Problem 5: Forecasting the demand for bread

2011 21st June

Fernando Baladrón Laura Barrigón Ángeles Garrido Alejandro González Verónica Hernández

0. Summary

1. Introduction

The proposed problem by Neometrics consists on forecasting the daily demand for bread. The solutions of this problem derive a huge economic impact on the income of distribution companies. Because of that it’s a very complicated problem.

For resolving this problem we have the data for several points of sales, in every one we have the date, the shipped values and returned values. We have created the variable “sold” which derives from subtract the returned values from the shipped ones. We also have the data for the advertisement pressure.

The particularity of these series analysis is that it exists “censored data” or “Out of stock”, i. e. we have some dates in every point of sale which his shipped value is greater than zero and his returned value is zero and it exists too values which the variable “sold” is zero, i. e. the days the stores were closed. The censured data means that they have sold everything, so we don’t know if they would have sold more. Therefore, we must take it into account for predicting the sold.

The prediction desired is to know how many units should be sent to every point of sale in every day for the next 7 days.

For solving this problem we have used SAS and SPSS.

2. Descriptive analysis

We have 70 series one for each store. The first 65 series start 2-1-2004 and end 31-12-2009 so there are 2191 data for every series.

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.

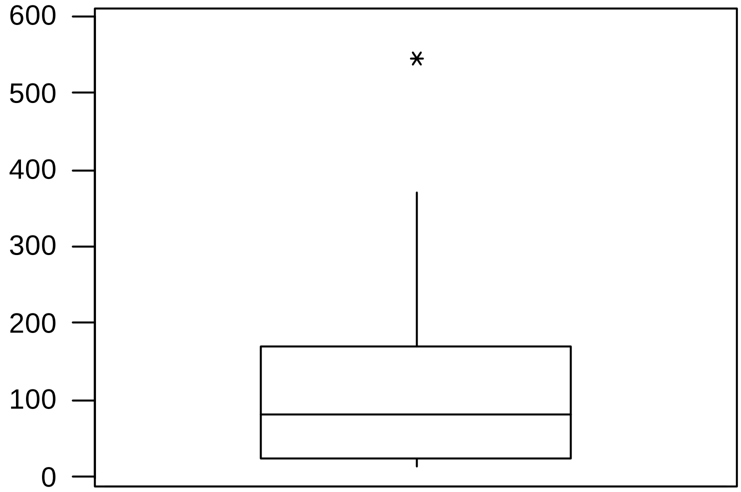
As we have said the interesting points of these set of series are the “censored data” and the “closed store values”. We will focus on the “censored data” afterwards, during the epigraph 4.

We have study several characteristics for the variable “sold”. The columns Monday to Sunday means how many zeros there are in the variable “sold” when it’s that day of the week. The column Total means how many days has the store closed. “No favorite day” means that the store doesn’t closed systematically the same day of the week and “Favorite day” just the opposite. “With holydays” means that it exists a continued period of holidays and “No holydays” the other way round.

id	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	total	No favorite day	Favorite day	No holidays	Withholidays
1	2	2	1	2	7	2	2	18	1	0	1	0
2	32	31	28	27	34	31	35	218	1	0	0	1
3	2	2	1	2	7	2	2	18	1	0	1	0
4	6	5	8	8	18	2	2	49	1	0	1	0
5	10	10	8	9	14	9	9	69	1	0	1	0
6	193	18	16	19	26	21	21	314	0	1	0	1
7	4	2	1	3	7	2	2	21	1	0	1	0
8	19	15	13	13	20	16	16	112	1	0	1	0
9	2	4	2	3	8	2	2	23	1	0	1	0
10	3	4	4	4	9	3	3	30	1	0	1	0
11	40	46	33	41	48	27	46	281	1	0	0	1
12	4	2	1	3	7	4	4	25	1	0	1	0
13	4	3	1	2	7	3	4	24	1	0	1	0
14	34	30	24	31	31	37	37	224	1	0	0	1
15	22	18	21	24	29	26	26	166	1	0	0	1
16	20	20	19	20	25	20	20	144	1	0	0	1
17	19	14	14	15	18	18	91	189	1	0	0	1
18	3	2	1	2	7	2	2	19	1	0	1	0
19	22	18	17	18	23	18	18	134	1	0	0	1
20	17	15	15	18	24	21	20	130	1	0	0	1
21	14	12	10	11	16	12	12	87	1	0	1	0

22	29	22	21	23	28	28	27	178	1	0	0	1
23	3	2	1	2	7	2	2	19	1	0	1	0
24	3	3	2	2	7	2	2	21	1	0	1	0
25	3	2	2	2	7	2	2	20	1	0	1	0
26	50	47	44	45	49	48	50	333	1	0	0	1
27	35	32	31	33	39	38	40	248	1	0	0	1
28	12	13	12	13	17	12	12	91	1	0	1	0
29	3	2	3	3	8	2	3	24	1	0	1	0
30	15	13	13	12	17	13	13	96	1	0	1	0
31	48	36	31	36	41	41	313	546	0	1	0	1
32	2	4	1	3	7	3	2	22	1	0	1	0
33	24	12	8	11	17	9	277	358	0	1	0	1
34	26	19	17	16	26	20	147	271	0	1	0	1
35	13	50	11	11	9	4	4	102	1	0	1	0
36	2	3	1	4	7	3	3	23	1	0	1	0
37	3	2	1	2	7	2	4	21	1	0	1	0
38	18	21	16	18	23	17	17	130	1	0	0	1
39	9	10	9	10	14	11	10	73	1	0	1	0
40	4	3	2	3	8	5	3	28	1	0	1	0
41	34	25	26	26	30	33	33	207	1	0	0	1
42	15	14	13	15	15	14	10	96	1	0	1	0
43	2	3	3	2	8	3	2	23	1	0	1	0
44	3	3	3	5	10	2	2	28	1	0	1	0
45	3	2	1	3	7	2	3	21	1	0	1	0
46	24	25	24	27	32	28	28	188	1	0	0	1
47	4	2	1	2	7	2	2	20	1	0	1	0
48	25	24	23	24	32	23	23	174	1	0	0	1
49	2	2	1	4	9	2	2	22	1	0	1	0
50	23	20	19	20	26	21	21	150	1	0	0	1
51	3	2	1	2	8	2	2	20	1	0	1	0
52	16	13	12	12	17	14	15	99	1	0	1	0
53	6	5	4	5	10	5	5	40	1	0	1	0
54	23	8	5	6	10	5	5	62	1	0	1	0
55	10	270	9	22	18	4	3	336	0	1	0	1
56	18	16	12	16	20	13	16	111	1	0	1	0
57	7	5	1	5	9	2	5	34	1	0	1	0
58	9	7	6	7	11	9	11	60	1	0	1	0
59	14	10	10	7	13	4	3	61	1	0	1	0
60	3	4	2	6	8	2	2	27	1	0	1	0
61	16	16	14	15	20	15	15	111	1	0	1	0
62	6	5	3	5	9	4	4	36	1	0	1	0
63	25	24	23	26	30	26	24	178	1	0	0	1
64	7	7	3	6	10	3	4	40	1	0	1	0
65	11	13	7	17	14	4	4	70	1	0	1	0
66	14	10	9	10	16	14	13	86	1	0	1	0
67	23	20	15	16	22	250	25	371	0	1	0	1
68	17	15	14	18	17	13	15	109	1	0	1	0
69	2	2	2	1	5	1	0	13	1	0	1	0
70	14	11	9	13	16	13	14	90	1	0	1	0

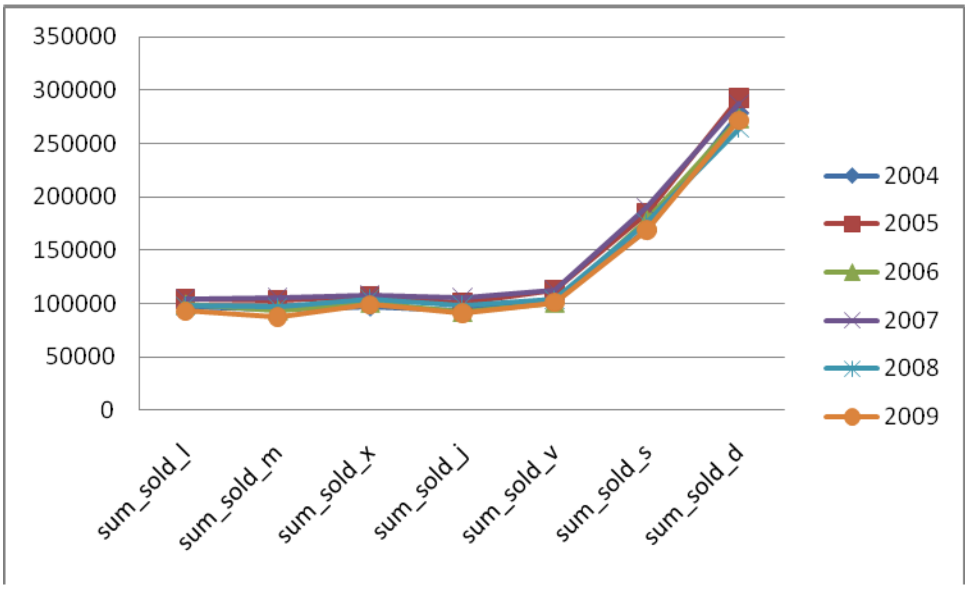
With a boxplot chart we study the variable “Total”. It represents the median, first quartile and the third quartile. The whisker represents the values except the outliers which is the asterisk. That asterisk belongs to the series 31. We can appreciate that there is only a few of them the ones which have more than 180 days for holidays.



In the next table we can see some mains statics for the variable total.

N	Mean	Median	TrMean	StDev	SE Mean	Minimum	Maximum	Q1	Q3
70	111,2	79,5	98,4	110	13,1	13	546	23,8	168

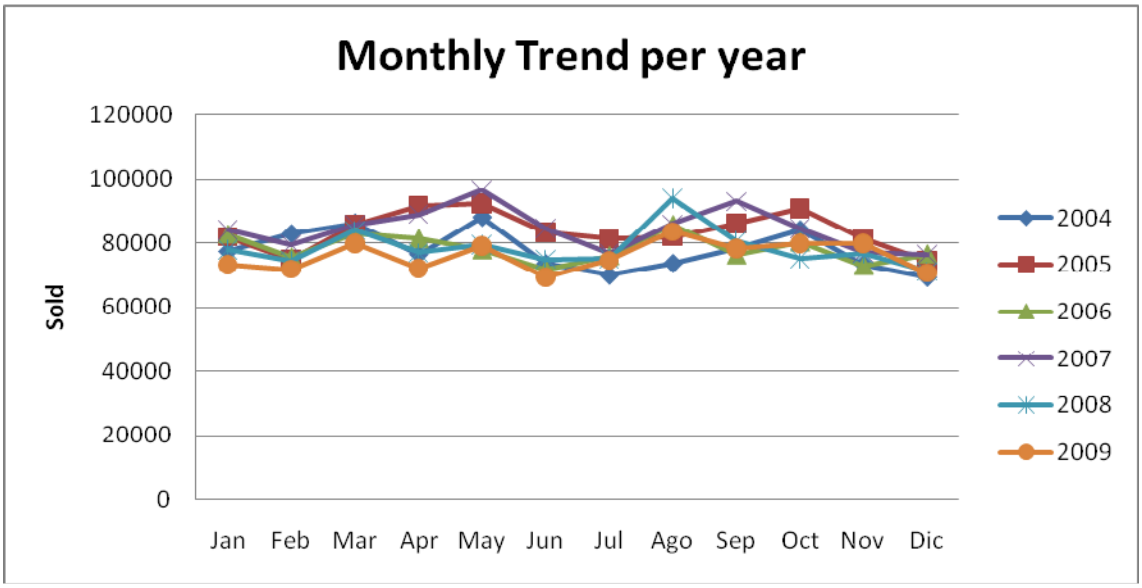
In these two graphs we can see the favorite day of the week chosen for close by the stores. On the next one we show the days in which the stores sell more.



