Problem 5:
Forecasting the demand for bread

neometrics

V UCM Modelling Week

Master in Mathematical Engineering UCM

0. Summary

neometrics

Problem 5:

Forecasting the demand for bread

1. Introduction

➢ The problem proposed by Neometrics consists on **forecasting the daily demand for bread**.

➢ To resolve this problem we have the data for several points of sales, in every one we knows the **date**, the **shipped values** and **returned values**. We have defined "**sold**" variable:

Sold = Shipped – return

Also, we have **advertising pressure**.

➢ The particularity of these series analysis is that:

- It exists "**censured data**" or "**out of stock**": in every point of sale, there are some dates which his shipped value is greater than zero and his returned value is zero.

out of stock cost

- It exists values which the sales were zero, i. e. stores were closed (holidays or unexpected close).

return cost

neometrics

Problem 5:

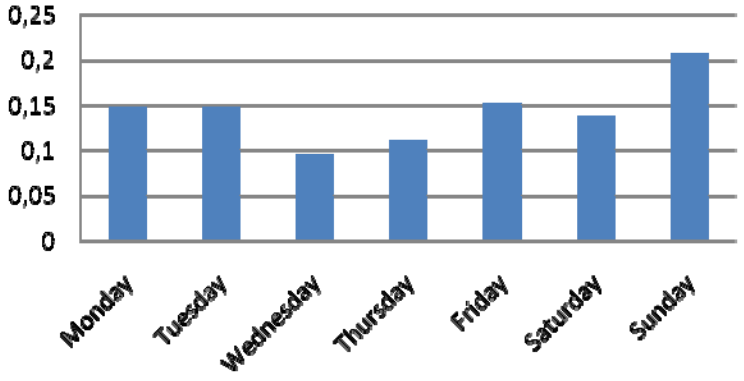
Forecasting the demand for bread

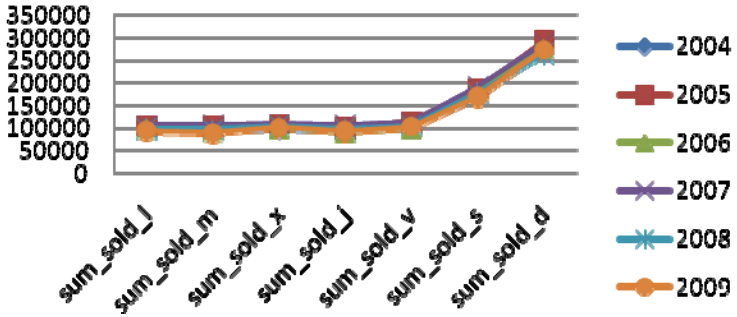
2. Descriptive analysis

- **70 series** one for each store.

- The first 65 series start 2-1-2004 and end 31-12-2009.

- The 66th series starts 17-5-2004 and ends 31-12-2009.
- The 67th series starts 7-3-2005 and ends 31-12-2009.
- The 68th series starts 8-5-2005 and ends 31-12-2009.
- The 69th series starts 1-4-2006 and ends 31-12-2009.
- The 70th series starts 8-5-2005 and ends 31-12-2009.





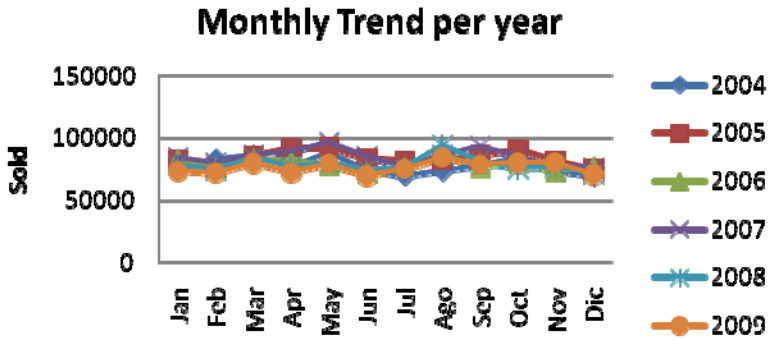
neometrics

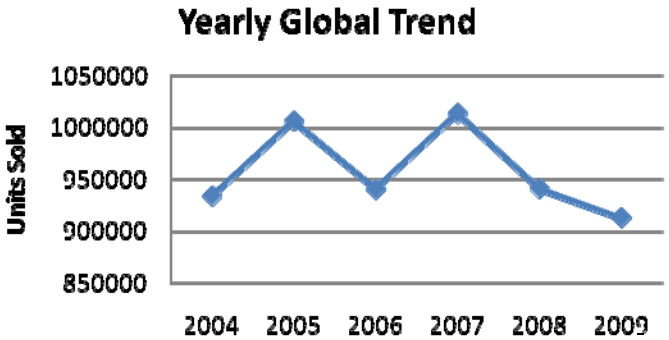
Problem 5:

Forecasting the demand for bread

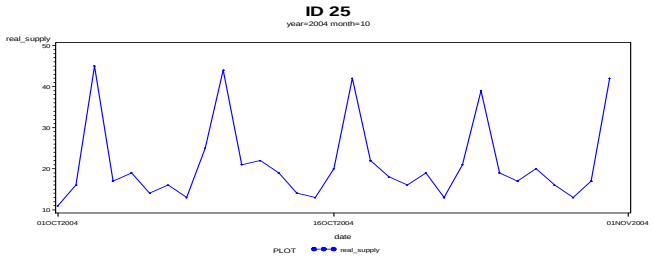
2. Descriptive analysis

- **Trend**





- **Seasonality**: Weekly



- **Cycle**: NO

neometrics

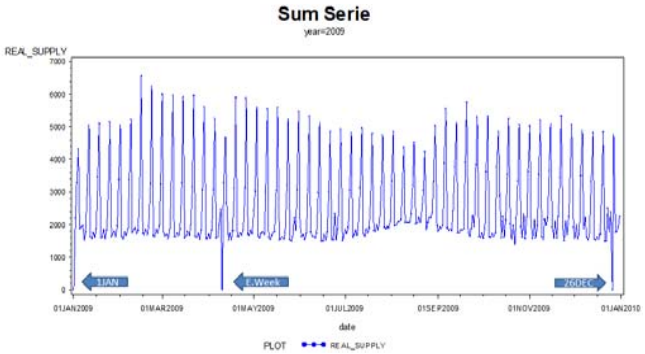
Problem 5:

Forecasting the demand for bread

2. Descriptive analysis

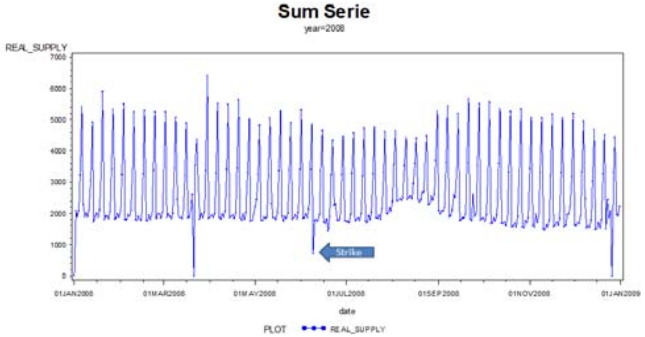
Other interesting points of these series are:

- Normally, all the series close the same days: Easter Week, New Year, Christmas,…



year = 2009

- The transport strike (09-06-2008)



year = 2008

neometrics

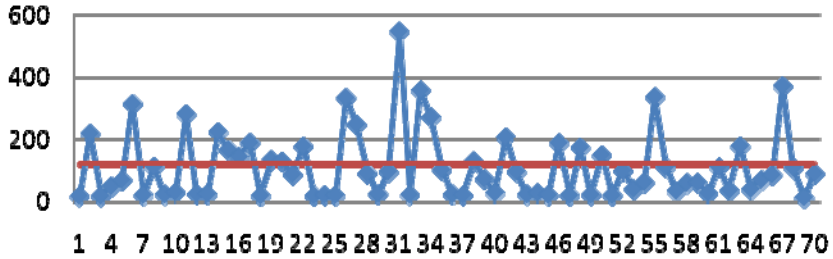
Problem 5:

Forecasting the demand for bread

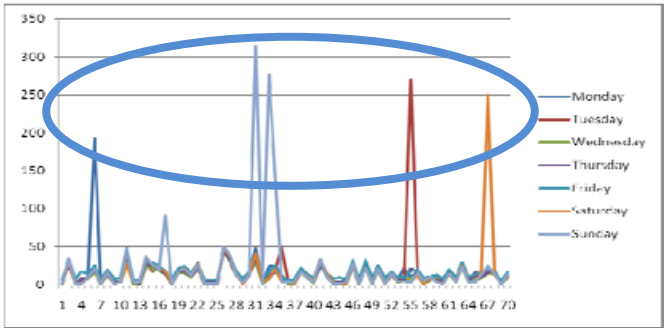
3. Classification

We have decided to make the following classification of the series:

• Points of sales **closed on holidays** (August) **or not**.



