"Dimitrie Cantemir" Christian University
Knowledge Horizons - Economics
Volume 7, No. 4, pp. 18–22
P-ISSN: 2069-0932, E-ISSN: 2066-1061

www.orizonturi.ucdc.ro

FORECASTING SEASONAL FACTORS METHOD vs. REGRESSION METHOD WITH MS EXCEL

Petru BALOGH[1], Pompiliu GOLEA[2]

[1] *"Dimitrie Cantemir" Christian University, E-mail: petrubalogh@yahoo.ro*
[2] *"Dimitrie Cantemir", Christian University, E-mail: golea_p@yahoo.com*

Abstract *Predicting sales for highly seasonal products is very different compared to products who sell regularly throughout the year. In this paper we analyze the results from the seasonal factors method and from the regression method. The example used will be predicting sales of bottled water in Romania. The sales prediction will be made for the previous year, so that the results can be compared with the actual sales numbers for bottled water. MS Excel software was used due to its accessibility. The authors recommend the regression method.*

Key words:
Forecasting, Seasonal factor, Regression, Comparative analysis
JEL Codes:
C13

1. Introduction

Markets that are very seasonal in nature don't allow any mistakes. Product sales must follow a very well thought-out plan because with these types of markets there isn't enough time to make adjustments, a wrong prediction can be very costly.

Sales forecasting of actual sales as well as potential sales is usually the responsability of the marketing department.

In figure 1 we show the main methods used for sales forecasting.

.

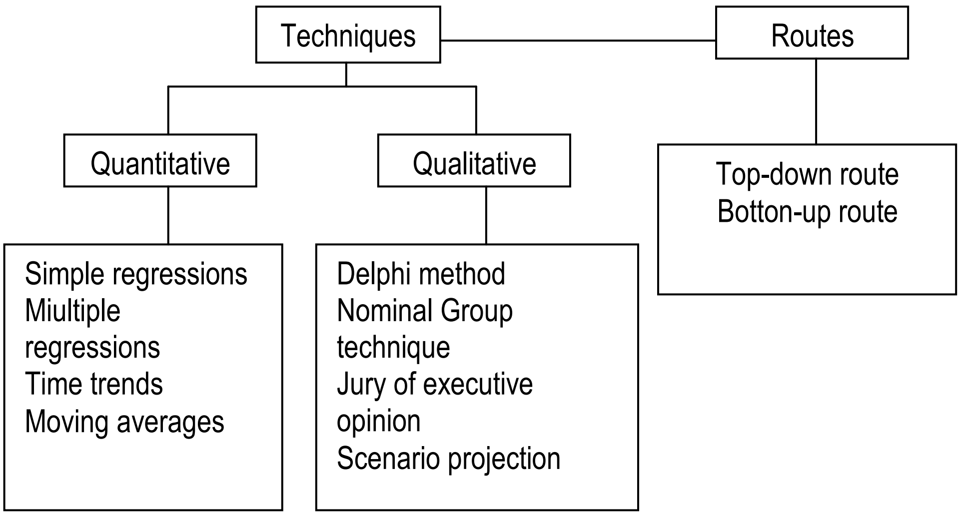

Figure 1. Forecasting methods classificasions (Source: 1)

Correctly determining the basis for solving a marketing problem relies on predicting the changes that can occur in the external factors that affect sales. These factors can be: economical, technical, political, demographic etc.

There are many methods that can be used for making these kinds of predictions. Some of them aren't very rigorous from a statistical and mathematical standpoint but they are still used because of their ease of use quality, flexibility, as well as an acceptable accuracy but only in relation to their cost.

2. Methodology of research

This paper compares the results from the seasonality factor method with those from the multiple regression method.

The data for bottled water consumption for years n-2 and n-1 are shown in table 1.

The data for year n, for the months of january, february, march and april are actual sales data that will be compared with the predition results for the same months, using the two methods mentioned above.

Table 1. Bottled water consumption [%]

month/year	n-2	n-1	n
January	49	55	54
February	54	53	56
March	53	54	57
April	53	57	57
May	58	57	
June	54	61	
July	57	58	
August	59	58	
September	54	56	
October	56	55	
November	55	57	
December	54	52	

Source: DaedalusMillwardBrown

The seasonality factor (index) method entails:

- Calculating an average between the years n-2 and n-1, for each period (month in this example)
- Calculating an average of all sales
- Assuming the average of all sales as an average of future sales (if the possible factors that can influence the results are known, this value can be modified)
- We multiply the seasonality factor with the average of all sales.