Then we will see the analysis of nonstationary structure:

- **Trend of the series:**
  - **Monthly**

In the next graphic we can appreciate there is not trend:

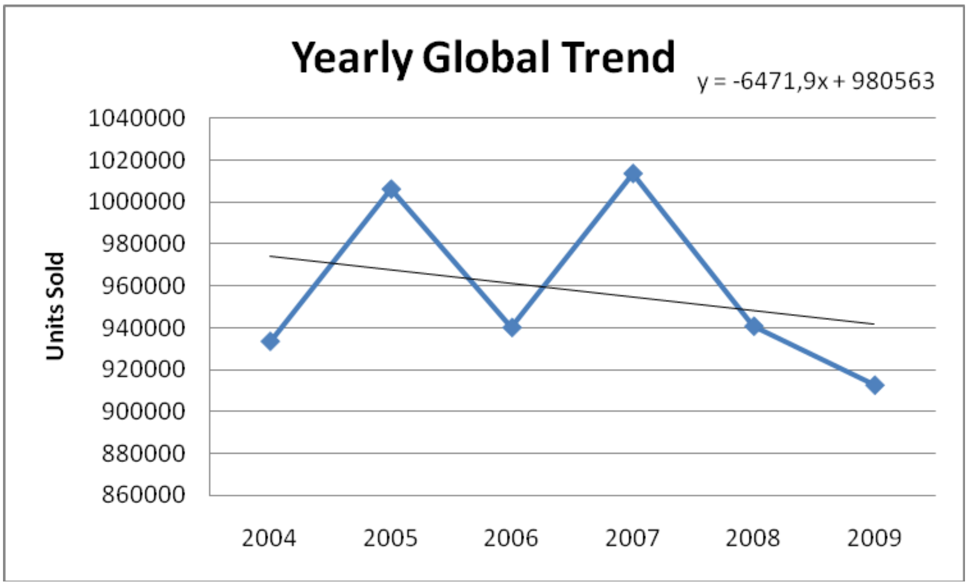


Just to take a look, the 2006's data, for example, is:

year	date	sold
2006	Jan	82761
2006	Feb	75582
2006	Mar	82914
2006	Apr	81437
2006	May	77911
2006	Jun	71878
2006	Jul	76028
2006	Ago	85772
2006	Sep	76287
2006	Oct	80296
2006	Nov	72905
2006	Dic	76550

- **Yearly**

We can consider there is no trend because it only decreases the 6%.



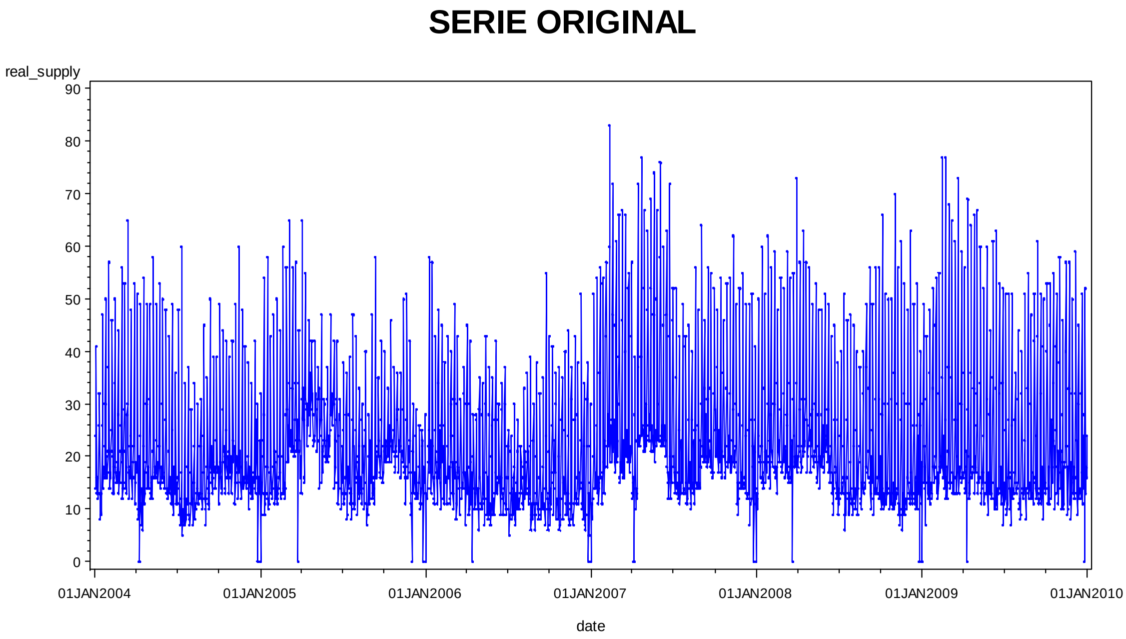
The data used in this case:

year	sold
2004	933731
2005	1006082
2006	940321
2007	1013523
2008	940936
2009	912875

We have chosen four series, following the criteria that we specify on the 3rd epigraph.

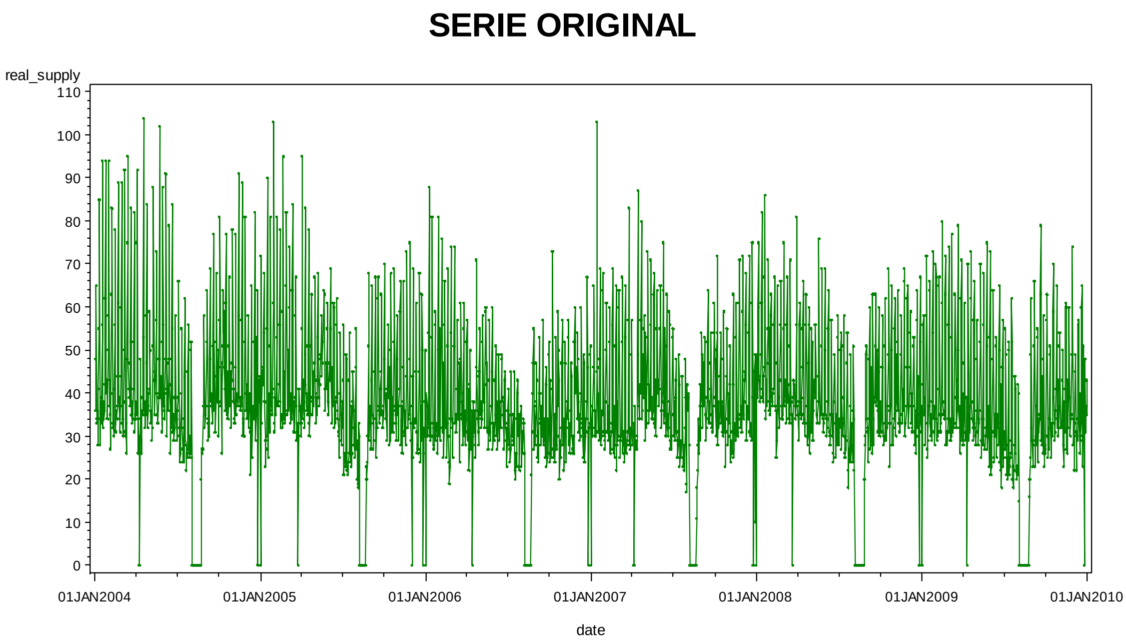
Those series are 25, 38, 34 and 55. We will see the plots of those series in which we can appreciate that there is not significant trend:

SERIE 25



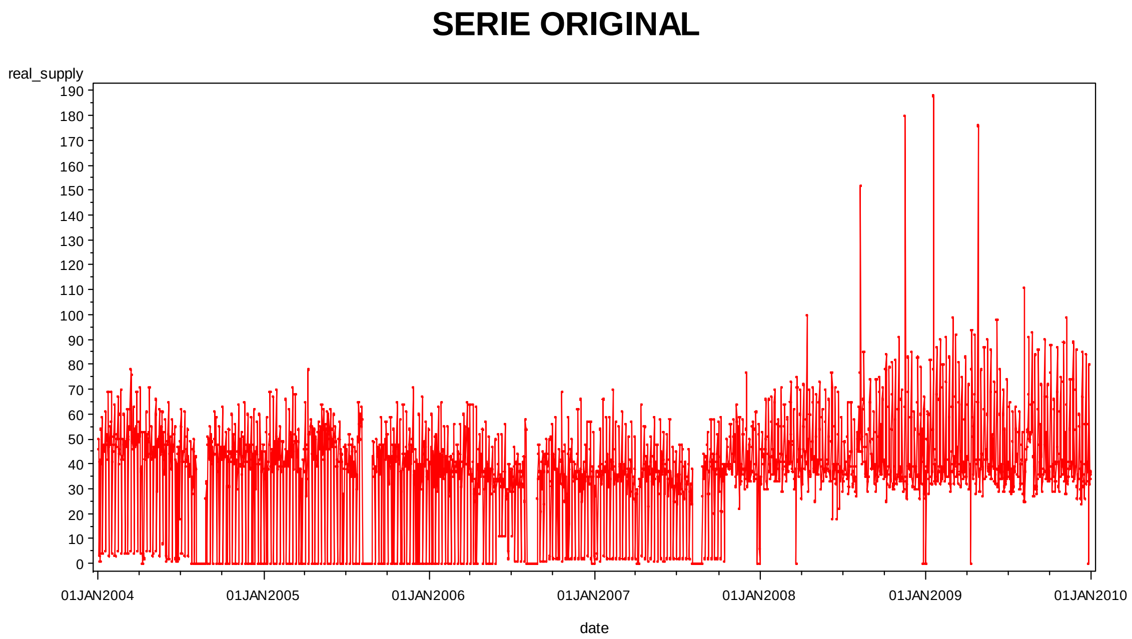
We can see a softly increase of the variability. And we can see they only have 20 days of holidays.

SERIE 38



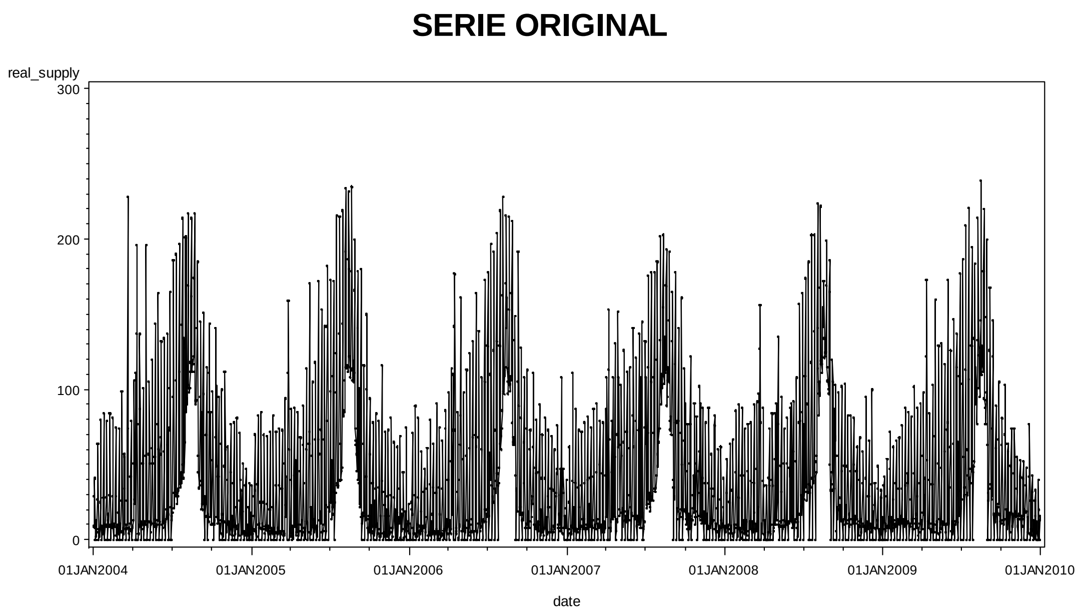
On these one we can appreciate the seasonality and that they don't have too many days off (only 130 in 6 years).

