UNIL | Université de Lausanne
HEC Lausanne

A methodology to understand and to forecast the total demand of well-established products

Grandadam Patrik
5th of July 2019

Faculty of Business and Economics (HEC Lausanne)
Master in Management, Orientation in Business Analytics

Executive Summary

OBJECTIVES

- Propose a methodology to forecast the future demand
- Implement an automated tool
- Propose accurate forecasts

METHODOLOGY

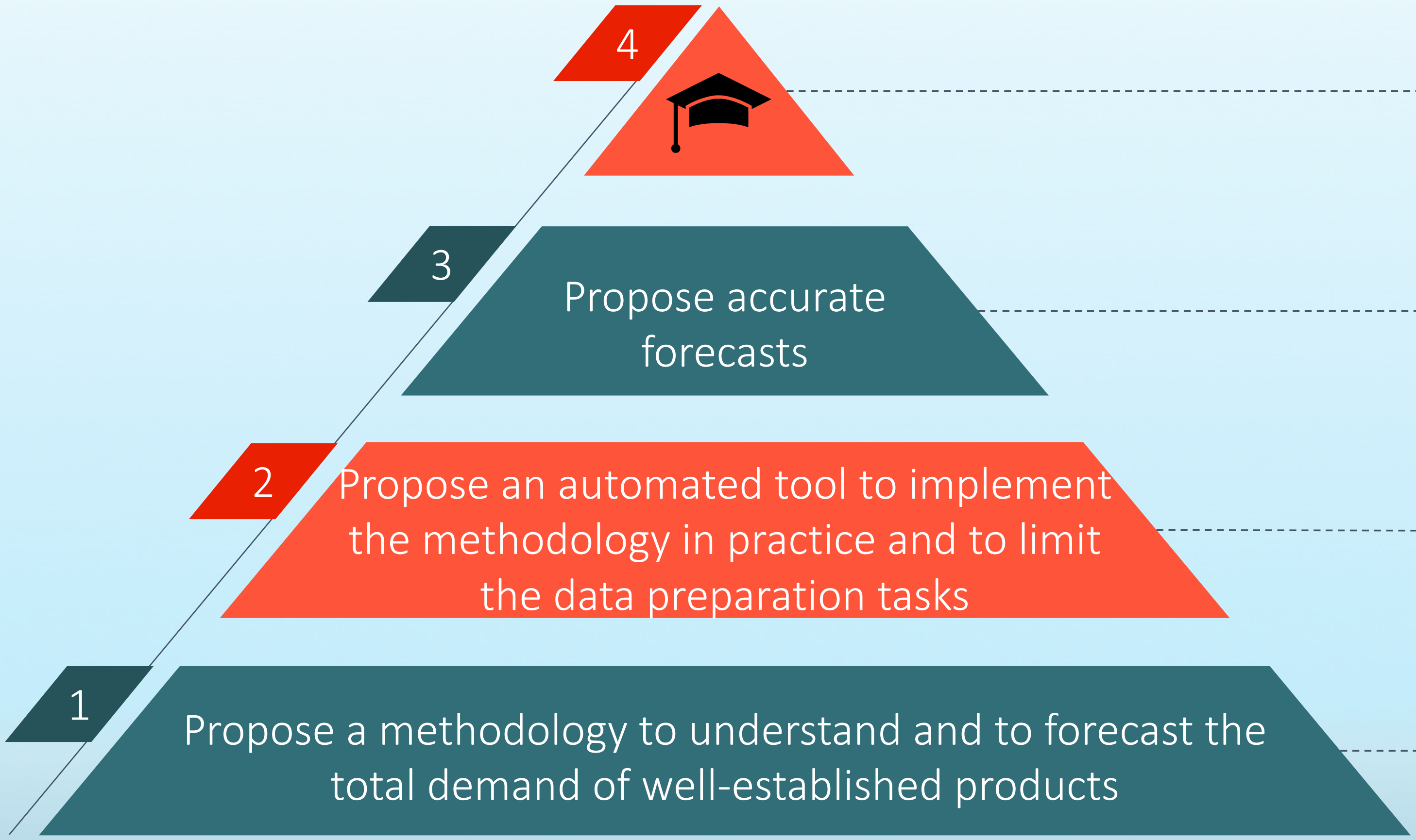
- A four step methodology
  - Take historical data
  - Generate the historical baseline
  - Forecast the future baseline
  - Forecast the final sales

RESULTS

- A precise methodology
- An automated tool taking the form of R functions regrouped in a package
- A possibility to use advanced forecasting methods without any specific knowledge
- A promising forecasting accuracy on the testing set

UNIL | Université de Lausanne
HEC Lausanne

Objectives



Show an academical maturity by using concepts taught in multiple university courses

The forecast resulting from the methodology should be accurate and reflecting the future

The methodology should be easily implementable in a business context

The main objective is to present the methodology that should be followed to increase the likelihood of making accurate forecasts

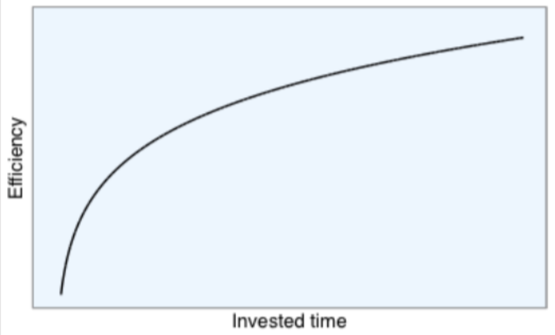
Trade-offs regarding the automated tool

UNIL | Université de Lausanne
HEC Lausanne

Time Vs Efficiency

- “Unlimited” potential improvements...
- ...but the improvements become smaller and smaller

➢ Seek for an appropriate level of efficiency



Time Vs Accuracy

- “Unlimited” number of potential models to test…
- …but each model takes time to be computed

➢ Test only models that are known to provide good forecasting accuracies

Automation Vs Control

- Loss of control for automated processes
- Inputting parameters is a manual work

➢ Anticipation of the user’s needs

➢ Propose default values that can be changed

➢ Include intermediate results in the outputs of the function

Time Vs future time

Automation
≈
Investment

- Time invested today to win time tomorrow

➢ Implement feature if

time won tomorrow

>

time spent today

UNIL | Université de Lausanne
HEC Lausanne

Thesis added value

1. Presentation of a methodology that uses advanced statistical methods and that can be implemented in a business context

2. Implementation of a promotion detection algorithm to spot statistically past promotions, to remove their effects and to generate a baseline

3. Automation of the full process to do accurate forecasts using functions that split the data, tests multiple models, selects the best one, forecasts using this model and retrieves the multiple graphs

4. Creation of an R package that regroups the functions, document them and provide examples of their way of working

A lot of concepts and of models where developed by Rob J Hyndman in his book *Forecasting: Principles and Practice* (Hyndman & Athanasopoulos (2018)).
My work hasn't the pretention to improve any of the models nor to come up with new statistical concepts but focuses on the automation and business implementation.

