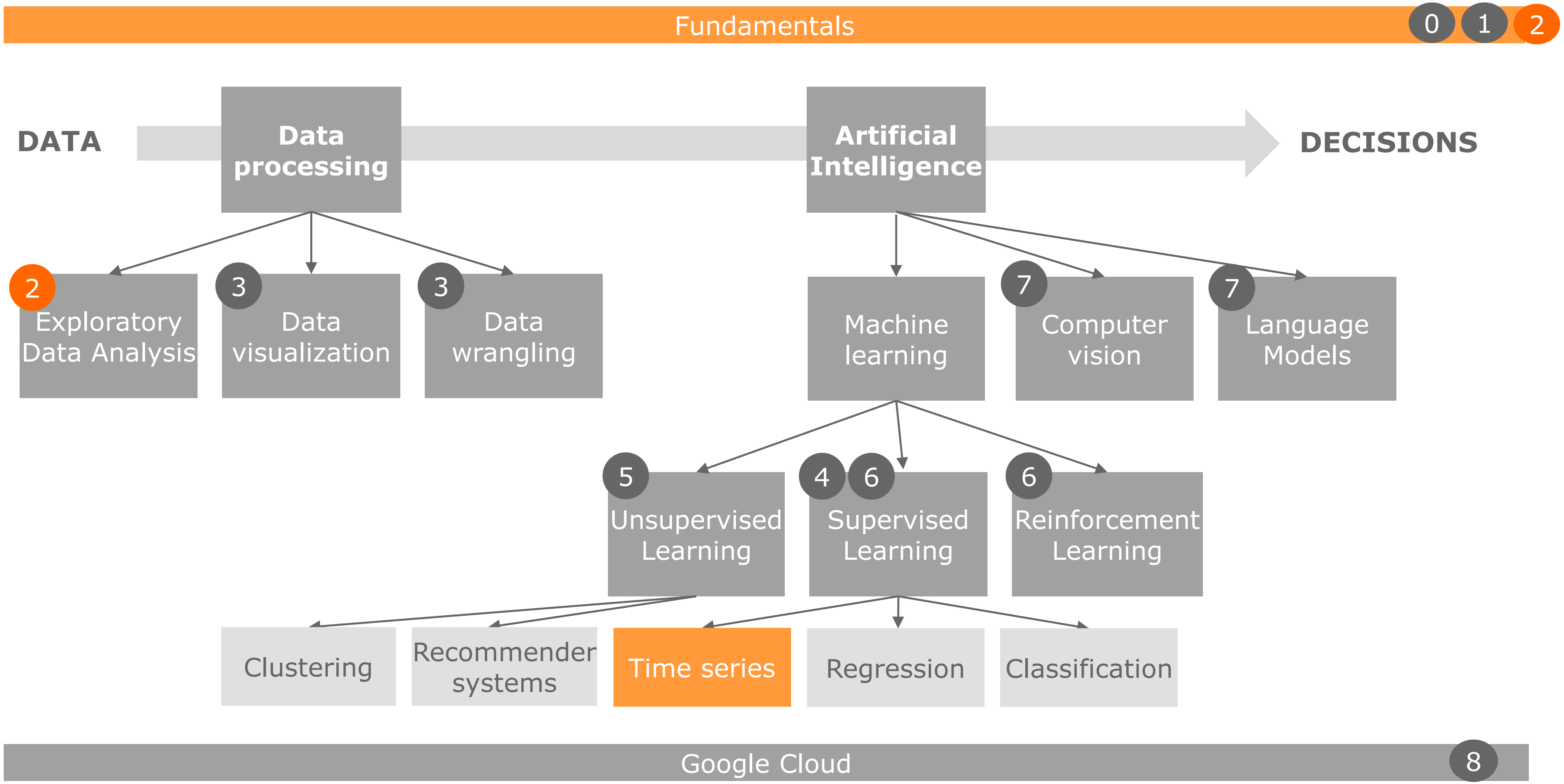
Semi-supervised learning



Unsupervised learning



Academia DTI



Fundamentals 0 1 2

DATA

Data processing

Artificial Intelligence

DECISIONS

2 Exploratory Data Analysis

3 Data visualization

3 Data wrangling

Machine learning

7 Computer vision

7 Language Models

5 Unsupervised Learning

4 6 Supervised Learning

6 Reinforcement Learning

Clustering

Recommender systems

Time series

Regression

Classification

Google Cloud 8

Agenda

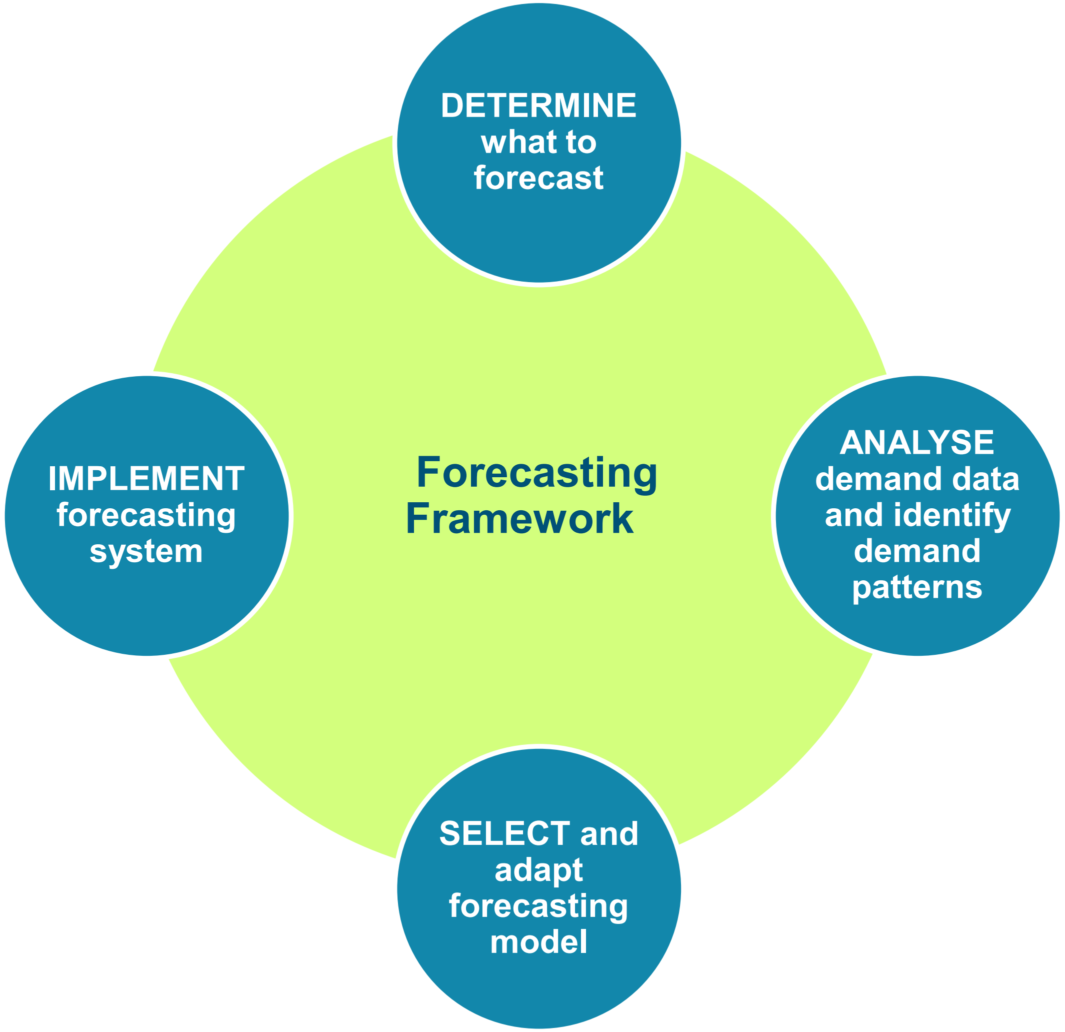
- **The Role of Time Series**
- Time Series Methods: Exponential Smoothing
- Accuracy, Outliers and Aggregation
- Hybrid Methods
- Box & Jenkins

Why is Forecasting Important?

- Defining more precisely how ***UNCONTROLLABLE VARIABLES*** (relevant for decision process at stake) will behave in the future
- Lags in Decision Making: If we could always adjust instantaneously and costlessly to new conditions there would be no need for forecasts
- It is not an exact science; one must blend **experience, judgment, and technical expertise**

The six Steps in the Forecasting process

- **Step 1** Identify the goal of the forecast
- **Step 2** Establish a time horizon and the lag
- **Step 3** Select a forecasting technique
- **Step 4** Conduct the forecast (analyze data)
- **Step 5** Determine its accuracy
- **Step 6** Monitor the forecast



Forecasting Framework

DETERMINE what to forecast

ANALYSE demand data and identify demand patterns

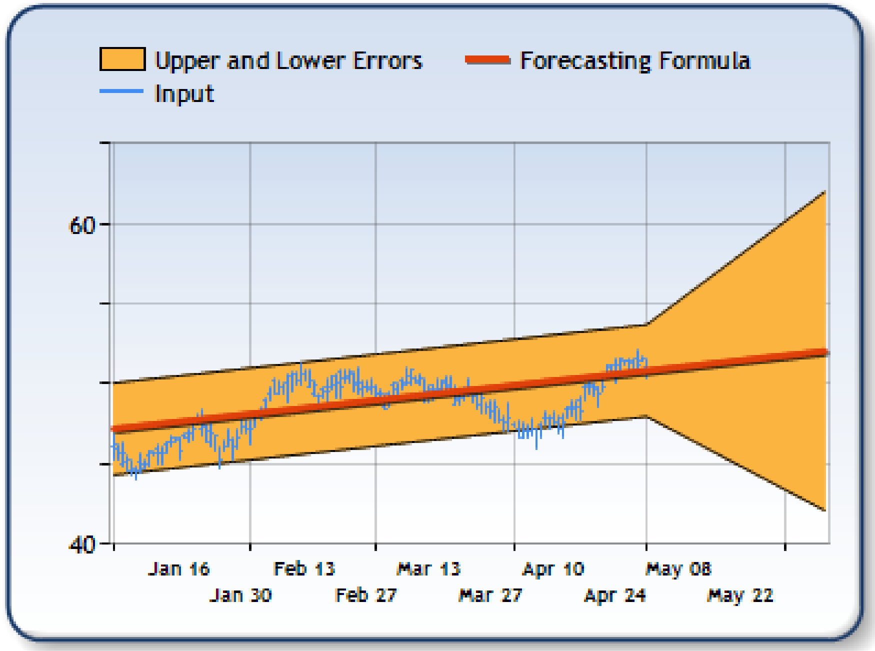
SELECT and adapt forecasting model

IMPLEMENT forecasting system

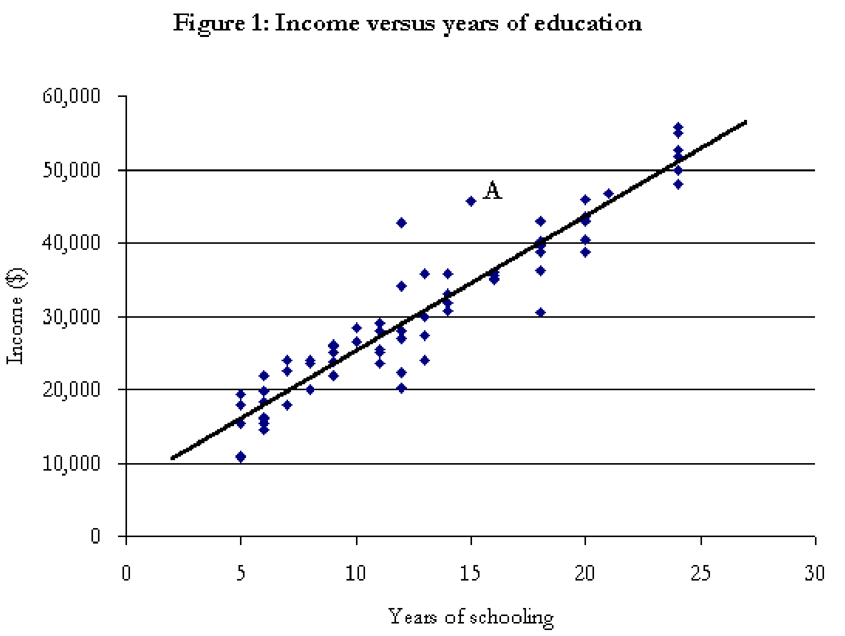
Forecasting Quantitative Methods

"It is far better to foresee even without certainty than not to foresee at all. "

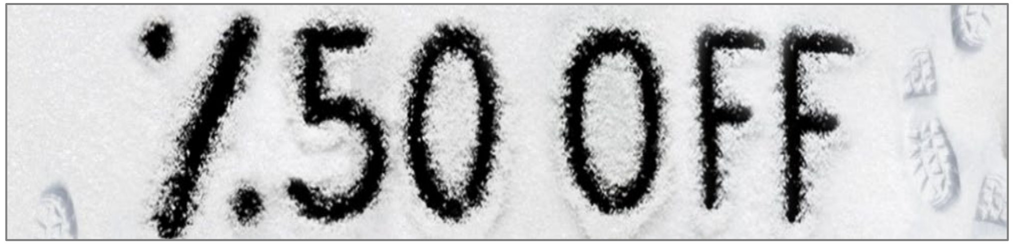
Direct Extrapolation Methods



Casual Methods



Forecasting Quantitative Methods



Source: Invent Analytics

Agenda

The Role of Time Series

Time Series Methods: Exponential Smoothing

Accuracy, Outliers and Aggregation

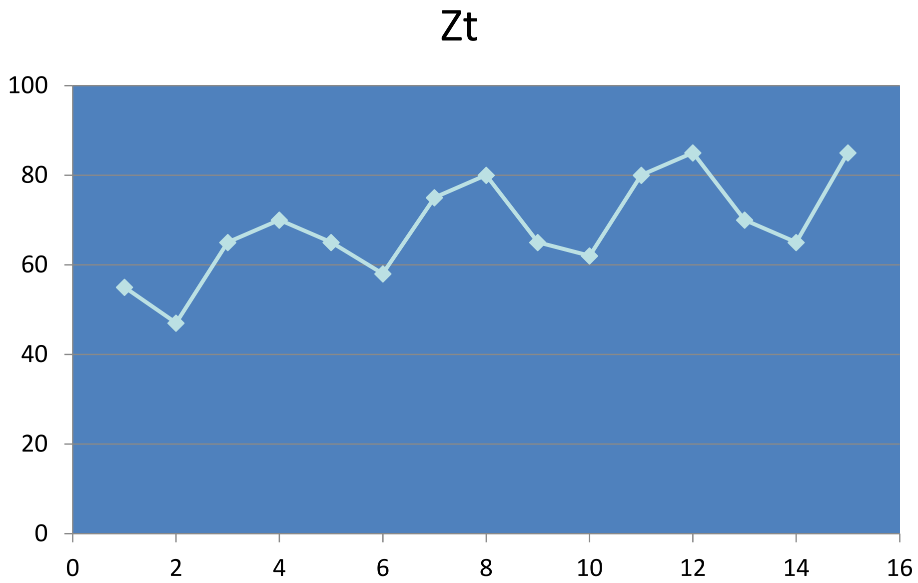
Hybrid Methods

Box & Jenkins

Exercise

Consider the following time series:

t	*Zt*
1	55
2	47
3	65
4	70
5	65
6	58
7	75
8	80
9	65
10	62
11	80
12	85
13	70
14	65
15	85
16	90

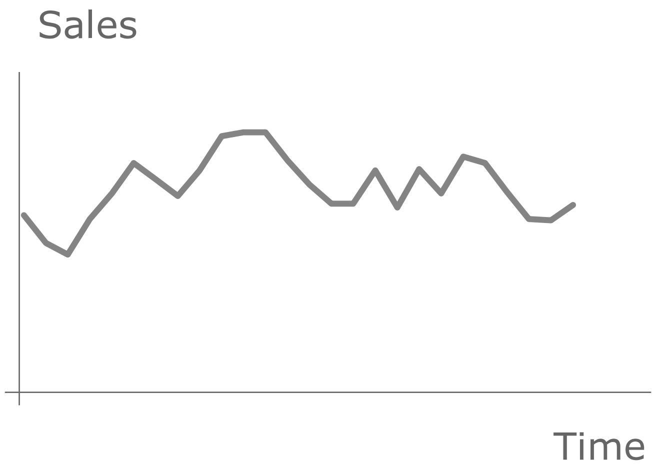


- What is the forecast for t=17?

A time series is composed of 3 main components: level, trend and seasonality

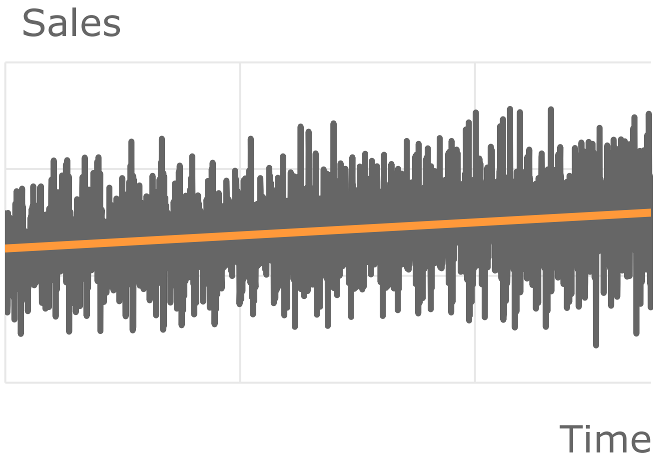
Decomposition models

Level



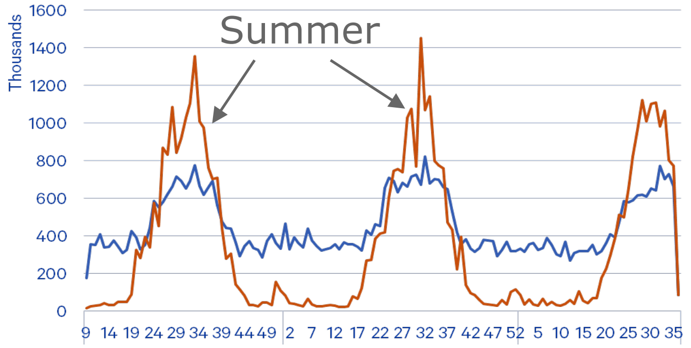
Product expected average sales

Trend



Time series evolution over time

Seasonality



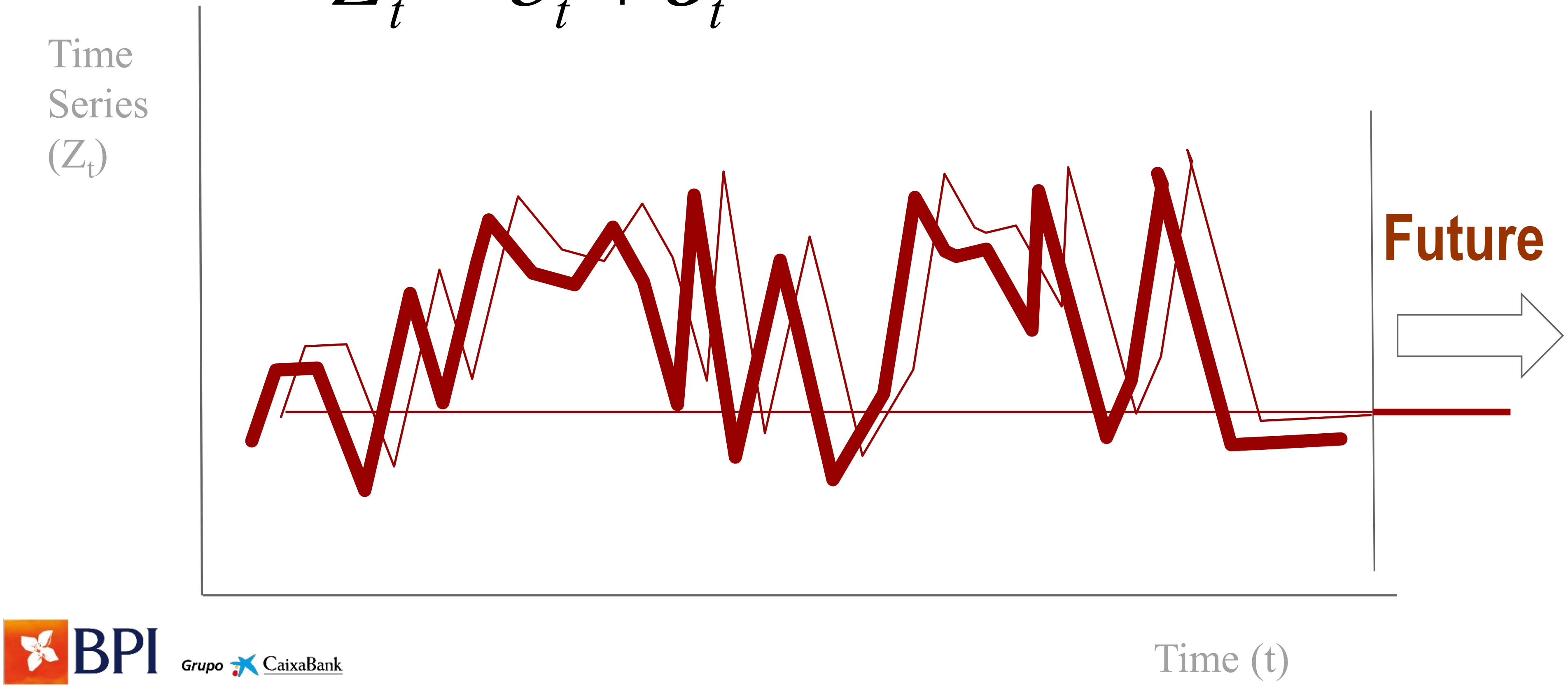
- **Sales patterns** on different periods – year, month, week

A decomposition model projects the impact of past **trend** and **seasonality** (additive or multiplicative) on a current sales **level** in order to forecast future sales

Exponential Smoothing Methods

Series Without Trend and Without Seasonality

$$Z_t = \upsilon_t + \varepsilon_t$$



Exponential Smoothing Methods

Local Stationary Time Series

Simple Moving Average Method

The level is estimated based on the last N obervations

$$n_t = (Z_t + Z_{t-1} + \cdots + Z_{t-N+1}) / N$$

The only parameter to fix is the number of terms of the moving average (N).

