Forecasting with R
A practical workshop

International Symposium on Forecasting 2016
19th June 2016

Nikolaos Kourentzes

nikolaos@kourentzes.com
http://nikolaos.kourentzes.com

Fotios Petropoulos

fotpetr@gmail.com
http://fpetropoulos.eu

Lancaster University Management School | Lancaster Centre for Forecasting

Cardiff University | Cardiff Business School

Forecasting Society

About us

Nikos

- Associate Professor at Lancaster University
- Member of the Lancaster Centre for Forecasting
- Research interests: temporal aggregation and hierarchies, model selection and combination, intermittent demand, promotional modelling and supply chain collaboration
- Forecasting blog: http://nikolaos.kourentzes.com

Fotios

- Assistant Professor at Cardiff University
- Forecasting Support Systems Editor of *Foresight*
- Director of the International Institute of Forecasters
- Research interests: behavioural aspects of forecasting and improving the forecasting process, applied in the context of business and supply chain

Nikos and Fotios are the founders of the **Forecasting Society** (www.forsoc.net)

Outline of the workshop

1. Overview of R Studio
2. Introduction to R
3. Time series exploration

   Time series components, decomposition, ACF/PACF functions, …
4. Forecasting for fast demand

   Naïve, Exponential Smoothing, ARIMA, MAPA, Theta, evaluation, …
5. Forecasting for intermittent demand

   Croston's method, SBA, TSB, temporal aggregation, classification, …
6. Forecasting with causal methods

   Simple and multiple regression, residual diagnostics, selecting variables, …
7. Advanced methods in forecasting

   Hierarchical forecasting, ABC-XYZ analysis, LASSO

Have fun and enjoy your day!

Section 1

1. **Overview of R Studio**
2. Introduction to R
3. Time series exploration

   Time series components, decomposition, ACF/PACF functions, …
4. Forecasting for fast demand

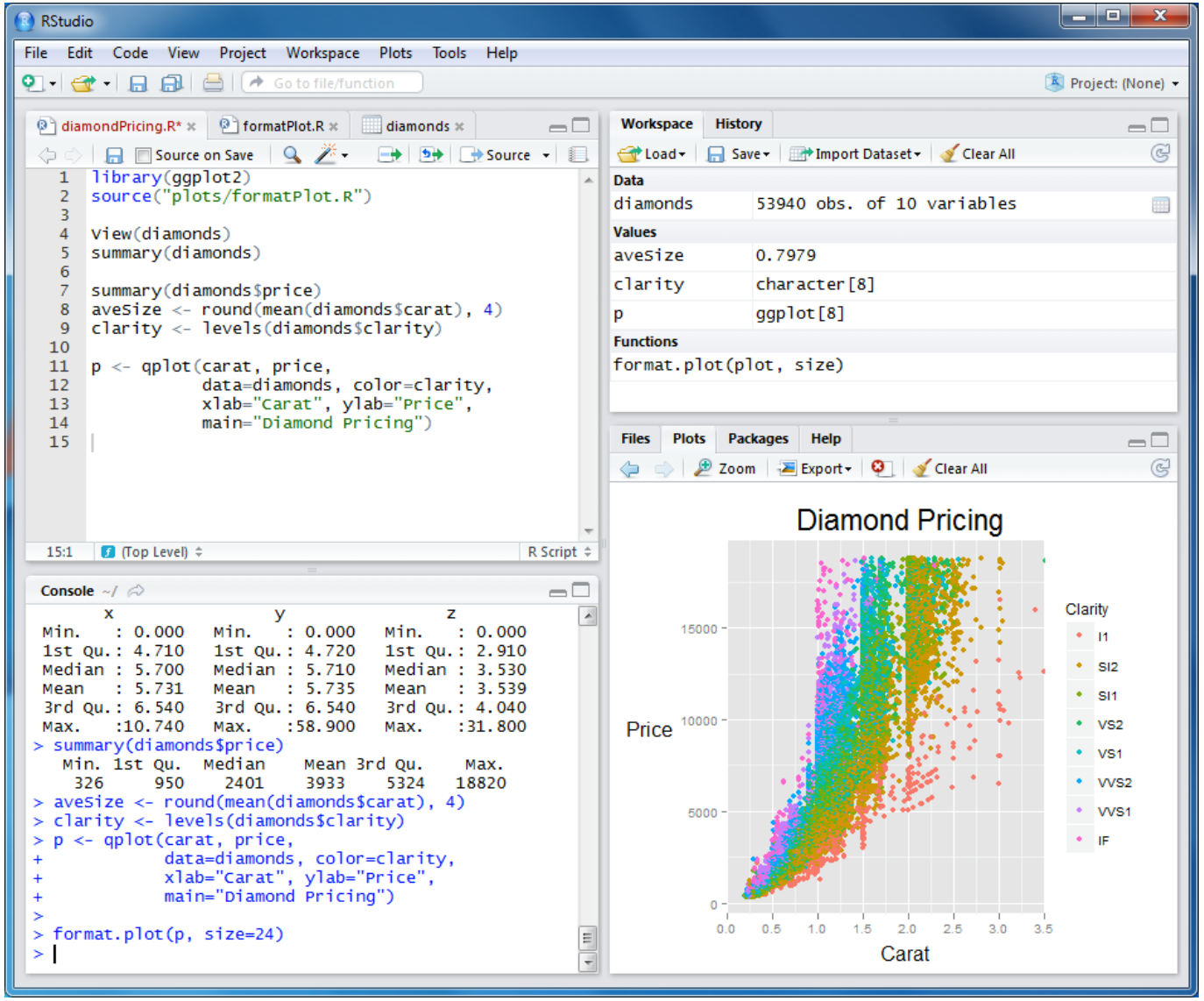
   Naïve, Exponential Smoothing, ARIMA, MAPA, Theta, evaluation, …
5. Forecasting for intermittent demand

   Croston's method, SBA, TSB, temporal aggregation, classification, …
6. Forecasting with causal methods

   Simple and multiple regression, residual diagnostics, selecting variables, …
7. Advanced methods in forecasting

   Hierarchical forecasting, ABC-XYZ analysis, LASSO

Overview of RStudio



Section 2

1. ~~Overview of R Studio~~
2. **Introduction to R**
3. Time series exploration
   Time series components, decomposition, ACF/PACF functions, …
4. Forecasting for fast demand
   Naïve, Exponential Smoothing, ARIMA, MAPA, Theta, evaluation, …
5. Forecasting for intermittent demand
   Croston's method, SBA, TSB, temporal aggregation, classification, …
6. Forecasting with causal methods
   Simple and multiple regression, residual diagnostics, selecting variables, …
7. Advanced methods in forecasting
   Hierarchical forecasting, ABC-XYZ analysis, LASSO

Section 3

1. ~~Overview of R Studio~~
2. ~~Introduction to R~~
3. **Time series exploration**
   Time series components, decomposition, ACF/PACF functions, …
4. Forecasting for fast demand
   Naïve, Exponential Smoothing, ARIMA, MAPA, Theta, evaluation, …
5. Forecasting for intermittent demand
   Croston's method, SBA, TSB, temporal aggregation, classification, …
6. Forecasting with causal methods
   Simple and multiple regression, residual diagnostics, selecting variables, …
7. Advanced methods in forecasting
   Hierarchical forecasting, ABC-XYZ analysis, LASSO

Section 4

~~1. Overview of R Studio~~

~~2. Introduction to R~~

~~3. Time series exploration~~

~~Time series components, decomposition, ACF/PACF functions, …~~

4. Forecasting for fast demand

Naïve, Exponential Smoothing, ARIMA, MAPA, Theta, evaluation, …

5. Forecasting for intermittent demand

Croston's method, SBA, TSB, temporal aggregation, classification, …

6. Forecasting with causal methods

Simple and multiple regression, residual diagnostics, selecting variables, …

7. Advanced methods in forecasting

Hierarchical forecasting, ABC-XYZ analysis, LASSO

Exponential Smoothing (ets)

The state space implementation of exponential smoothing considers the following combinations of error, trend and seasonality:

- Error: **A**dditive or **M**ultiplicative
- Trend: **N**one, **A**dditive or **M**ultiplicative (damped or not)
- Season: **N**one, **A**dditive or **M**ultiplicative

The usual notation is ETS(Error, Trend, Season), so for instance:

- ETS(A,N,N) has additive errors, no trend and no season → SES
- ETS(M,M,M) has all components multiplicatively

Exponential Smoothing (ets)

We typically optimise ETS using MLE or equivalently minimise the augmented sum of squared errors criterion:

$$\mathcal{S}(\boldsymbol{\theta}, \boldsymbol{x}_0) = [exp(\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{x}_0))]^{1/n} = \left|\prod_{t=1}^{n} r(\boldsymbol{x}_{t-1})\right|^{2/n} \sum_{t=1}^{n} e_t^2$$

For additive errors $r(\boldsymbol{x}_{t-1}) = 1$, so this is equal to the well known MSE:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2$$

This is used to optimise both the smoothing parameters and the initial values.

Exponential Smoothing (ets)

Having a likelihood allows us to use information criteria to select the best ETS model out of the 30 possible alternatives. A common choice is Akaike's Information Criterion:

$$\text{AIC} = -2ln(\mathcal{L}) + 2k$$

Given that time series often have limited sample size a better selection is to use AICc that is corrected for sample size. This is the default option in the forecast package.

ARIMA (auto.arima)

The function auto.arima allows automatic specification of SARIMA models. This is done as follows:

- Test for stationarity in a seasonal context using OCSB (up to 1 seasonal difference)
- Test for stationarity using KPSS (up to 2 differences)
- Difference appropriately based on the test results
- Start from a reasonable AR and MA order and search neighbouring specifications (max AR & MA order: 5, max SAR & SMA order: 2)
- Compare alternative models using AICc (default) and pick best.