SERIE 34



In these series we can see that in the first four years they have holidays very often but then they stop to have holidays and the sells increase.

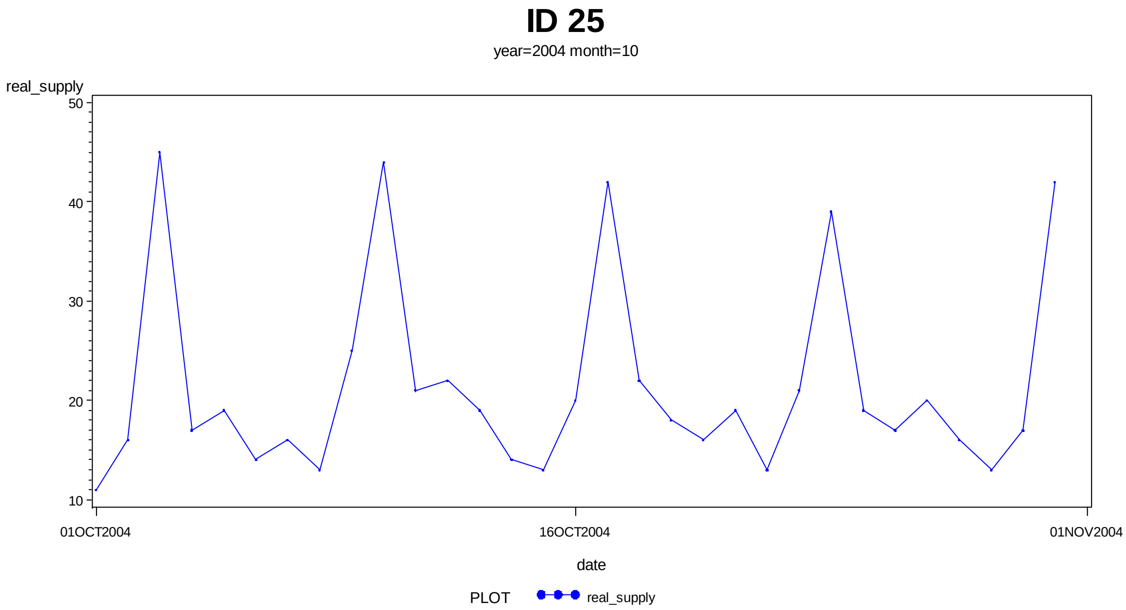
SERIES 55



These series its one with lots of holidays and in august they sells the maximums.

- **Seasonality**:

We note that the seasonality of the series is weekly, because we can see the pattern in here.

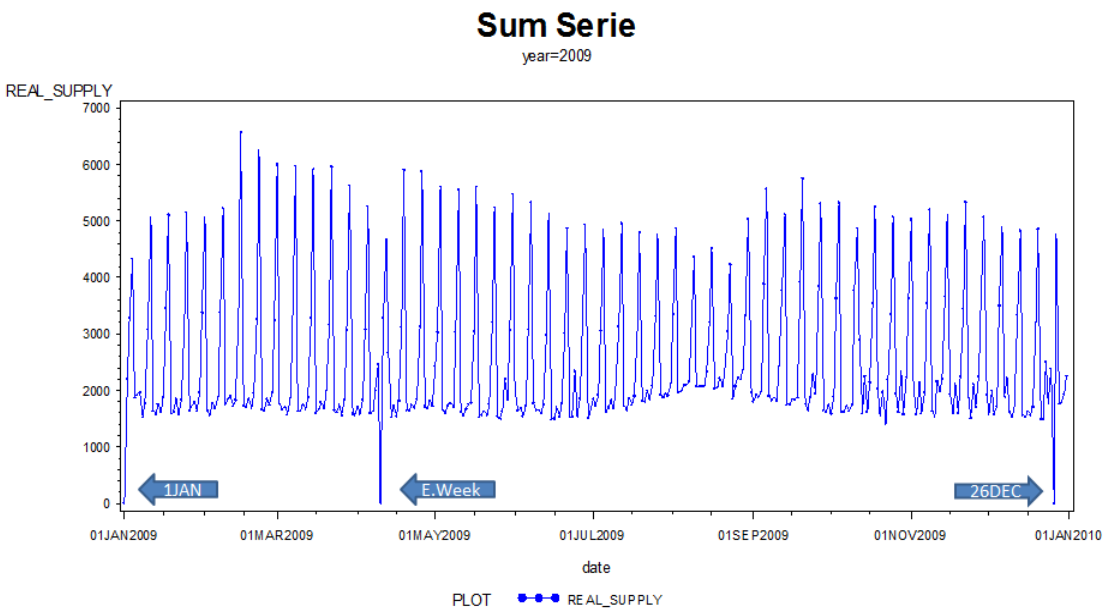


- **Cycle:**

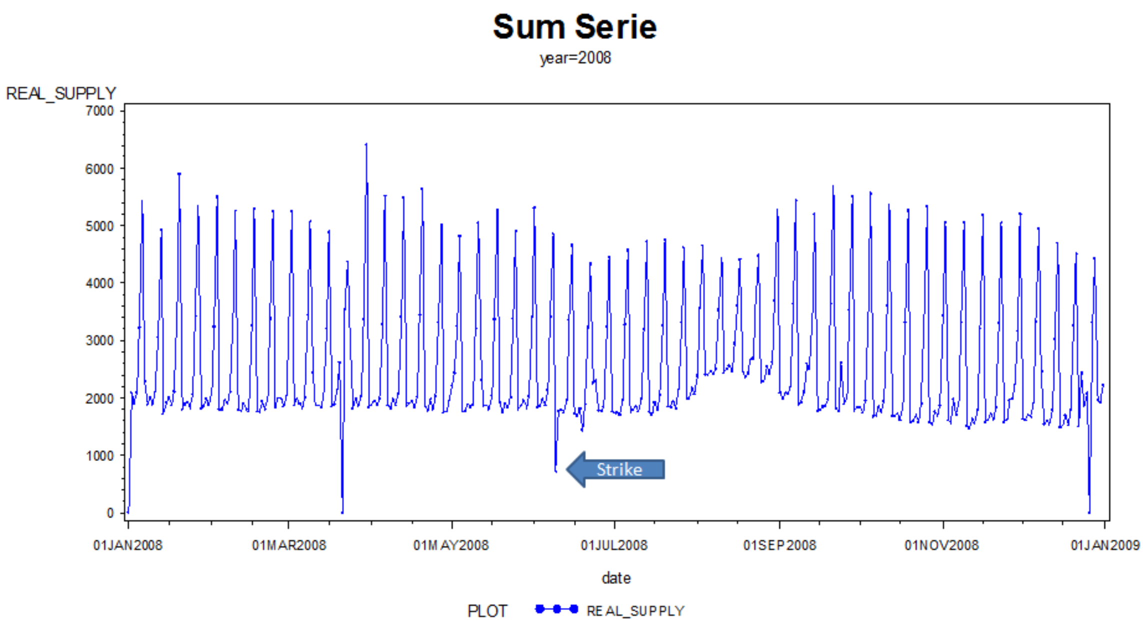
The series have no cycles.

Other interesting points of these series are:

- Normally closed series all the same holidays (all of them closed January the 1st, one day of the Easter holidays and December the 26th)
  We have summed the 70 series for study all of them together:



- The transport strike (09-06-2008)



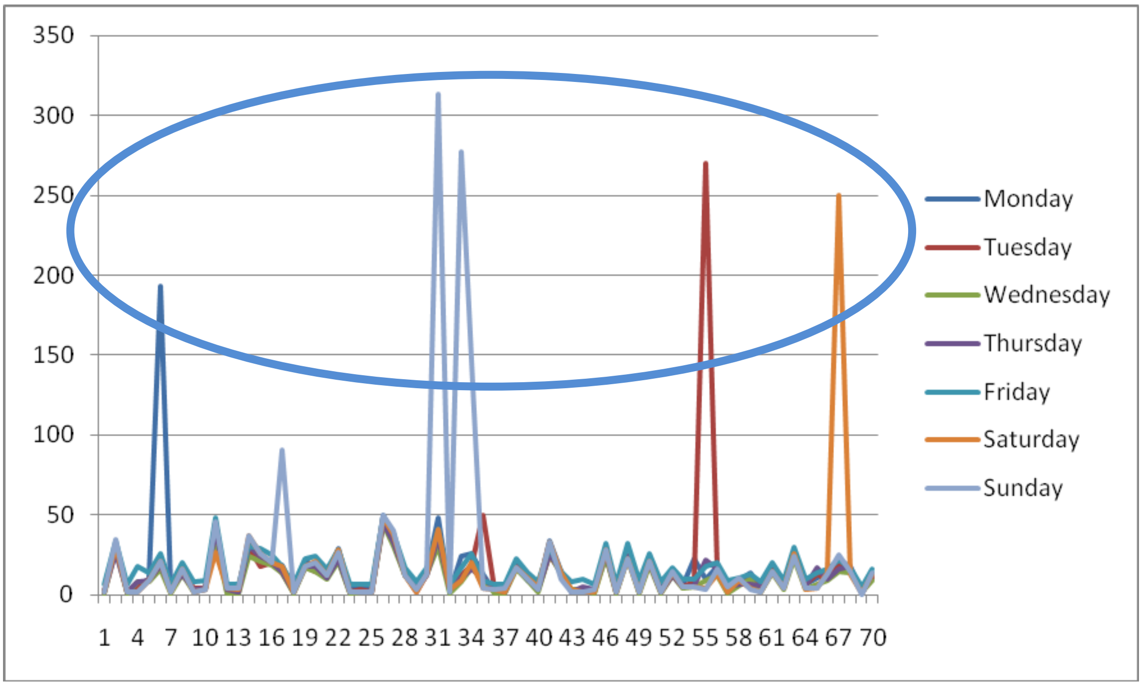
3. Classification

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

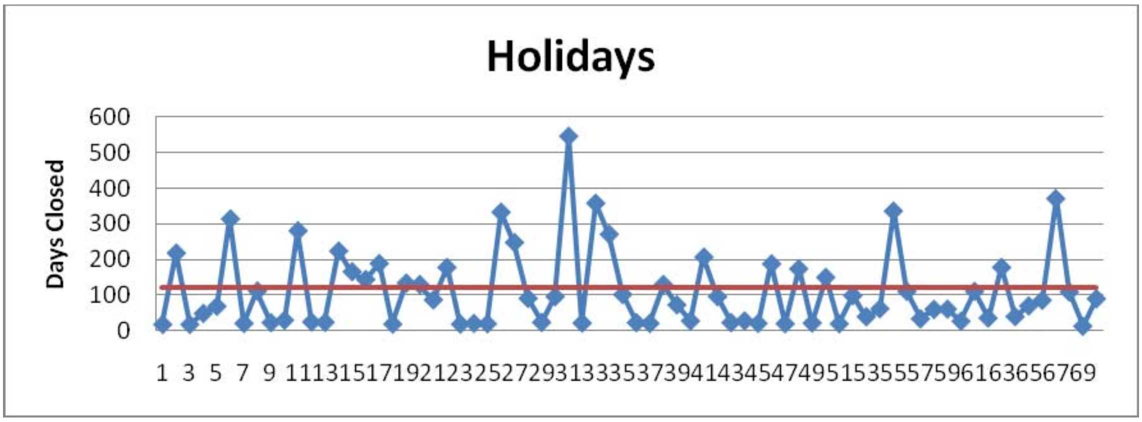
- Points of sales closed on holidays (mainly August) or not.
- Points of sales closed systematically the same day of the week or not.