One could compute the value of N that minimizes the Mean Square Error of One Step Forecasts.

$$\text{EQM} = \frac{1}{T} \cdot \sum_{t=N+1}^{N+T} e_t^2 = \frac{1}{T} \cdot \sum_{t=N+1}^{N+T} \left[Z_t - \hat{Z}_{t-1}(1)\right]^2$$

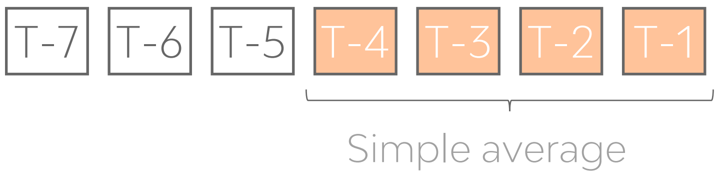
Exponential smoothing shows valuable advantages when compared to moving average

Moving averaging models

Illustration

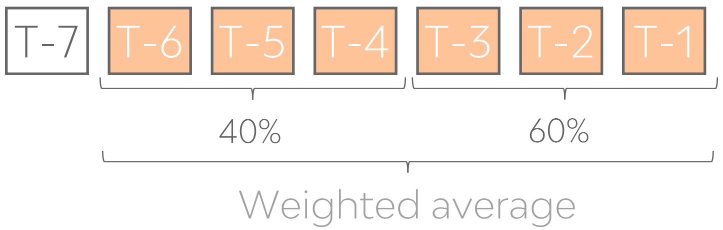
Pros/Cons

Moving average



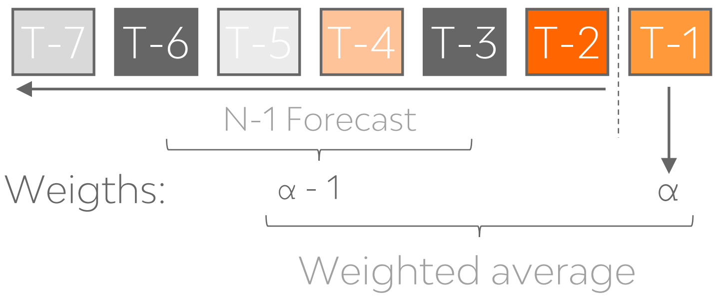
- ⊕ Simple method with rolling horizon
- ⊖ Does not differentiate last N observations
- ⊖ Ignores previous observations

Weighted moving average



- ⊕ Most recent observations are “heavier” (weight)
- ⊖ Differentiates last N observations randomly
- ⊖ Ignores previous observations

Exponential Smoothing



- ⊕ “Corrects”, at each observation, the forecast value
- ⊕ Most recent observations are exponentially “heavier”
- ⊕ Uses all past observations
- ⊕ Only requires the last observation and the last forecast to be kept
- ⊖ Implies parametrizing α

Exponential Smoothing Methods

Series Without Trend and Without Seasonality

Forecasts

$$\hat{Z}_t(1) = \hat{Z}_t(2) = \ldots = \hat{Z}_t(k) = \ldots = n_t = \hat{\upsilon}_t$$

$\hat{Z}_t(k)$: Forecast of Z_{t+k} made in instant *t*, after knowing the data Z_t .

n_t : Estimative of time series level in instant *t*.

Simple Exponential Smoothing

$$n_t = \alpha \cdot Z_t + (1-\alpha) \cdot n_{t-1} \qquad (0 \le \alpha \le 1)$$

Since $\hat{Z}_t(1) = n_t$

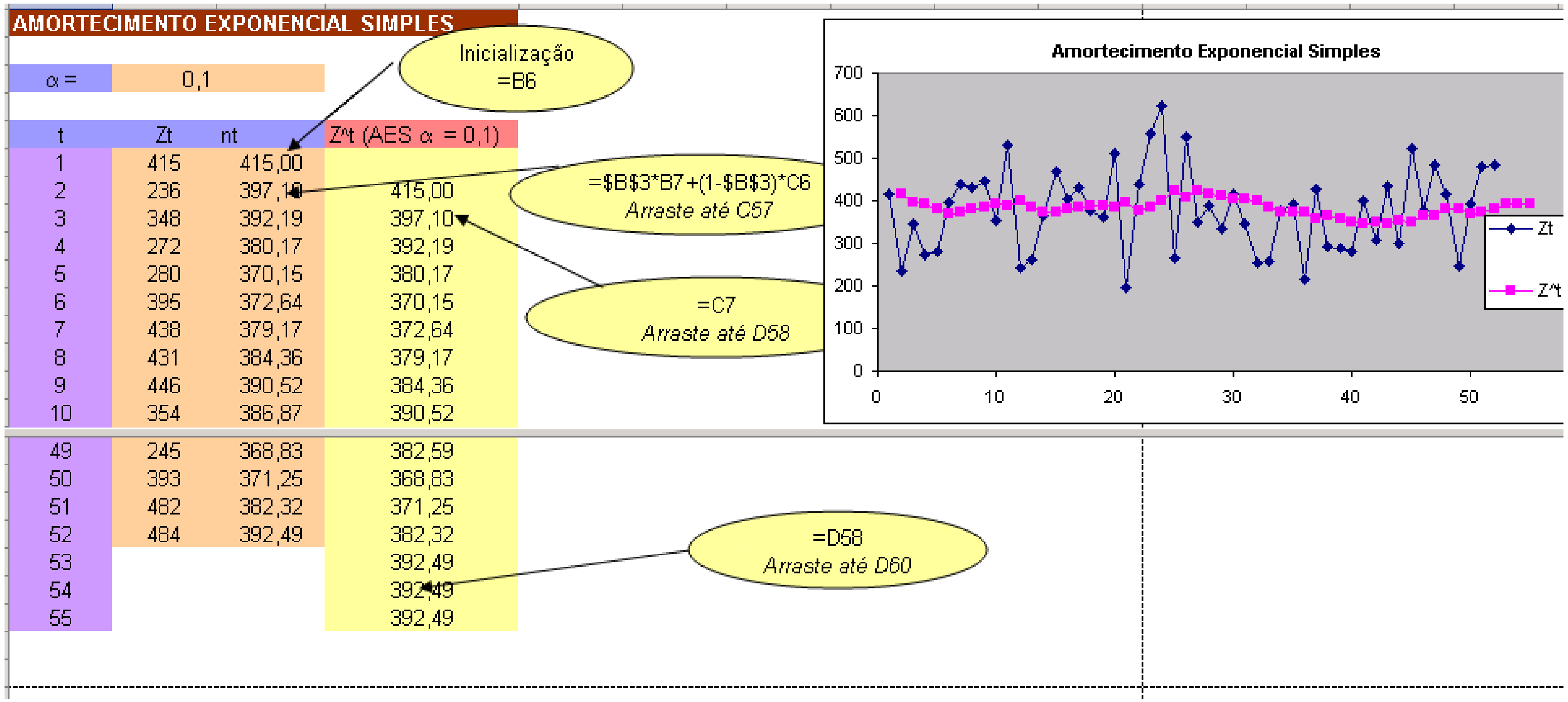
$$\begin{aligned} \hat{Z}_t(1) &= \alpha \cdot Z_t + (1-\alpha) \cdot \hat{Z}_{t-1}(1) \\ &= \hat{Z}_{t-1}(1) + \alpha \cdot \left[Z_t - \hat{Z}_{t-1}(1) \right] \\ &= \hat{Z}_{t-1}(1) + \alpha \cdot e_t \qquad (0 \le \alpha \le 1) \end{aligned}$$

α : Smoothing rate

e_t : Forecast error in instant *t*

Exponential Smoothing Methods

Simple Exponential Smoothing



AMORTECIMENTO EXPONENCIAL SIMPLES

α = 0,1

t	Zt	nt	Z^t (AES α = 0,1)
1	415	415,00	
2	236	397,10	415,00
3	348	392,19	397,10
4	272	380,17	392,19
5	280	370,15	380,17
6	395	372,64	370,15
7	438	379,17	372,64
8	431	384,36	379,17
9	446	390,52	384,36
10	354	386,87	390,52
49	245	368,83	382,59
50	393	371,25	368,83
51	482	382,32	371,25
52	484	392,49	382,32
53			392,49
54			392,49
55			392,49

Inicialização
=B6

=B3*B7+(1-B3)*C6
Arraste até C57

=C7
Arraste até D58

=D58
Arraste até D60

Smoothing factors determine how reactive forecast model is towards demand variations

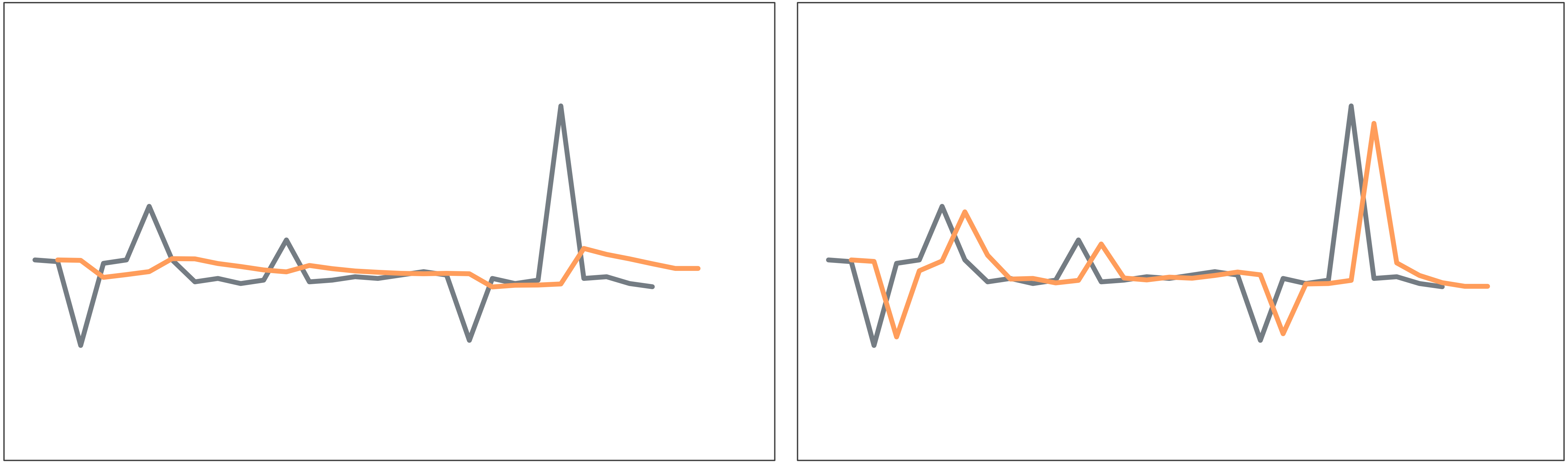
Exponential smoothing models

Less Reactive

More Reactive

0 ◀ Smoothing Factor ▶ 1

Example



Exponential Smoothing Methods

Local stationary Time Series: Tracking Signal

Adaptive Exponential Smoothing de Trigg & Leach

$$TS_t = \frac{EA_t}{EAA_t}$$

$$EA_t = \beta \cdot e_t + (1-\beta) \cdot EA_{t-1} \qquad \text{(Smoothed Error)}$$

$$EAA_t = \beta \cdot |e_t| + (1-\beta) \cdot EAA_{t-1} \qquad \text{(Absolute Smoothed Error)}$$

$$\mathrm{e}_{\mathrm{t}} = Z_t - \hat{Z}_{t-1}(1) \qquad \text{(Forecast Error)}$$

β : *smoothing rate* (constant) of forecast error and absolute forecast error

$$\alpha_{\mathrm{t}} = |TS_t|$$

Naive Forecasts (equal to alfa=1)

✓A naive forecast for any period equals the previous period’s actual value

✓ Low cost, easy to prepare, easy to understand, but less accurate forecasts

✓ Can be applied to seasonal or trend data

Examples:

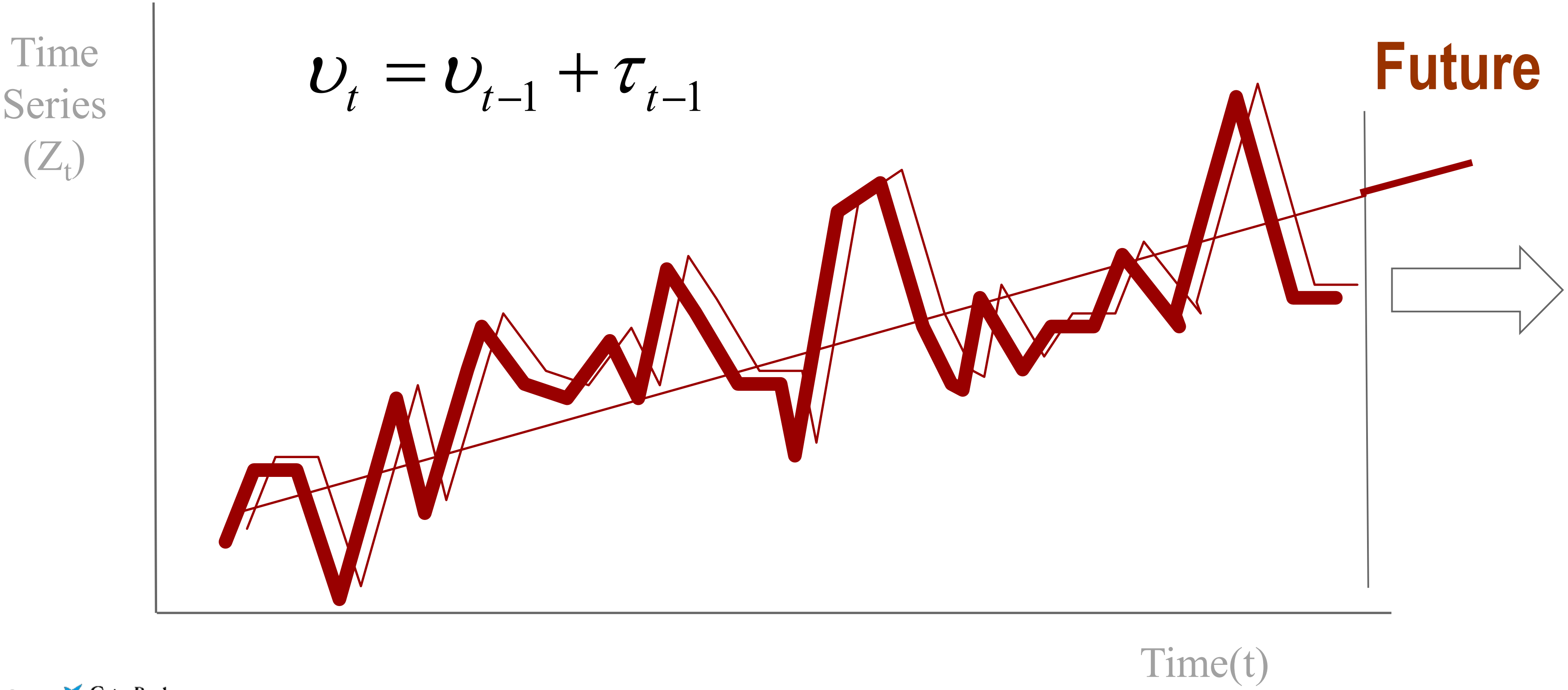
- If last week’s demand was 50 units, the naive forecast for the coming week is 50 units.
- If seasonal pattern exists, the naive forecast for *next January* would equal the actual demand for *January of this year*.

Exponential Smoothing Methods

Series With Trend and Without Seasonality

$$Z_t = \upsilon_t + \varepsilon_t$$

$$\upsilon_t = \upsilon_{t-1} + \tau_{t-1}$$



Exponential Smoothing Methods
Series With Trend and Without Seasonality

Forecasts

$$\hat{Z}_{\mathrm{t}}(k) = n_t + b_t \cdot k$$

$\hat{Z}_t(k)$: Forecast of $Z_{\mathrm{t+k}}$ made in instant *t*, after knowing the data Z_{t}

$n_t = \hat{\upsilon}_t$: Estimate of the level of the series in instant *t*.