TBATS (tbats)

TBATS uses Box-Cox transformation, exponential smoothing, trigonometric seasonality and ARMA errors:

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega}-1}{\omega}; & \omega \neq 0 \\ \log y_t & \omega = 0 \end{cases}$$

Box-Cox transform

$$y_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-1}^{(i)} + d_t$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$$

Deterministic and stochastic trend

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t$$

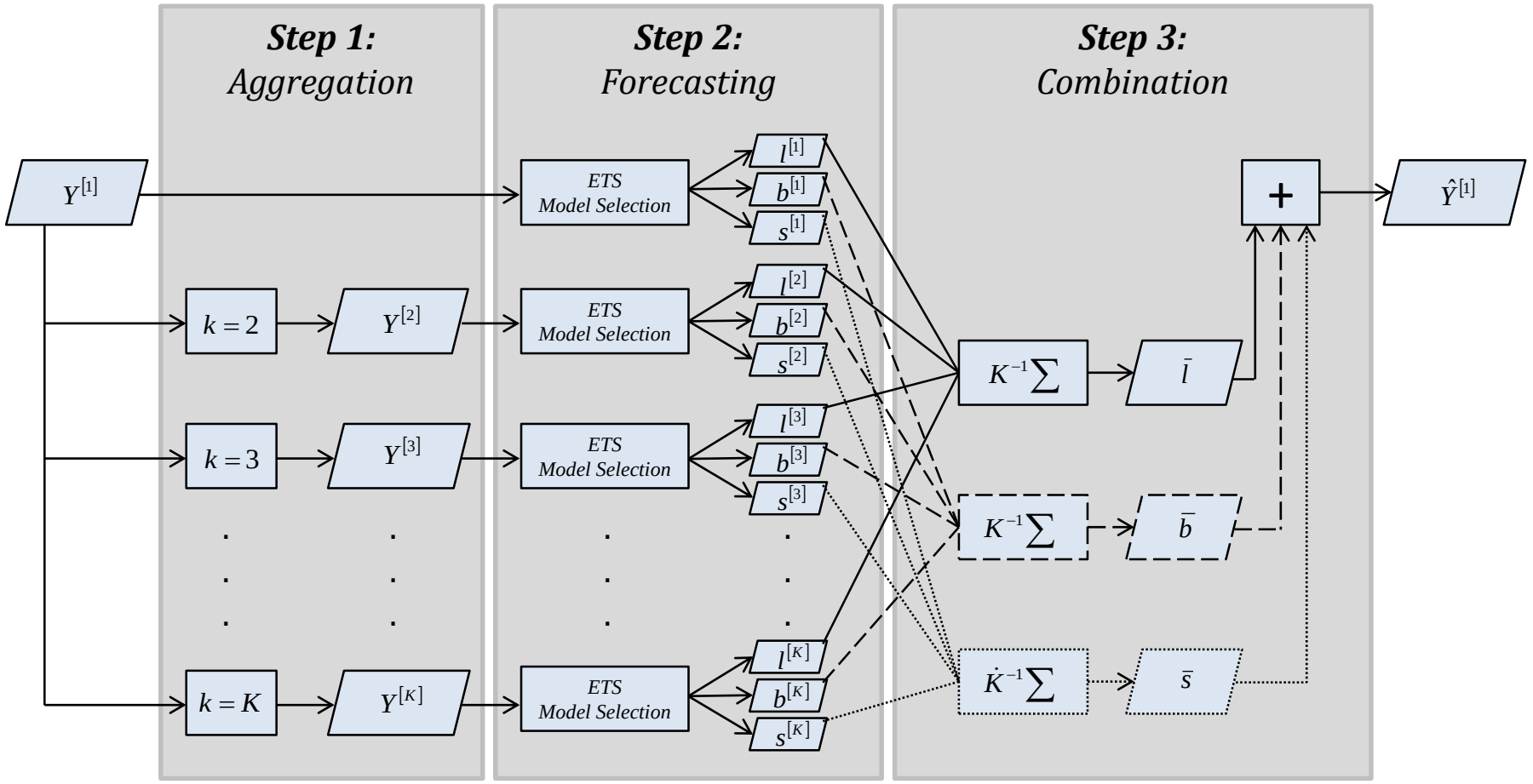
$$s_{j,t}^{*(i)} = -s_{j,t-1} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t$$

Trigonometric seasonlity

$$d_t = \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t,$$

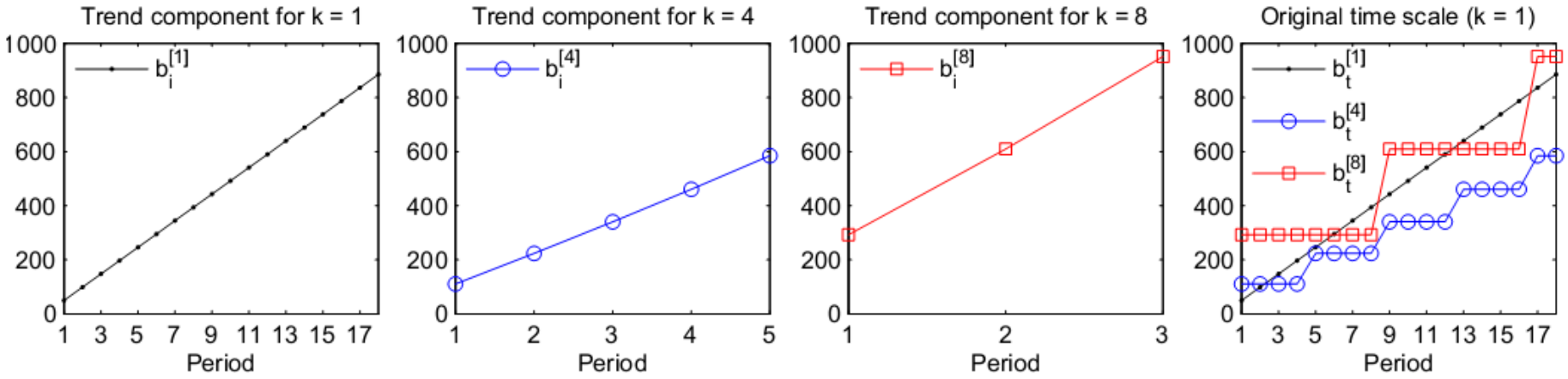
ARMA errors

Multiple Aggregation Prediction Algorithm (mapa)

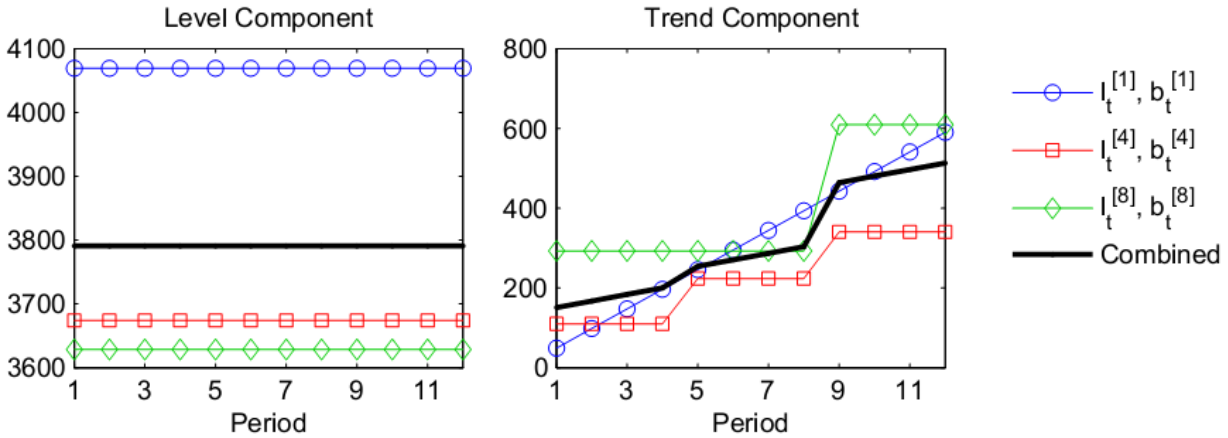


Multiple Aggregation Prediction Algorithm (mapa)

Transform states to additive and to original sampling frequency



Combine states (components)



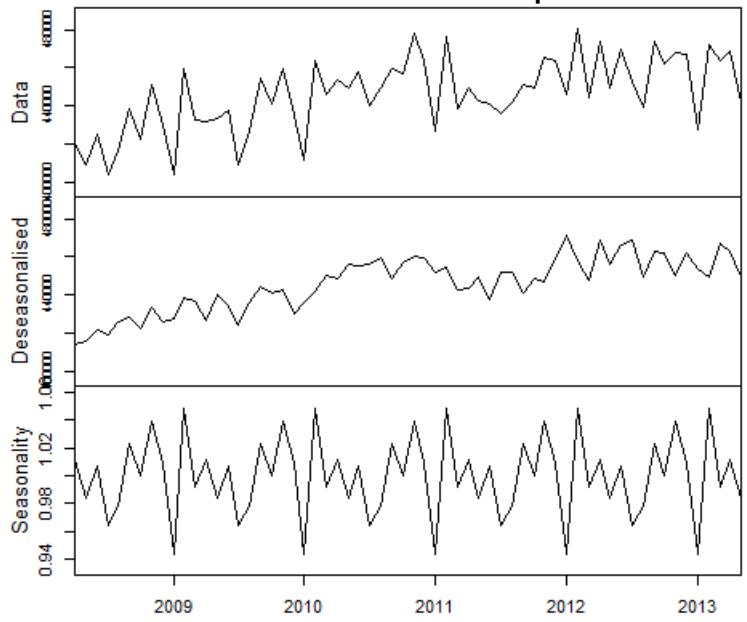
Produce forecasts

$$\hat{y}^{[1]}_{t+h^{[1]}} = \bar{l}_{t+h^{[1]}} + \bar{b}_{t+h^{[1]}} + \bar{s}_{t-S+h^{[1]}}$$