METHODOLOGY

Methodology

Step 1

Taking historical sales data

- Collecting historical sales data
- Uploading the data on R

Step 2

Generating the historical baseline

- Spotting statistically external impactors and removing their effects
- Applying a smoother to cancel phasing effects

baseline()

Step 3

Forecasting the future baseline

- Taking the historical baseline and training machine learning and time series models
- Selecting the model that leads to the best accuracy on the testing set

predict_baseline()
or
baseline() + predict_sales()

Step 4

Forecasting the future final sales

- Taking the forecasted baseline and manually adding the external impactors
- **Alternative:** directly predicting the final sales without using the baseline methodology

predict_baseline() + impactors
or
predict_sales()

R implementation

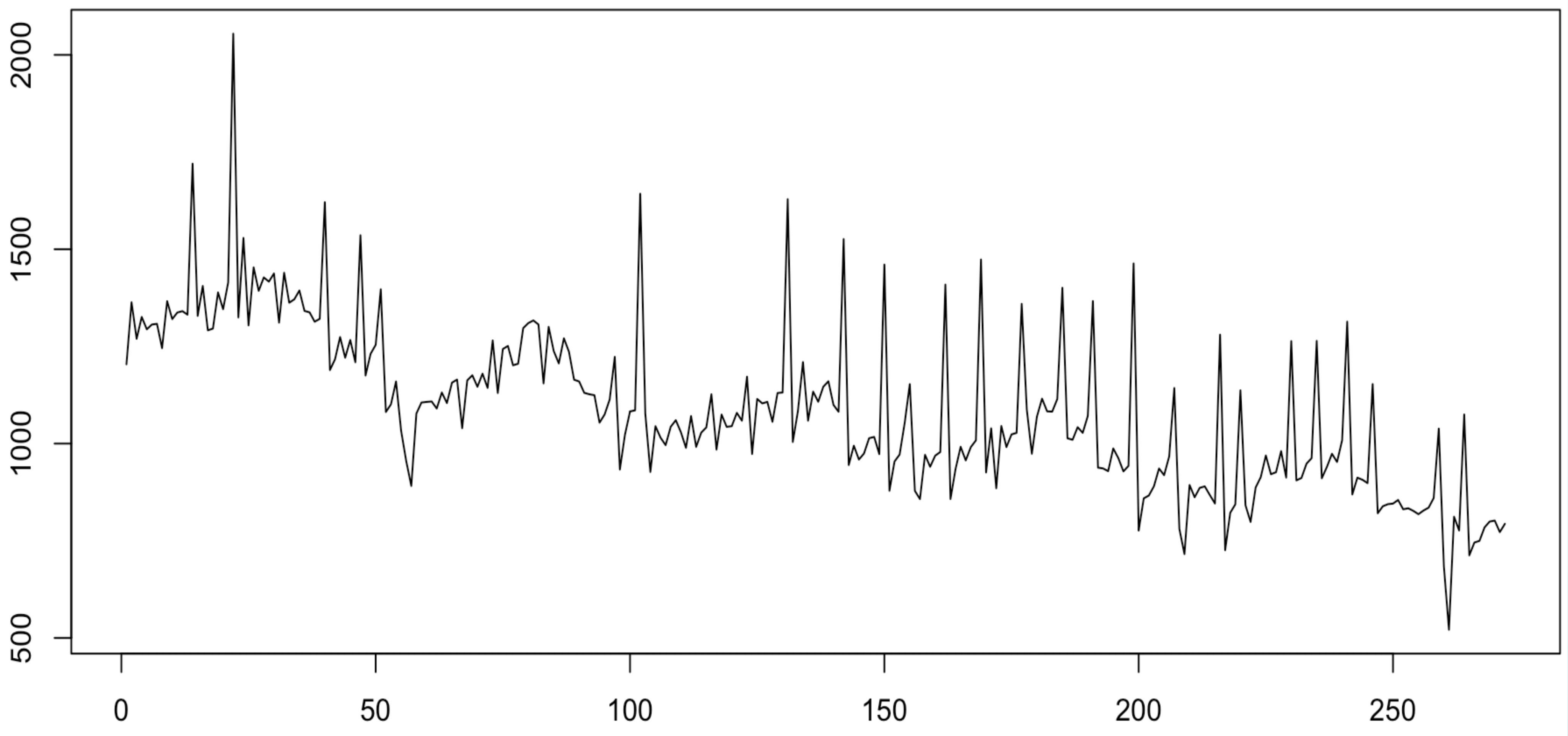
UNIL | Université de Lausanne
HEC Lausanne

Methodology

Step 1	Step 2	Step 3	Step 4
Taking historical sales data	**Generating the historical baseline**	**Forecasting the future baseline**	**Forecasting the future final sales**
Collecting historical sales data	Spotting statistically external impactors and removing their effects	Taking the historical baseline and training machine learning and time series models	Taking the forecasted baseline and manually adding the external impactors
Uploading the data on R	Applying a smoother to cancel phasing effects	Selecting the model that leads to the best accuracy on the testing set	**Alternative:** directly predicting the final sales without using the baseline methodology
	baseline()	*predict_baseline()* or baseline() + predict_sales()	*predict_baseline()* + impactors or *predict_sales()*