$b_t = \hat{\tau}_t$: Estimate of the trend of the series

Linear Holt Exponential Smoothing

$$n_t = \alpha \cdot Z_t + (1-\alpha) \cdot (n_{t-1} + b_{t-1}) \qquad 0 \le \alpha \le 1$$

$$b_t = \beta \cdot (n_t - n_{t-1}) + (1-\beta) \cdot b_{t-1} \qquad 0 \le \beta \le 1$$

$$\hat{Z}_{\mathrm{t}}(k) = n_t + b_t \cdot k$$

Initialization:

$$\mathrm{b}_1 = 0$$

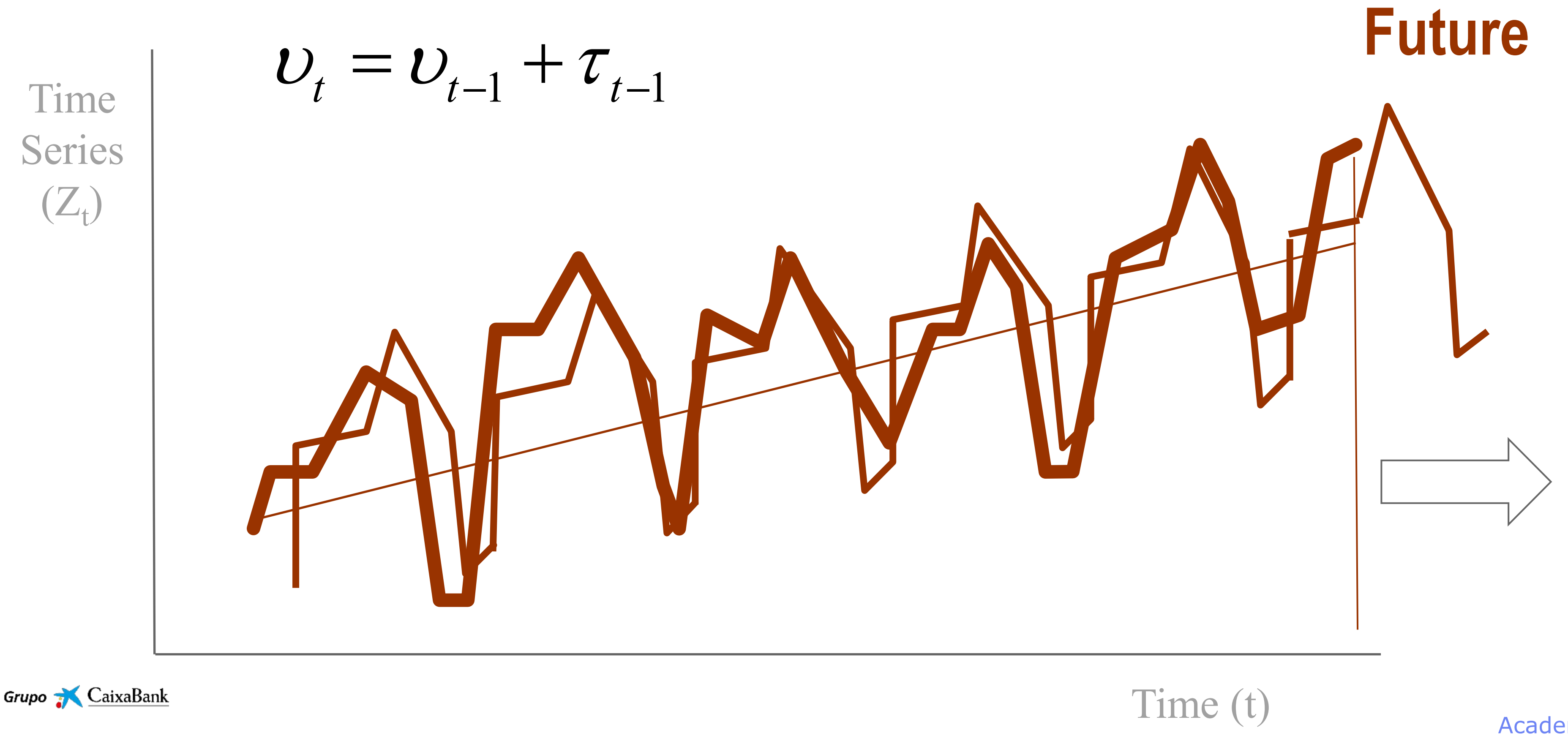
$$\mathrm{n}_1 = Z_1$$

Exponential Smoothing Methods

Series With Trend and Seasonality

$$Z_t = \upsilon_t + \phi_t + \varepsilon_t$$

$$\upsilon_t = \upsilon_{t-1} + \tau_{t-1}$$

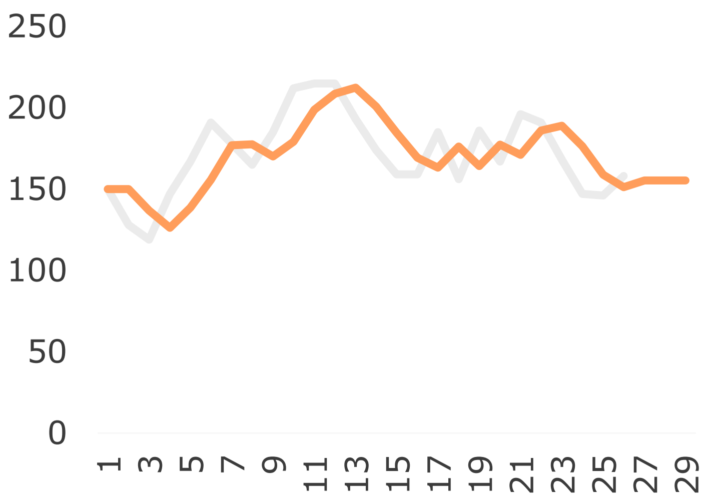


Besides level, ES models may consider trend and seasonality as well

Exponential smoothing models

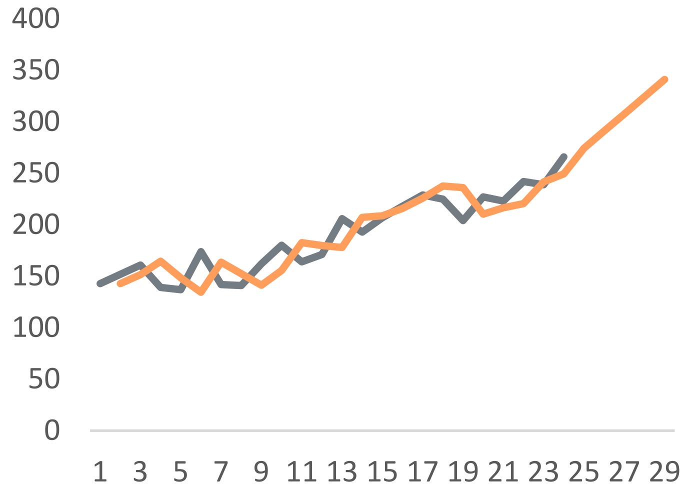
— Sales — Forecast

Simple Exponential Smoothing (*Simple ES*)



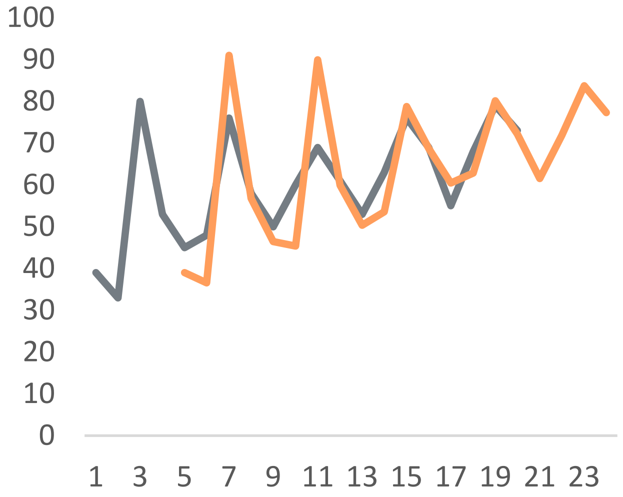
Level (α)

Holt's Method (*Trend ES*)



Level (α) | Trend (β)

Holt Winters' Method (*Seasonal ES*)



Level (α) | Trend (β) | Seasonality (γ)

Exponential Smoothing Methods
HoltWinters: Series With Trend and Seasonality

Forecasts

$$\hat{Z}_{\mathrm{t}}(k) = n_t + b_t \cdot k + f_{t+k-s}$$

$\hat{Z}_{\mathrm{t}}(k)$	Forecast of $Z_{\mathrm{t+k}}$ made in instant *t*, after knowing the data Z_{t}
$n_t = \hat{\upsilon}_t$	Estimate level of series in instant *t*.
$b_t = \hat{\tau}_t$	Estimate of the trend of the series.
$f_{t+k-s} = \hat{\phi}_{t+k}$	Estimate of the seasonal component for instant *t*+*k*.

Holt-Winters Method – Additive Model

$$n_t = \alpha \cdot (Z_t - f_{t-s}) + (1-\alpha) \cdot (n_{t-1} + b_{t-1}) \qquad 0 \le \alpha \le 1$$

$$b_t = \beta \cdot (n_t - n_{t-1}) + (1-\beta) \cdot b_{t-1} \qquad 0 \le \beta \le 1$$

$$f_t = \gamma \cdot (Z_t - n_t) + (1-\gamma) \cdot f_{t-s} \qquad 0 \le \gamma \le 1$$

$$\hat{Z}_{\mathrm{t}}(k) = n_t + b_t \cdot k + f_{t+k-s} \text{ , for } k = 1, 2, ..., \mathrm{s}$$

$$\hat{Z}_{\mathrm{t}}(k) = n_t + b_t \cdot k + f_{t+k-2s} \text{ , for } k = s+1, \mathrm{s}+2, ..., 2\mathrm{s}$$

Initialization:

$i)\quad n_s = Z^* = \frac{1}{s} \cdot \sum_{t=1}^{s} Z_t$

$ii)\quad b_s = 0$

$iii)\quad f_j = Z_j - Z^* \quad (j = 1, ..., \mathrm{s})$

Croston Method for Intermittent Demand

When a product experiences several periods of zero demand. Often demand is small, and sometimes highly variable in size.

Separates intermittent data into two components

Non-zero demand time

Inter-demand interval time series

Both series are forecasted independently using single exponential smoothing (SES)

Demand Size

Inter-arrival time

Dividing resulting estimates produces the final forecast

Average Demand per time period

$$If\ z_t = 0,\ then$$

$$z'_t = z'_{t-1},$$

$$p'_t = p'_{t-1}$$

$$Otherwise$$

$$z'_t = \alpha z_t + (1-\alpha) z'_{t-1}$$

$$p'_t = \alpha p_t + (1-\alpha) p'_{t-1}$$

$$where\ 0 < \alpha < 1$$

Interpreting Croston Results

Sample outputs from Croston

Demand	Interval
3.97	1.41

Represents Demand Size

Represents Inter-arrival time

In the next 1.4 days we are likely to get demand for 3.9 units of the forecasted

Assuming we are forecasting for next 5 days, below will be the output for average demand for these 5 days.

2.81	2.81	2.81	2.81	2.81

3.97 / 1.41

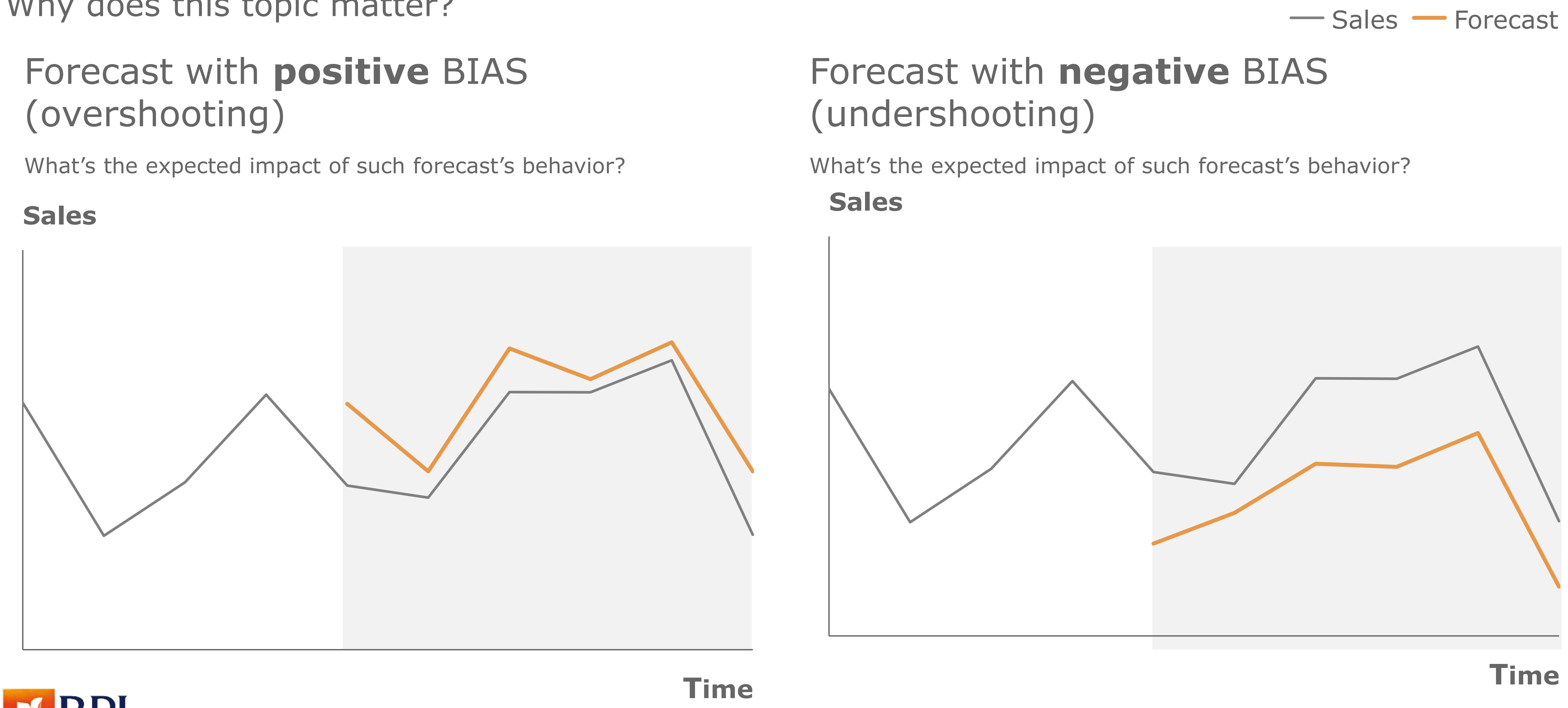
For next 5 days, we are likely to get demand for 14 (2.81 * 5)

Agenda

- The Role of Time Series
- Time Series Methods: Exponential Smoothing
- **Accuracy, Outliers and Aggregation**
- Hybrid Methods
- Box & Jenkins

Forecast's error impacts the entire planning activities and lead to future uncertainty

Why does this topic matter?



Forecast errors impacts companies in different ways

Example of a "fresh" food retailer

Stockouts

Lack of Visibility

Depreciated sales

Shrinkage

Sales Underestimation

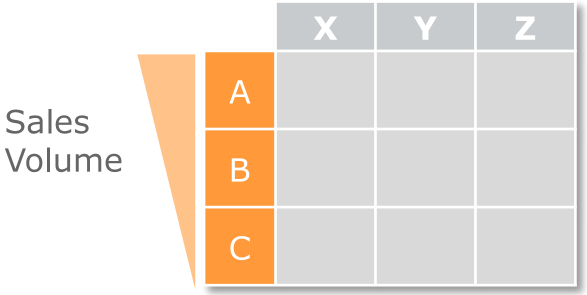
Sales Overestimation

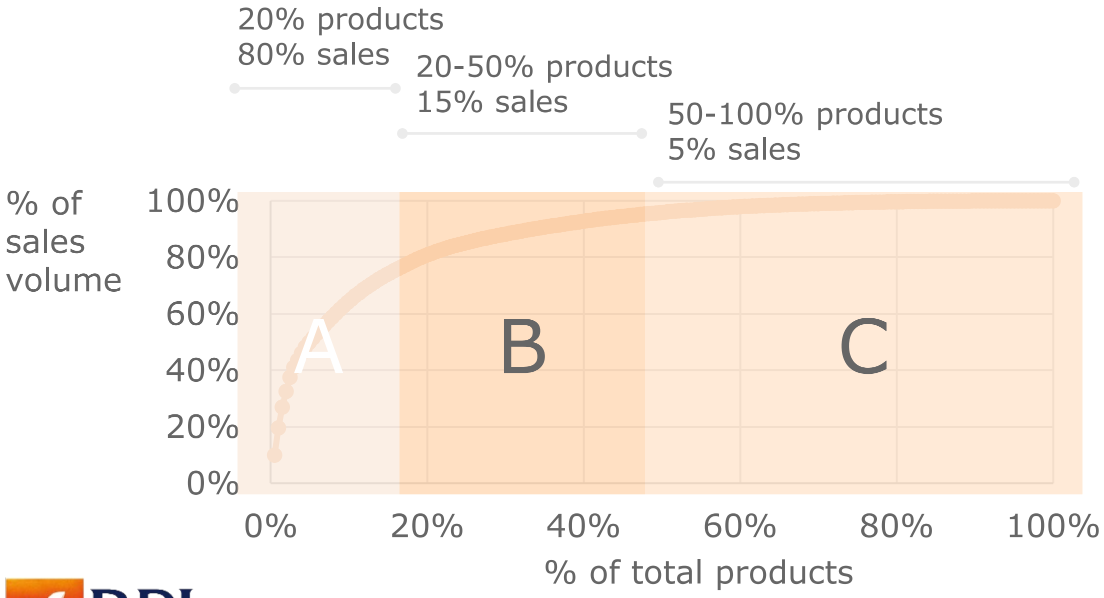
- **Lost sales** potential
- **Reduced** customer loyalty and **satisfaction**

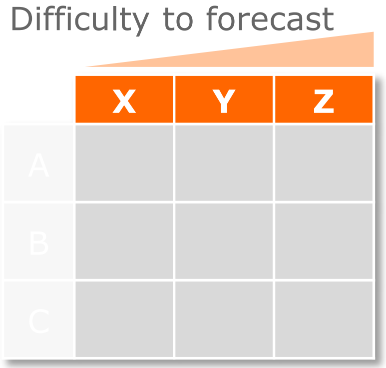
- Unnecessary **stock** level **increase**
- **Increase** in capital, transportation and warehousing **costs** regarding **stock**
- **Higher risk of expiring** validity (shrinkage)

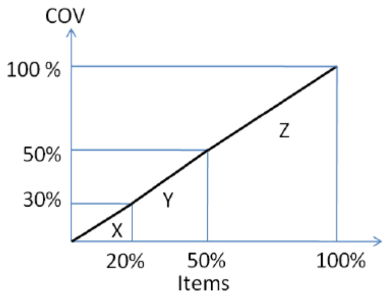
An ABC/XYZ analysis allows to differentiate products based on their relevance in sales and difficulty to forecast

Which products should be forecasted by an analytical model?



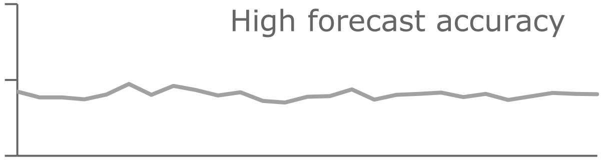






$$COV = \frac{\text{Standard Deviation of Demand}}{\text{Mean of Demand}}$$

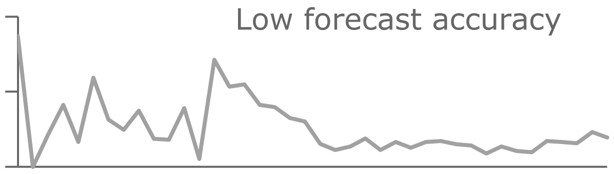
X Relatively stable, small fluctuations or with a clear pattern


High forecast accuracy

Y More significant fluctuations

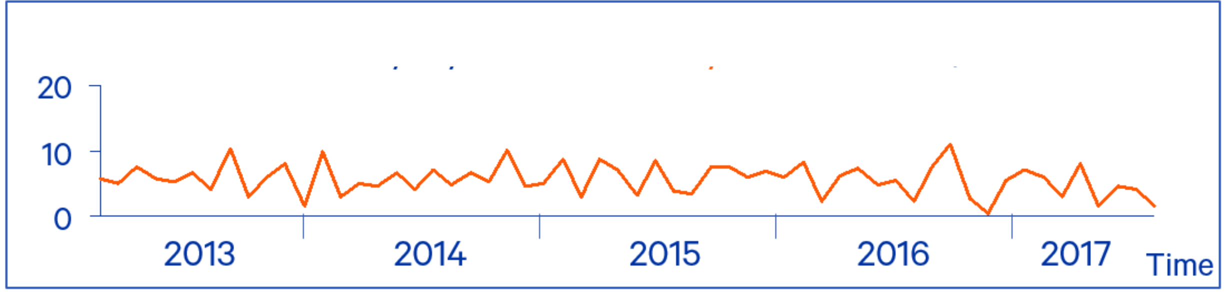
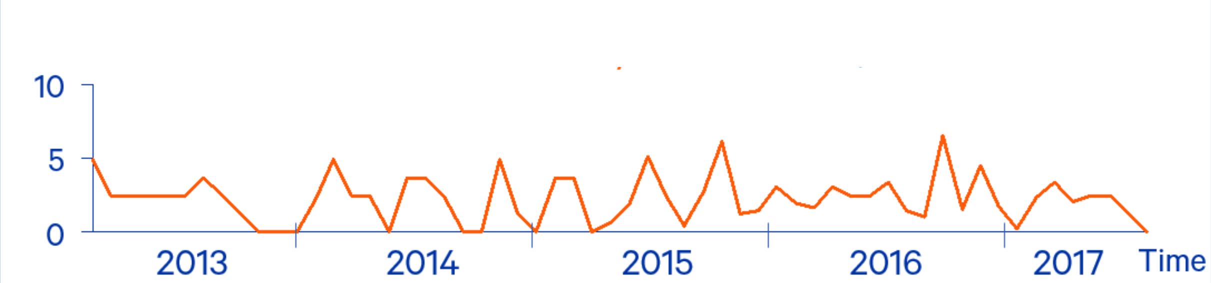
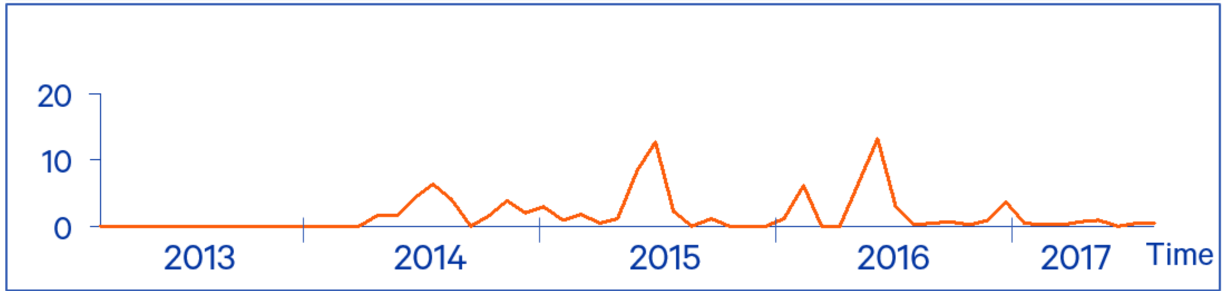