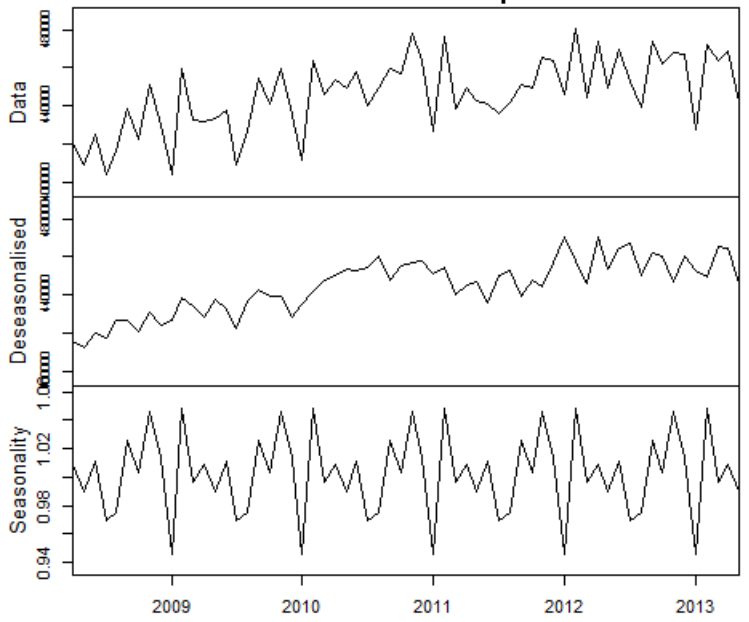
Theta method (theta)

First a time series is decomposed using **classical multiplicative decomposition**:

Deterministic decomposition



Stochastic decomposition



In TStools to allow the seasonal pattern to evolve a pure seasonal model is used instead:

$$s_t = s_{t-m} + \gamma \left(\frac{y_t}{w_t} - s_{t-m} \right)$$

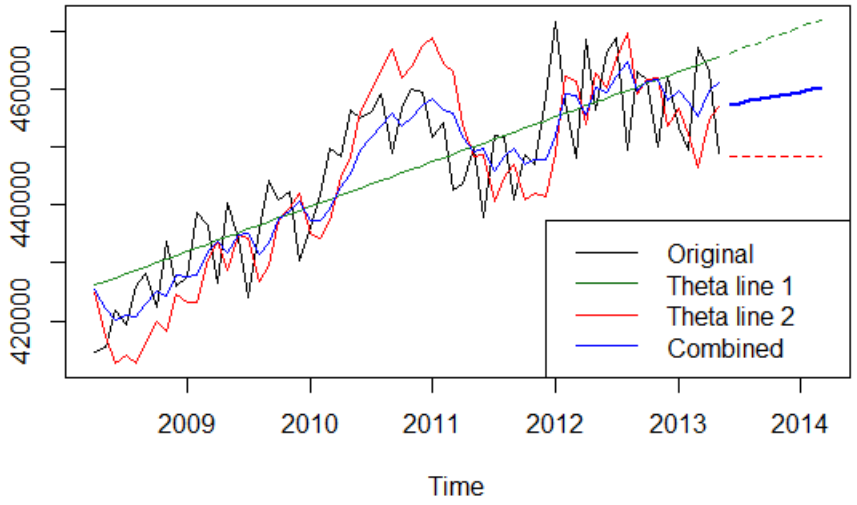
Obviously when γ → 0 then it is the deterministic case.

Theta method (theta)

Then the deseasonalised time series is broken down in two lines:

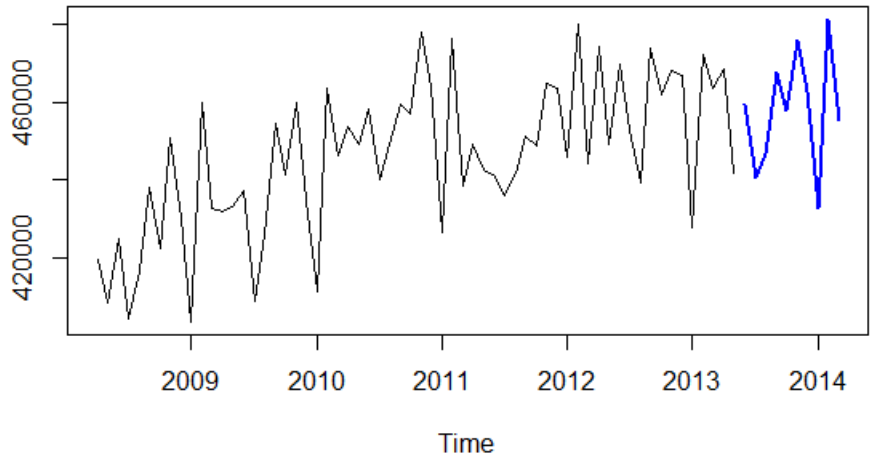
- a linear trend → long term trend
- 2 x (deseasonalised data - linear trend) → inflate variability

Each series is forecasted separately using linear regression and single exponential smoothing and their forecast is then combined:



Theta method (theta)

Finally the forecast of the deseasonalised time series is re-seasonalised with the indices calculated previously to give the final forecast:

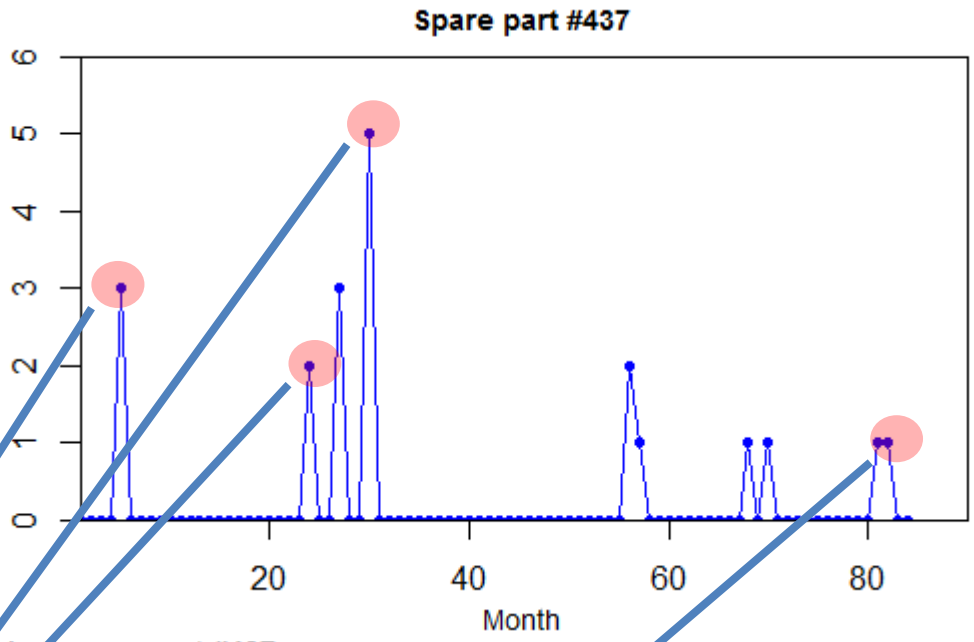


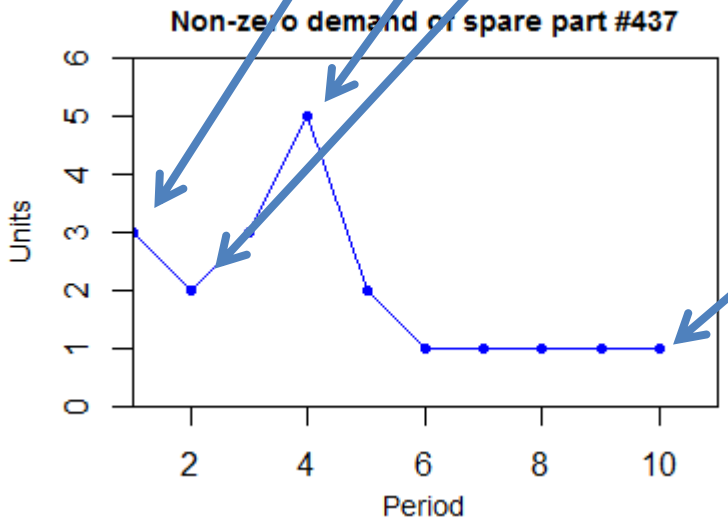
Section 5

1. ~~Overview of R Studio~~
2. ~~Introduction to R~~
3. ~~Time series exploration~~
   ~~Time series components, decomposition, ACF/PACF functions, …~~
4. ~~Forecasting for fast demand~~
   ~~Naïve, Exponential Smoothing, ARIMA, MAPA, Theta, evaluation, …~~
5. **Forecasting for intermittent demand**
   Croston's method, SBA, TSB, temporal aggregation, classification, …
6. Forecasting with causal methods
   Simple and multiple regression, residual diagnostics, selecting variables, …
7. Advanced methods in forecasting
   Hierarchical forecasting, ABC-XYZ analysis, LASSO