R implementation

Example: Step 1



Simple plot of historical sales data of a product that is subject to promotions

UNIL | Université de Lausanne
HEC Lausanne

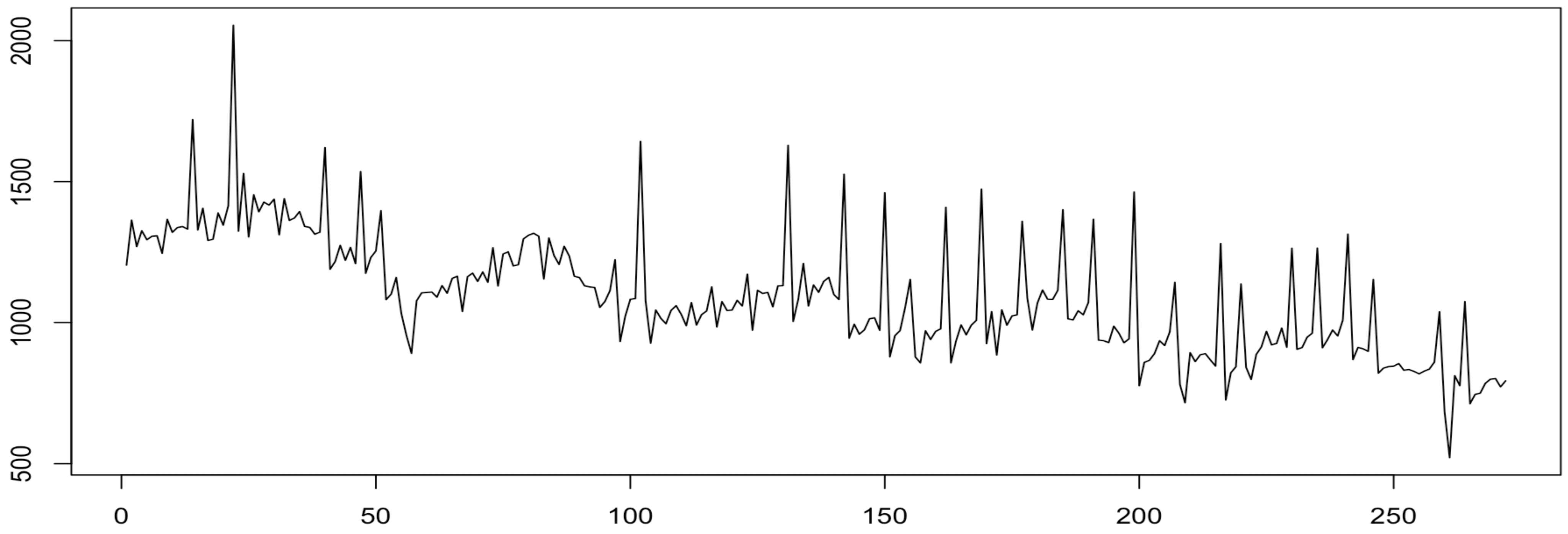
Methodology

Step 1	Step 2	Step 3	Step 4
Taking historical sales data	Generating the historical baseline	Forecasting the future baseline	Forecasting the future final sales
Collecting historical sales data	Spotting statistically external impactors and removing their effects	Taking the historical baseline and training machine learning and time series models	Taking the forecasted baseline and manually adding the external impactors
Uploading the data on R	Applying a smoother to cancel phasing effects	Selecting the model that leads to the best accuracy on the testing set	**Alternative:** directly predicting the final sales without using the baseline methodology
	baseline()	*predict_baseline()* or baseline() + predict_sales()	*predict_baseline()* + impactors or *predict_sales()*

R implementation

Example: Step 2

Baseline: The expected value of a time series given that it is not subject to any promotion, external impactor or phasing effect.



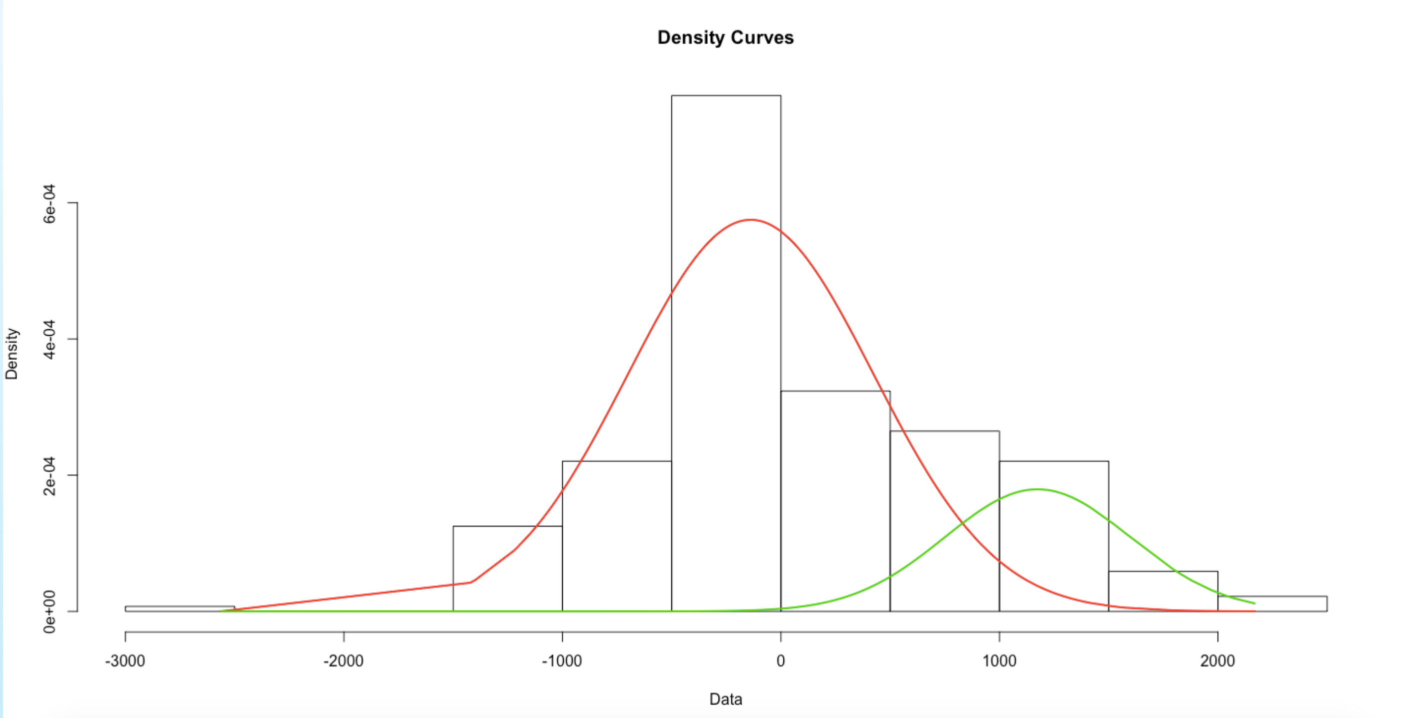
Residuals: The difference between the actuals and their moving median.