Medium forecast accuracy[1]

Z Very intense and intermittent fluctuations, no clear pattern


Low forecast accuracy

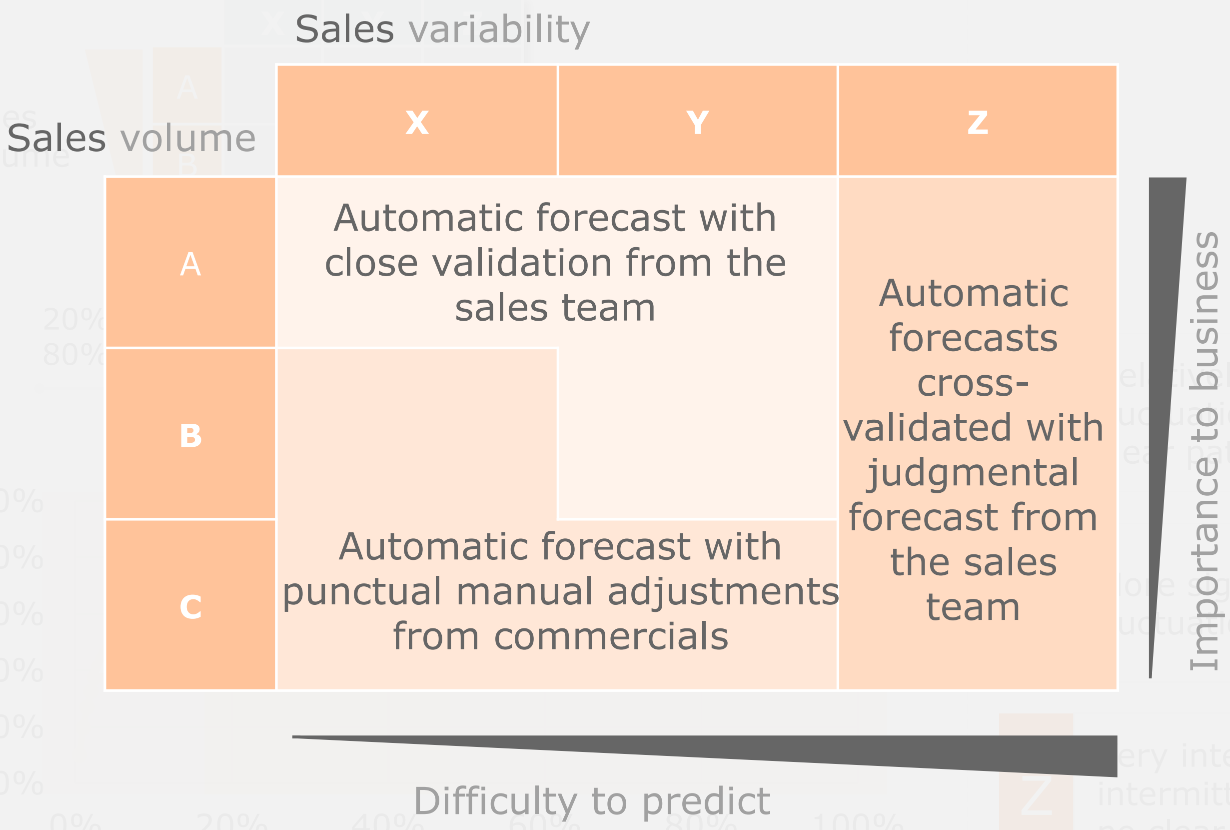
An XYZ analysis differentiates products over their variability and thus the difficulty of forecast generation

Example#1

	% Products	% Sales	Criteria	
X	12%	54%	Relatively stable with small fluctuations (CV[1] < 70%)	
Y	13%	19%	Fluctuations are common and more impactful (70% < CV[1] < 90%)	
Z	75%	27%	Intermittent or extremely fluctuant (CV[1] > 90%)	

An ABC/XYZ analysis allows to differentiate products based on their relevance in sales and difficulty to forecast

Forecast evaluation



Automatic forecast with close validation from the sales team

Automatic forecast with punctual manual adjustments from commercials

Automatic forecasts cross-validated with judgmental forecast from the sales team

Products with high demand (A or B) and little to moderate variation of demand over time (X or Y) are **generally easier to predict**, with greater accuracy and business impact

Exercise ABC/XYZ: Italian Retailer

Open Sales Data.xls

Commercial inputs are increasingly important for Y and Z products

Example#2

ABC/XYZ analysis

% analyzed SKUs 60,3%

% in total sales 96,6%

% of SKUs

Sales Volume

	X	Y	Z
A	13,0%	10,6%	4,1%
B	6,8%	11,1%	7,9%
C	13,6%	11,7%	21,2%

Difficulty to forecast

% of sales

	X	Y	Z
A	45,3%	26,7%	8,2%
B	4,3%	6,2%	4,3%
C	1,3%	1,1%	2,5%

Approved (system) forecast accuracy[1]

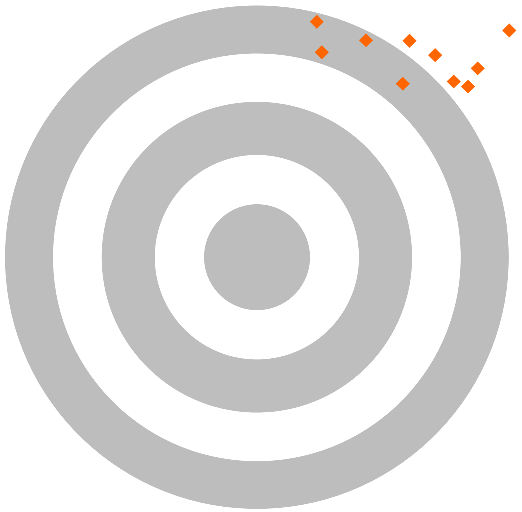
	X	Y	Z
A	89,2% (86,4%)	81,0% (69,7%)	66,2% (54,9%)
B	82,0% (82,1%)	76,5% (74,6%)	49,8% (45,0%)
C	85,7% (81,9%)	73,8% (77,9%)	52,5% (50,6%)
	+2,58pp	+9,09pp	+7,87pp

Commercial validation impact (weighted by sales)

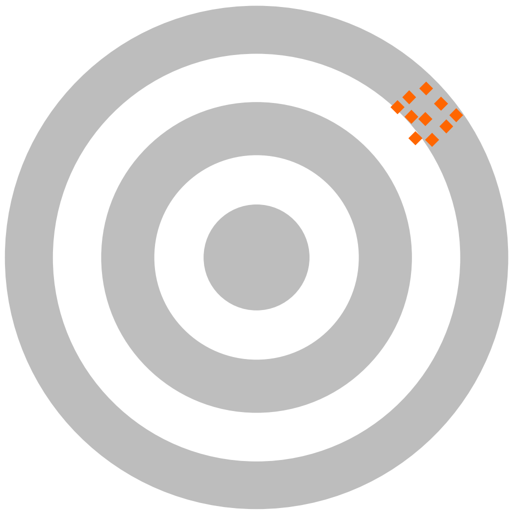
[1] accuracy measured at SKU x month level, system accuracy in brackets

Evaluating forecast’s performance is crucial in order to fine-tune the predictive models

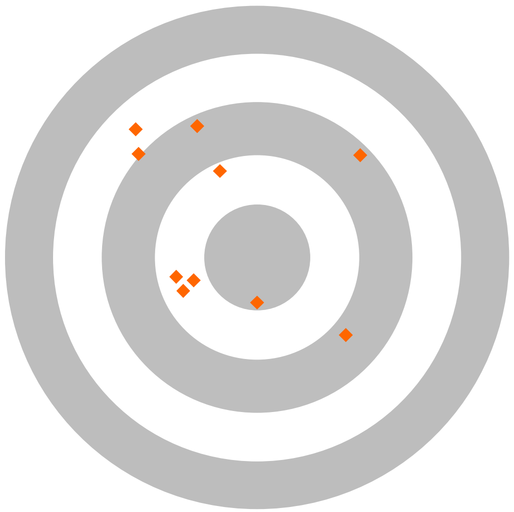
Forecast evaluation



High bias,
High variability



High bias,
Low variability



Low bias,
High variability

Low bias,
Low variability

Bias: systematic overestimation (or underestimation)

Variability: random error (unpredictable)

There is a variety of metrics that can be used to evaluate forecast's precision, both relative and absolute

Absolute Measures

$$e_t = \hat{Z}_t - Z_t$$

$$ME = \frac{\sum_{t=1}^{n} e_t}{n}$$ Mean Error

$$MAE = \frac{\sum_{t=1}^{n} |e_t|}{n}$$ Mean Absolute Error

$$MSE = \frac{\sum_{t=1}^{n} e_t^2}{n}$$ Mean Squared Error

Relative Measures

$$PE_t = \left(\frac{\hat{Z}_t - Z_t}{Z_t}\right) x\ 100$$ Percentual Error

$$BIAS(MPE) = \frac{\sum_{t=1}^{n} PE_t}{n}$$ Mean Percentual Error

$$\boldsymbol{MAPE} = \frac{\sum_{t=1}^{n} |PE_t|}{n}$$ Mean Absolute Percentual Error

MAPE and **BIAS are the most common metrics** to evaluate forecast's precision. These metrics are weighted by the sales in order to guarantee that forecasting errors in high seller products contribute more significantly for the global error

Metrics may be weighted given the importance of each product

Formula

Example

Simple Average

$$BIAS(EPM) = \frac{\sum_{i}^{N} BIAS_i}{N}$$

$$MAPE = \frac{\sum_{i}^{N} MAPE_i}{N}$$

	i=1	*i=2*	*i=3 (N)*
	A	B	C
Vendas:	10 000	500	20
MAPE:	10%	70%	100%

MAPE=60%

Weighted Average

$$BIAS(EPM) = \frac{\sum_{i}^{N} (\hat{Z}_i - Z_i)}{\sum_{i}^{N} Z_i}$$

$$MAPE = \frac{\sum_{i}^{N} \left|\hat{Z}_i - Z_i\right|}{\sum_{i}^{N} Z_i}$$

	A	B	C
Vendas:	10 000	500	20
MAPE:	10%	70%	100%

MAPE=13%

Note: Z_t – Real sales at period t; $\hat{Z}_t$ - Sales forecast, at *t-1*, for period *t*

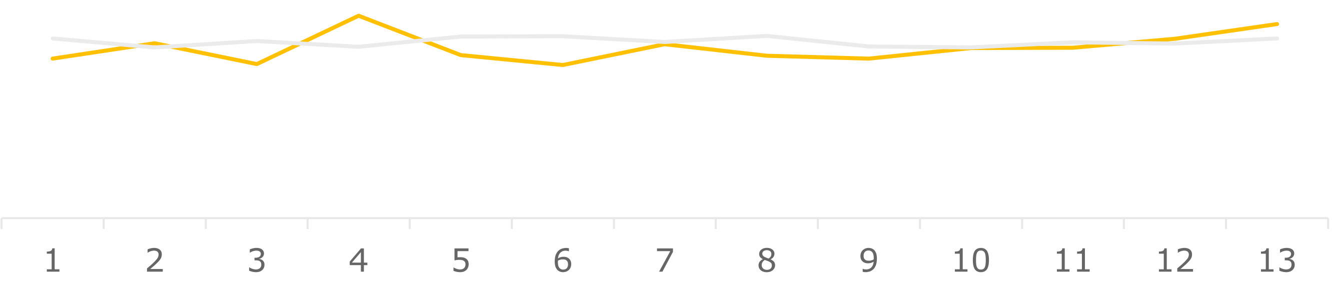
MAPE and BIAS provide distinct, rich information at different aggregation levels

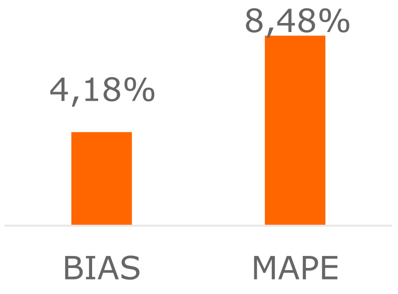
Forecast evaluation

MAPE - Mean absolute percentual error evaluates the mean deviation at a detailed level

BIAS - Forecast bias identifies systematic deviations, either underestimation or overestimation

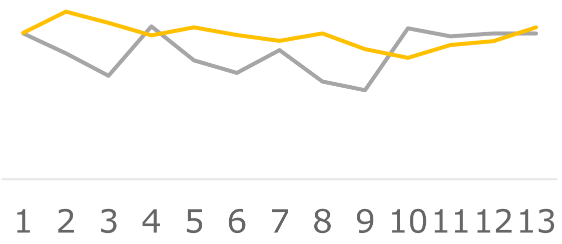
Aggregated level (e.g. group of products)



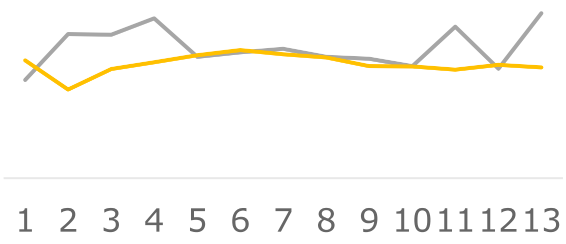


Detailed level (e.g. product)

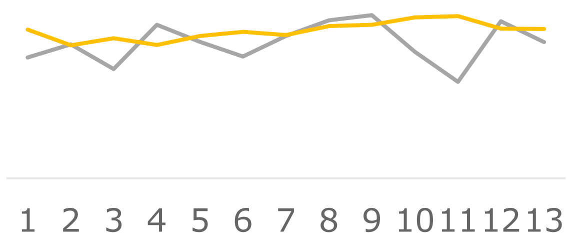
MAPE = 22.4%
BIAS =+ 16.7%

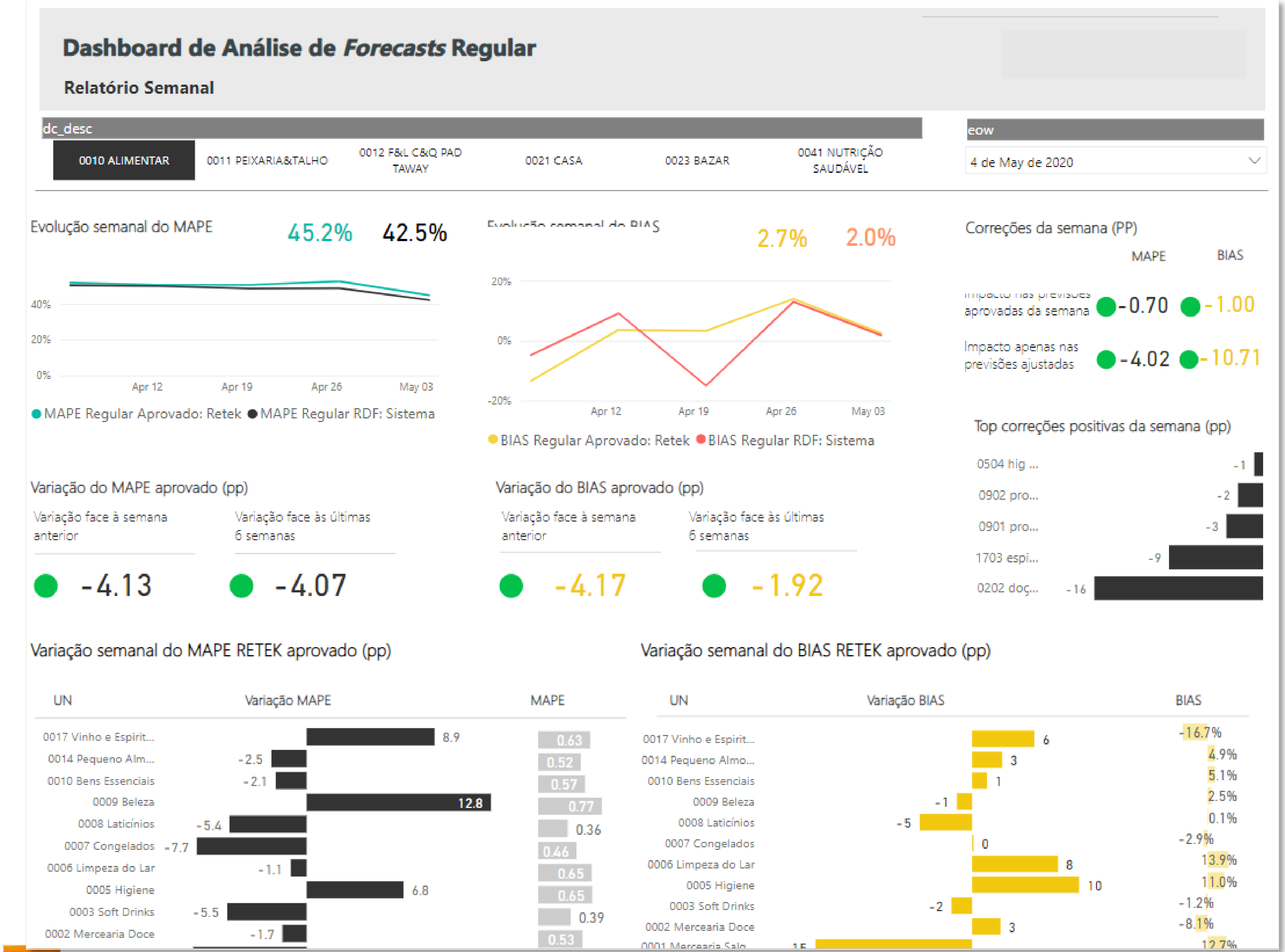


MAPE = 14.5%
BIAS =- 10.5%



MAPE = 16.1%
BIAS =+ 11.8%





Dashboard de Análise de *Forecasts* Regular

Relatório Semanal

4 de May de 2020

Evolução semanal do MAPE

45.2% 42.5%

Evolução semanal do BIAS

2.7% 2.0%

Correções da semana (PP)

	MAPE	BIAS
Impacto nas previsões aprovadas da semana	● -0.70	● -1.00
Impacto apenas nas previsões ajustadas	● -4.02	● -10.71

Top correções positivas da semana (pp)

0504 hig ...	-1
0902 pro...	-2
0901 pro...	-3
1703 espi...	-9
0202 doç...	-16

Variação do MAPE aprovado (pp)

Variação face à semana anterior	Variação face às últimas 6 semanas
● -4.13	● -4.07

Variação do BIAS aprovado (pp)

Variação face à semana anterior	Variação face às últimas 6 semanas
● -4.17	● -1.92

Variação semanal do MAPE RETEK aprovado (pp)

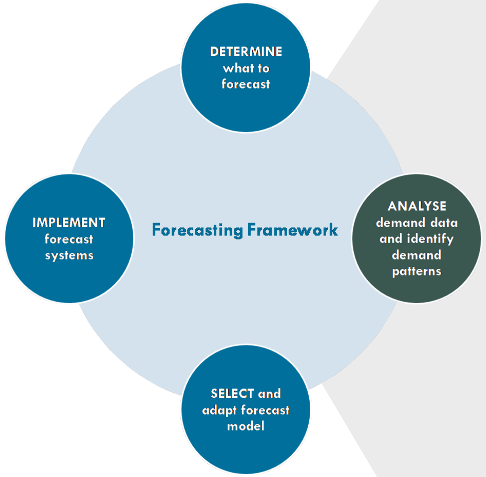
UN	Variação MAPE	MAPE
0017 Vinho e Espirit...	8.9	0.63
0014 Pequeno Alm...	-2.5	0.52
0010 Bens Essenciais	-2.1	0.57
0009 Beleza	12.8	0.77
0008 Laticínios	-5.4	0.36
0007 Congelados	-7.7	0.46
0006 Limpeza do Lar	-1.1	0.65
0005 Higiene	6.8	0.65
0003 Soft Drinks	-5.5	0.39
0002 Mercearia Doce	-1.7	0.53