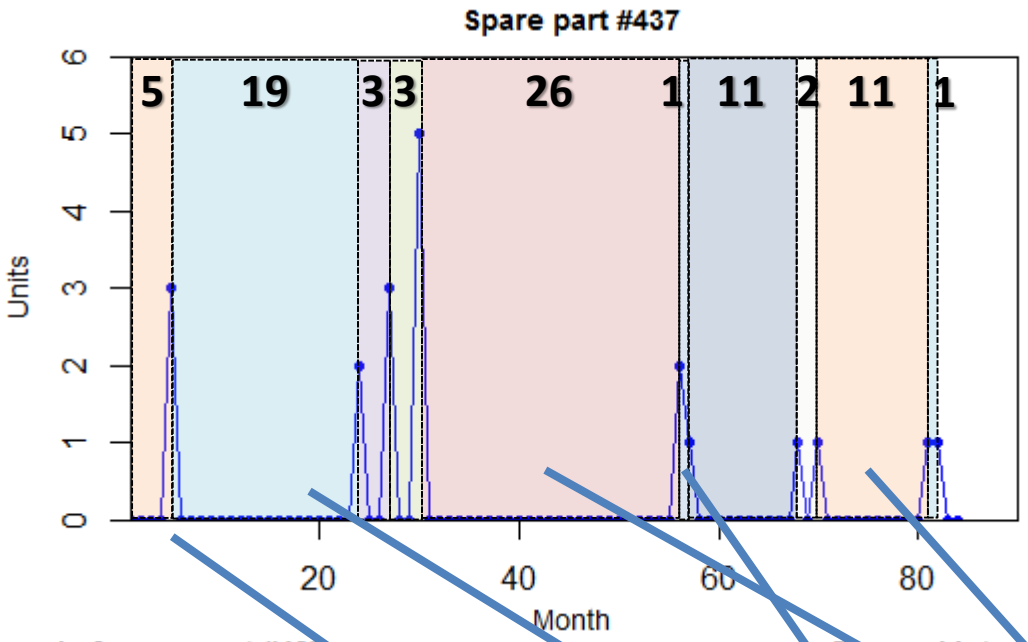
Croston's method

From the original series we first construct a non-zero demand series (z)

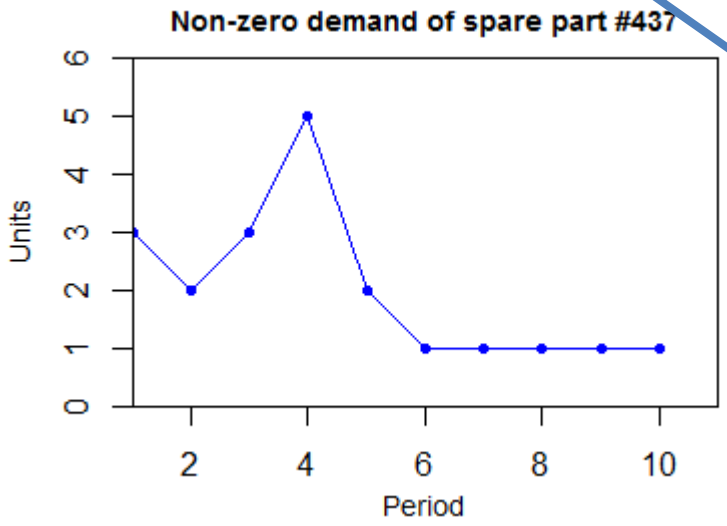


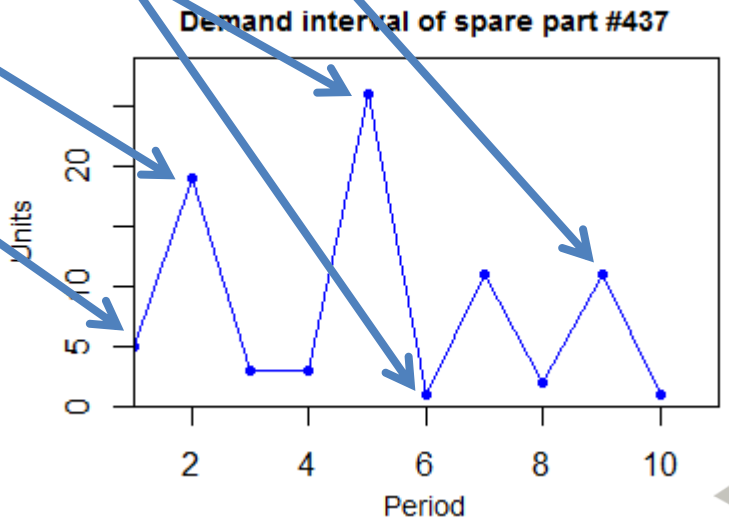


Croston's method

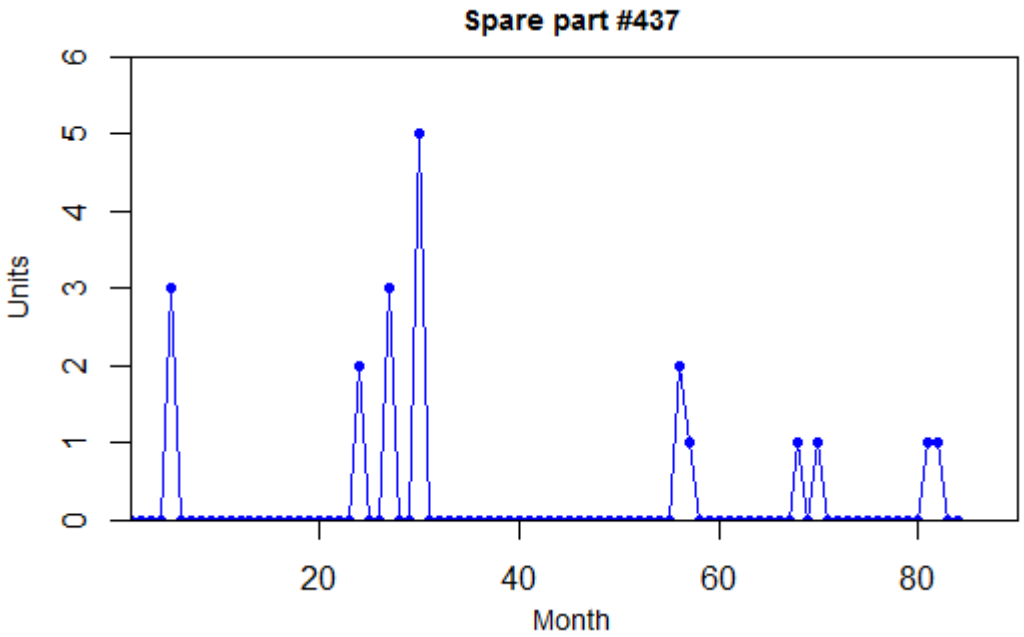


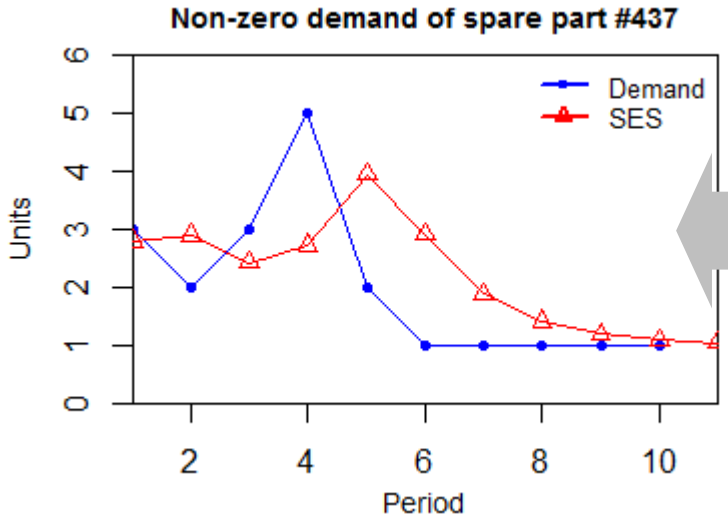
The we create an interval series by counting every how many periods there is demand (x).