Graphically it's easy to see this classification between id's:

Points of sales closed the same day of the week



Points of sales closed on holidays



4. Data imputation for “closed store”and censured values

There are two kinds of data that we have to impute. These data correspond with the following situations:

- 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: 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.

We think that the data imputation is very important to avoid the model learns from wrong data. Our goal is to predict sales regardless of whether the point of sales was open or closed.

Methods of data imputation

Censured values

The store has sold *M* packs of bread, the day *D* and the day of the week is W. The returned has been 0.

If the real sales are unknown because the store sold all the bread shipped then we can impute these data with the following methods:

1. Average of previous data.

We consider the average of the sales greater or equal than M, previously sold to date D and which the day of week is exactly W.

2. Average of previous data setting a time horizon.

We consider the average of the sales greater or equal than M, previously sold to date D and within a month and which the day of week is exactly W.

3. Average of *k* previous data without setting a time horizon.

We consider the average of the *k* sales greater or equal than M, previously sold to date D and which the day of week is exactly W.

The first method is not appropriate in our problem because we have a history of 6 years. Therefore, the situation could be very different in 2010 than in 2004.

The second method is appropriate, however it runs the risk of not finding any data for the period specified (limit time horizon: monthly, bimonthly).

The third method is quite appropriate. If the series have significant trend, applying this method trend is lost. As shown above, the series haven't trend so significant. For that reason, we have decided to apply the third method setting k=3. That is, we do the average of the tree sales more recently whose verify the conditions.

If sales do not verify the previous conditions, then we have to use an alternative. We propose the following:

- Average increase of sales that have already been estimated.

We calculated the average increase in sales whit the sales that have already been estimated.

Closed store

The store has received *M* packs of bread, the day *D* and the day of the week is W and it has returned M packs of bread.

If the real sales are unknown because the store was closed then we can impute these data with the following methods:

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.

Once again we have decided to apply the third method

If sales do not verify the previous conditions, then we have to use an alternative. We propose the following:

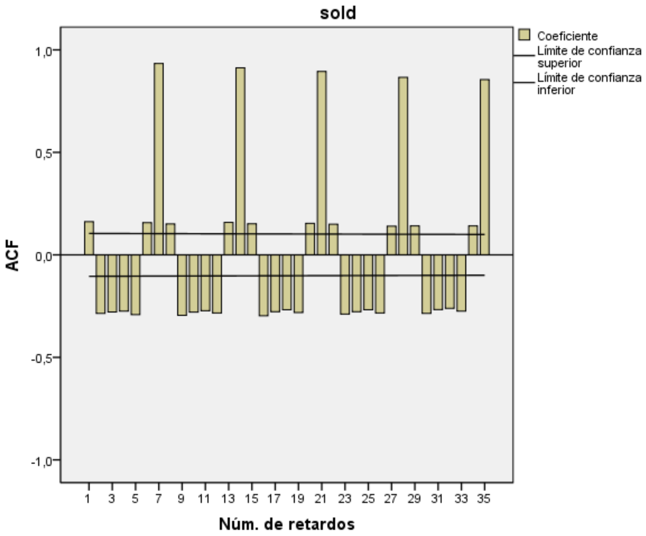
- The next sale in the same day

We take the next sale greater or equal than M sold in the same day of the week and whose returned was greater than zero. So, we are imputing with a sale similar and close in time.

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.

The first autocorrelation function is:



It's seasonal with period 7.

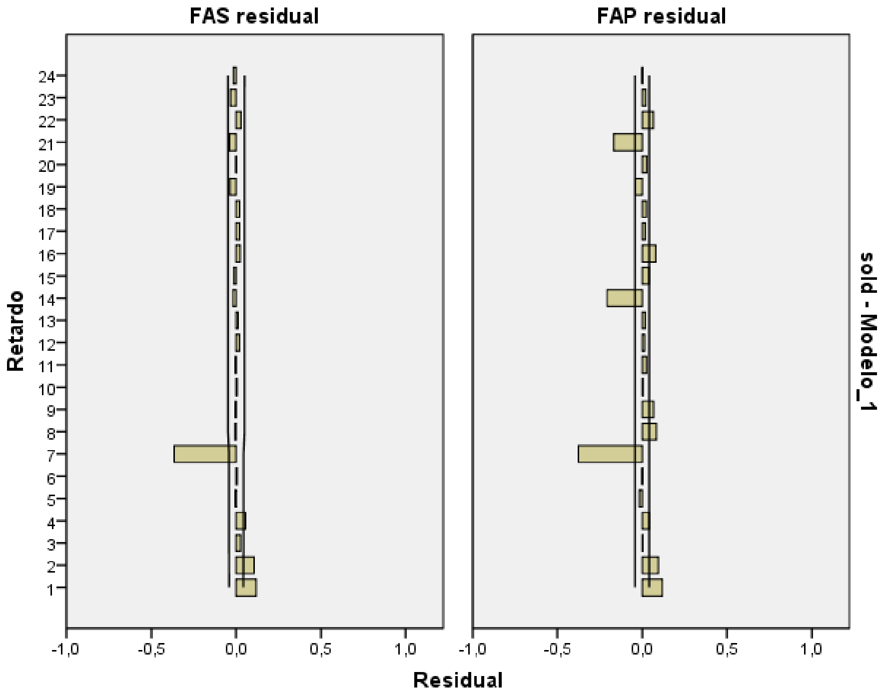
We take a difference with lag 7: **ARIMA(0,0,0)x(0,1,0)**

Ajuste del modelo

Estadístico de ajuste	Media	Mínimo	Máximo	Percentil 5	10	25	50	75	90	95
R-cuadrado estacionaria	-7,654E-15	-7,654E-15	-7,654E-15	-7,654E-15	-7,654E-15	-7,654E-15	-7,654E-15	-7,654E-15	-7,654E-15	-7,654E-15
R-cuadrado	,851	,851	,851	,851	,851	,851	,851	,851	,851	,851
RMSE	5,100	5,100	5,100	5,100	5,100	5,100	5,100	5,100	5,100	5,100
MAPE	19,598	19,598	19,598	19,598	19,598	19,598	19,598	19,598	19,598	19,598
MaxAPE	233,388	233,388	233,388	233,388	233,388	233,388	233,388	233,388	233,388	233,388
MAE	3,714	3,714	3,714	3,714	3,714	3,714	3,714	3,714	3,714	3,714
MaxAE	34,997	34,997	34,997	34,997	34,997	34,997	34,997	34,997	34,997	34,997
BIC normalizado	3,262	3,262	3,262	3,262	3,262	3,262	3,262	3,262	3,262	3,262

Parámetros del modelo ARIMA

				Estimación	ET	t	Sig.
sold-Modelo_1	sold	Sin transformación	Constante	,003	,110	,030	,976
			Diferenciación estacional	1			



The FAC also display a "spike" in the delay 7. And the FACP is becoming smaller in delays multiples of 7: Set seasonal MA.

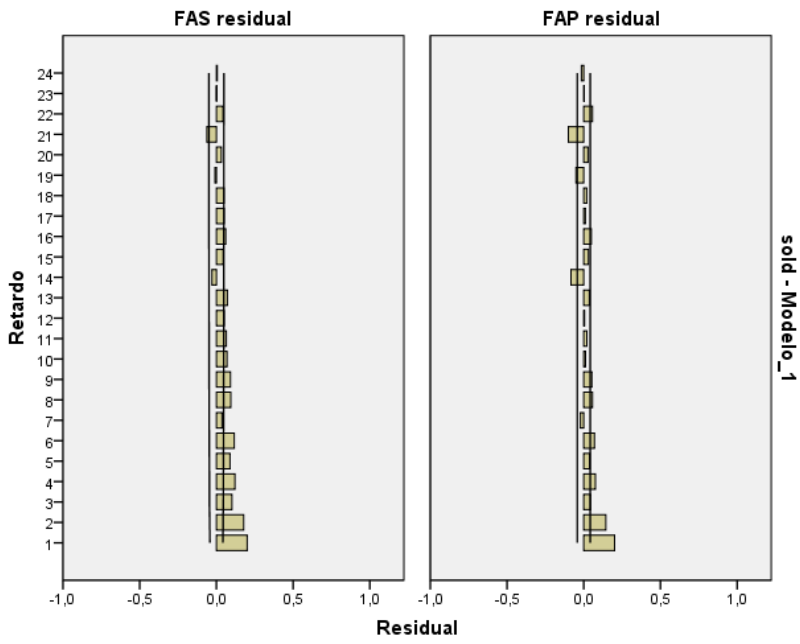
ARIMA(0,0,0)x(0,1,1):

Ajuste del modelo

Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil 5	10	25	50	75	90	95
R-cuadrado estacionaria	,183	.	,183	,183	,183	,183	,183	,183	,183	,183	,183
R-cuadrado	,878	.	,878	,878	,878	,878	,878	,878	,878	,878	,878
RMSE	4,612	.	4,612	4,612	4,612	4,612	4,612	4,612	4,612	4,612	4,612
MAPE	17,911	.	17,911	17,911	17,911	17,911	17,911	17,911	17,911	17,911	17,911
MaxAPE	203,372	.	203,372	203,372	203,372	203,372	203,372	203,372	203,372	203,372	203,372
MAE	3,361	.	3,361	3,361	3,361	3,361	3,361	3,361	3,361	3,361	3,361
MaxAE	33,205	.	33,205	33,205	33,205	33,205	33,205	33,205	33,205	33,205	33,205
BIC normalizado	3,064	.	3,064	3,064	3,064	3,064	3,064	3,064	3,064	3,064	3,064

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
sold-Modelo_1	sold	Sin transformación	Constante		,000	,048	-,004	,997
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,514	,019	27,692	,000



We think about adjust an AR(1).

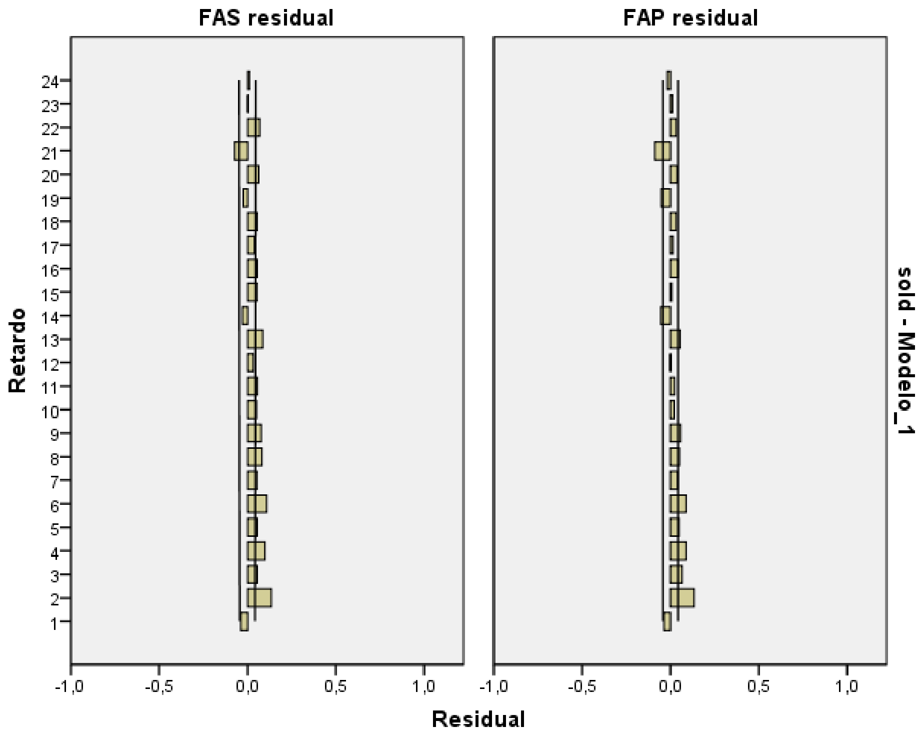
ARIMA (1,0,0)x(0,1,1)

Ajuste del modelo

Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil 5	10	25	50	75	90	95
R-cuadrado estacionaria	,220	.	,220	,220	,220	,220	,220	,220	,220	,220	,220
R-cuadrado	,884	.	,884	,884	,884	,884	,884	,884	,884	,884	,884
RMSE	4,508	.	4,508	4,508	4,508	4,508	4,508	4,508	4,508	4,508	4,508
MAPE	17,431	.	17,431	17,431	17,431	17,431	17,431	17,431	17,431	17,431	17,431
MaxAPE	201,855	.	201,855	201,855	201,855	201,855	201,855	201,855	201,855	201,855	201,855
MAE	3,274	.	3,274	3,274	3,274	3,274	3,274	3,274	3,274	3,274	3,274
MaxAE	32,773	.	32,773	32,773	32,773	32,773	32,773	32,773	32,773	32,773	32,773
BIC normalizado	3,022	.	3,022	3,022	3,022	3,022	3,022	3,022	3,022	3,022	3,022

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
sold-Modelo_1	sold	Sin transformación	Constante		-,001	,052	-,021	,983
			AR	Retardo 1	,225	,021	10,561	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,585	,018	32,832	,000



We think about adjust a regular MA.