Variação semanal do BIAS RETEK aprovado (pp)

UN	Variação BIAS	BIAS
0017 Vinho e Espirit...	6	-16.7%
0014 Pequeno Almo...	3	4.9%
0010 Bens Essenciais	1	5.1%
0009 Beleza	-1	2.5%
0008 Laticínios	-5	0.1%
0007 Congelados	0	-2.9%
0006 Limpeza do Lar	8	13.9%
0005 Higiene	10	11.0%
0003 Soft Drinks	-2	-1.2%
0002 Mercearia Doce	3	-8.1%

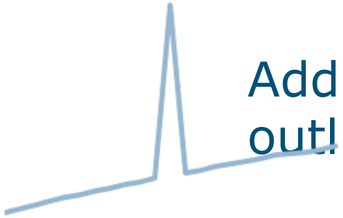
Gorgeous Ice Cream Case Study

Outliers: atypical values caused by "rare" events or data entry errors

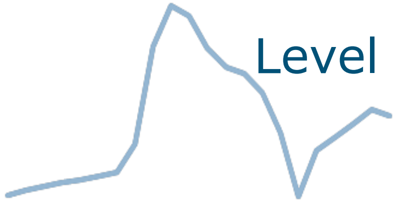


Analyze demand data and identify demand patterns

- Outliers
  https://www.youtube.com/watch?v=UIAuPDpnwfY



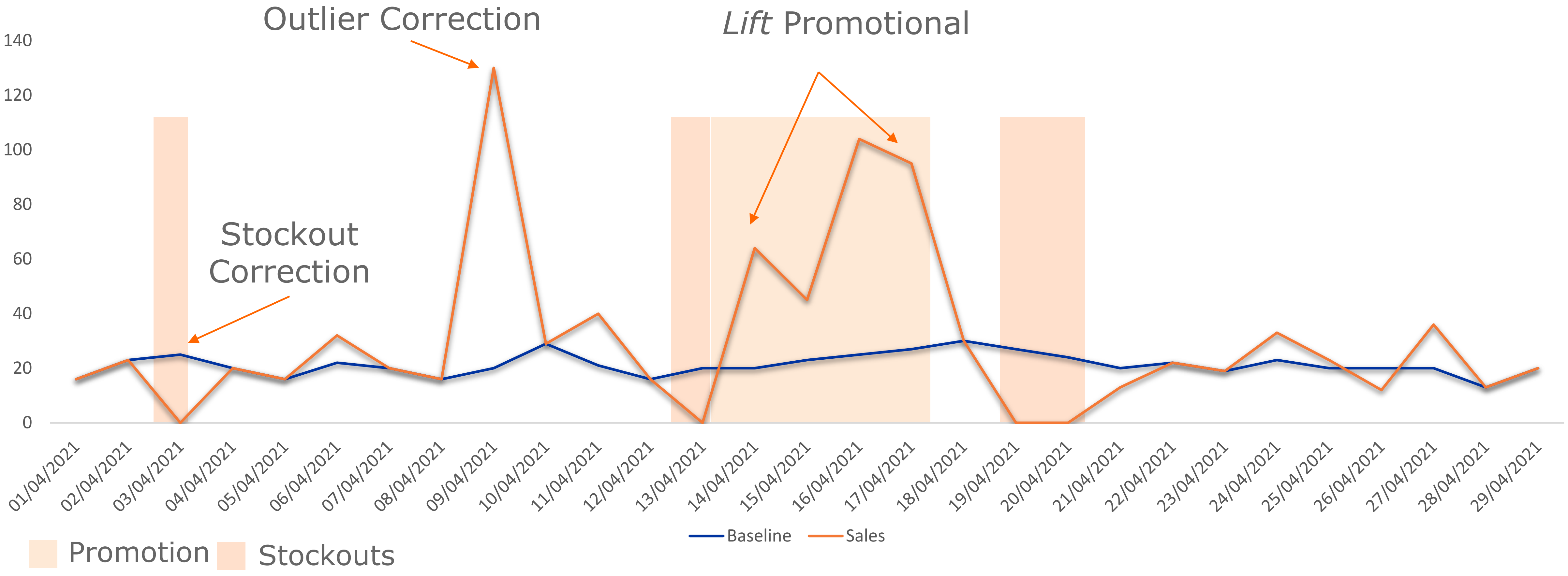
Additive outliers



Level shifts

Transitory changes

Before generating forecast, the historical baseline should be created



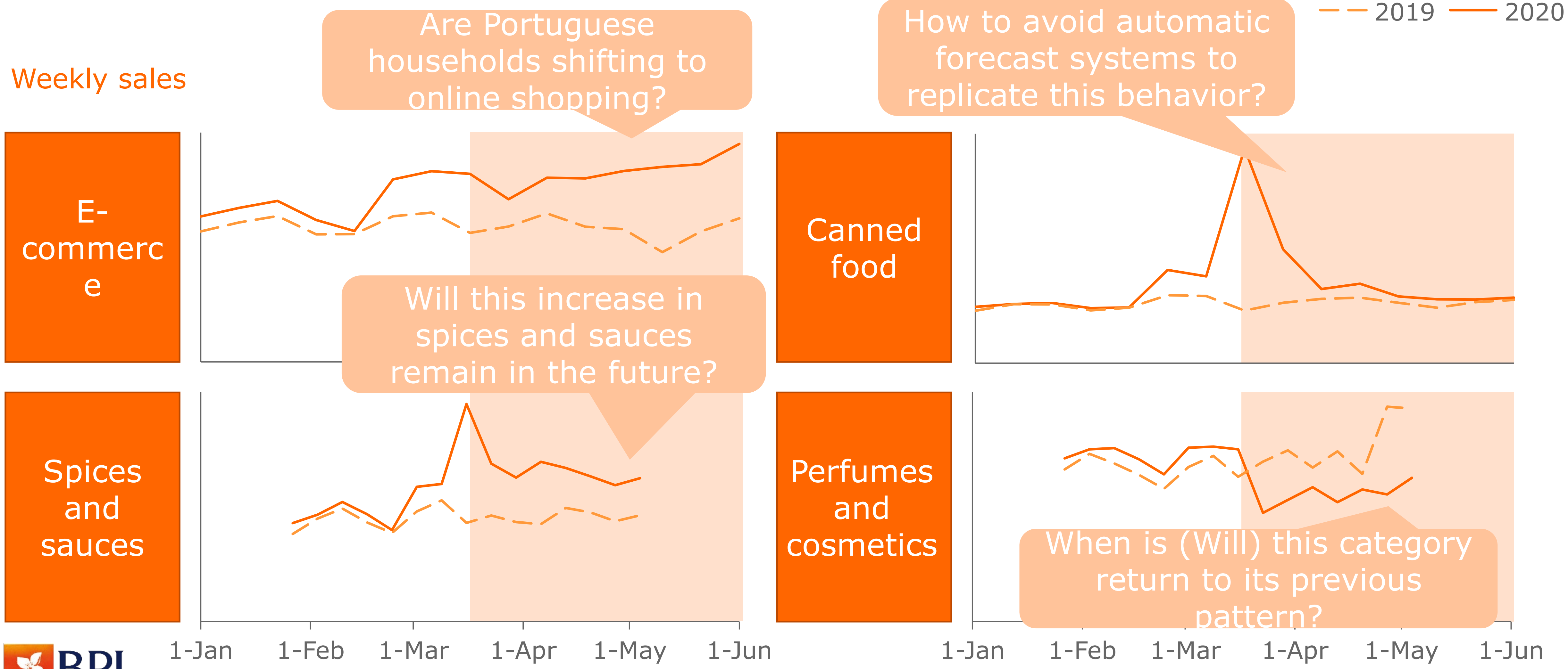
Outlier Correction

Lift Promotional

Stockout Correction

Many sectors have recently experienced dramatic changes on demand - relying solely on history is no longer valid

Demand Planning



Are Portuguese households shifting to online shopping?

How to avoid automatic forecast systems to replicate this behavior?

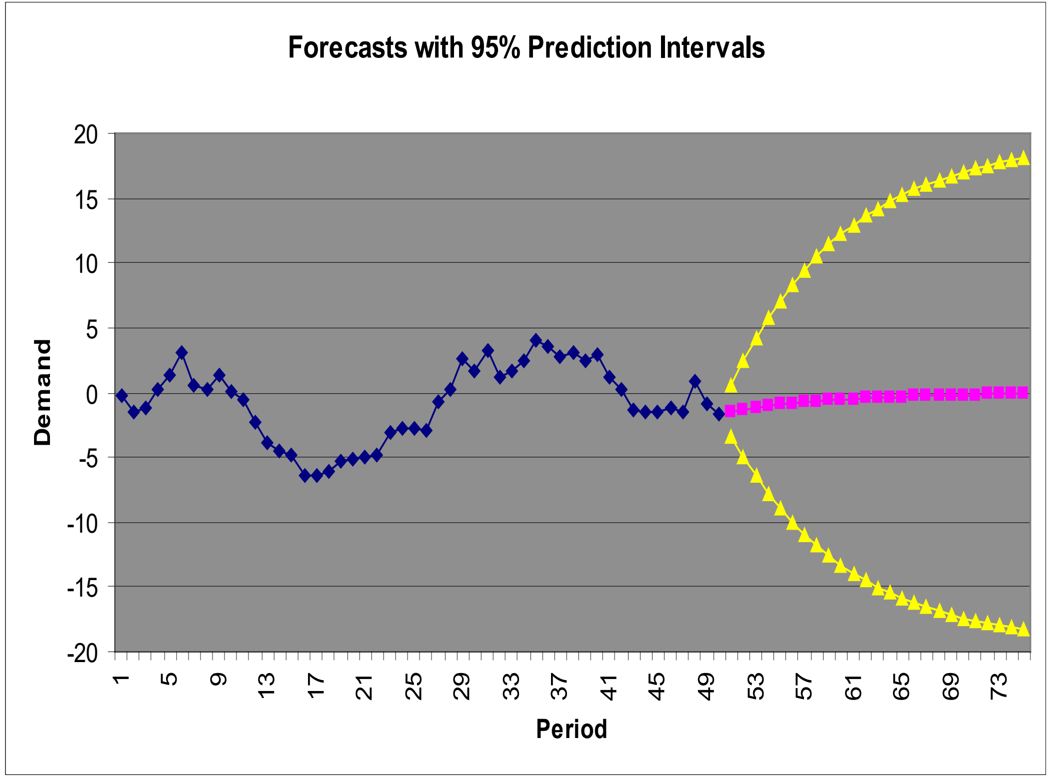
Will this increase in spices and sauces remain in the future?

When is (Will) this category return to its previous pattern?

Prediction Intervals

We can create prediction intervals of the basic form:

"There is a xx% probability that the actual future value will fall in the range Forecast ± K (example: K=1.96*σ(forecast error), for 95% of confidence)

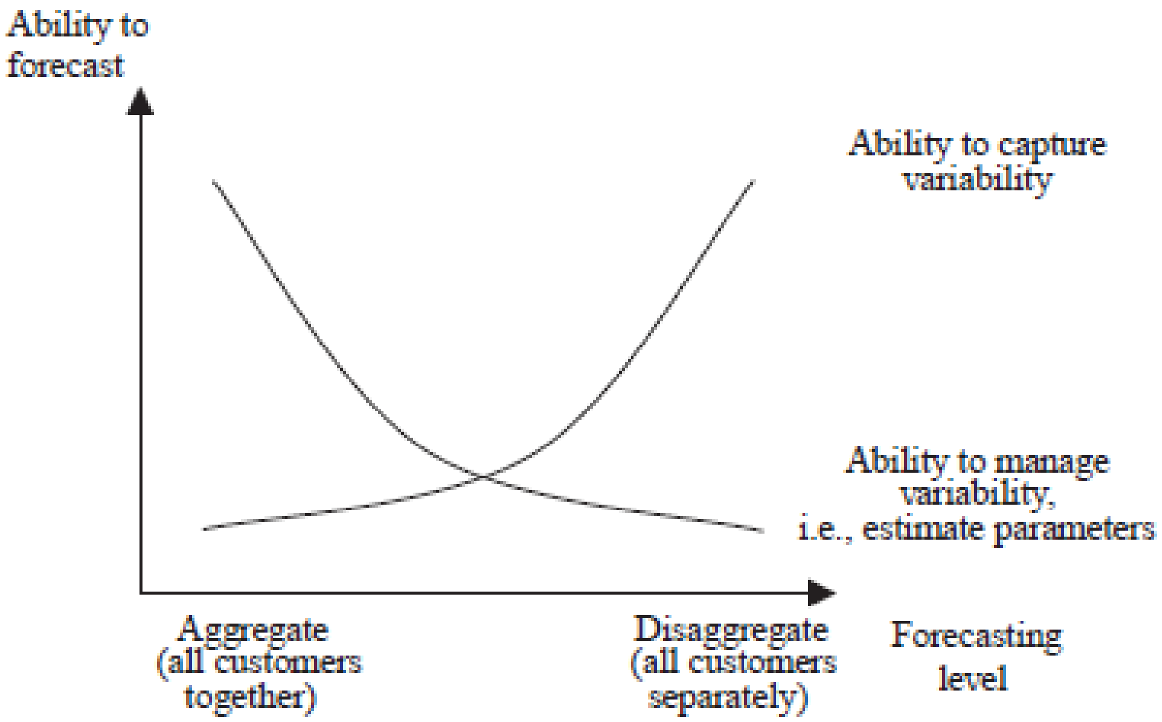


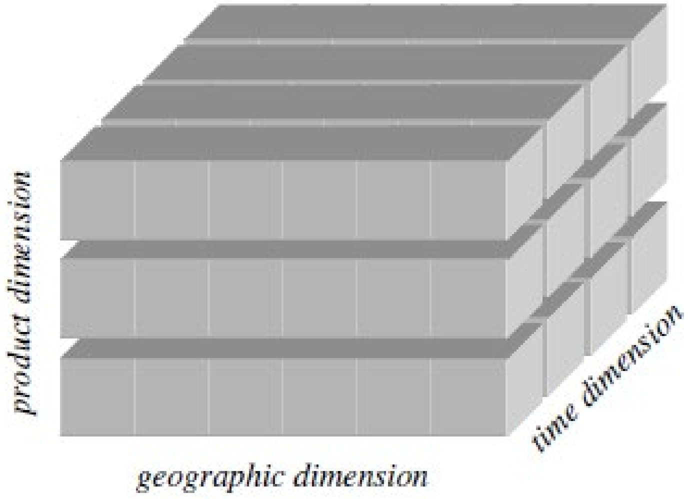
Carlson Department Store Case Study

(1) the amount of sales Carlson would have made if the hurricane had not struck;

(2) whether Carlson is entitled to any compensation for excess sales from increased business activity after the storm

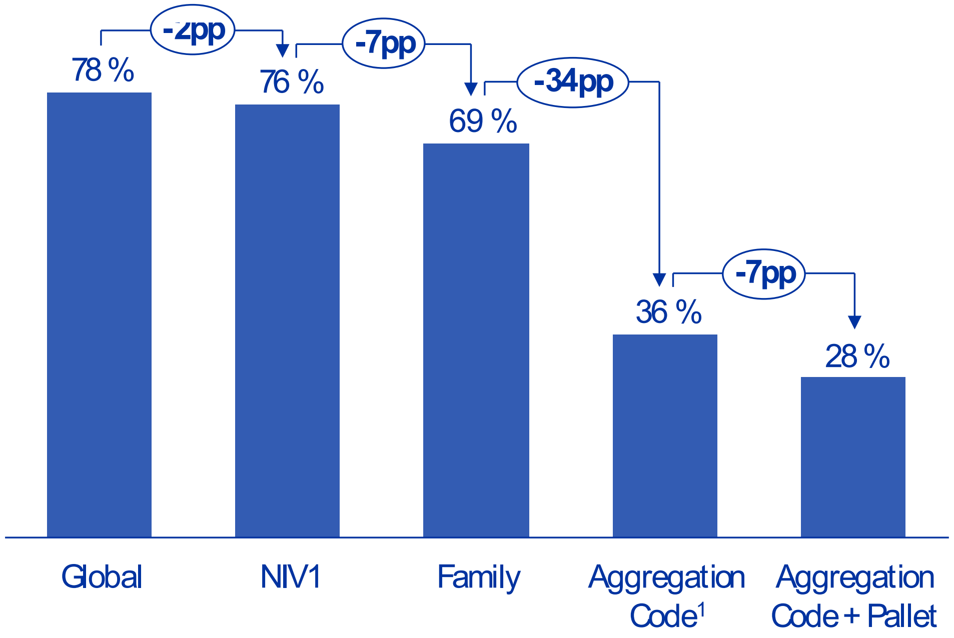
Impact of aggregation level on forecasting performance

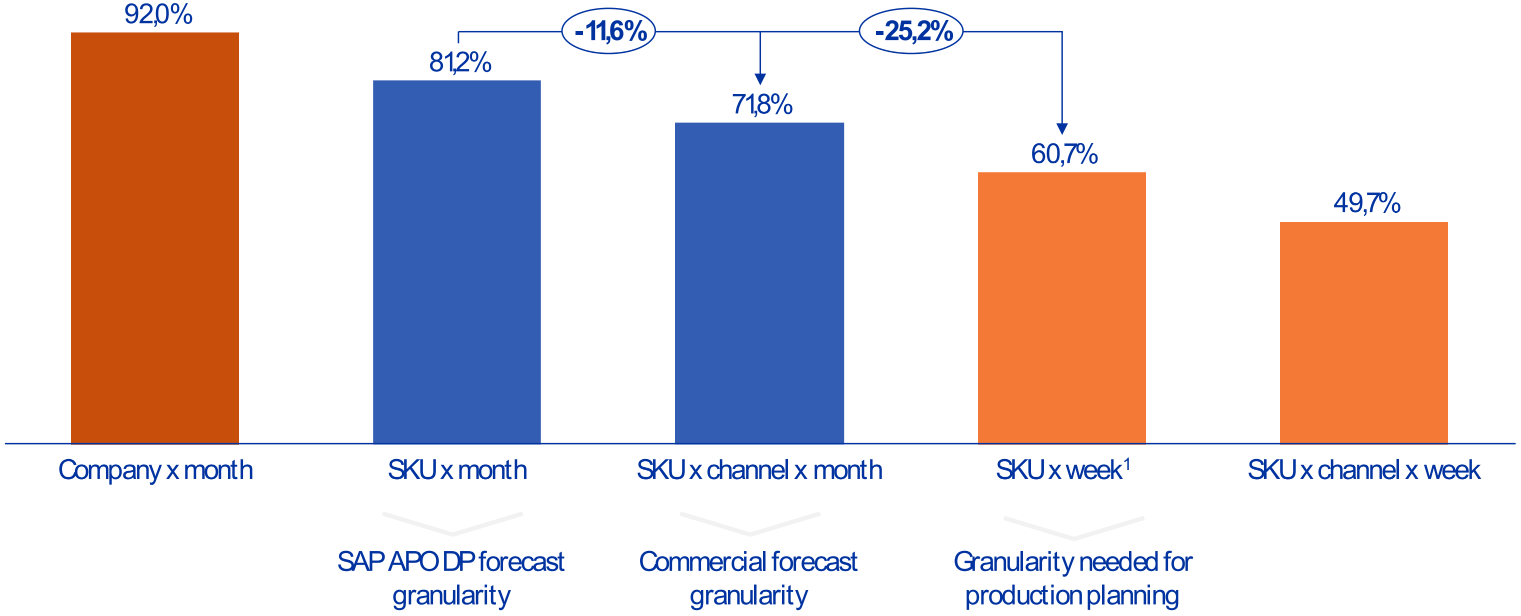




When comparing top-down with bottom-up forecasting processes it is often appropriate to measure forecasting accuracy both at aggregate and disaggregate level