Step 2 – Generating the baseline

1) Removing external impactors effects

→ Spotting the external impactors (peaks of demand) that have a significant impact statistically (*mixtools* package)

→ Removing their effect



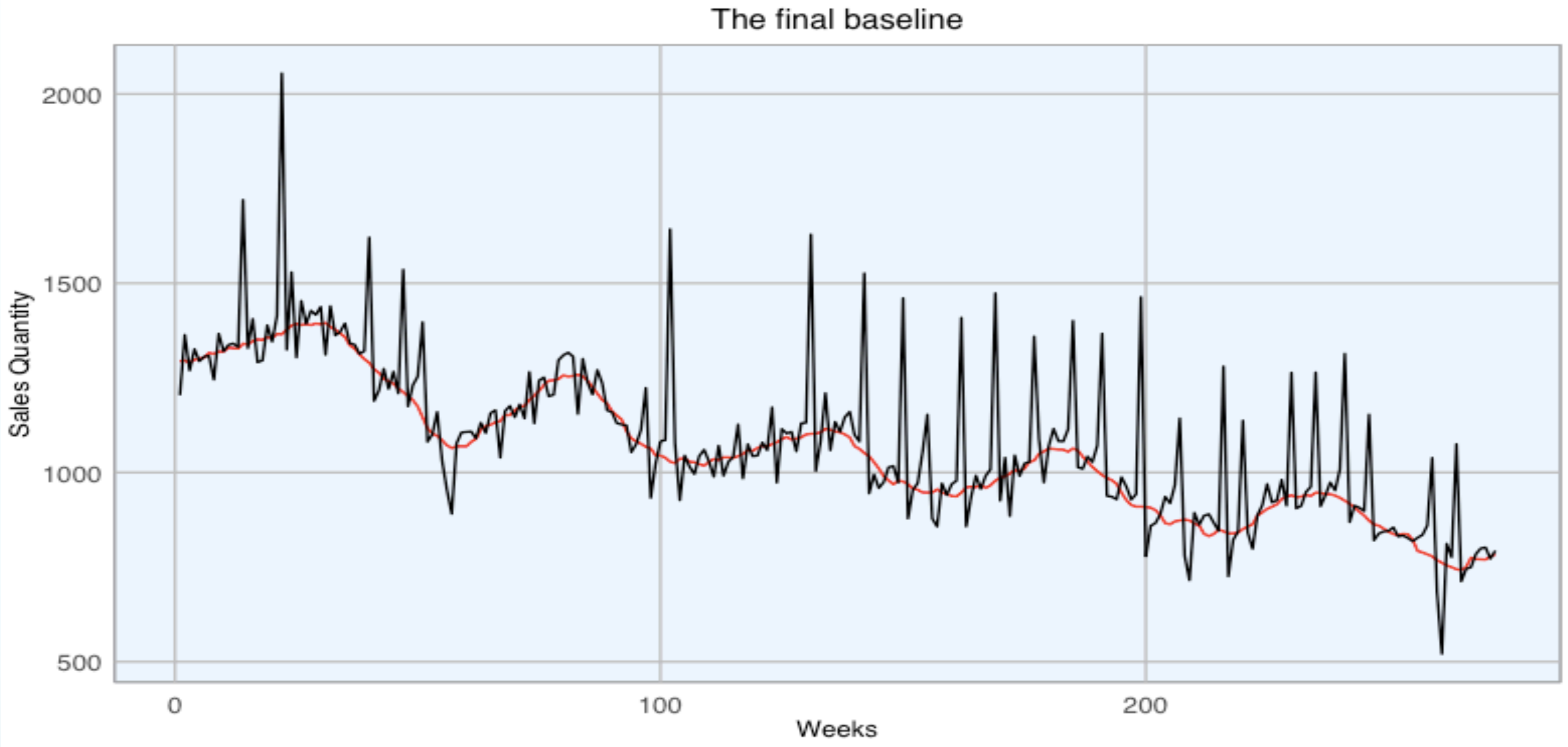
2) Apply a smoother to remove phasing effects

This process is automated in the *baseline()* function

Note that the graph is not the one of the product taken as example. It has been used here for illustration purposes only

UNIL | Université de Lausanne
HEC Lausanne

Example: Step 2



The result of the *baseline()* function applied to the historical data of a product facing promotions

UNIL | Université de Lausanne
HEC Lausanne

Methodology

Step 1	Step 2	Step 3	Step 4
Taking historical sales data	**Generating the historical baseline**	**Forecasting the future baseline**	**Forecasting the future final sales**
Collecting historical sales data	Spotting statistically external impactors and removing their effects	Taking the historical baseline and training machine learning and time series models	Taking the forecasted baseline and manually adding the external impactors
Uploading the data on R	Applying a smoother to cancel phasing effects	Selecting the model that leads to the best accuracy on the testing set	**Alternative:** directly predicting the final sales without using the baseline methodology
	baseline()	*predict_baseline()* or baseline() + predict_sales()	*predict_baseline()* + impactors or *predict_sales()*