• Points of sales **closed the same day of the week or not**.



neometrics

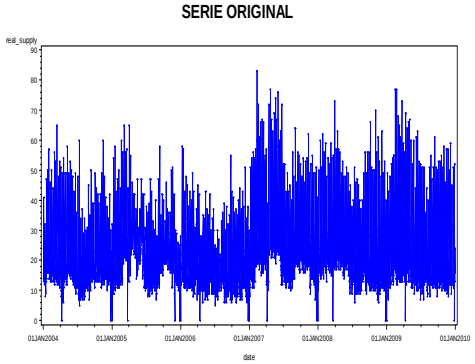
Problem 5:

Forecasting the demand for bread

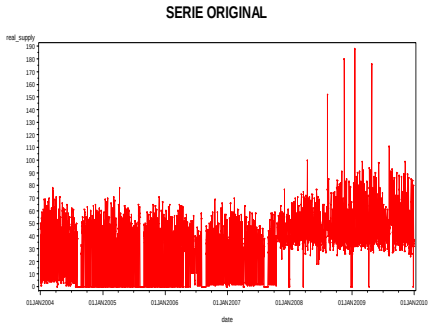
3. Classification

We have chosen four representative series

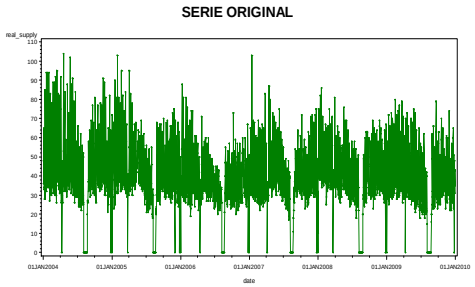
- Series 25: Only have 20 days of holidays.



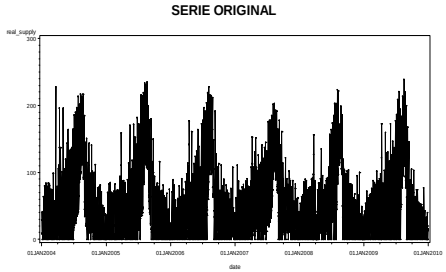
- Series 34: In the first four years they have holidays but then they stop to have holidays and the sells increase



- Series 38: Don't have too many days off.



- Series 55: With a lot of holidays



neometrics

Problem 5:

Forecasting the demand for bread

4. Data imputation for missing and censured values

There are two kinds of data that we have to impute:

• **Censured values**: the total shipped bread was sold. Then the returned was zero. In this case, we don't know the real sold that the point of sales would have had.

• **Closed store values**: the points of sales were closed then the shipped and returned values were identical. The reasons can be:

-Period of holidays

-Only one day off.

4. Data imputation for missing and censured values

Methods of data imputation:

- **Censured values:**
  1. Average of previous data.
  2. Average of previous data setting a time horizon.
  3. **Average of *k* previous data without setting a time horizon.**

The third method is quite appropriate. If sales do not verify the conditions then we have to use an **alternative method**. We propose the following:

Average increase of sales that have already been estimated.

- **Closed store:**
  1. Average of previous data
  2. Average of previous data setting a time horizon.
  3. **Average of *k* previous data without setting a time horizon.**

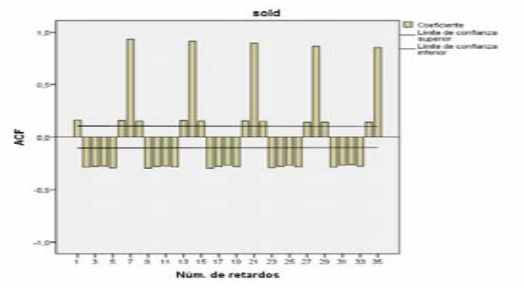
We have decided to apply the third method. If sales do not verify the conditions then we have to use an **alternative method**. We propose the following:

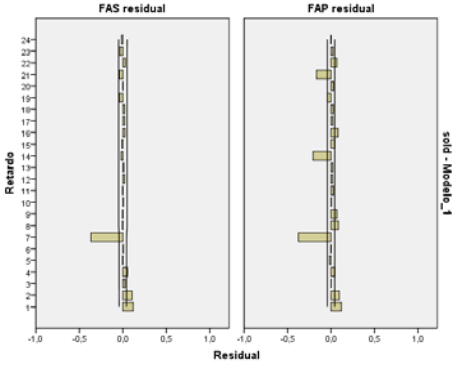
The next sale in the same day of the week.