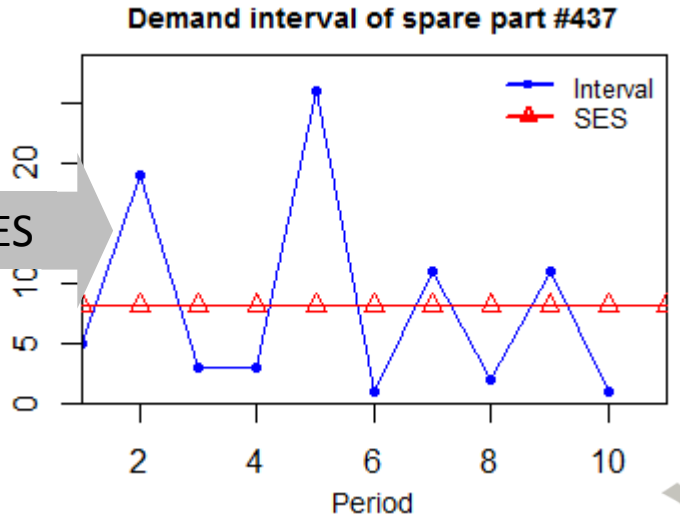


Croston's method



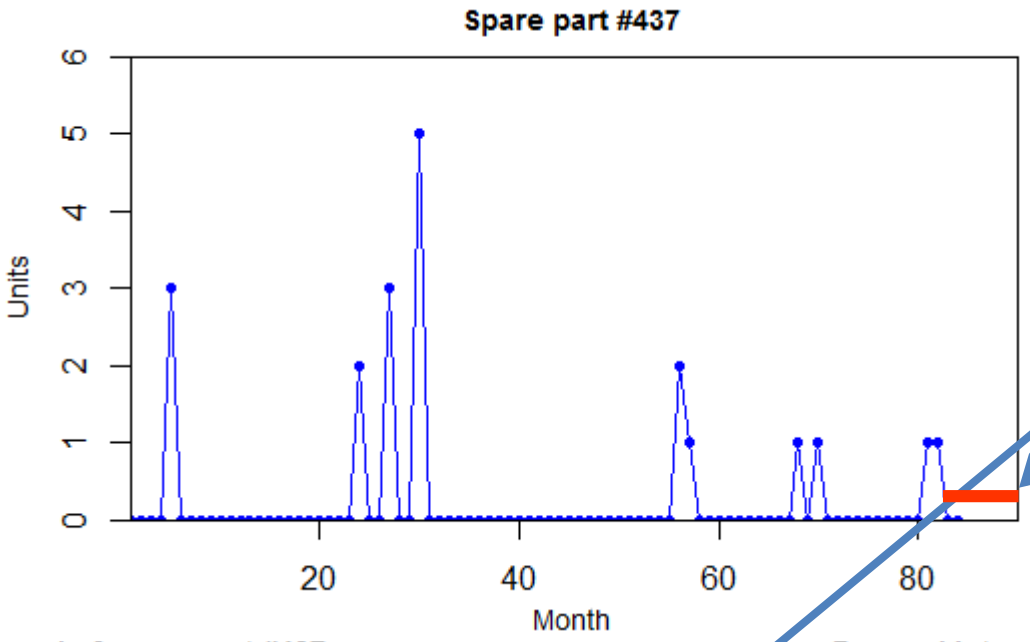


Forecast with SES



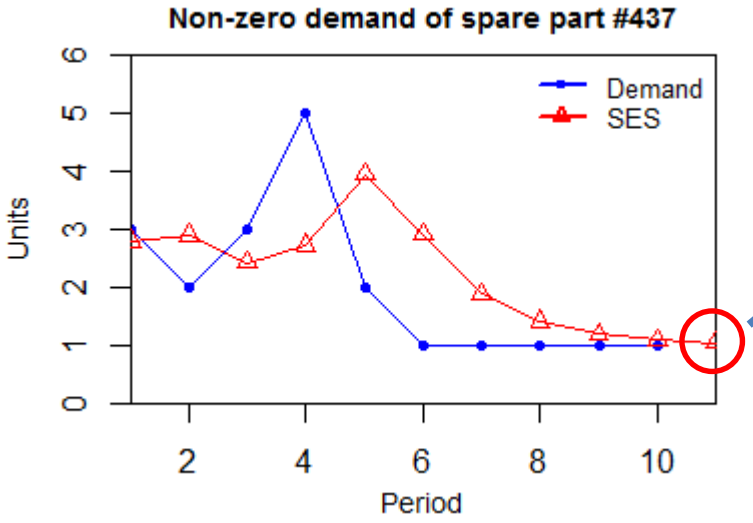
Croston's method

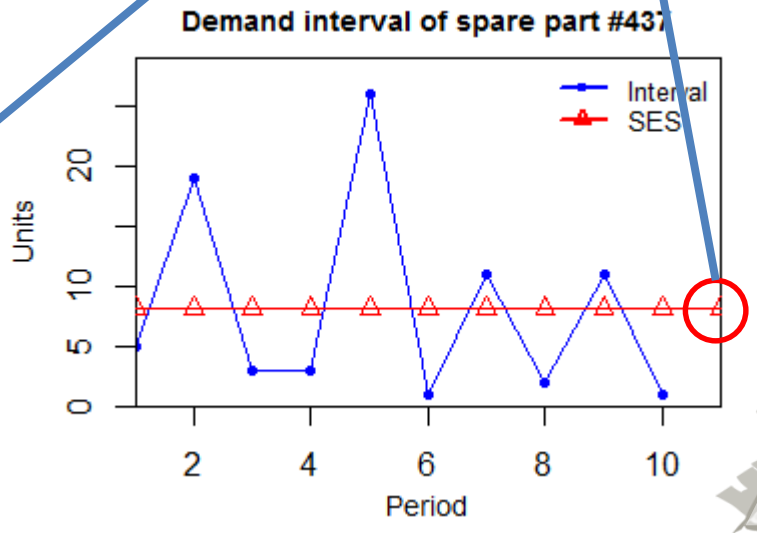
We divide the estimated demand and interval to produce the Croston forecast



Demand

Interval





SBA

Syntetos and Boylan [2005] proposed an approximation that corrects the inversion bias in Croston's method.

Croston

$$\hat{c}_t = \frac{\hat{z}_t}{\hat{x}_t}$$

Smooth demand size

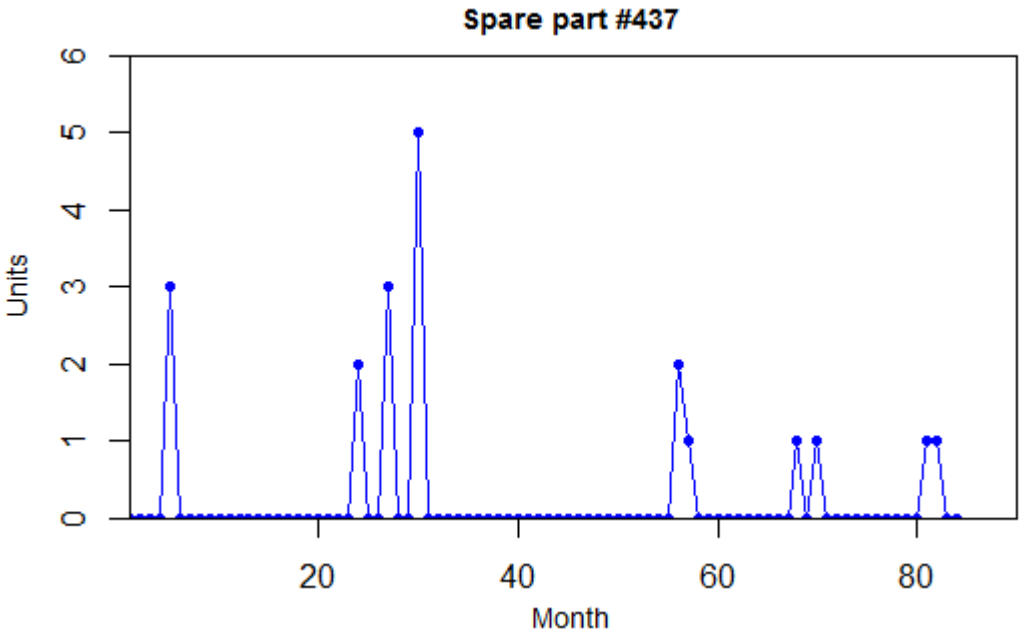
Smooth demand interval

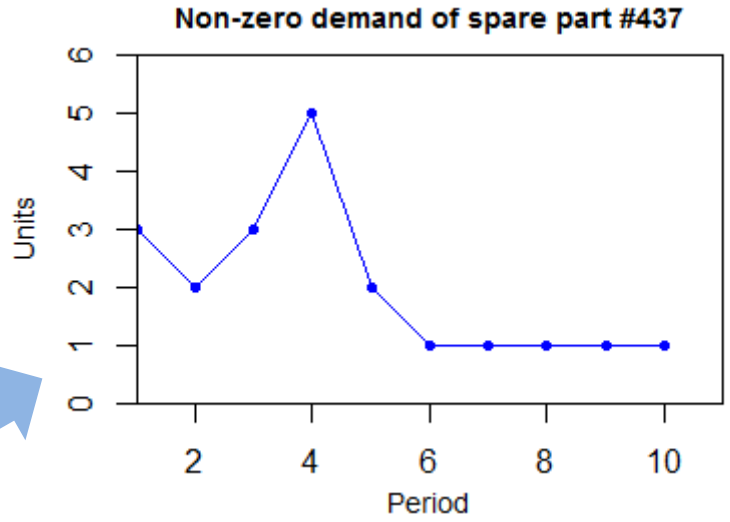
SBA

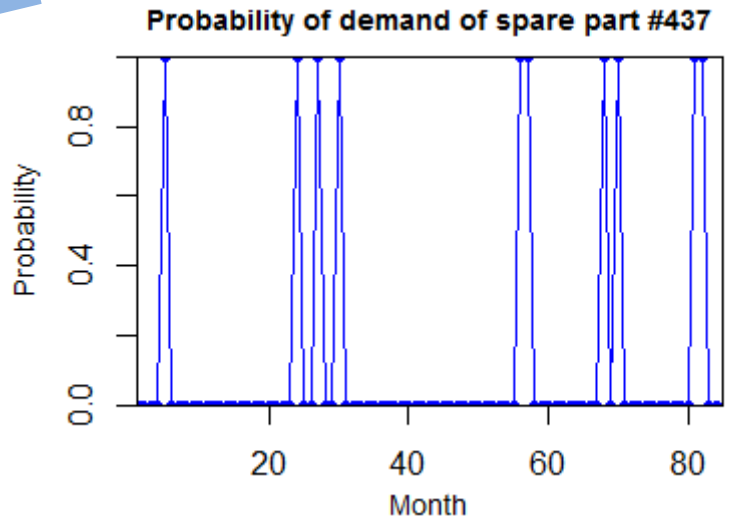
Smoothing parameter of intervals

$$\hat{c}'_t = \left(1 - \frac{a_x}{2}\right) \frac{\hat{z}_t}{\hat{x}_t} = \left(1 - \frac{a_x}{2}\right) \hat{c}$$