R implementation

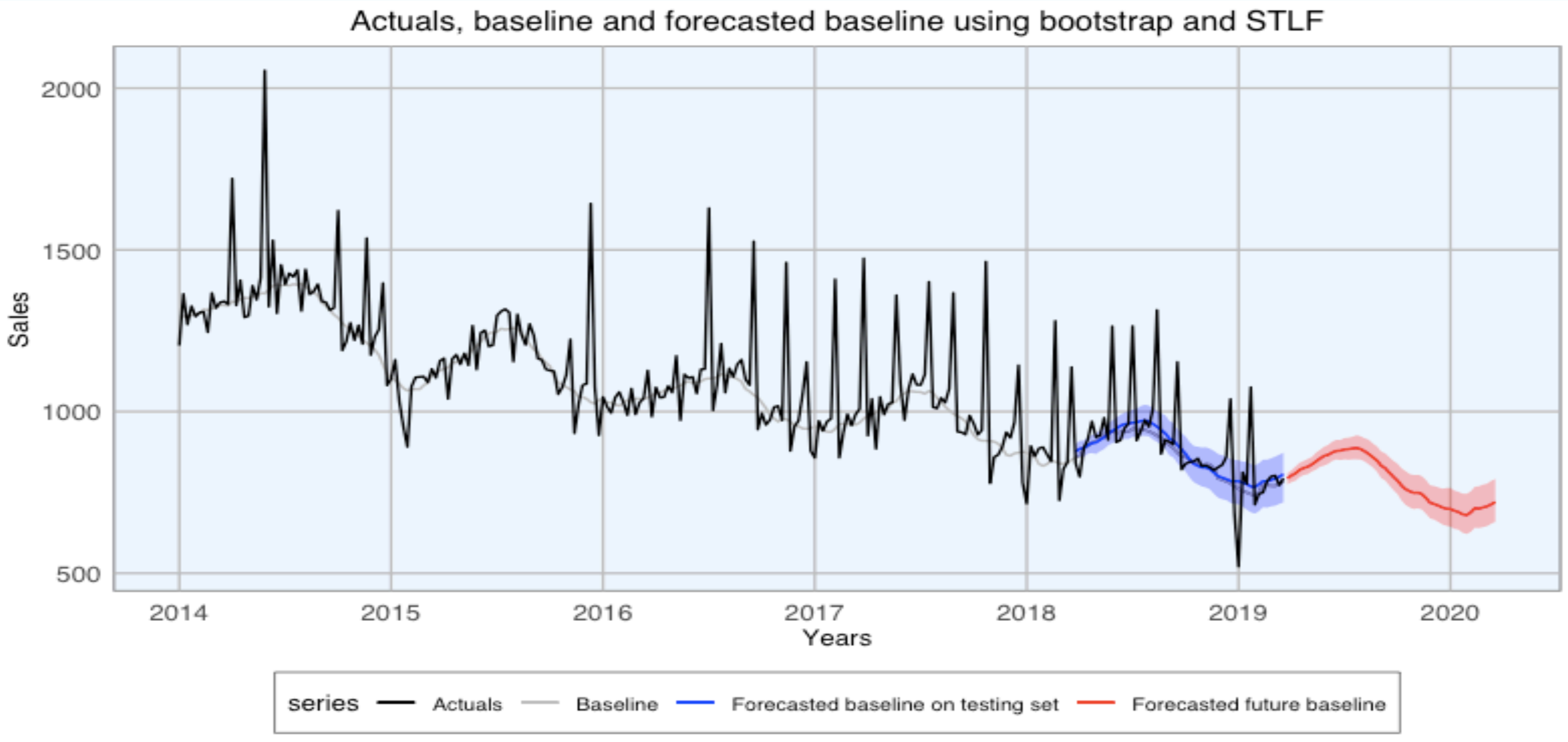
UNIL | Université de Lausanne
HEC Lausanne

Step 3 – Forecasting the future baseline

1. Take the historical baseline
2. Create a training set and a testing set
3. Train multiple time series and machine learning models
4. Evaluate the accuracy of the trained model on the testing set
5. Select the model that performs the best to predict the future baseline

This process is automated in the *predict_baseline()* function

Example: Step 3



The result of the *predict_baseline()* function

UNIL | Université de Lausanne
HEC Lausanne

Methodology

Step 1	Step 2	Step 3	Step 4
Taking historical sales data	Generating the historical baseline	Forecasting the future baseline	Forecasting the future final sales
Collecting historical sales data	Spotting statistically external impactors and removing their effects	Taking the historical baseline and training machine learning and time series models	Taking the forecasted baseline and manually adding the external impactors
Uploading the data on R	Applying a smoother to cancel phasing effects	Selecting the model that leads to the best accuracy on the testing set	**Alternative:** directly predicting the final sales without using the baseline methodology
	baseline()	*predict_baseline()* or baseline() + predict_sales()	*predict_baseline()* + impactors or *predict_sales()*

R implementation

Step 4 – Forecasting the final sales

Option 1: Manually adding impactors to the baseline

1. Estimate the effects of external impactors based on:
   - Managerial inputs
   - Residuals (difference between the actuals and the baseline)
   - Linear regressions
2. Add these effects to the baseline

Option 2: Directly predicting the final sales

1. Forgetting about the baseline methodology
2. Directly predict the final sales of a product by only performing the "Step 3" presented previously