5. Preliminary model for the expected sales

We will explain the model that we have adjusted for the SERIES 25, It is the best model for the four series that we have chosen.

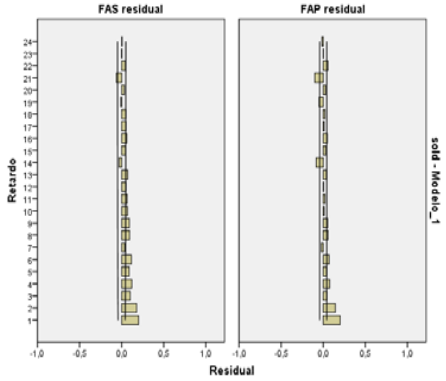
We take a difference with lag 7





also display a "spike" in the delay 7. And the PACF is becoming smaller
i sonal MA.

ARIMA(0,0,0)x(0,1,1). We think about adjust an AR(1).



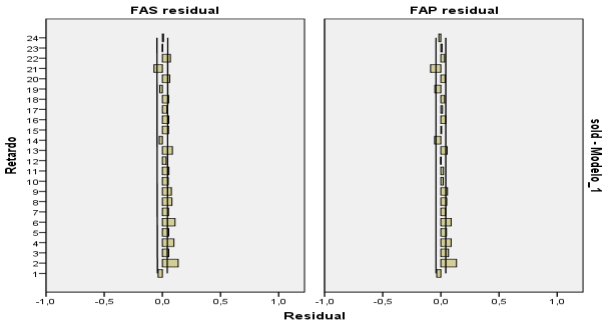
Problem 5:

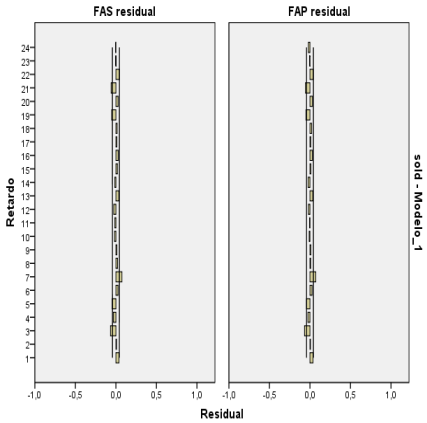
Forecasting the demand for bread

neometrics

5. Preliminary model for the expected sales

ARIMA (1,0,0)x(0,1,1). We think about adjust a regular MA.





ARIMA(1,0,1)(0,1,1)

Final model: **ARIMA(1,0,1)x(0,1,1)**

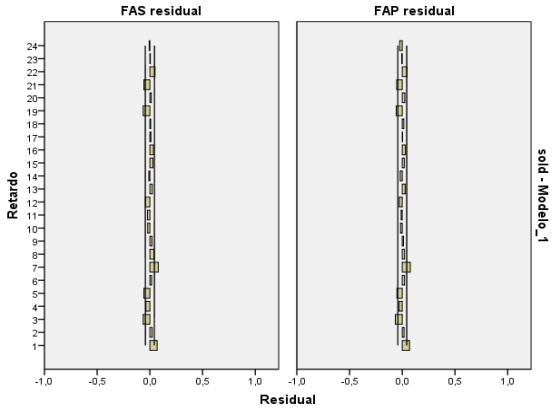
neometrics

Problem 5:

Forecasting the demand for bread

5. Preliminary model for the expected sales

Now we adjust an exponential smoothing: Simple exponential smoothing.