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

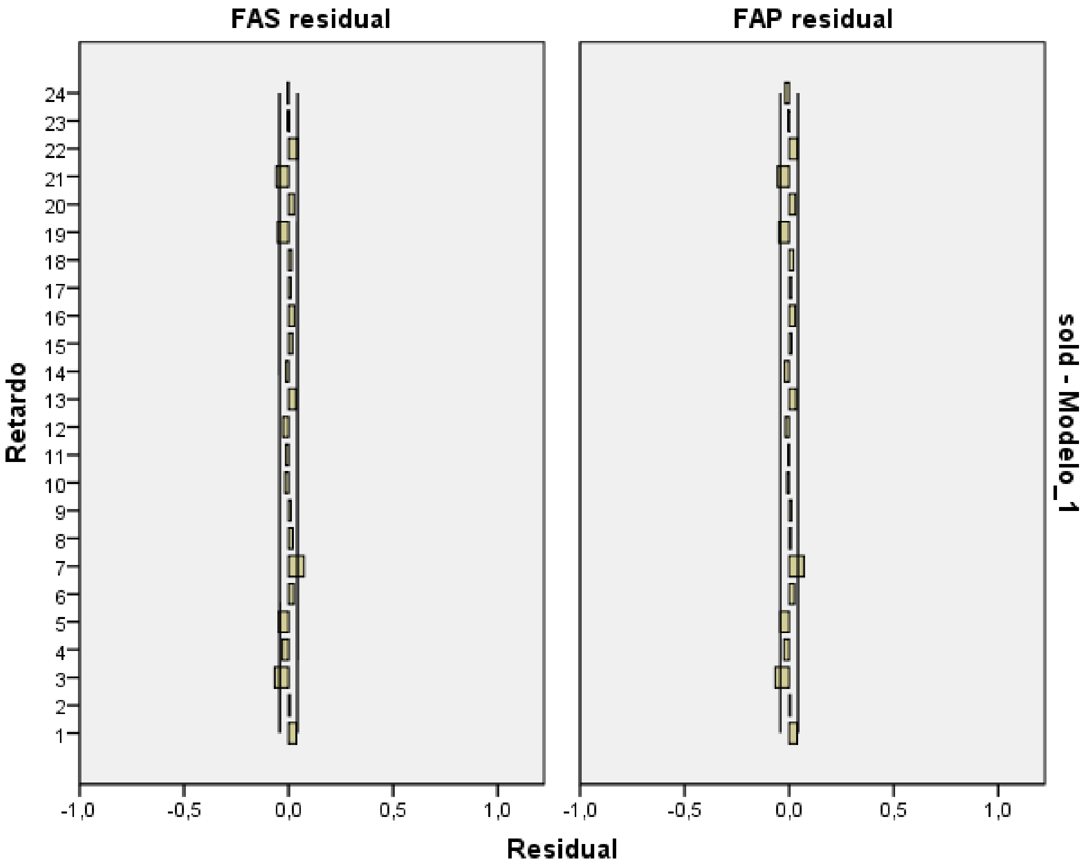
Ajuste del modelo

Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil 5	10	25	50	75	90	95
R-cuadrado estacionaria	,262	.	,262	,262	,262	,262	,262	,262	,262	,262	,262
R-cuadrado	,890	.	,890	,890	,890	,890	,890	,890	,890	,890	,890
RMSE	4,384	.	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384	4,384
MAPE	16,902	.	16,902	16,902	16,902	16,902	16,902	16,902	16,902	16,902	16,902
MaxAPE	161,133	.	161,133	161,133	161,133	161,133	161,133	161,133	161,133	161,133	161,133
MAE	3,186	.	3,186	3,186	3,186	3,186	3,186	3,186	3,186	3,186	3,186
MaxAE	30,579	.	30,579	30,579	30,579	30,579	30,579	30,579	30,579	30,579	30,579
BIC normalizado	2,970	.	2,970	2,970	2,970	2,970	2,970	2,970	2,970	2,970	2,970

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
sold-Modelo_1	sold	Sin transformación	Constante		,000	,093	-,005	,996
			AR	Retardo 1	,942	,014	68,922	,000

	MA	Retardo 1	,807	,023	34,812	,000
	Diferenciación estacional		1			
	MA, estacional	Retardo 1	,707	,017	42,649	,000



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

- Now we adjust an exponential smoothing: Simple exponential smoothing.

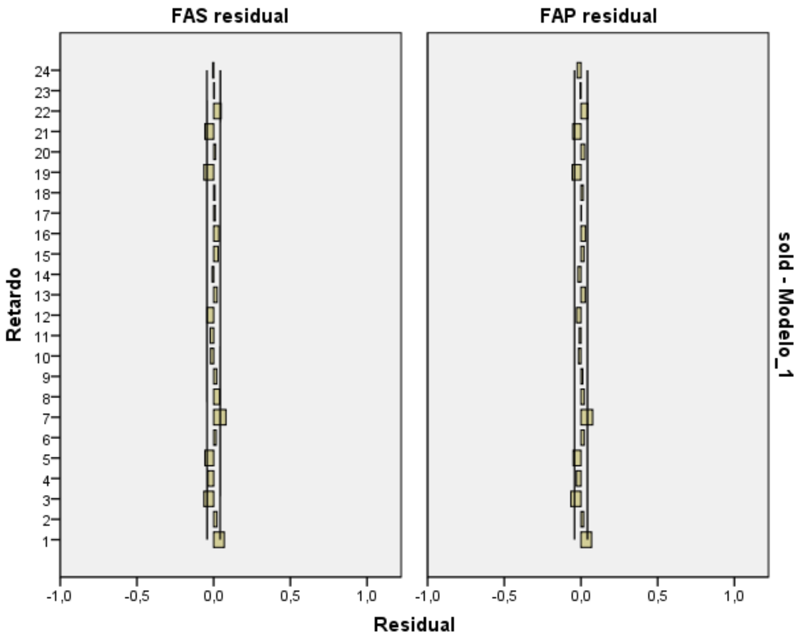
Ajuste del modelo

Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil 5	10	25	50	75	90	95
R-cuadrado estacionaria	,580	.	,580	,580	,580	,580	,580	,580	,580	,580	,580
R-cuadrado	,890	.	,890	,890	,890	,890	,890	,890	,890	,890	,890
RMSE	4,383	.	4,383	4,383	4,383	4,383	4,383	4,383	4,383	4,383	4,383
MAPE	16,921	.	16,921	16,921	16,921	16,921	16,921	16,921	16,921	16,921	16,921
MaxAPE	159,059	.	159,059	159,059	159,059	159,059	159,059	159,059	159,059	159,059	159,059
MAE	3,188	.	3,188	3,188	3,188	3,188	3,188	3,188	3,188	3,188	3,188

MaxAE	30,689	.	30,689	30,689	30,689	30,689	30,689	30,689	30,689	30,689	30,689
BIC normalizado	2,963	.	2,963	2,963	2,963	2,963	2,963	2,963	2,963	2,963	2,963

Parámetros del modelo de suavizado exponencial

Modelo			Estimación	ET	t	Sig.
sold-Modelo_1	Sin transformación	Alpha (Nivel)	,109	,010	10,498	,000
		Delta (Estación)	,306	,017	18,128	,000



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): $(1-\Phi_1 * B)(1-B_7)X(t)=(1-\theta_1 * B)(1-\Theta_7 * B_7)\varepsilon(t)$

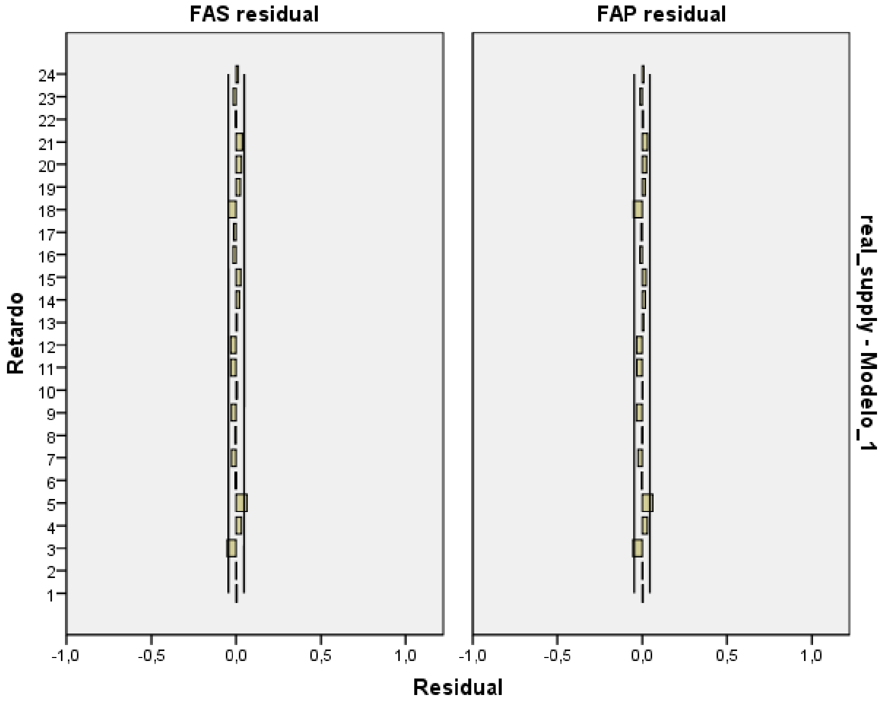
SERIE 34

The parameters for serie number 34 are:

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
real_supply-Modelo_1	real_supply	Sin transformación	AR	Retardo 1	,687	,072	9,530	,000
			MA	Retardo 1	,535	,084	6,388	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,782	,015	51,941	,000

And the FAS and FAP are:



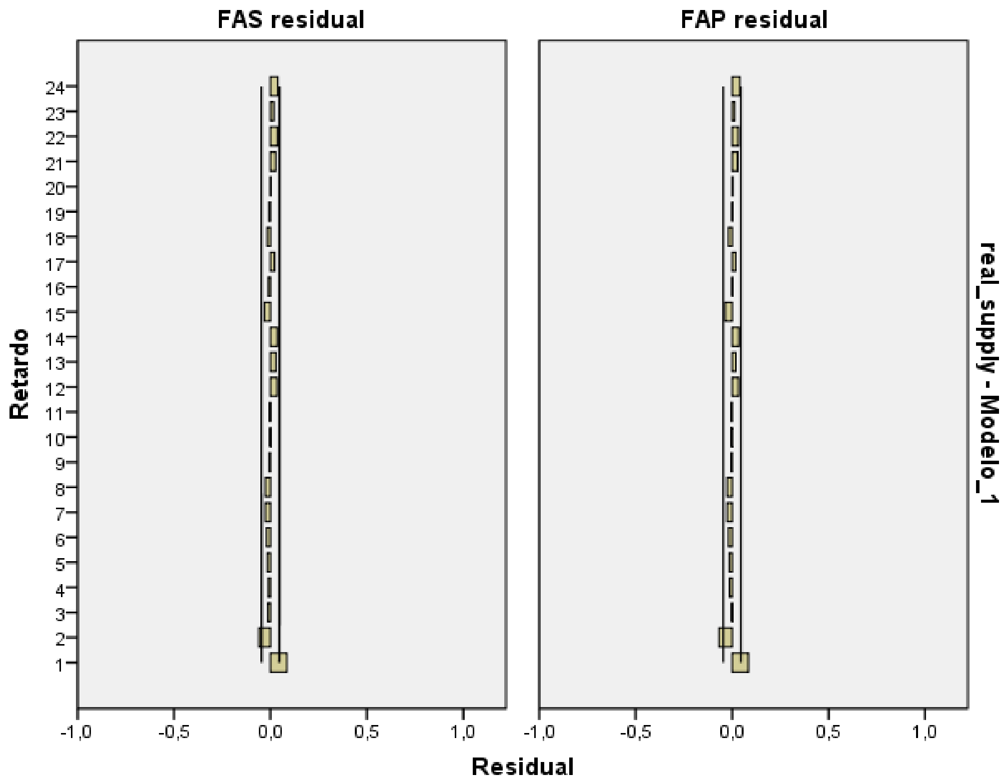
We can see that the general model adjusts well.

SERIE 38:

With the serie number 38, the parameters estimates for the ARIMA model are:

Parámetros del modelo ARIMA

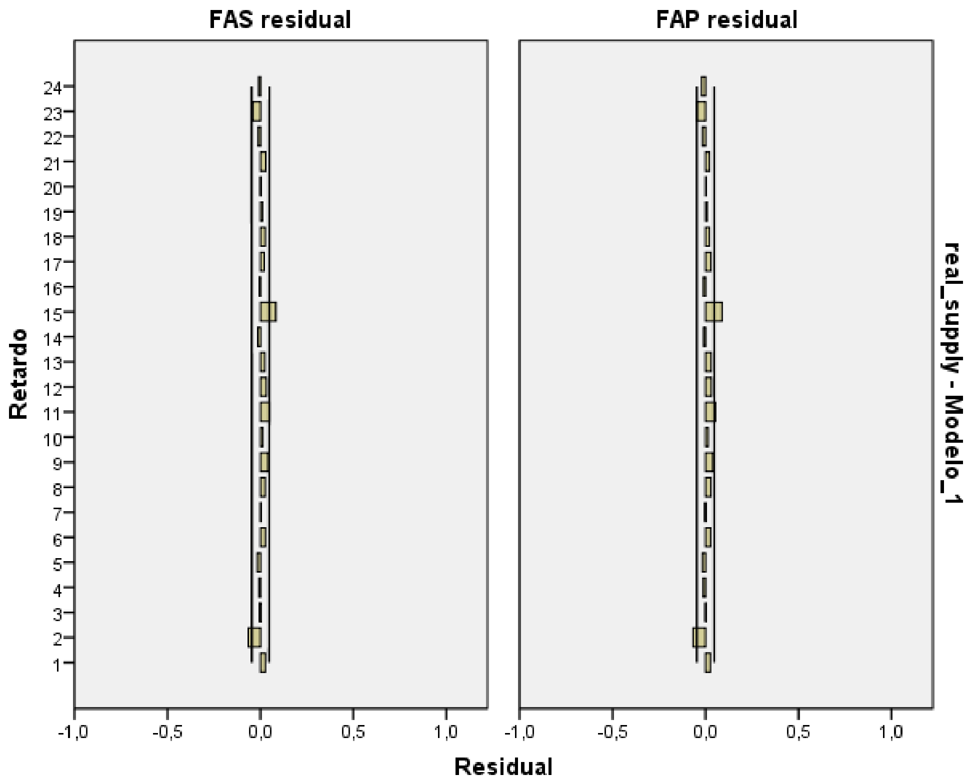
					Estimación	ET	t	Sig.
real_supply-Modelo_1	real_supply	Sin transformación	AR	Retardo 1	,942	,015	61,915	,000
			MA	Retardo 1	,810	,026	31,600	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,875	,012	70,720	,000



SERIE 55:

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
real_supply-Modelo_1	real_supply	Sin transformación	AR	Retardo 1	,805	,024	33,449	,000
			MA	Retardo 1	,393	,037	10,694	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,500	,021	23,474	,000



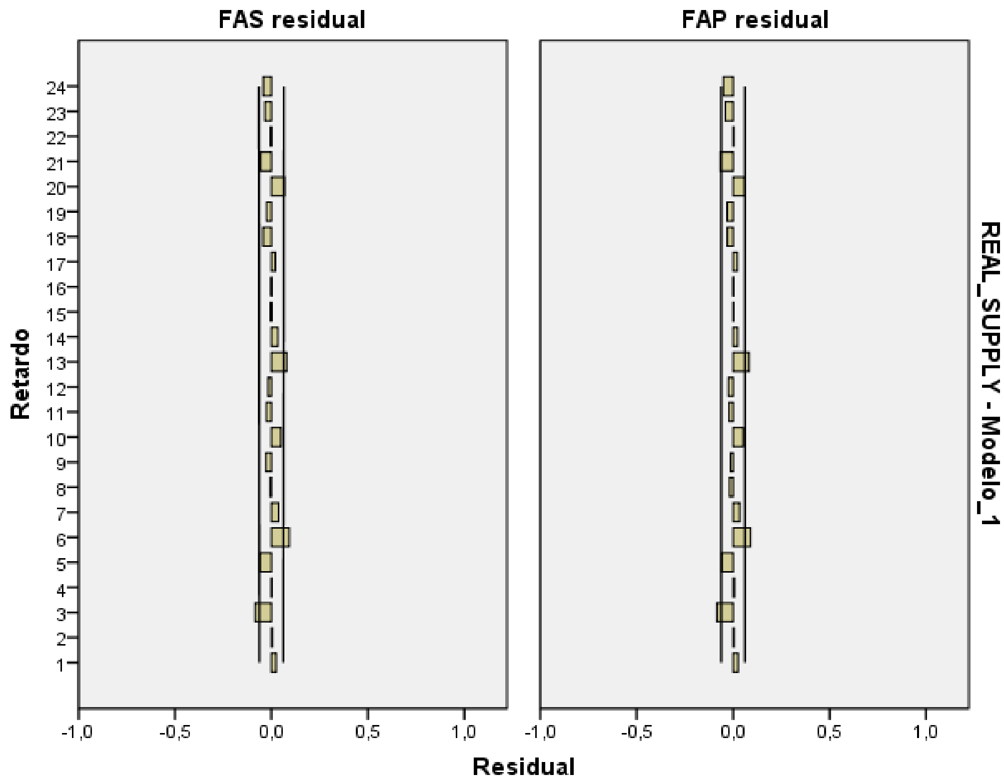
In this case, the model adjusts well too.

SERIE GLOBAL:

With the sum serie:

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
REAL_SUPPLY-Modelo_1	REAL_SUPPLY	Sin transformación	AR	Retardo 1	,939	,023	41,461	,000
			MA	Retardo 1	,825	,036	22,979	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,660	,026	25,358	,000



Even now, the model adjusts well too.

6. Intervention analysis

WITH THE ADVERSTISING PRESSURE

We choose the model ARIMA. We include with SPSS the variable advertising pressure.

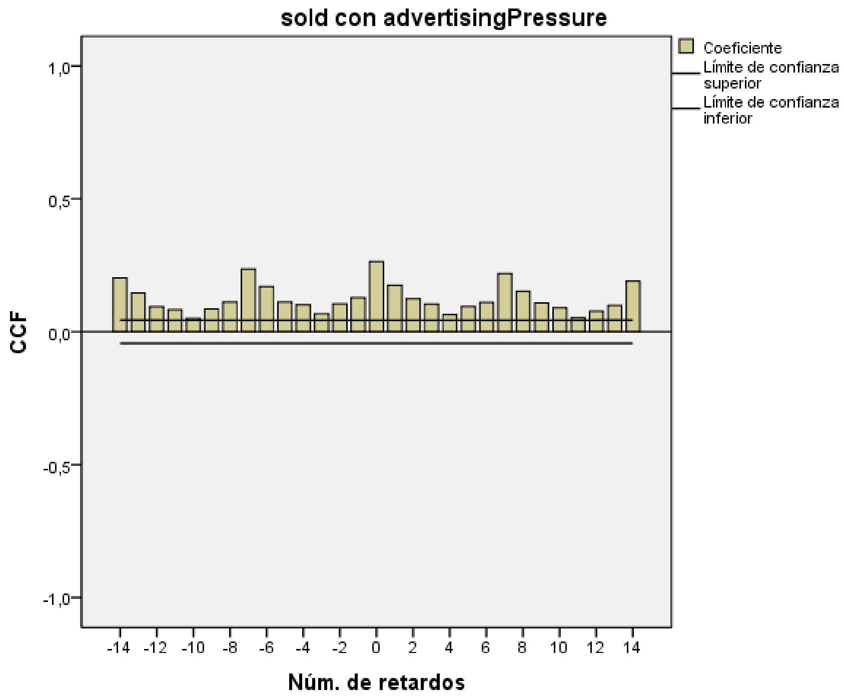
Cross correlation function:

Cross correlation

Par de series:sold con advertisingPressure

Retardo	Correlación cruzada	Típ. Error[a]
-5	,112	,022
-4	,102	,022
-3	,068	,022
-2	,105	,022
-1	,128	,022
0	*,264*	*,022*
1	,175	,022
2	,124	,022
3	,104	,022
4	,064	,022
5	,095	,022

a. Basado en los supuestos de que las series no presentan correlación cruzada y que una de las series es ruido blanco.