This process is automated in the *predict_sales()* function

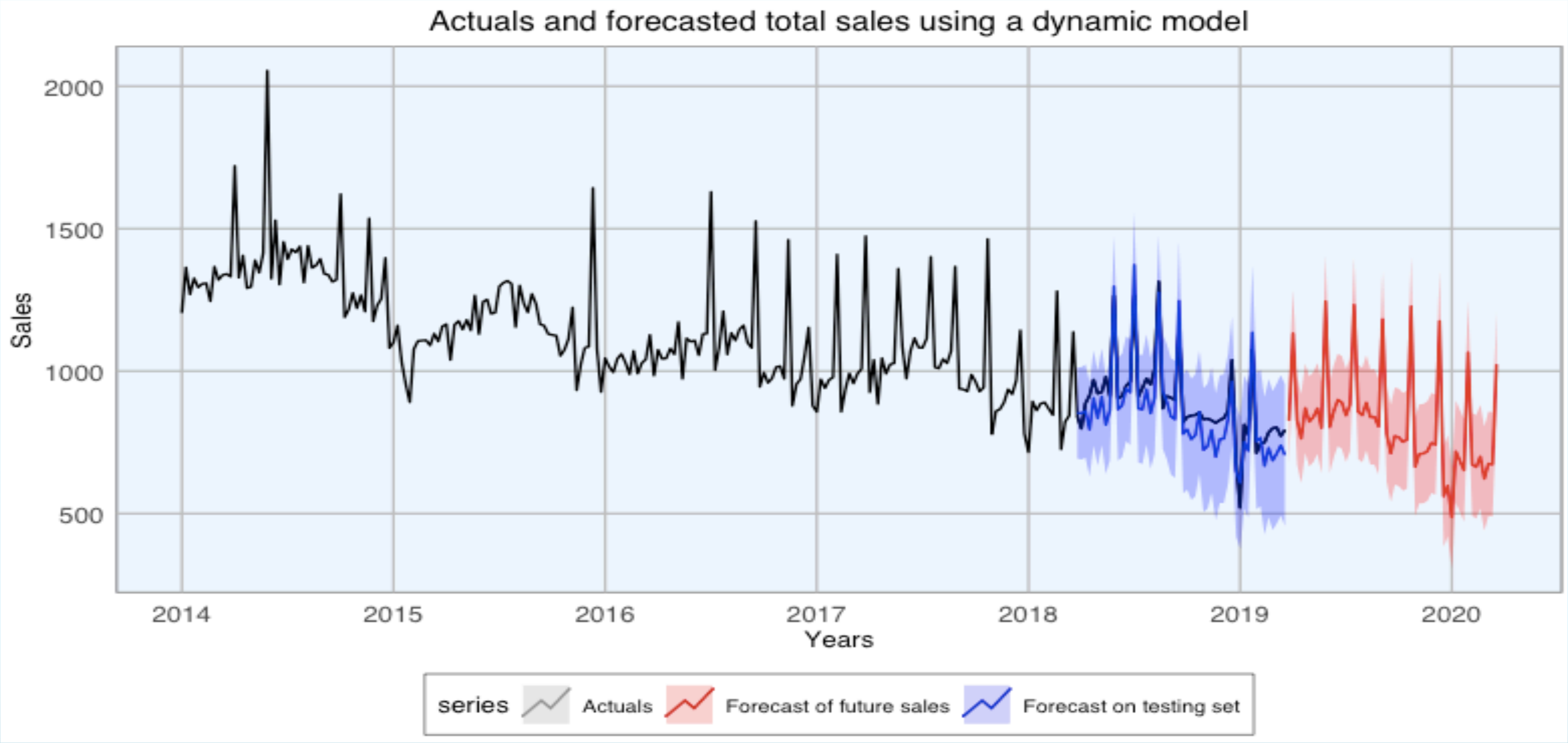
Option 1 – pros:
Clear and precise
Helps to understand external events / promotions impacts / stocking effects / account switching

Option 2 – pros:
Completely automated

Option 1 – cons:
Additional manual work

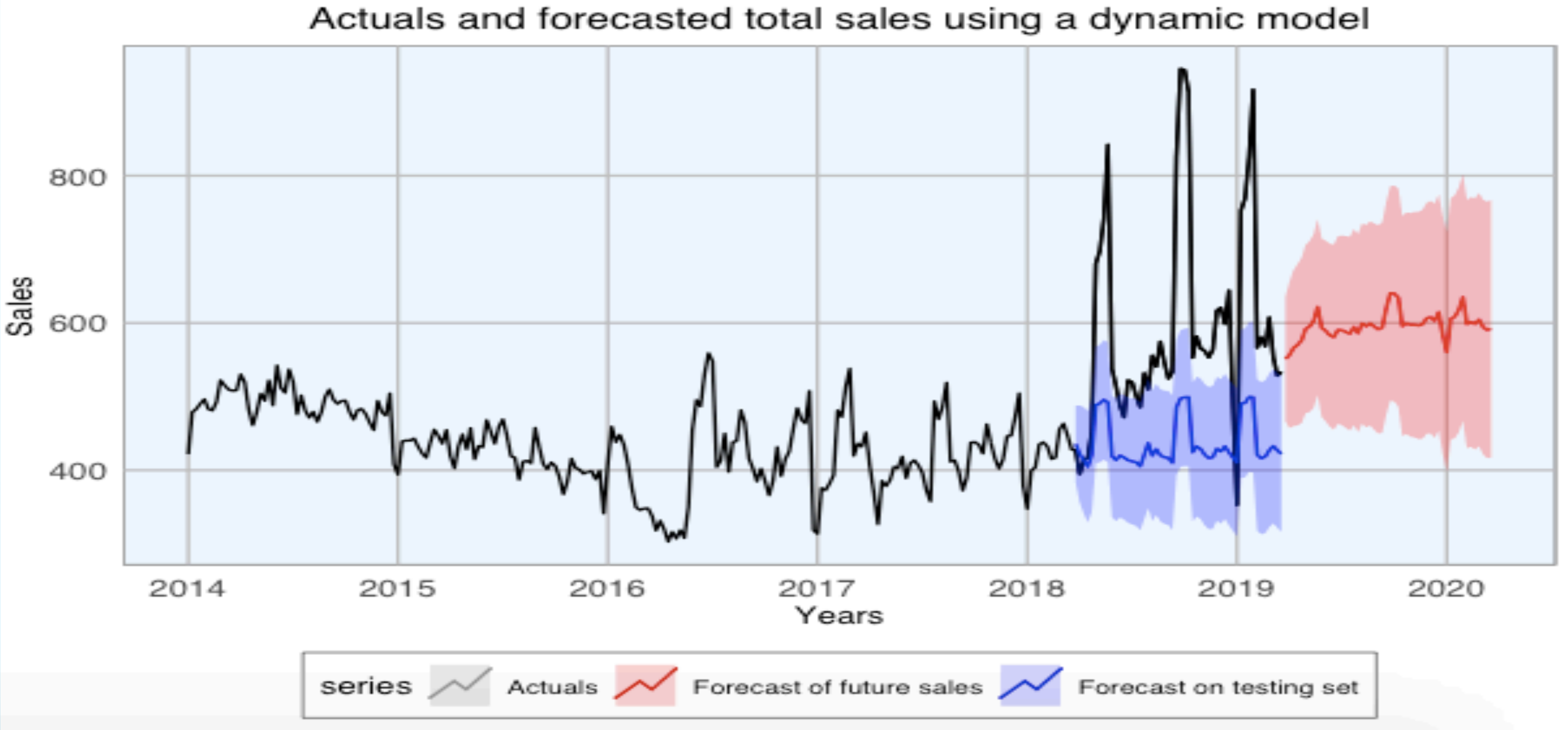
Option 2 – cons:
Can lead to inaccurate results
Doesn't smooth the results
Restricts the analysis to a dynamic model