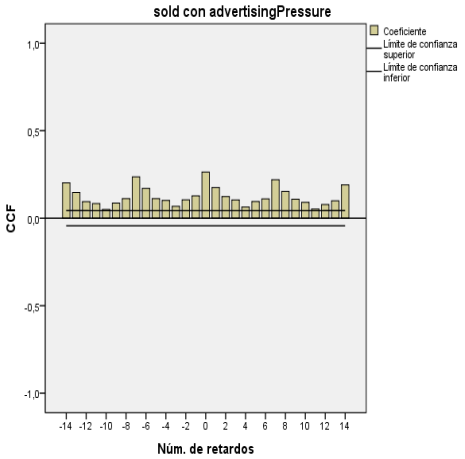
Comparing the two models:

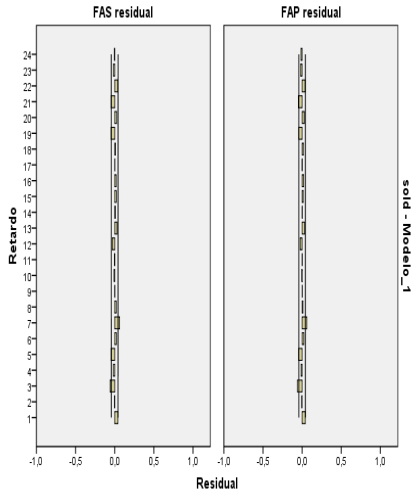
- ARIMA(1,0,1)x(0,1,1) → R-square: 89% BIC: 2,970
- Exponential smoothing → R-square: 89% BIC: 2,963

The final model is ARIMA(1,0,1)x(0,1,1): $(1-\Phi 1B)(1-B^{*7})X(t)=(1-\theta 1B)(1-\Theta 7B^{*7})\varepsilon(t)$

6. Intervention analysis

WITH THE ADVERSTISING PRESSURE





The final model for all series is:

$$(1-\Phi 1B)\,(1-B^{*7})\,X(t) = (1-\theta 1B)\,(1-\Theta 7B^{*7})\,\varepsilon(t) + 0{,}022\,(1-B^{*7})\,Ad(t)$$

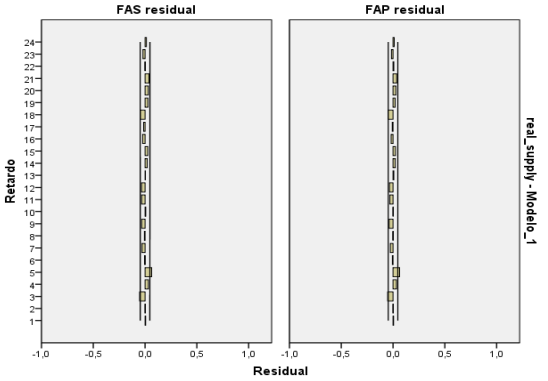
neometrics

Problem 5:

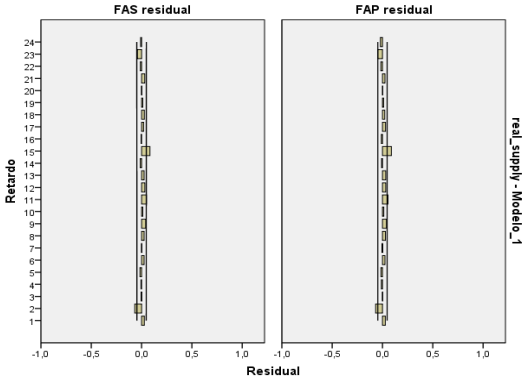
Forecasting the demand for bread

6. Intervention analysis

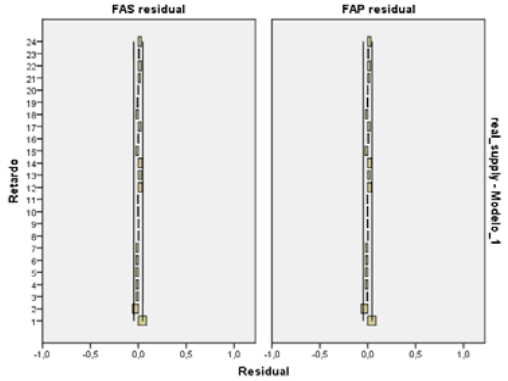
- **SERIE 34:**



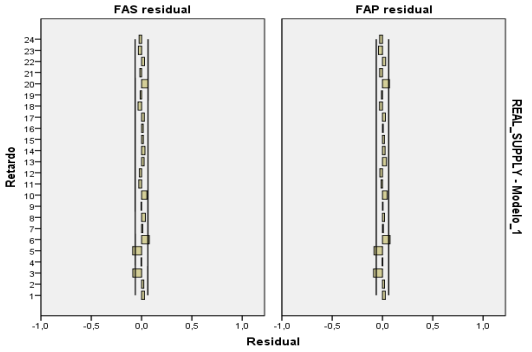
- **SERIE 55:**



- **SERIE 38:**



- **SERIE GLOBAL:**



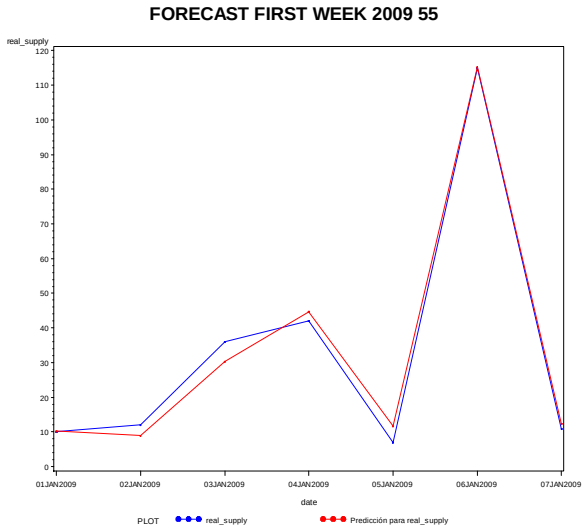
neometrics

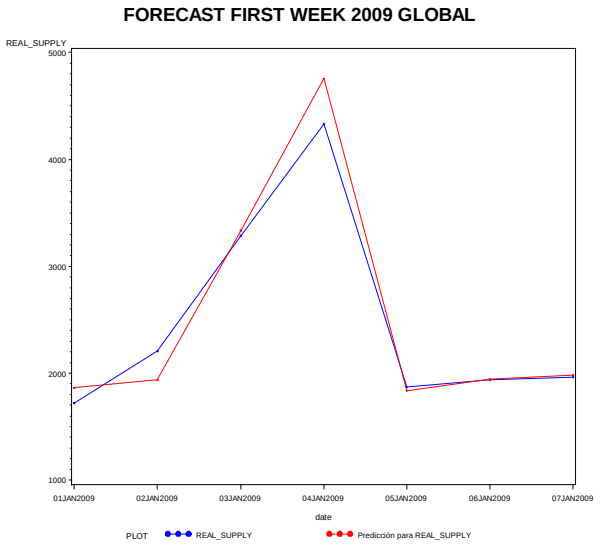
Problem 5:

Forecasting the demand for bread

7. Model validation

The model has been created to forecast one week.





neometrics

Problem 5:

Forecasting the demand for bread

8. Determination of the daily demand

- Obviously, **we'll never send more units of bread** of the historical **maximum** of the point of sale.

- Once we know the expected sales, it will be analyze the **benefits of sending** one unit more, two units more, tree units more and so on.

- For doing that from historical data, we estimate the probability of sell expected sales plus one unit more conditioned to expected sales have been sold. After that, we estimate the probability of sell expected sales plus two units more conditioned to expected sales plus one unit more have been sold and so on. Finally, we estimate the probability of sell maximum sales conditioned to one unit less than the maximum has been sold.

- The **benefits of sending k units of bread more than expected sales** will be calculated as the benefit of selling one unit of bread (0.6 €) times the probability of selling extra unit less the cost of producing the unit (0.2 €) and less the distribution cost of every unit (0.1 €) less the cost of expired product (0.12 €).

- Obviously, the benefits of sending k units of bread more than the expected sales must be greater than zero. In that case, we will determine to send k units more than expected sales.

9. Conclusions

- GLOBAL MODEL
  - With the given series, we could have made a different model for each one, however we had assumed that this is only a **sample** of all data amount.
  - In practice what the company requires is a **global model** that works for all of them. And we have seen that it fits well on all series chosen.
- HOLIDAYS
  - Instead of forecasting the real series we had chosen to **forecast imputed series** (holidays or censured values).
  - It means that we have forecasted even the days that we know store is closed (registers type: shipped=0 returned=0).
  - Assuming that the distribution company knows when stores are closed and it's not going to deliver nothing to it.
  - We think it's a better way to do it, because if the store finally decides to open we have estimated sales for that day.

THE END

Problem 5:
Forecasting the demand for bread

THANK YOU !!

neometrics

V UCM Modelling Week

Master in Mathematical Engineering UCM