Forecast Accuracy at different aggregation levels



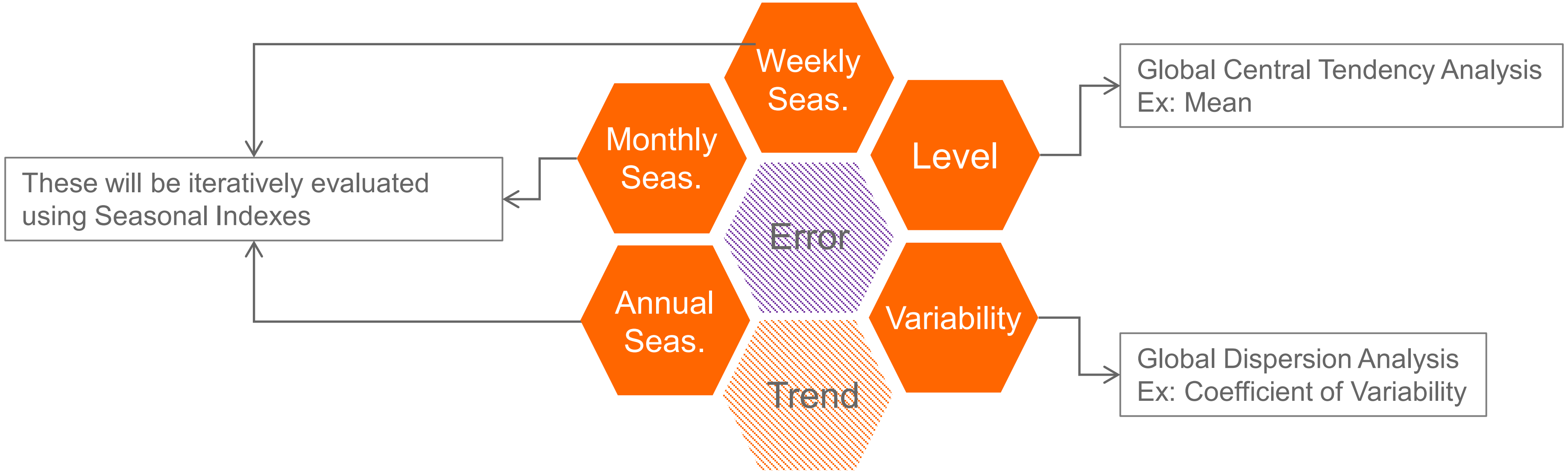


Impact of aggregation level on forecasting performance

Aggregation Policies

Clustering Process

- Select & Calculate Explanatory Variables – Item Clustering
  - The set of explanatory variables are the inputs for the clustering process to determine which observations are alike and which aren't;
  - Therefore, these inputs should be based on the time series components:



These will be iteratively evaluated using Seasonal Indexes

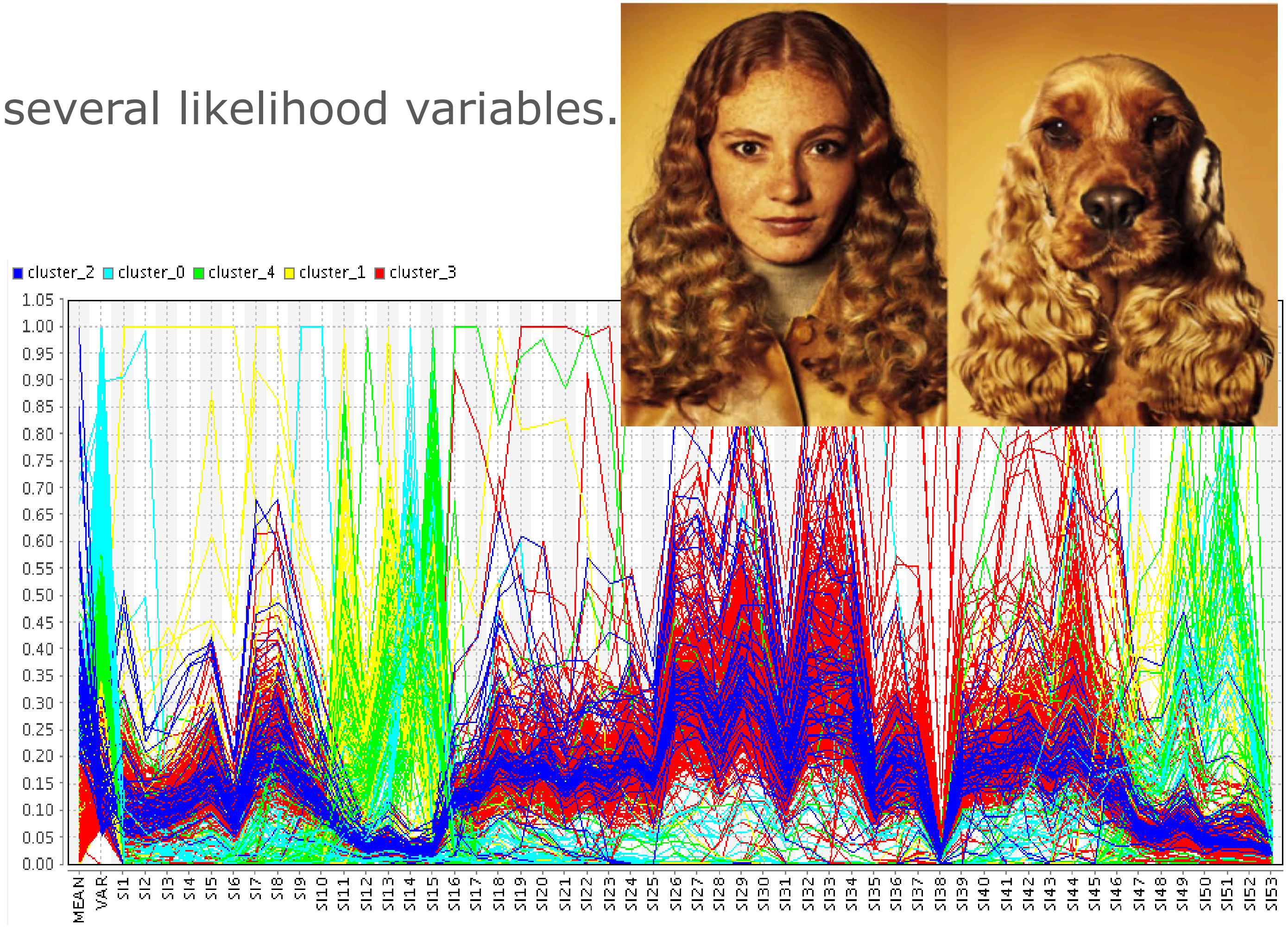
Global Central Tendency Analysis
Ex: Mean

Global Dispersion Analysis
Ex: Coefficient of Variability

Impact of aggregation level on forecasting performance

Aggregation Policies

Clustering according to several likelihood variables.



- ➢ Honey & Sweets
- ➢ Deserts

864 SKU's

- ✓ Cl.0 – 63 SKU's
- ✓ Cl.1 – 133 SKU's
- ✓ Cl.2 – 55 SKU's
- ✓ Cl. 3 – 463 SKU's
- ✓ Cl.4 – 150 SKU's

Impact of aggregation level on forecasting performance

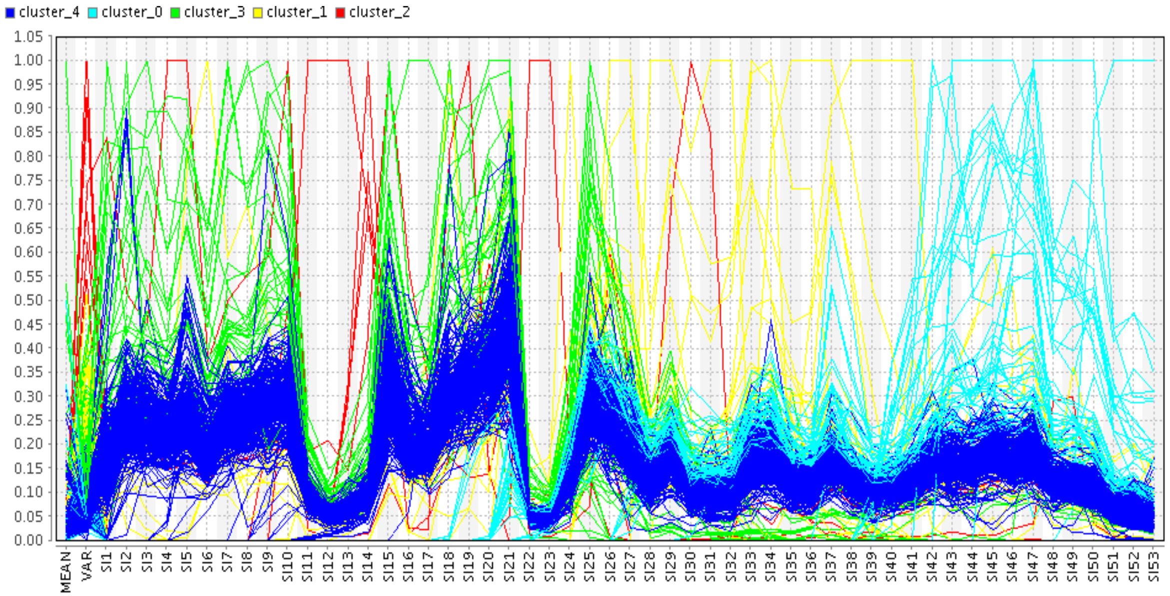
Aggregation Policies

Clustering according to several likelihood variables.

- Water
- Beverages
- Soda

K-Means
- K=5

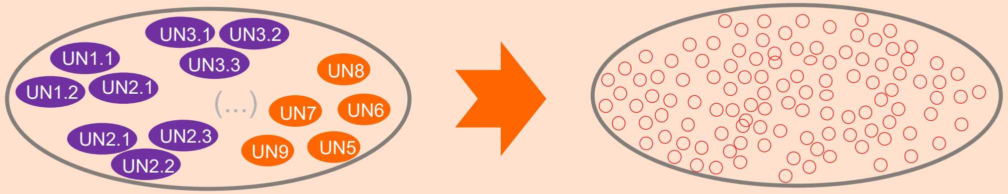
- Avg. Dist = -0.572
- Daves Bouldin = -1.277
- Example Dist. = 0.534



429 SKU’s

- ✓ Cl.0 – 49 SKU’s
- ✓ Cl.1 – 31 SKU’s
- ✓ Cl.2 – 7 SKU’s
- ✓ Cl. 3 – 36 SKU’s
- ✓ Cl.4 – 306 SKU’s

Impact of aggregation level on forecasting performance

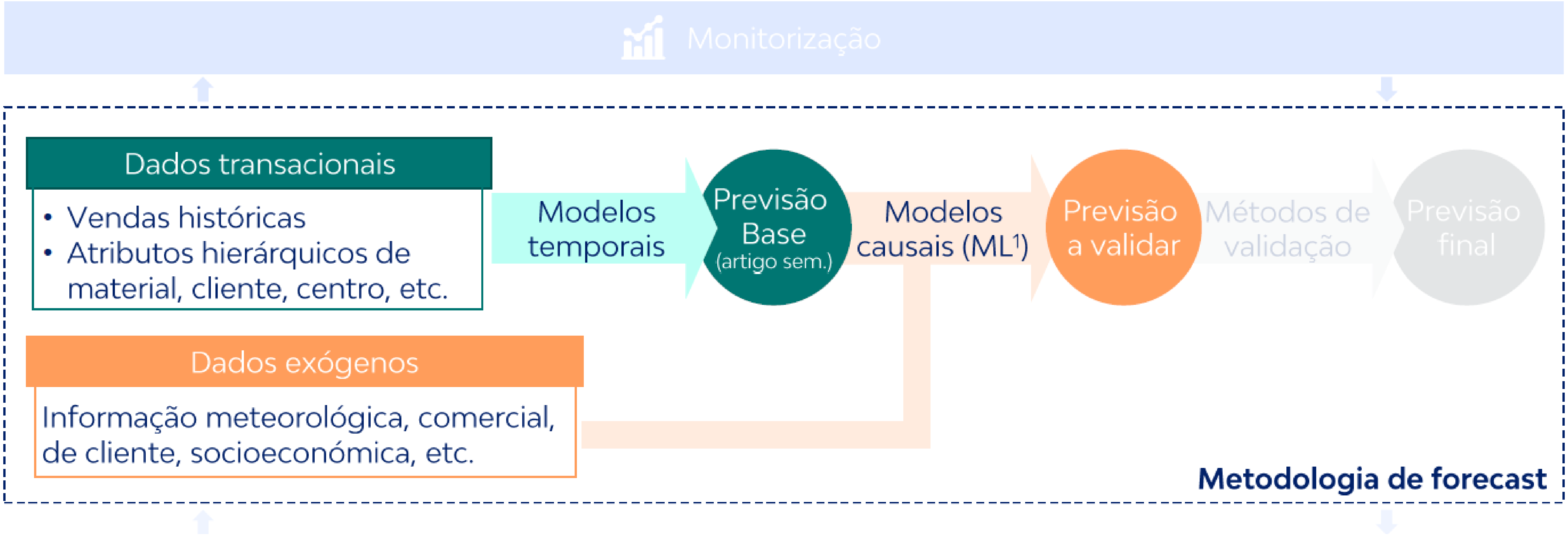


60K x 100 ITEM/LOCS **200~300 CLUSTERS**

Agenda

- The Role of Time Series
- Time Series Methods: Exponential Smoothing
- Accuracy, Outliers and Aggregation
- **Hybrid Methods**
- Box & Jenkins

Using advanced analytic models with many sources of insights to understand how demand evolve



Monitorização

Dados transacionais

- Vendas históricas
- Atributos hierárquicos de material, cliente, centro, etc.

Dados exógenos

Informação meteorológica, comercial, de cliente, socioeconómica, etc.

Modelos temporais

Previsão Base (artigo sem.)

Modelos causais (ML[1])

Previsão a validar

Métodos de validação

Previsão final

Metodologia de forecast

Using advanced analytic models with many sources of insights to understand how demand evolve

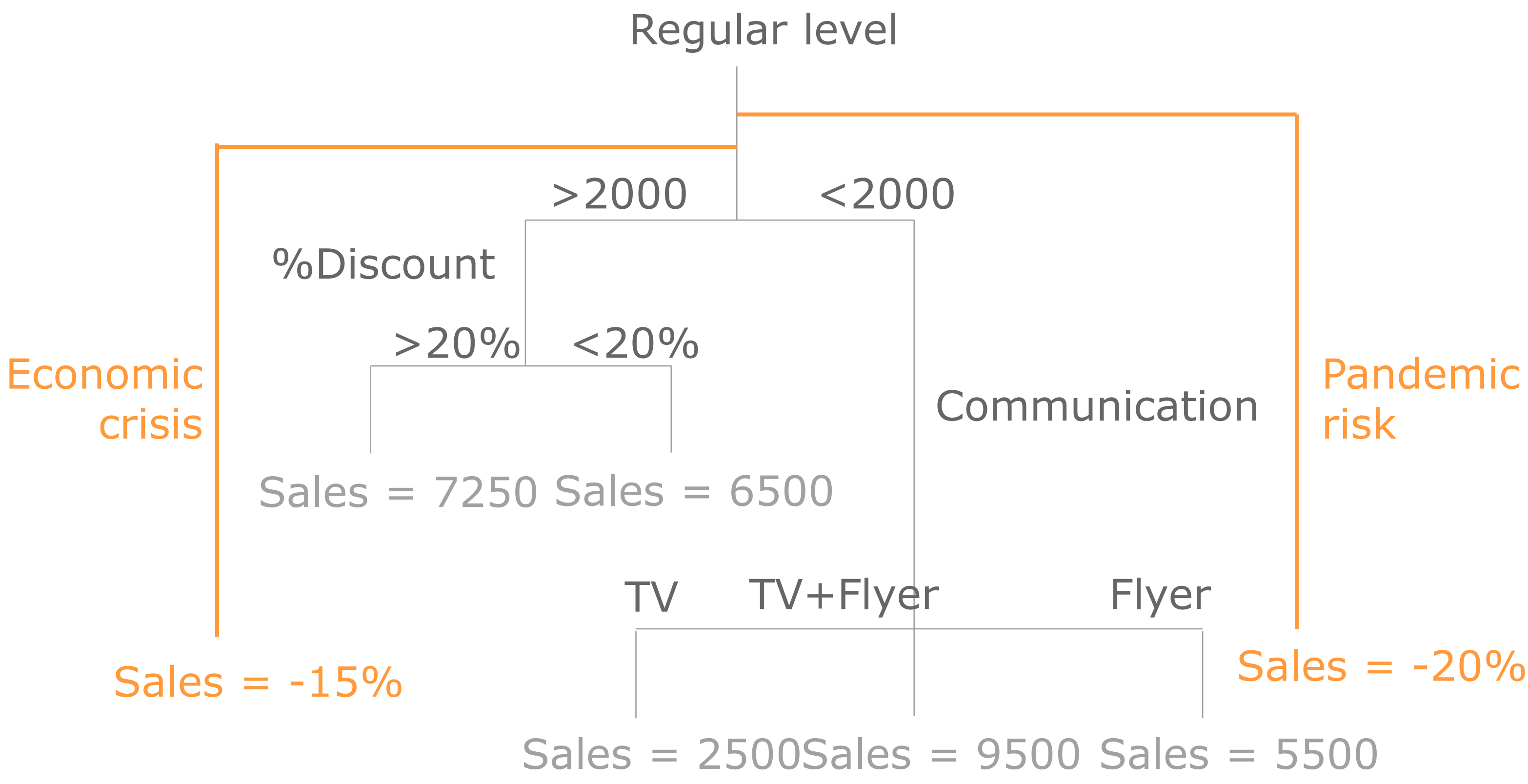
	Time-series	Machine learning (ML)		
	Holt-Winters + ensemble	GLM (Generalized linear model)	GBM (Gradient boosting machine)	RF[1] (Random Forest)
Vantagens	• Capturam os **efeitos temporais** de uma série de vendas - **nível, sazonalidade e tendência** • Maior **explicabilidade**	• Permitem a **incorporação de variáveis exógenas** para além da série de vendas, estimando a **importância de cada uma** para a previsão final • ***Self-learning* dos parâmetros** dos modelos com a alteração do histórico		
		Maior **explicabilidade**	Capturam **efeitos não-lineares e interação entre variáveis**	
Desvantagens	• **Não permitem a incorporação de variáveis exógenas** • Necessária **atualização manual** de parâmetros	Assumem a **linearidade de efeitos**	Podem levar a uma maior **complexidade de interpretação**	

Leading organizations are using advanced analytic models with many sources of insights to understand how demand evolve

Use advanced analytic models (e.g. Machine Learning) and **leverage internal and external sources** of information to **extract patterns and insights out of the past sales**

Price
Promotions
Weather
Consumer research
Social listening
Online Search Trends
Trends from other countries
Social economic scenario