Example: Step 4 – option 2



The result of the *predict_sales()* function

Example: Step 4 – option 2



The result of the *predict_sales()* function that doesn't lead to accurate results

RESULTS

Used data

Weekly scanning data

From 2014 To W12 of 2019

For each brand In each account

19 products subject to promotions

54 products not subject to promotions

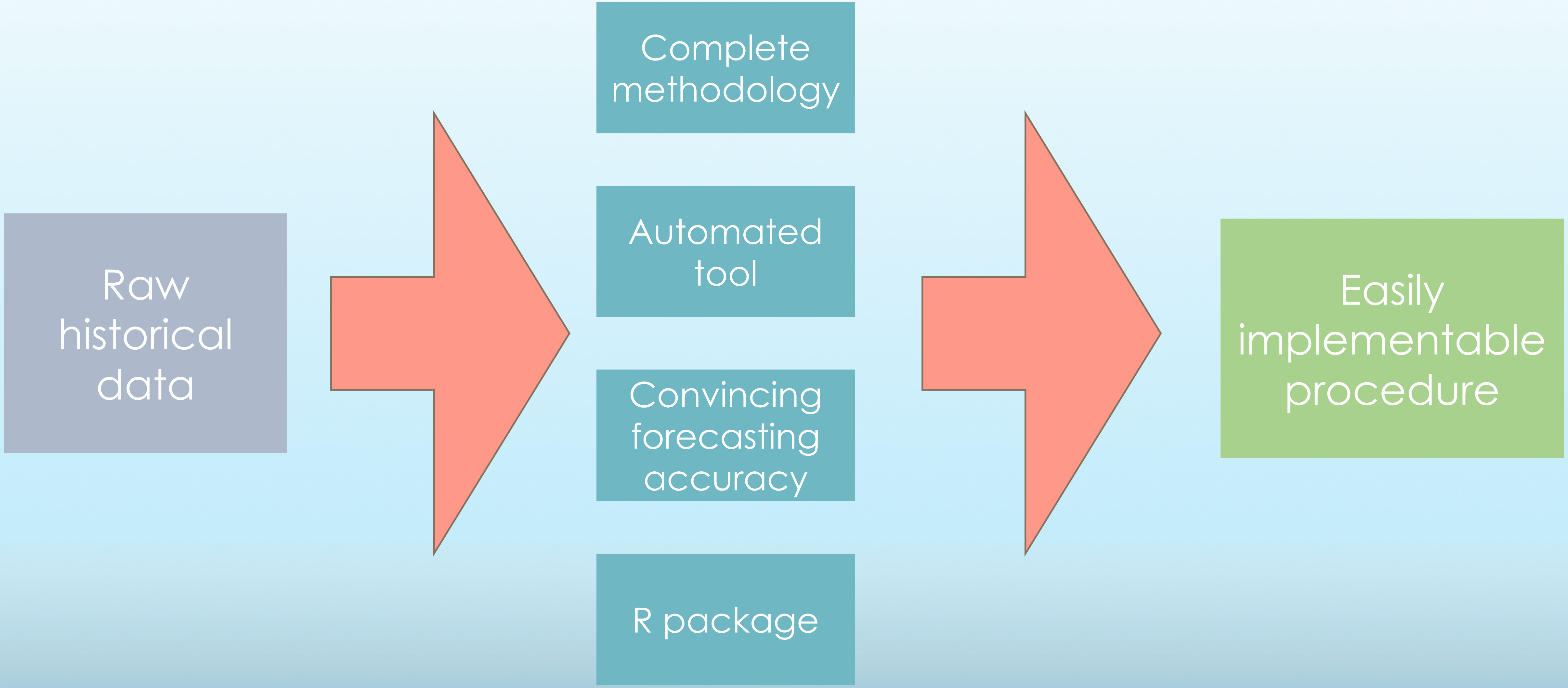
4 different scenarios

	1	2	3	4
Different scenarios	Promotions		No promotions	
	All automated	Baseline + impactors	Baseline (final sales)	Final sales
Used functions	**predict_sales(...)** promo_done = TRUE xreg = ...	**predict_baseline (...)** + manual impactors	**predict_baseline (...)**	**predict_sales(...)**
Obtained accuracy	89%	96%	97%	93%

CONCLUSION

UNIL | Université de Lausanne
HEC Lausanne

Outputs of the thesis



UNIL | Université de Lausanne
HEC Lausanne

Limitations and recommendations

POTENTIAL ISSUES

- Sufficient amount of data should be provided
- The tool has been built to take into account all kind of seasonality but has been tested mainly on weekly data

POSSIBLE IMPROVEMENTS

→ Implementation of a continuous improvement policy

TIME CONSUMPTION

- The functions of the package run the whole methodology
- It can take multiple minutes to be ran for each product

POSSIBLE IMPROVEMENTS

→ Identify and use only the best model
→ Explore new ways to reduce the time consumption

ACCURACY ESTIMATION

- The accuracy estimation might be biased
- It selects the best model and takes this accuracy for granted