TSB Method

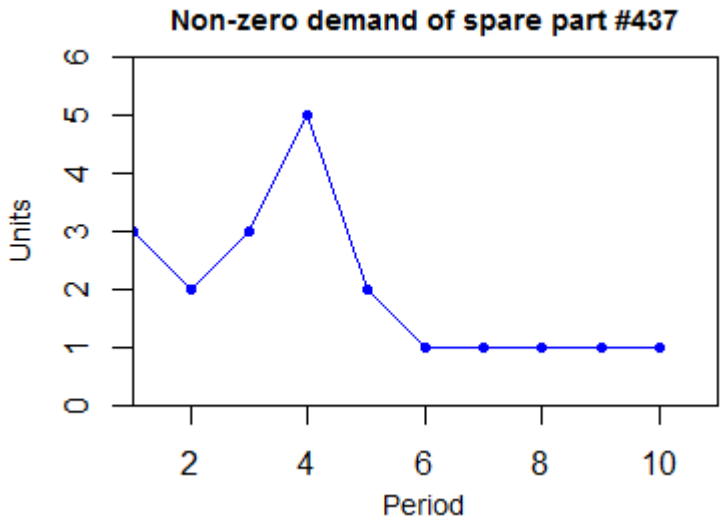


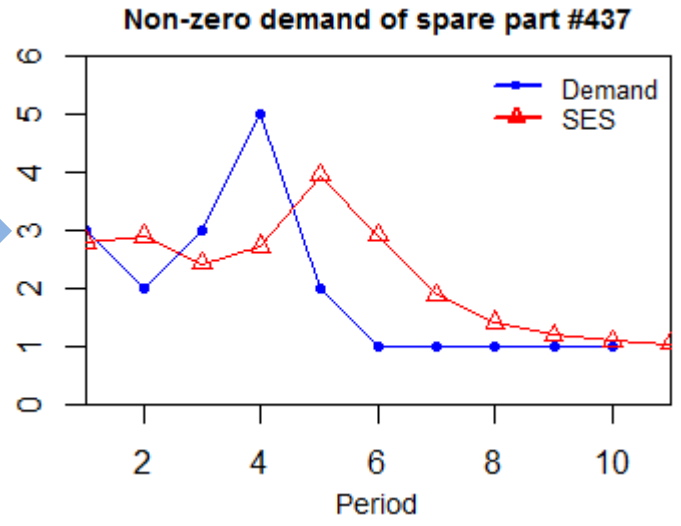




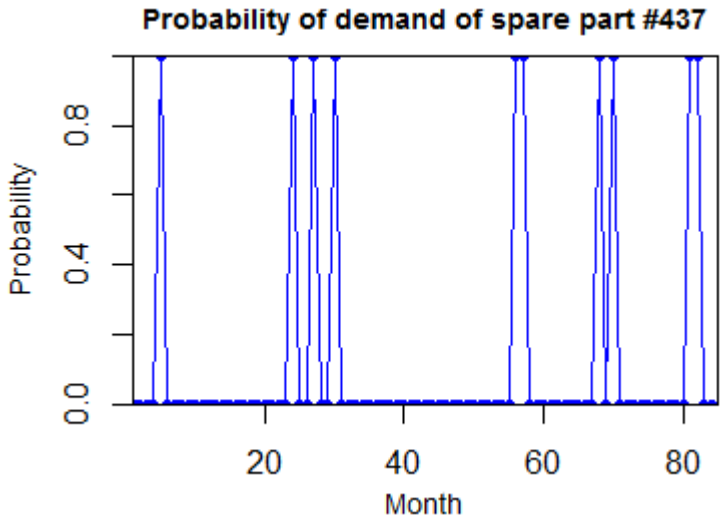
The demand probability is equal to 1 when demand occurred. This series is as long as the original series

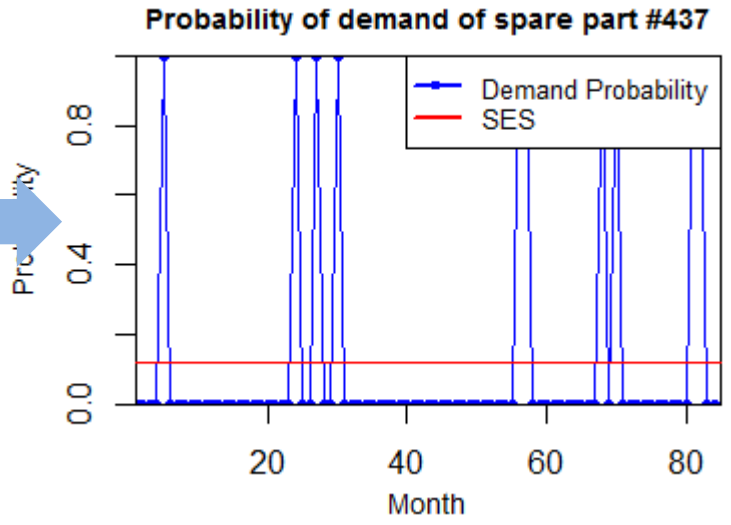
TSB Method





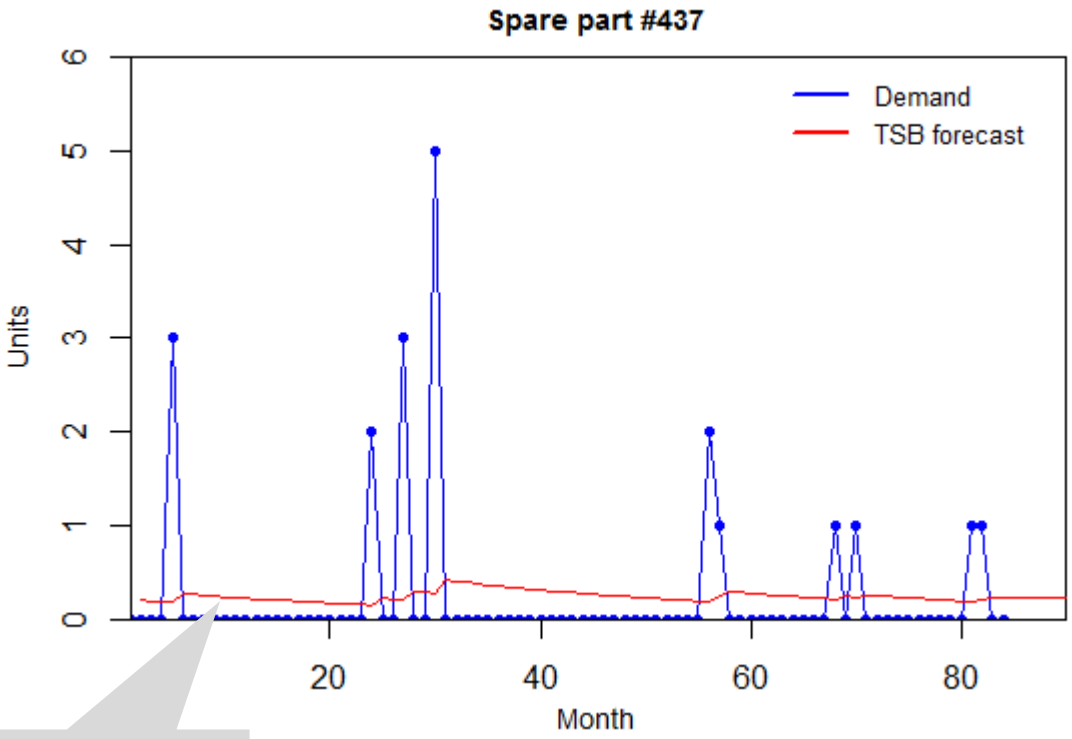
$$\hat{f}_t = \hat{z}_t \hat{d}_t$$





The forecast is the product of the demand and probability estimates

TSB Method



The decline in the forecast is because TSB models the obsolescence of the item.

Classification

For an ID time series we can calculate the non-zero demand (**z**) and the demand interval (**x**). Using these we can define:

$$p = \bar{x}$$

$$v = \left(\frac{s_z}{\bar{z}} \right)^2$$

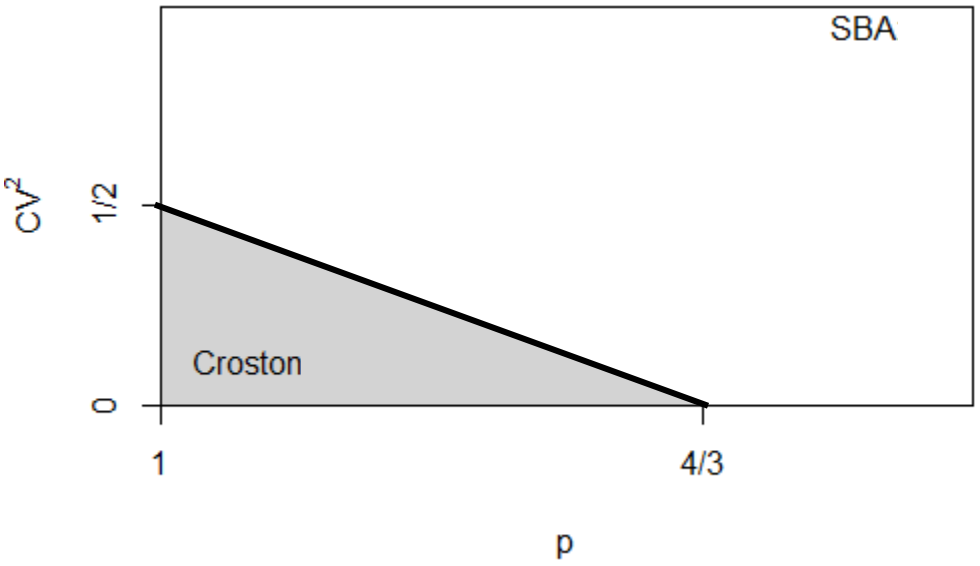
Average demand interval

Coefficient of variation of non-zero demand squared

Using these we can classify the time series into groups better modelled with Croston's method or with SBA.

Classification

Coefficient of variation of non-zero demand squared



Average demand interval

Time series with low variability of demand and relatively low intermittency should be forecasted with Croston's method. The rest should be forecasted with SBA.