ILLUSTRATIVE



Regular level

>2000
<2000

%Discount

>20%
<20%

Sales = 7250 Sales = 6500

Communication

TV
TV+Flyer
Flyer

Sales = 2500Sales = 9500 Sales = 5500

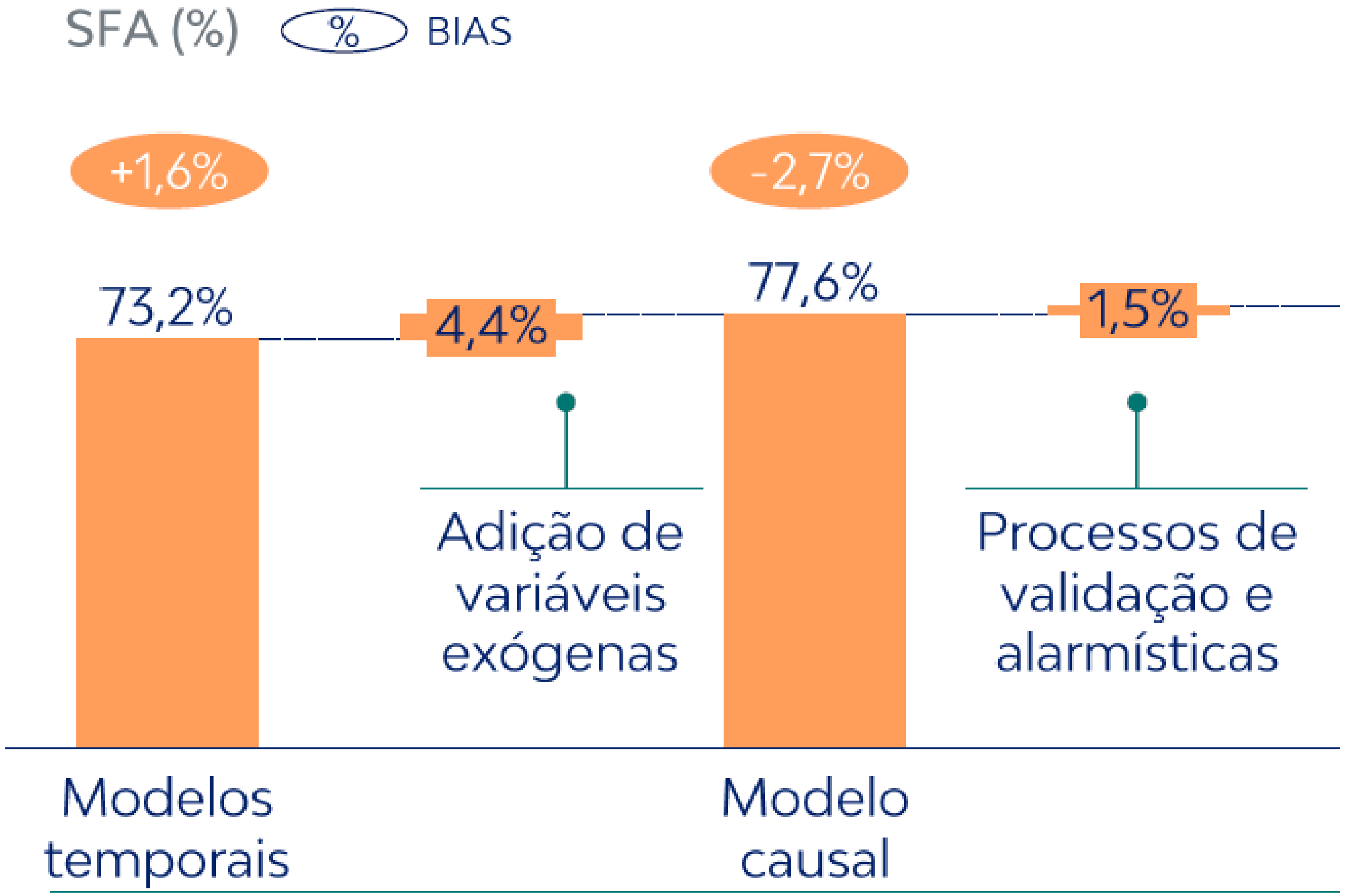
Economic crisis

Sales = -15%

Pandemic risk

Sales = -20%

Using advanced analytic models with many sources of insights to understand how demand evolve



Agenda

- The Role of Time Series
- Time Series Methods: Exponential Smoothing
- Accuracy, Outliers and Aggregation
- Hybrid Methods
- **Box & Jenkins**

B&J: Base Idea of the method

WHATEVER THE FORECASTING MODELS, THE FORECASTS ARE OBTAINED FROM THE AVAILABLE OBSERVATIONS, ACCORDING TO A FUNCTION OF TYPE

$$\hat{Z}_t(k) = f(Z_t, Z_{t-1}, Z_{t-2}, \ldots)$$

IN THE PREVIOUS METHODS, EACH OF THEM BASED ON ITS OWN, THE ADOPTION OF A METHOD INVOLVED THE ESTABLISHMENT OF THE BASIC STRUCTURE OF THE RELATIONSHIP BETWEEN FORECASTS AND AVAILABLE DATA

$$\hat{Z}_t(1) = \alpha \cdot Z_t + \alpha \cdot (1-\alpha) \cdot Z_{t-1} + \cdots$$

FOR EXAMPLE, IN THE SIMPLE EXPONENTIAL SMOOTHING METHOD ONLY ONE DEGREE OF FREEDOM WAS AVAILABLE TO ADAPT THE MODEL TO REALITY: THE CHOICE OF α

B&J: Base Idea of the method

IN THE APPROACH PROPOSED BY BOX AND JENKINS

- A WIDE RANGE OF ALTERNATIVE MODELS IS CONSIDERED

- ON THE BASIS OF HISTORICAL OBSERVATIONS, IT IS CALCULATED THAT DIFFERENT STATISTICS ARE CARRIED OUT ON THE BASIS OF THESE DIFFERENT TESTS THAT ALLOW YOU TO CHECK WHICH OF THE MODELS IT IS THE ONE THAT BEST ADAPTS TO THE AVAILABLE DATA

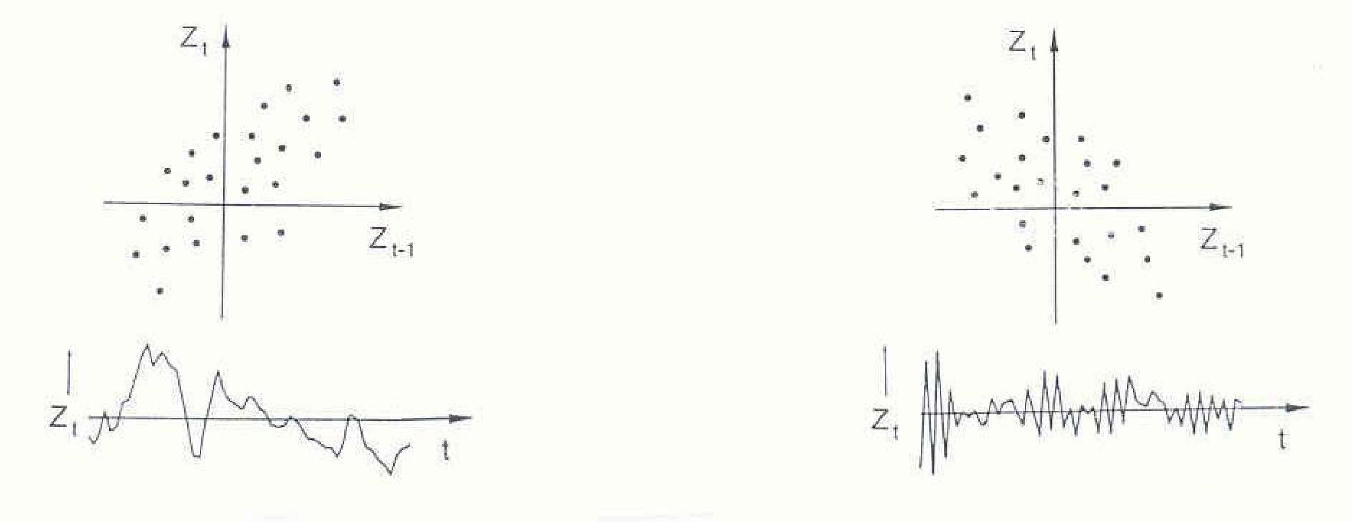
B&J Models applicable to stationary time series

Autoregressive Models

- THE SIMPLEST MODEL THAT TRANSLATES THE EXISTENCE OF CORRELATION BETWEEN THE SUCCESSIVE TERMS OF A SERIES IS

$$\boxed{Z_t = \phi_1 \cdot Z_{t-1} + E_t \qquad \text{COM } E_t \sim IN(0,\sigma_E^2)}$$

$\phi_1 > 0$ $\phi_1 < 0$



B&J Models applicable to stationary time series

Autoregressive Models

/…

ANOTHER, MORE COMPLICATED MODEL FOR TRANSLATING THE EXISTENCE OF CORRELATION BETWEEN THE SUCCESSIVE TERMS OF A SERIES WOULD BE

$$\boxed{Z_t = \phi_1 \cdot Z_{t-1} + \phi_2 \cdot Z_{t-2} + E_t} \quad \text{COM } E_t \sim IN(0, \sigma_E^2)$$

IN GENERAL, AN AUTOREGRESSIVE MODEL OF ORDER p CAN BE DEFINED AS FOLLOWS:

$$AR(p): \boxed{Z_t = \phi_1 \cdot Z_{t-1} + \phi_2 \cdot Z_{t-2} + \cdots + \phi_p \cdot Z_{t-p} + E_t} \quad \text{COM } E_t \sim IN(0, \sigma_E^2)$$

THE DESIGNATION OF THE MODEL - AUTOREGRESSIVE - HAS ITS ORIGIN IN THE FACT THAT IT IS A REGRESSION OF THE VARIABLE Z_t IN FUNCTION OF ITSELF, BUT WITH A LAG 1, 2, … , p …/

B&J Models applicable to stationary time series

Autoregressive Models

A MORE SUCCINCT WAY OF REPRESENTING AN AR(p) MODEL IS ACHIEVED BY USING THE 'BACK SHIFT OPERATOR'

IT IS AN OPERATOR THAT, WHEN APPLIED TO THE REALIZATION OF ANY VARIABLE AT THE TIME t A TRANSFORMS IT INTO THE REALIZATION OF THAT VARIABLE IN t-1

$$B Z_t = Z_{t-1}$$

$$B^2 Z_t = B(B Z_t) = B Z_{t-1} = Z_{t-2}$$

$$(\cdots)$$

$$\boxed{B^j Z_t = Z_{t-j}}$$

B&J Models applicable to stationary time series

Autoregressive Models

/…

USING OPERATOR B, THE AR(p) TURNS INTO

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + E_t$$

$$Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \cdots - \phi_p Z_{t-p} = E_t$$

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) Z_t = E_t$$

$$\text{AR}\,(p): \quad \phi(B) Z_t = E_t$$

WHERE Φ(B) REPRESENTS A POLYNOMIAL OF DEGREE p IN B

B&J Models applicable to stationary time series

Moving Average Models

AUTOREGRESSIVE MODELS CAN BE MANIPULATED AND EXPRESSED THE VARIABLE Z_t AS A FUNCTION OF ERRORS RECORDED AT TIME t OR AT EARLIER INSTANTS

EXAMPLE

AR(1):

$$\begin{aligned} Z_t &= \phi_1 Z_{t-1} + E_t \\ &= \phi_1 \cdot (\phi_1 Z_{t-2} + E_{t-1}) + E_t \\ &= \phi_1^2 Z_{t-2} + \phi_1 E_{t-1} + E_t \\ &= \phi_1^2 \cdot (\phi_1 Z_{t-3} + E_{t-2}) + \phi_1 E_{t-1} + E_t \\ &= E_t + \phi_1 E_{t-1} + \phi_1^2 E_{t-2} + \phi_1^3 E_{t-3} + \cdots \end{aligned}$$

B&J Models applicable to stationary time series

Moving Average Models

- …/

IN A MOVING AVERAGE MODEL OF ORDER q, THE VARIABLE Z_t IS EXPRESSED AS A FUNCTION OF THE ERRORS RECORDED AT THE TIME t AND IN THE PREVIOUS MOMENTS

A MODEL OF THE MOVING AVERAGE TYPE OF ORDER q IS THEN GIVEN BY

$$\begin{aligned}\mathbf{MA(q):}\quad Z_t &= E_t + \theta_1 \cdot E_{t-1} + \theta_2 \cdot E_{t-2} + \cdots + \theta_q \cdot E_{t-q} \\ &= \left(1 + \theta_1 \cdot B + \theta_2 \cdot B^2 + \cdots + \theta_q \cdot B^q\right) \cdot E_t \\ &= \theta(B) \cdot E_t\end{aligned}$$

WHERE q(B) REPRESENTS A POLYNOMIAL OF DEGREE q IN B

AN AR MODEL WITH A FINITE NUMBER OF TERMS (AND THEREFORE WITH A FINITE NUMBER OF PARAMETERS) IS EQUIVALENT TO AN MA MODEL WITH AN INFINITE NUMBER OF PARAMETERS

B&J Models applicable to stationary time series

Mixed Models

-

MIXED MODELS ARE MODELS THAT SIMULTANEOUSLY HAVE AN AR(P) COMPONENT AND AN MA(q) COMPONENT, BEING DENOTED MODELS(p,q)

$$\mathbf{ARMA(p,q)}: \quad \phi(B)\cdot Z_t = \theta(B)\cdot E_t$$

$$\left(1-\phi_1\cdot B-\phi_2\cdot B^2-\cdots-\phi_p\cdot B^p\right)\cdot Z_t = \left(1+\theta_1\cdot B+\theta_2\cdot B^2+\cdots+\theta_q\cdot B^q\right)\cdot E_t$$

ARMA(p,q) MODELS (WITH A FINITE NUMBER OF PARAMETERS) ARE ALSO CONVERTIBLE INTO AR OR MA MODELS WITH INFINITE NUMBER OF PARAMETERS

B&J Models applicable to stationary time series

Model Selection

THE AIM OF THIS PHASE IS, ON THE ONE HAND, TO IDENTIFY, AMONG THE BJ MODELS, WHICH ONE(S) BEST SUITS EACH SERIES AND, ON THE OTHER HAND, TO PRELIMINARILY ESTIMATE THE PARAMETERS OF THE IDENTIFIED MODEL(S)

THE IDENTIFICATION OF THE MODEL(S) IS CARRIED OUT ON THE BASIS OF THE BEHAVIOR OF THE AUTOCORRELATION AND PARTIAL AUTOCORRELATION

$$\hat{\phi}_{kk} \sim N\left(\mu = 0,\ \sigma^2 \approx \frac{1}{N}\right),\ \text{PARA } k > p \quad \Rightarrow \quad \text{MODELO AR(p)}$$

$$r_k \sim N\left[\mu = 0,\ \sigma^2 \approx \frac{1}{N} \cdot \left(1 + 2 \cdot \sum_{i=1}^{q} r_i^2\right)\right],\ \text{PARA } k > q \quad \Rightarrow \quad \text{MODELO MA(q)}$$

$$\text{NEM } \hat{\phi}_{kk} \text{ NEM } r_k \text{ SOFREM UM «CUT OFF»} \quad \Rightarrow \quad \text{MODELO ARMA}$$

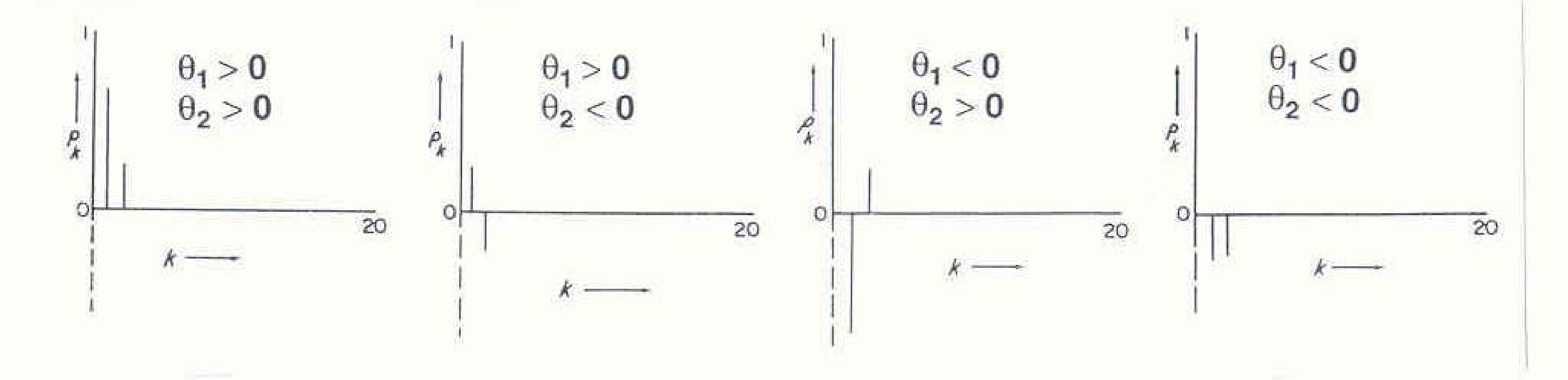
B&J Models applicable to stationary time series

- q .../

Autocorrelation Function (f.a.c.)

IN GENERAL, FOR A MA(q) MODEL, THE AUTOCORRELATIONS CANCEL EACH OTHER OUT FOR k>q; KNOWN AS 'CUT OFF', INSTEAD OF EXPONENTIAL DECAY

AS AN EXAMPLE, FOR A MODEL MA(2), THE PATTERNS THAT THE AUTOCORRELATION FUNCTION CAN DISPLAY ARE PRESENTED



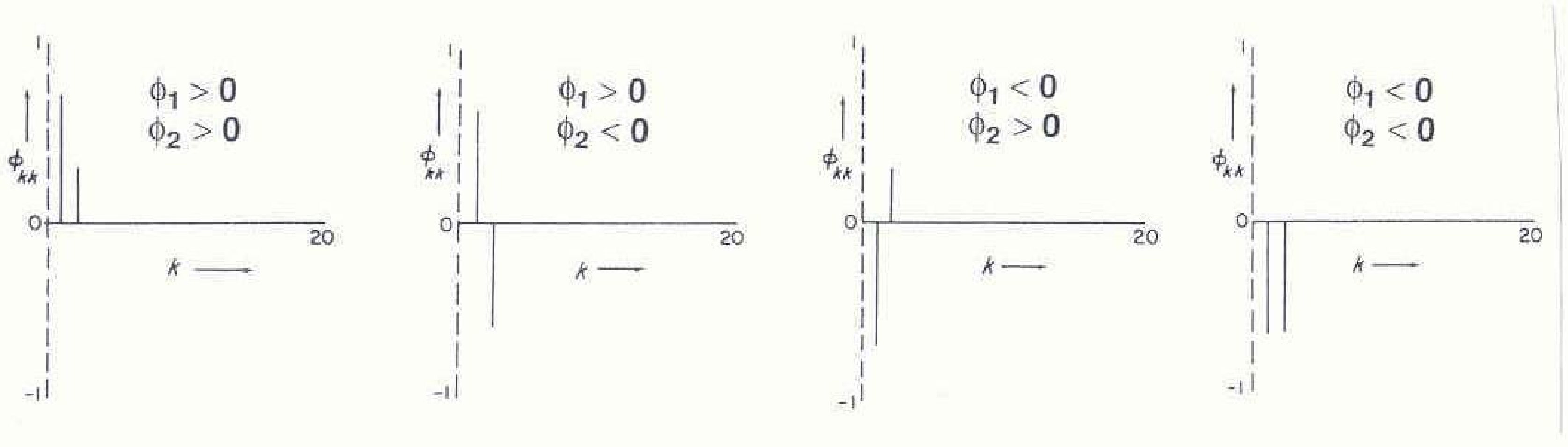
B&J Models applicable to stationary time series

• q .../

Partial Autocorrelation Function (f.a.c.p.)