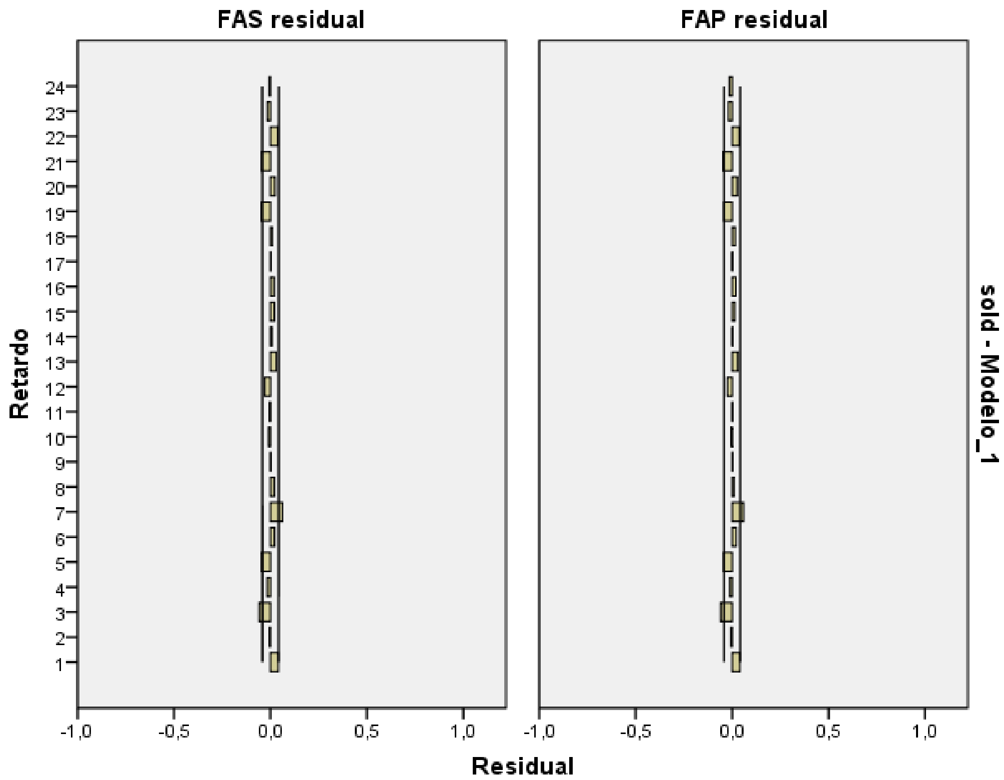


Model adjust

Estadístico de ajuste	Media	ET	Mínimo	Máximo	Percentil 5	10	25	50	75	90	95
R-cuadrado estacionaria	,335	.	,335	,335	,335	,335	,335	,335	,335	,335	,335
R-cuadrado	,901	.	,901	,901	,901	,901	,901	,901	,901	,901	,901
RMSE	4,161	.	4,161	4,161	4,161	4,161	4,161	4,161	4,161	4,161	4,161
MAPE	16,306	.	16,306	16,306	16,306	16,306	16,306	16,306	16,306	16,306	16,306
MaxAPE	148,488	.	148,488	148,488	148,488	148,488	148,488	148,488	148,488	148,488	148,488
MAE	3,064	.	3,064	3,064	3,064	3,064	3,064	3,064	3,064	3,064	3,064
MaxAE	20,053	.	20,053	20,053	20,053	20,053	20,053	20,053	20,053	20,053	20,053
BIC normalizado	2,866	.	2,866	2,866	2,866	2,866	2,866	2,866	2,866	2,866	2,866

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
sold-Modelo_1	sold	Sin transformación	AR	Retardo 1	,930	,018	50,536	,000
			MA	Retardo 1	,818	,028	29,560	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,738	,016	46,598	,000
	advertisingPressure	Sin transformación	Numerador	Retardo 0	,022	,001	15,766	,000
			Diferenciación estacional		1			



The parameter estimation of the variable advertising pressure is 0,022, i.e., an order to increase sales twenty units per 1000 units of “pressure advertising" inverted. And it is significant coefficient.

The final model for all series is:

$(1-\Phi_1 * B)(1-B_7) X(t)=(1-\theta_1 * B)(1-\Theta_7 * B_7)\varepsilon(t)+0{,}022(1-B_7)Ad(t)$

Being Ad(t) the variable advertisement pressure.

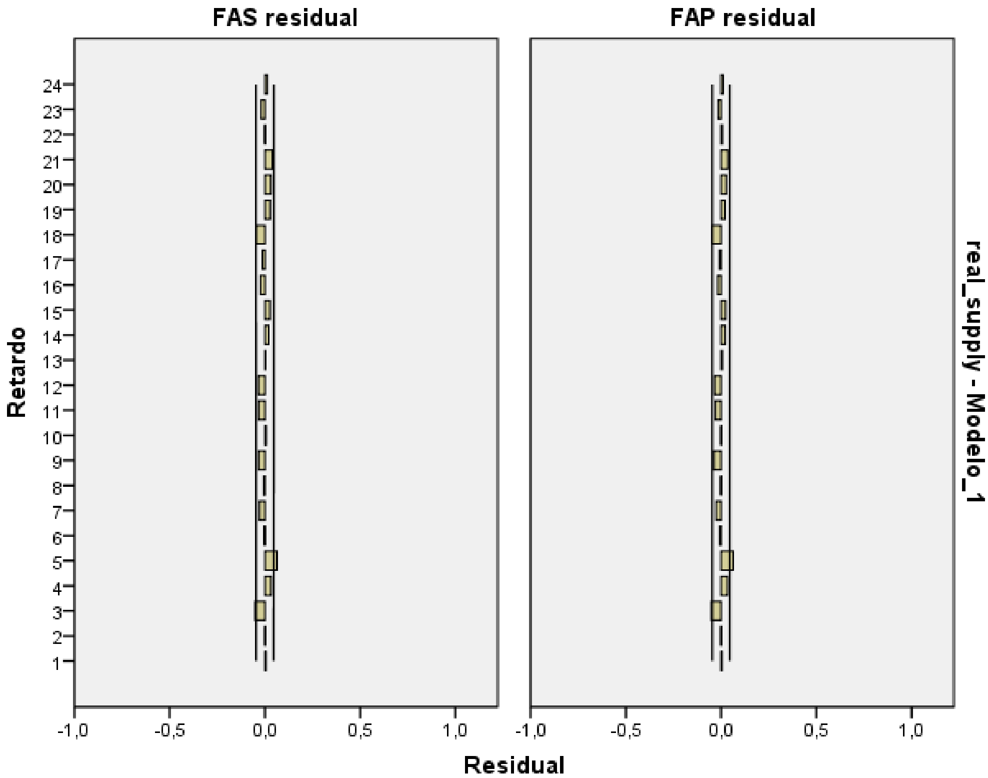
SERIE 34

If we include the advertising pressure variable, the parameters change a little bit.

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
real_supply-Modelo_1	real_supply	Sin transformación	AR	Retardo 1	,664	,080	8,263	,000
			MA	Retardo 1	,518	,092	5,646	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,783	,015	52,069	,000
	advertisingPresure	Sin transformación	Numerador	Retardo 0	,008	,003	2,856	,004
			Diferenciación estacional		1			

Advertising pressure is significant into the model, so we have to include it.



Now, the model ARIMA(1,0,1)X(0,1,1) adjusts well.

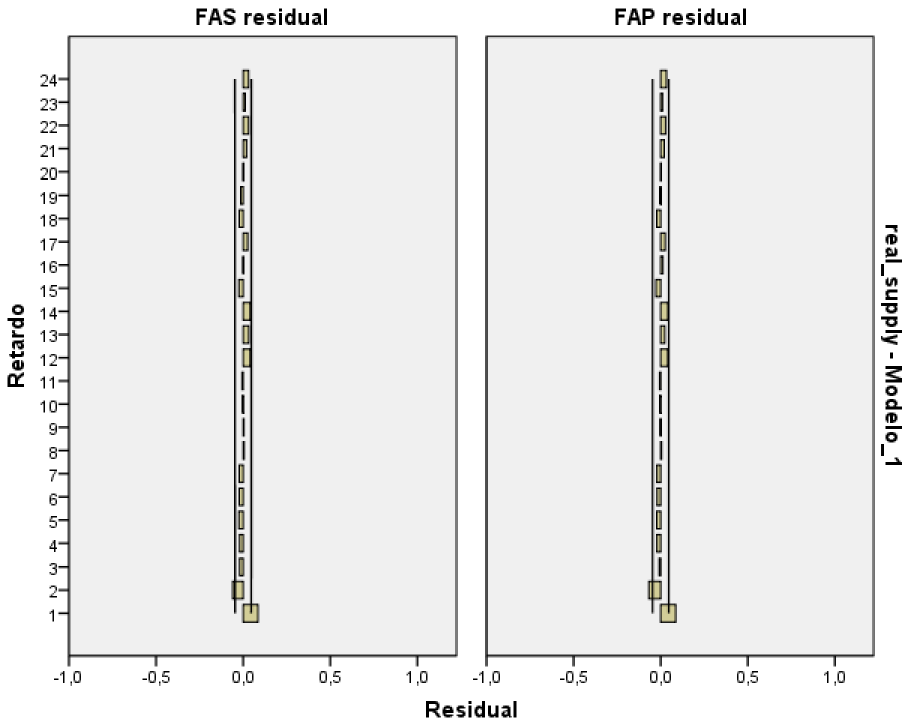
SERIE 38:

Including advertising pressure, we can see that it's significant. We have to include it into the model.

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
real_supply-Modelo_1	real_supply	Sin transformación	AR	Retardo 1	,910	,026	35,161	,000
			MA	Retardo 1	,799	,036	21,966	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,862	,013	66,123	,000
	advertisingPressure	Sin transformación	Numerador	Retardo 0	,022	,002	9,056	,000
			Diferenciación estacional		1			

Finally, the FAS and FAP functions are:

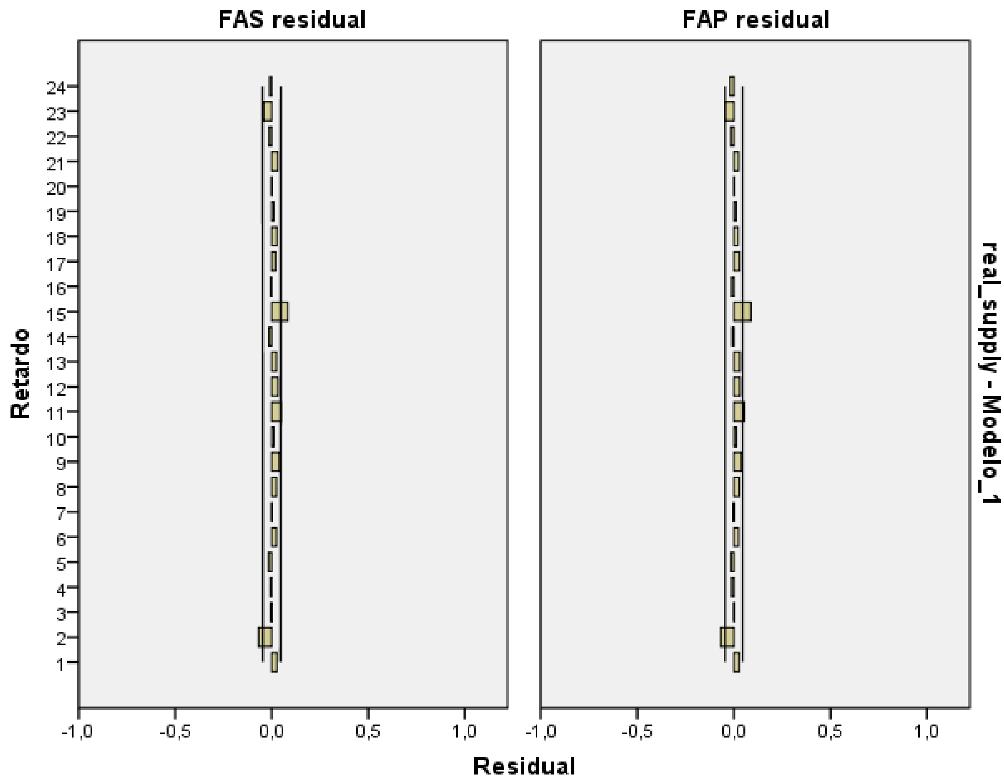


SERIE 55:

Including advertising pressure we can see the next table for the parameters of the model:

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
real_supply-Modelo_1	real_supply	Sin transformación	AR	Retardo 1	,809	,024	34,050	,000
			MA	Retardo 1	,393	,036	10,792	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,502	,021	23,540	,000
	advertisingPressure	Sin transformación	Numerador	Retardo 0	,008	,007	1,137	,256
			Diferenciación estacional		1			