Section 6

~~1. Overview of R Studio~~

~~2. Introduction to R~~

~~3. Time series exploration~~

~~Time series components, decomposition, ACF/PACF functions, …~~

~~4. Forecasting for fast demand~~

~~Naïve, Exponential Smoothing, ARIMA, MAPA, Theta, evaluation, …~~

~~5. Forecasting for intermittent demand~~

~~Croston's method, SBA, TSB, temporal aggregation, classification, …~~

6. Forecasting with causal methods

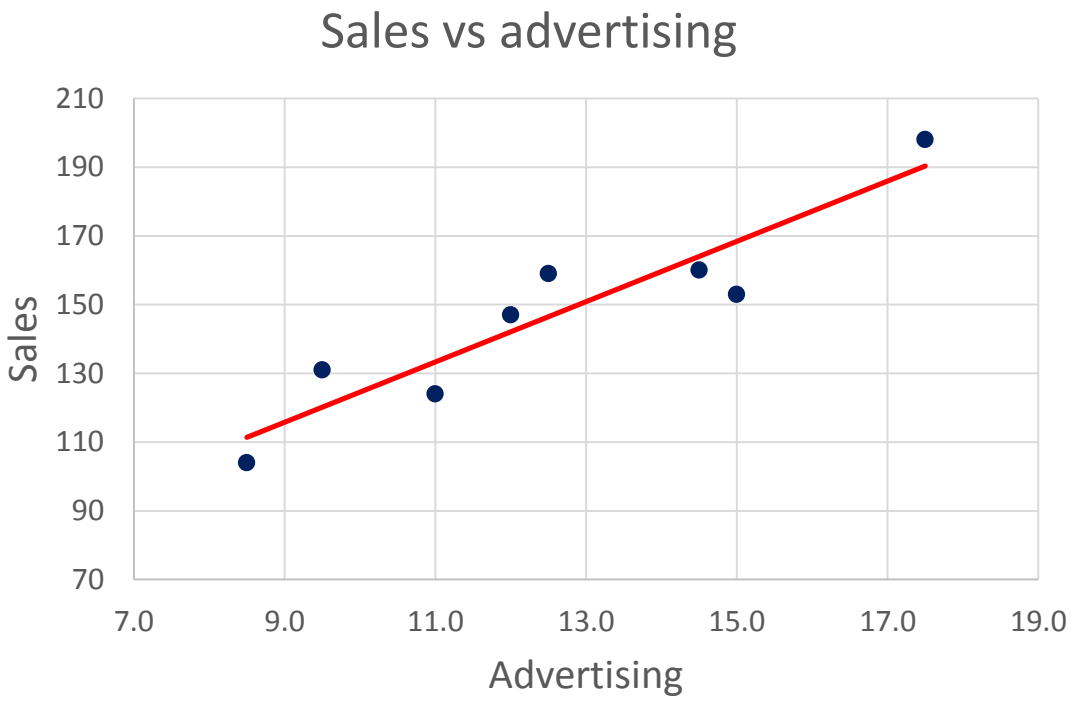
Simple and multiple regression, residual diagnostics, selecting variables, …

7. Advanced methods in forecasting

Hierarchical forecasting, ABC-XYZ analysis, LASSO

Simple regression

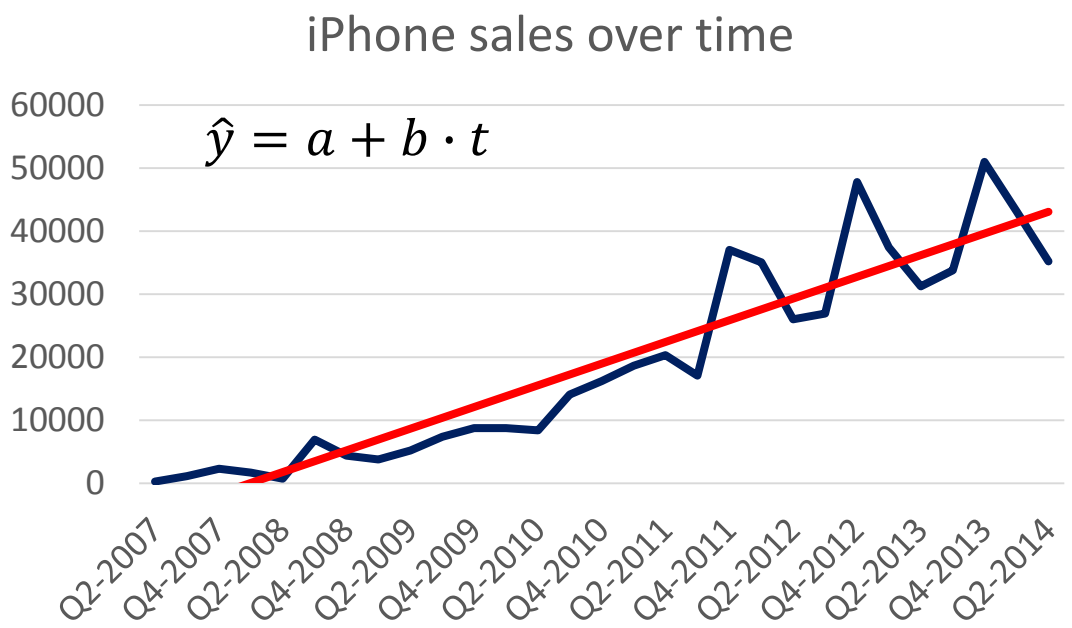
Period	Advertising (x)	Sales (y)
1	15.0	153
2	17.5	198
3	12.0	147
4	8.5	104
5	9.5	131
6	12.5	159
7	14.5	160
8	11.0	124



$$\hat{y} = a + b \cdot x$$

Linear regression on trend

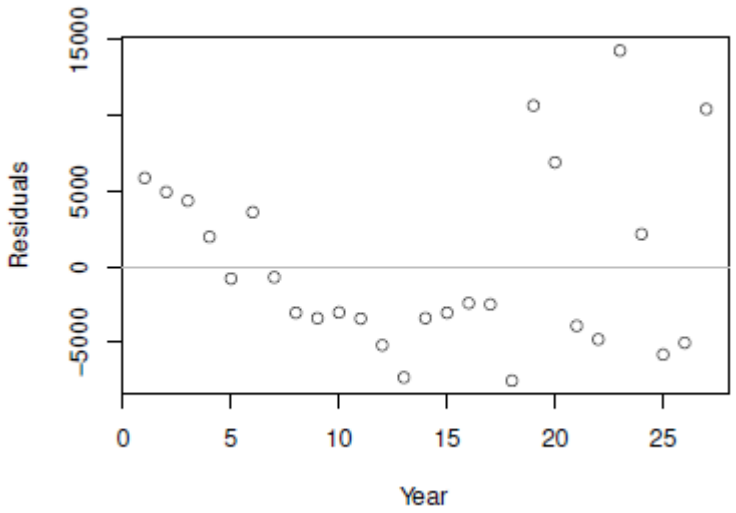
Time (t)	Period	Sales (y)
1	Q2-2007	270
2	Q3-2007	1119
3	Q4-2007	2315
4	Q1-2008	1703
5	Q2-2008	717
6	Q3-2008	6892
7	Q4-2008	4363
8	Q1-2009	3793
9	Q2-2009	5208
10	Q3-2009	7367
11	Q4-2009	8737
12	Q1-2010	8752
…	…	…

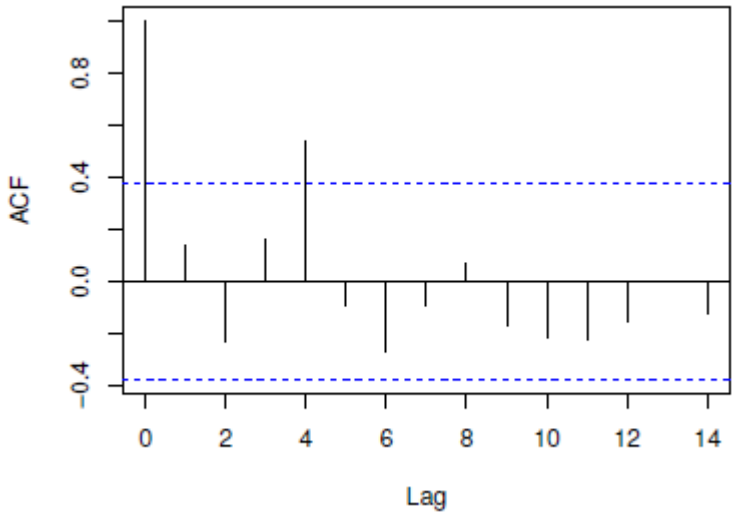


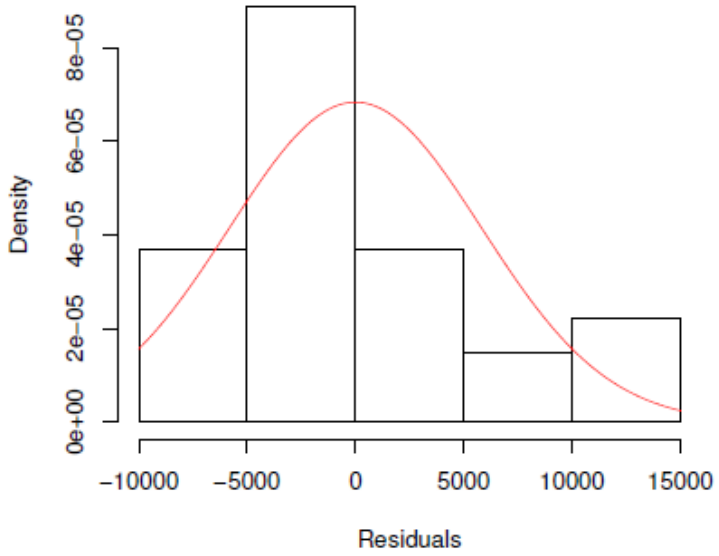
The residuals should:

- Have mean zero
- Not be autocorrelated
- Are unrelated to the predictor variable
- Be normally distributed
- Have constant variance