IN GENERAL, FOR A AR(p) MODEL, THE PARTIAL AUTOCORRELATIONS CANCEL EACH OTHER OUT FOR k>p; KNOWN AS 'CUT OFF', INSTEAD OF EXPONENTIAL DECAY

AS AN EXAMPLE, FOR A MODEL AR(2), THE PATTERNS THAT THE PARTIAL AUTOCORRELATION FUNCTION CAN DISPLAY ARE PRESENTED



B&J Models applicable to stationary time series

Model Selection

HYPOTHESIS TESTING OF AUTOCORRELATIONS

SUCCESSIVELY FOR q =0, 1, 2, ..., CHECK IF $\rho_k = 0$ PARA $k > q$

$$H_0: \rho_k = 0$$

$$H_1: \rho_k \neq 0$$

$$ET = \frac{r_k}{\sqrt{\frac{1}{N} \cdot \left(1 + 2 \cdot \sum_{i=1}^{q} r_i^2\right)}}$$

$$H_0 \text{ VERDADEIRA} \Rightarrow ET \sim N(0,1)$$

BASED ON THIS TEST, IDENTIFY WHICH AUTOCORRELATIONS ARE SIGNIFICANTLY DIFFERENT FROM ZERO

B&J Models applicable to stationary time series

Model Selection

HYPOTHESIS TESTING OF PARTIAL AUTOCORRELATIONS

VERIFY IF $\phi_{kk} = 0$ PARA $k > p$

$$H_0: \phi_{kk} = 0$$

$$H_1: \phi_{kk} \neq 0$$

$$ET = \frac{\hat{\phi}_{kk}}{\sqrt{\frac{1}{N}}}$$

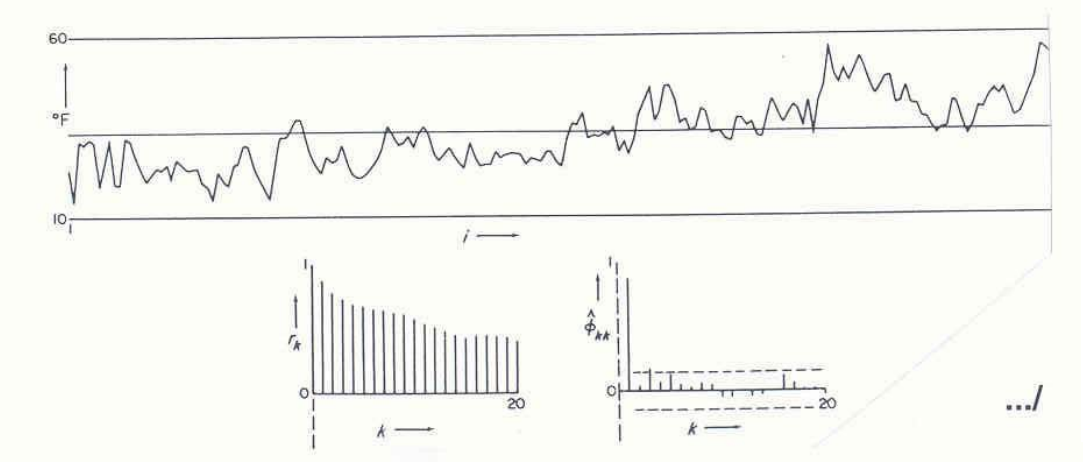
$$H_0 \text{ VERDADEIRA } \Rightarrow ET \sim N(0,1)$$

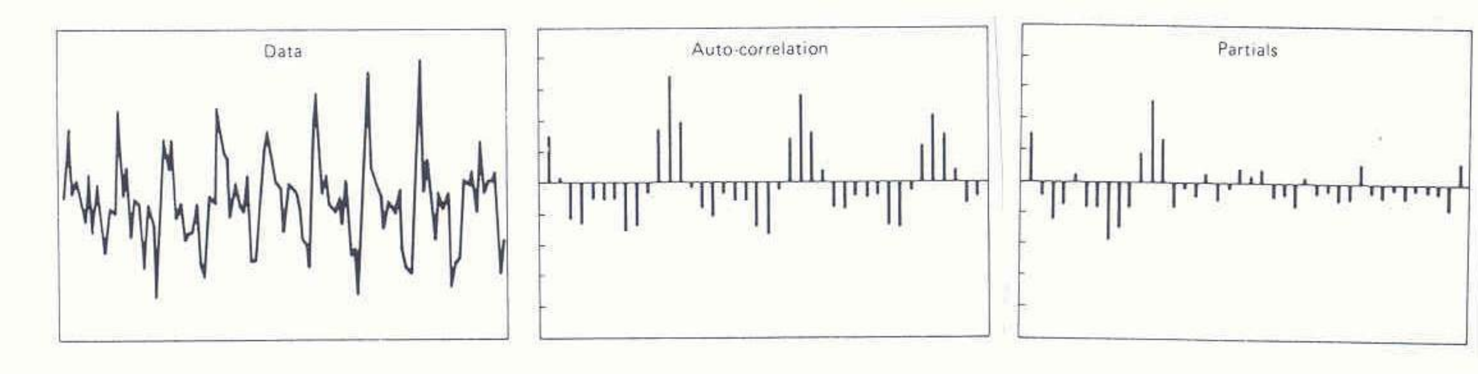
BASED ON THIS TEST, IDENTIFY WHICH PARTIAL AUTOCORRELATIONS ARE SIGNIFICANTLY DIFFERENT FROM ZERO

B&J Models applicable to non-stationary time series

Non-stationary signals

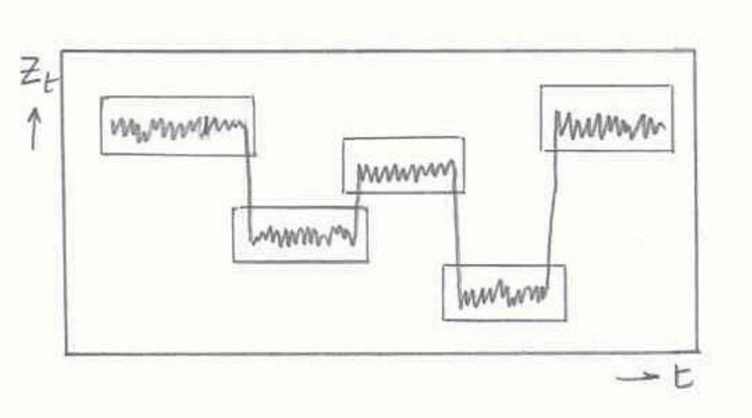
In case the non-stationarity origins from level and trend variations or seasonalities, it is revealed from the f.a.c. as it does not present cut-offs for low lag values.





B&J Models applicable to non-stationary time series

Local stationary time series with “jumps”



Locally

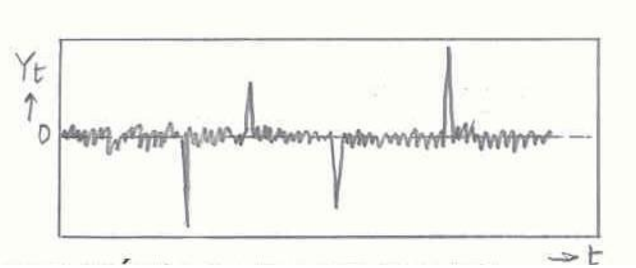
$$E(Z_t) \approx \text{CONSTANTE}$$

$$\Rightarrow \quad E(Z_t - Z_{t-1}) \approx 0$$

The differentiation of the original time series transforms it in a stationary time series

$$Y_t = Z_t - Z_{t-1} = \nabla Z_t$$

$$E(Y_t) \approx 0$$



AR(1)

$$Y_t = \phi_1 \cdot Y_{t-1} + E_t$$

$$Z_t - Z_{t-1} = \phi_1 \cdot (Z_{t-1} - Z_{t-2}) + E_t$$

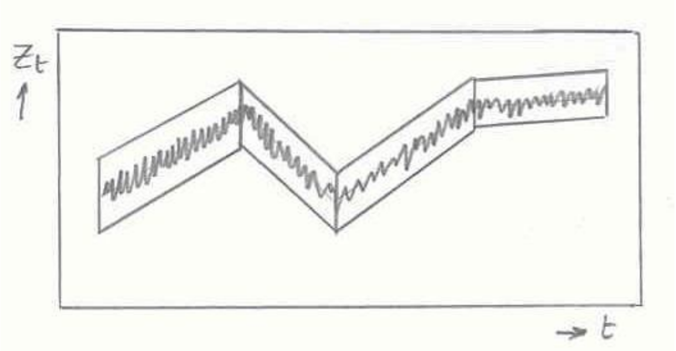
$$Z_t = (1 + \phi_1) \cdot Z_{t-1} - \phi_1 \cdot Z_{t-2} + E_t$$

$$Z_{t+1} = (1 + \phi_1) \cdot Z_t - \phi_1 \cdot Z_{t-1} + E_{t+1}$$

$$\hat{Z}_t(1) = (1 + \phi_1) \cdot z_t - \phi_1 \cdot z_{t-1} + E_{t+1} = (1 + \phi_1) \cdot z_t - \phi_1 \cdot z_{t-1}$$

B&J Models applicable to non-stationary time series

Time series with local linear trends



Locally

$$E(Z_t) \approx \alpha_0 + \alpha_1 \cdot t$$

$$\Rightarrow \quad E(Z_t - Z_{t-1}) \approx \alpha_1$$

$$\Rightarrow \quad E[(Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2})] \approx 0$$

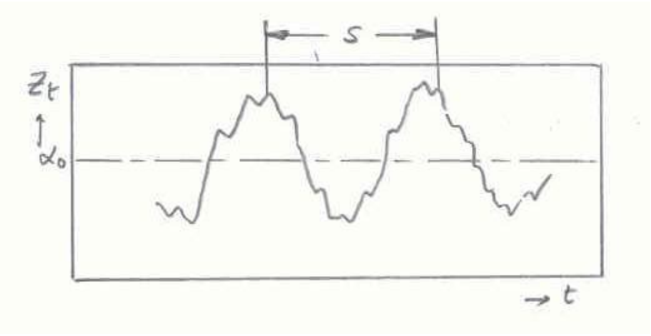
Double differentiation

$$Y_t = \nabla^2 Z_t = \nabla(Z_t - Z_{t-1}) = (Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2}) = Z_t - 2 \cdot Z_{t-1} + Z_{t-2}$$

$$E(Y_t) \approx 0$$

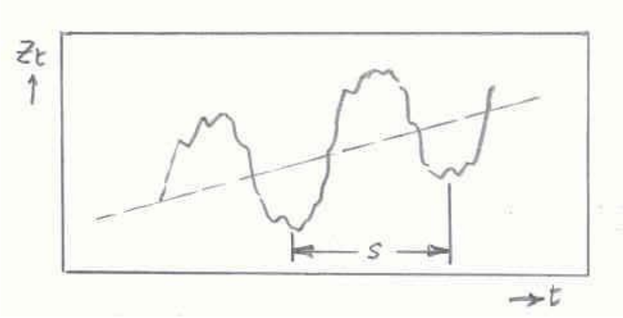
B&J Models applicable to non-stationary time series

Seasonal time series: Seasonal Autoregressive Integrated Moving Average



$$E(Z_t) \approx \alpha_0 + f_t \quad (f_t\text{: COMP. SAZONAL ADITIVA , COM PERIODO s)}$$

$$\Rightarrow \quad E(Z_t - Z_{t-s}) \approx 0 \qquad Y_t = \nabla_s Z_t = Z_t - Z_{t-s}$$



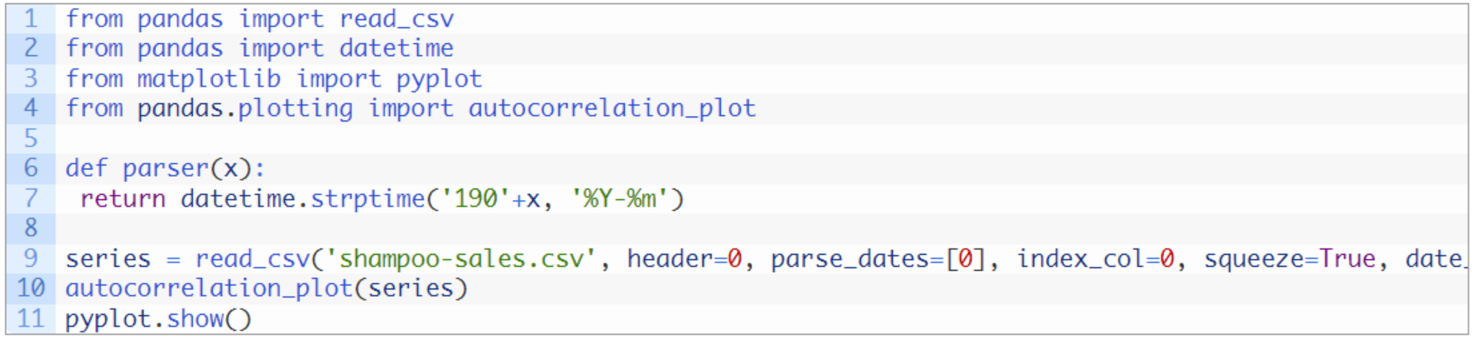
$$E(Z_t) \approx \alpha_0 + \alpha_1 \cdot t + f_t$$

$$\Rightarrow \quad E(Z_t - Z_{t-s}) \approx s \cdot \alpha_1 \qquad Y_t = \nabla \nabla_s Z_t = (Z_t - Z_{t-s}) - (Z_{t-1} - Z_{t-s-1})$$

$$\Rightarrow \quad E[(Z_t - Z_{t-s}) - (Z_{t-1} - Z_{t-s-1})] \approx 0$$

$$\text{SARIMA}(p,d,q) \times (P,D,Q): \quad \Phi_P(B^T) \cdot \phi_p(B) \cdot \nabla^d \cdot \nabla_s^D Z_t = \Theta_Q(B^T) \cdot \theta_q(B) \cdot E_t$$

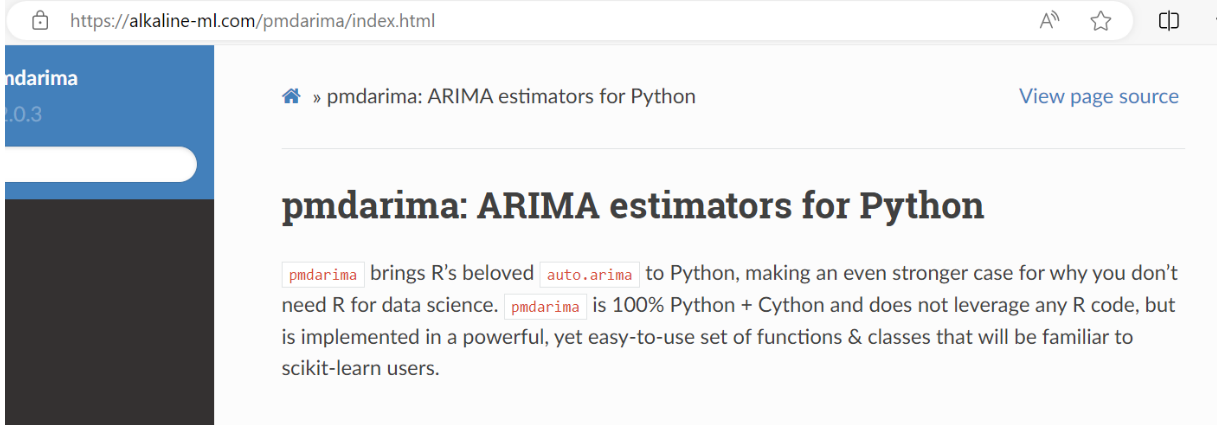
(S)ARIMA for Python



```
from pandas import read_csv
from pandas import datetime
from matplotlib import pyplot
from pandas.plotting import autocorrelation_plot

def parser(x):
 return datetime.strptime('190'+x, '%Y-%m')

series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True, date_
autocorrelation_plot(series)
pyplot.show()
```

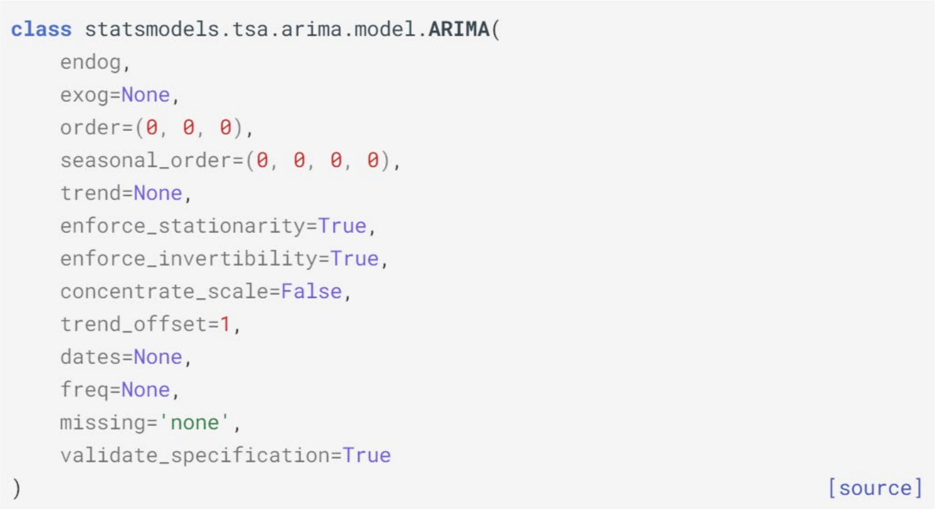


» pmdarima: ARIMA estimators for Python

pmdarima: ARIMA estimators for Python

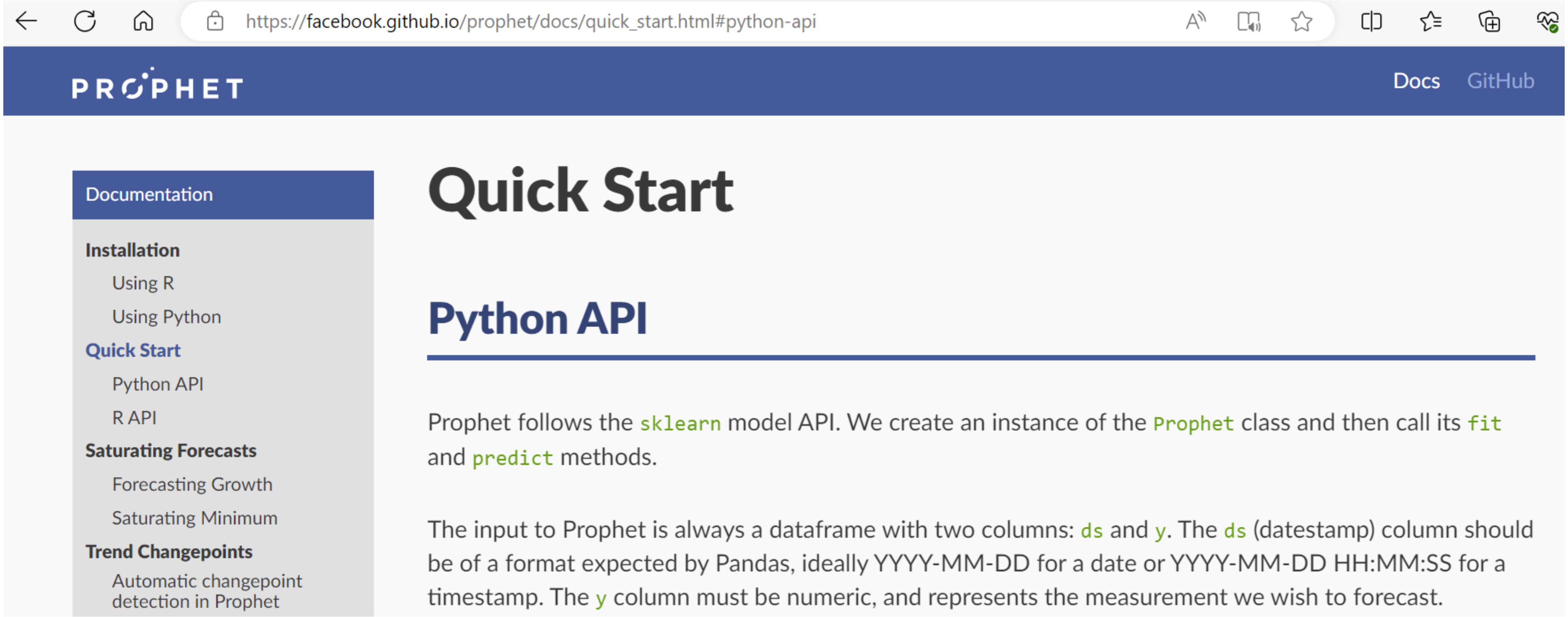
`pmdarima` brings R's beloved `auto.arima` to Python, making an even stronger case for why you don't need R for data science. `pmdarima` is 100% Python + Cython and does not leverage any R code, but is implemented in a powerful, yet easy-to-use set of functions & classes that will be familiar to scikit-learn users.

statsmodels.tsa.arima.model.ARIMA



```
class statsmodels.tsa.arima.model.ARIMA(
    endog,
    exog=None,
    order=(0, 0, 0),
    seasonal_order=(0, 0, 0, 0),
    trend=None,
    enforce_stationarity=True,
    enforce_invertibility=True,
    concentrate_scale=False,
    trend_offset=1,
    dates=None,
    freq=None,
    missing='none',
    validate_specification=True
)
```

Prophet - procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality,



Quick Start

Python API

Prophet follows the `sklearn` model API. We create an instance of the `Prophet` class and then call its `fit` and `predict` methods.

The input to Prophet is always a dataframe with two columns: `ds` and `y`. The `ds` (datestamp) column should be of a format expected by Pandas, ideally YYYY-MM-DD for a date or YYYY-MM-DD HH:MM:SS for a timestamp. The `y` column must be numeric, and represents the measurement we wish to forecast.

Make change happen