POSSIBLE IMPROVEMENTS

→ Creating a second testing set

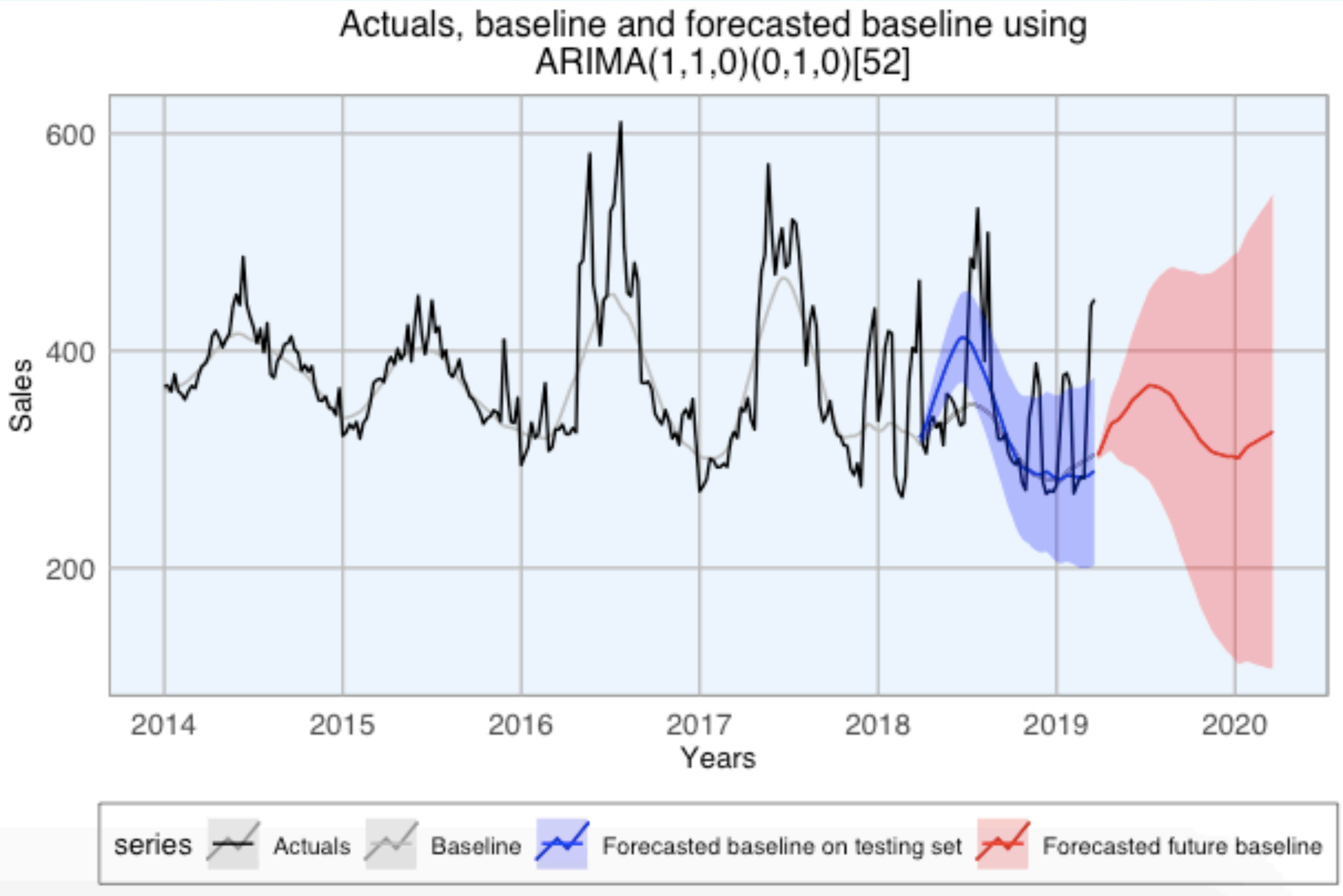
USER INTERFACE

- The user has to run the analysis on R
- Basic knowledge is assumed

POSSIBLE IMPROVEMENTS

→ Propose an adequate user interface (Shiny App, website, etc.)

Limitations



UNIL | Université de Lausanne
HEC Lausanne

THANK YOU

UNIL | Université de Lausanne
HEC Lausanne

References

Benaglia, T., Chauveau, D., Hunter, D., & Young, D. 2009. Mixtools: An r package for analyzing finite mixture models. ***Journal of Statistical Software***, 32(6): 1–29.
Bron, L., Grandadam, P., Kammoun, A., & Rouatbi, A. 2018. Baseline project on dolce gusto products. ***Unpublished Work, University of Lausanne***.
Chatfield, C. 2000. ***Time-series forecasting***. Chapman; Hall/CRC.
Chavez-Demoulin, V., & Davison, A. 2012. Modelling time series extremes. ***REVSTAT-Statistical Journal***, 10(ARTICLE): 109–133.
Chavez-Demoulin, V., Embrechts, P., & Sardy, S. 2014. Extreme-quantile tracking for financial time series. ***Journal of Econometrics***, 181(1).
Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. ***International Journal of Forecasting***, 5(4): 559–583.
Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. 1990. STL: A seasonal-trend decomposition. ***Journal of Official Statistics***, 6(1): 3–73.
Douglas, P. H. 1976. The cobb-douglas production function once again: Its history, its testing, and some new empirical values. ***Journal of Political Economy***, 84(5): 903–915.
George, B. 1994. ***Time series analysis: Forecasting & control, 3/e***. Pearson Education India. Hyndman, R. J., & Athanasopoulos, G. 2018. ***Forecasting: Principles and practice***. OTexts.
Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., et al. 2019. Package "forecast". ***Online] Https://Cran. R-Project. Org/Web/Packages/Forecast/Forecast. Pdf***.
Hyndman, R. J., Khandakar, Y., & others. 2007. ***Automatic time series for forecasting: The forecast package for r***. Monash University, Department of Econometrics; Business Statistics . . . . Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. 2008. ***Forecasting with exponential smoothing: The state space approach***. Springer Science & Business Media.
Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? ***Journal of Econometrics***, 54(1-3): 159–178.
Porteus, E. L. 2002. ***Foundations of stochastic inventory theory***. Stanford University Press.