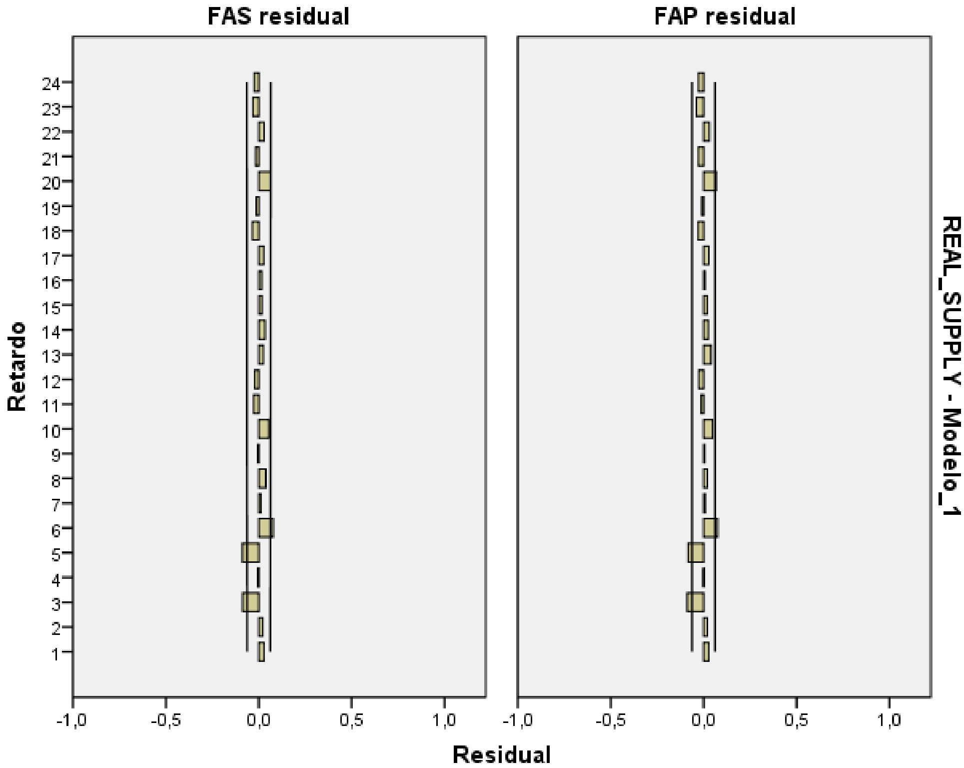
Now, the variable advertising pressure is not significant for any level.

SERIE GLOBAL:

Including advertising pressure, the parameters estimates are:

Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
REAL_SUPPLY-Modelo_1	REAL_SUPPLY	Sin transformación	AR	Retardo 1	,951	,022	44,061	,000
			MA	Retardo 1	,865	,034	25,791	,000
			Diferenciación estacional		1			
			MA, estacional	Retardo 1	,635	,027	23,689	,000
	advertisingPressure	Sin transformación	Numerador	Retardo 0	1,639	,124	13,255	,000
			Diferenciación estacional		1			

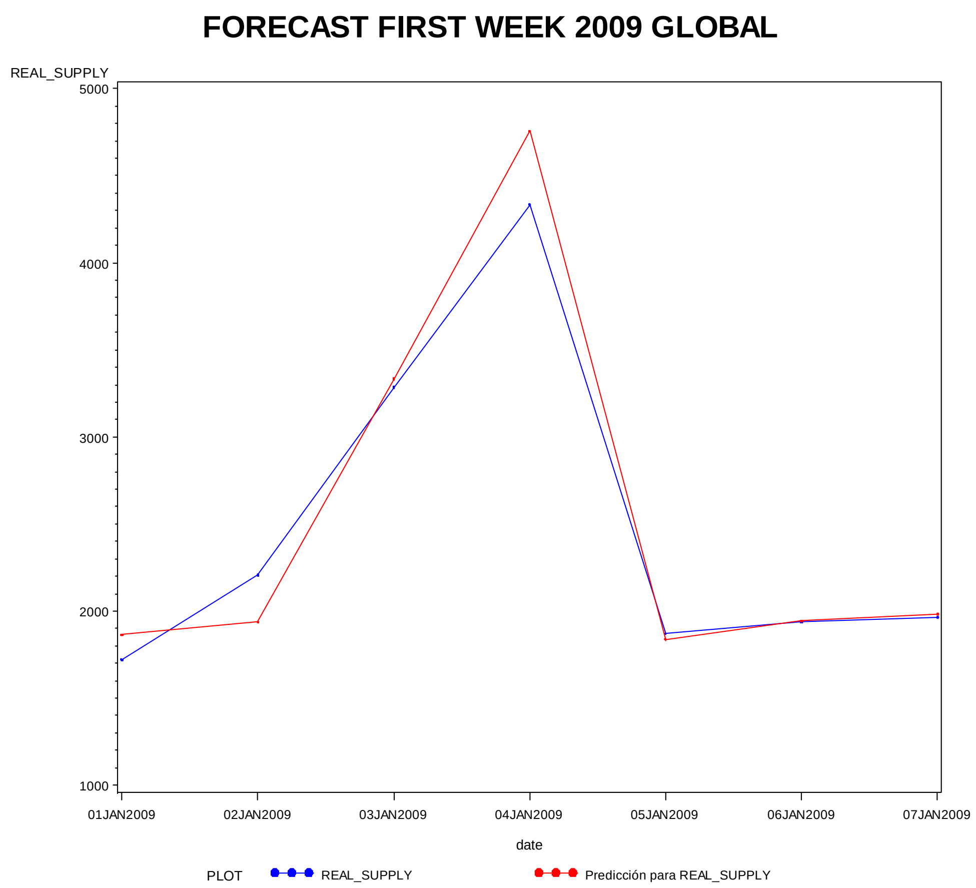


The variable advertising pressure is significant into the model.

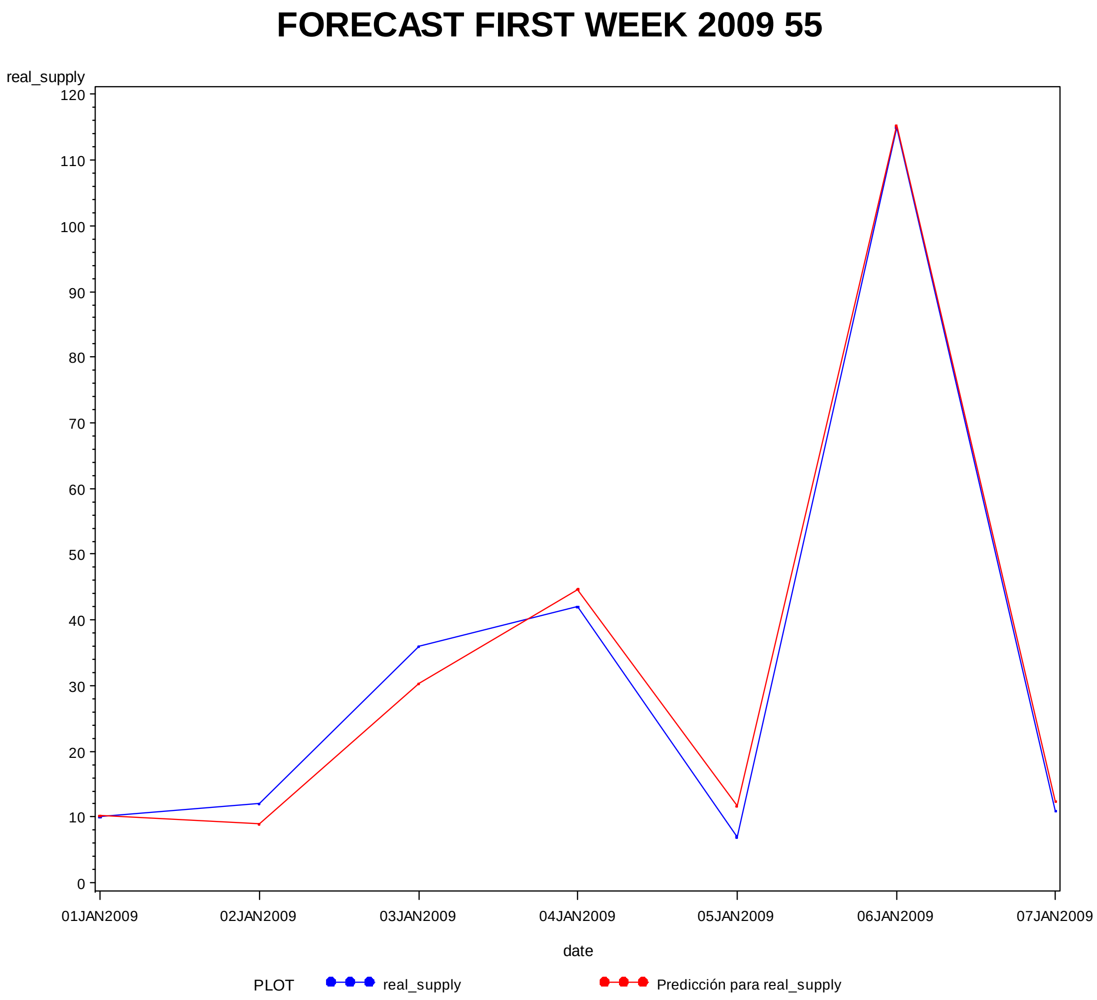
7. Model validation

The model has been created to forecast one week. The validation set chosen was the first week of 2009. In the next graphic you can see the forecasting. Take in mind that the blue serie (the one with real values) has imputed values, in this case, the real value of 01ENE is null in every id, but here is imputed as described previously.

date	REAL SUPPLY	Predicción para REAL_SUPPLY	Error std. de predicción	Límite de confianza inferior 95%	Límite de confianza superior 95%
01JAN2009	1721	1863.63	299.379	1276.86	2450.41
02JAN2009	2208	1939.53	301.324	1348.95	2530.12
03JAN2009	3286	3334.90	303.030	2740.97	3928.82
04JAN2009	4335	4758.32	304.527	4161.45	5355.18
05JAN2009	1872	1837.82	305.842	1238.38	2437.26
06JAN2009	1938	1943.92	306.998	1342.21	2545.62
07JAN2009	1966	1983.59	308.014	1379.89	2587.29



date	real_supply	Predicción para real_supply	Error std. de predicción	Límite de confianza inferior 95%	Límite de confianza superior 95%
01JAN2009	10	10.175	18.3043	-25.7005	46.051
02JAN2009	12	8.937	19.7923	-29.8554	47.729
03JAN2009	36	30.327	20.7014	-10.2468	70.901
04JAN2009	42	44.672	21.2709	2.9820	86.363
05JAN2009	7	11.745	21.6328	-30.6540	54.145
06JAN2009	115	115.237	21.8645	72.3831	158.091
07JAN2009	11	12.426	22.0137	-30.7199	55.572



8. Determination of the daily demand

Information on whether the point of sale is closed or not, will be known at the time of the prediction so, in that cases you must ignore the prediction we have made for the demand and no send at all.

It must be taken into account the loss associated to underestimate the demand and the cost of returns.

Once, the prediction of the bread sales is estimated by the fixed model we have to determine how many units of bread definitely we will send to the store. The decision will be taken after analyzing benefits and costs.

- First at all, for every store we have to know the maximum units of bread that has been sold at historical data. Obviously, we'll never send more units of bread than the historical maximum.

- 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. Therefore, we have to decide send the expected sales or the expected sales plus one unit more, two unts more,… or the historical maximum.

- 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 a 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 than maximize the benefits.

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 consequence, in practice what the company requires is a **global model** that works for all of them

HOLIDAYS

Instead of forecasting the real series we had chosen to forecast imputed series (holidays or censured values). That 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.

OTHER OPTIONS (7 SERIES INSTEAD OF 1)

As explained before, because of the big amount of data, we have chosen a global model. Other option that we contemplated was to make different series for each day of the week. We finally choose the global option, because the data imputation is easier this way. Specially, in the case of a store with 'favorite day'. We saw that the Sunday series (the weekday with the biggest closed ratio) was not very easy to fit well.

10. Appendix

We attache several SAS codes.