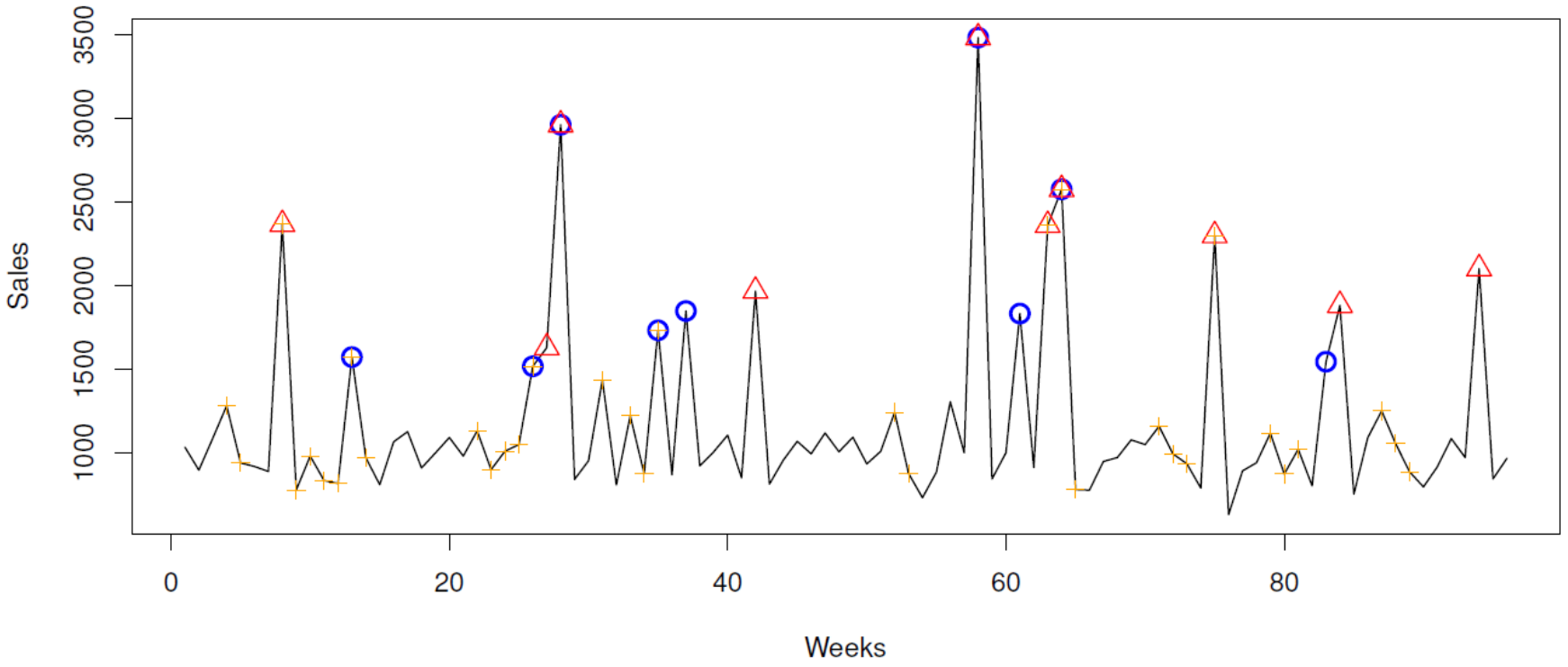
Residual diagnostics







Multiple regression



$$\hat{y} = b_0 + \sum_{i=1}^{3} b_i \, Promo_i + \sum_{j=1}^{3} b_{j+3} \, Promo\_lagged_j$$

Section 7

~~1. Overview of R Studio~~

~~2. Introduction to R~~

~~3. Time series exploration~~

~~Time series components, decomposition, ACF/PACF functions, …~~

~~4. Forecasting for fast demand~~

~~Naïve, Exponential Smoothing, ARIMA, MAPA, Theta, evaluation, …~~

~~5. Forecasting for intermittent demand~~

~~Croston's method, SBA, TSB, temporal aggregation, classification, …~~

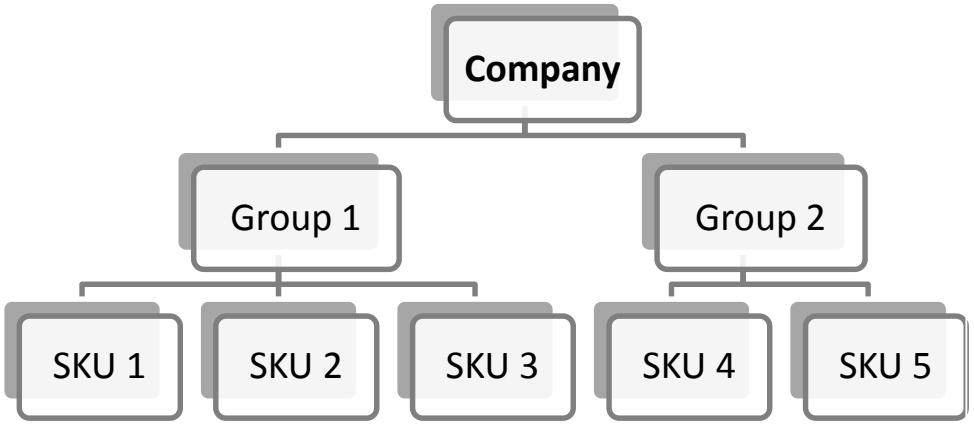
~~6. Forecasting with causal methods~~

~~Simple and multiple regression, residual diagnostics, selecting variables, …~~

7. Advanced methods in forecasting

Hierarchical forecasting, ABC-XYZ analysis, LASSO

Hierarchical forecasting



Hierarchies may refer to:

- Product types
- Geographical allocation
- Channels
- …

Problem: forecasts are different at each aggregation level!

Main approaches for reconciling hierarchical forecasts:

- **Top-down approach:** Forecast at the highest level and disaggregate using historical proportions
- **Bottom-up approach:** Forecast at the lowest level and aggregate the forecasts up to the required level
- **Middle-out approach**
- **Optimal approach:** optimally combines forecasts from each level

Shrinkage estimators

Let us consider the two regression models from before:

$$y_t = b_0 + b_1 X_{1,t} + b_2 X_{2,t}$$
$$y_t = c_0 + c_1 X_{1,t} + c_2 X_{2,t} + c_3 X_{3,t}$$

Two ideas:

- Instead of thinking X_3 being simply in or out of the model we can perceive it as a continuum, depending on the estimated coefficient c_3.
- Suppose we would keep the normalised coefficients small (close to zero) then the effect from variables would be minimal, i.e. our predicted variable would be less sensitive to changes in the explanatory variables.

  → If we are unsure about including a variable we could be more "conservative" and include it with a smaller coefficient.

Putting these together we get the so called **shrinkage estimators**.

Shrinkage estimators: LASSO

Although there are several one of the most popular ones is the:
Least Absolute Shrinkage and Selection Operator (LASSO)

The model is your conventional regression, the only difference is in how you estimate the coefficients.

Using *p* independent variables X, we model dependent variable *y* that has *n* observations:

$$y = \sum_{j=1}^{p} b_j X_j$$

But instead of OLS we use the lasso shrinkage estimator:

$$\min_{b \in \Re^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2 + \lambda \sum_{j=1}^{p} |b_j| \right\}$$

Mean squared error Shrinkage of b

Shrinkage estimators: LASSO

$$y = \sum_{j=1}^{p} b_j X_j$$

$$\min_{b \in \Re^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2 + \lambda \sum_{j=1}^{p} |b_j| \right\}$$

Mean squared error Shrinkage of b

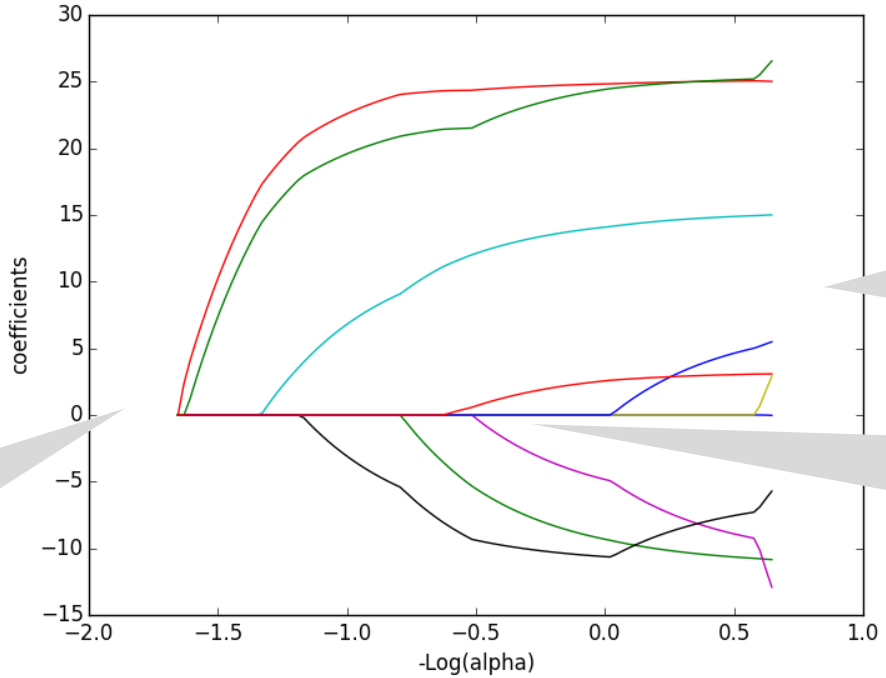
- As a variable is used more to fit better to the data, its coefficient will become bigger.
- As the coefficient becomes bigger the shrinkage penalty becomes bigger, pushing the coefficient to zero.
- Therefore lasso regression tries to keep variable coefficients small → it balances over and underfit.

Shrinkage estimators: the effect of λ

$$\min_{b \in \Re^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2 + \lambda \sum_{j=1}^{p} |b_j| \right\}$$

The parameter λ controls the amount of shrinkage:

- If λ = 0, lasso becomes OLS.
- There is a λ that all variables will be excluded from the model.



Low λ, coefficients are non-zero and large

Mid λ, only important coefficients are non-zero

Very high λ all variable coefficients are zero

How to find λ?

Finding the λ parameter is not a trivial problem. The most common approach is to use cross-validation and pick the λ that provides good cross-validated error.

What is cross-validation?

1. Take all the available in-sample data and split it into K parts (folds).
2. Fit the model in all 9 parts and test in the remaining one

3. Repeat until all K folds have been used as test…

4. Measure the total error across all “tests”. This is the cross-validated error.

The cross-validated error approximates the true prediction error and is more reliable than the in-sample fitting error.

Nikolaos Kourentzes
email: nikolaos@kourentzes.com
blog: http://nikolaos.kourentzes.com

Fotios Petropoulos
email: fotpetr@gmail.com
site: http://fpetropoulos.eu

Forecasting Society
www.forsoc.net

Lancaster Centre for Forecasting
www.forecasting-centre.com/

Lancaster University Management School | Lancaster Centre for Forecasting | Cardiff University Business School | Forecasting Society