Fig. 1 Calculating the seasonality coefficient using MSExcel

	A	C	D	E	F	G	H	I
1								
2	month/year	n-2	n-1	n		Seasonal auverage	Seasonal Factor	Expected year n
3	January	49	55	54		52	0,939052	52
4	February	54	53	56		53,5	0,96614	54
5	March	53	54	57		53,5	0,96614	54
6	April	53	57	57		55	0,993228	55
7	May	58	57			57,5	1,038375	
8	June	54	61			57,5	1,038375	
9	July	57	58			57,5	1,038375	
10	August	59	58			58,5	1,056433	
11	September	54	56			55	0,993228	
12	October	56	55			55,5	1,002257	
13	November	55	57			56	1,011287	
14	December	54	52			53	0,957111	
15								
16	Auverage	54,67	56,08					
17	Overal auverage	55,375						

Regression Forecasting Procedure

- time series is modeled as having k seasons (Here we illustrate k = 12 months);
- The combination of 0's and 1's for each of the dummy variables at each period indicate the season corresponding to the time series value.
  - Season 1: $S_1 = 1$, $S_2 = 0$,…………… $S_{12} = 0$

– Season 2: $S_1 = 0$, $S_2 = 1$, …………..$S_{12} = 0$

…………………………………………………..

– Season 12: $S_1 = 0$, $S_2 = 0$,………… $S12_3 = 1$

– Multiple regression is then done on with t, S_1, S_2, ….and S_{12} as the independent variables and the time series values y_t as the dependent variable.

$$Y_t = \alpha_0 + \alpha_1 \cdot t + \alpha_2 \cdot S_1 + \alpha_3 \cdot S_2 + \cdots + \alpha_{13} \cdot S_{12}.$$

To determine the α_0, α_1 ,… α_{13} coefficients we use the Data Analysis module of MSExcel.

L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Y_i	t	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
January	49	1	1	0	0	0	0	0	0	0	0	0	0	0
February	54	2	0	1	0	0	0	0	0	0	0	0	0	0
March	53	3	0	0	1	0	0	0	0	0	0	0	0	0
April	53	4	0	0	0	1	0	0	0	0	0	0	0	0
May	58	5	0	0	0	0	1	0	0	0	0	0	0	0
June	54	6	0	0	0	0	0	1	0	0	0	0	0	0
July	57	7	0	0	0	0	0	0	1	0	0	0	0	0
August	59	8	0	0	0	0	0	0	0	1	0	0	0	0
September	54	9	0	0	0	0	0	0	0	0	1	0	0	0
October	56	10	0	0	0	0	0	0	0	0	0	1	0	0
November	55	11	0	0	0	0	0	0	0	0	0	0	1	0
December	54	12	0	0	0	0	0	0	0	0	0	0	0	1
January	55	13	1	0	0	0	0	0	0	0	0	0	0	0
February	53	14	0	1	0	0	0	0	0	0	0	0	0	0
March	54	15	0	0	1	0	0	0	0	0	0	0	0	0
April	57	16	0	0	0	1	0	0	0	0	0	0	0	0
May	57	17	0	0	0	0	1	0	0	0	0	0	0	0
June	61	18	0	0	0	0	0	1	0	0	0	0	0	0
July	58	19	0	0	0	0	0	0	1	0	0	0	0	0
August	58	20	0	0	0	0	0	0	0	1	0	0	0	0
September	56	21	0	0	0	0	0	0	0	0	1	0	0	0
October	55	22	0	0	0	0	0	0	0	0	0	1	0	0
November	57	23	0	0	0	0	0	0	0	0	0	0	1	0
December	52	24	0	0	0	0	0	0	0	0	0	0	0	1

Fig. 2 Regression input

After selecting the Y_i (Input Y Range), t și S_i (Input X Range) and e variables, setting the trust level (95%) we get the α_i . coefficient.

Data analysis shows us that:

– Adjusted R Square=0,907;
– F>Fsignificance (1314,05>2,35E-14);
– $T_{stat} > p_{value}$ for all α_i coefficients.

This means that all the requirements for accepting the values of α_i coefficients are met.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,999678192
R Square	0,999356488
Adjusted R Square	0,907745385
Standard Error	2,077111923
Observations	24

"Strength" of regression

Low p-value for F-test

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	73701,54167	5669,35	1314,05	2,35454E-14
Residual	11	47,45833333	4,31439		
Total	24	73749			

Low p-values for all t-tests

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
X Variable 1	0,118055556	0,070664783	1,67064	0,12297	-0,037476582	0,273587693
X Variable 2	51,17361111	1,549799675	33,0195	2,3E-12	47,76252503	54,5846972
X Variable 3	52,55555556	1,573779434	33,3945	2,1E-12	49,09169038	56,01942073
X Variable 4	52,4375	1,600522228	32,7627	2,6E-12	48,91477433	55,96022567
X Variable 5	53,81944444	1,629892058	33,0203	2,3E-12	50,23207621	57,40681267
X Variable 6	56,20138889	1,661749639	33,8206	1,8E-12	52,5439026	59,85887518
X Variable 7	56,08333333	1,695954783	33,0689	2,3E-12	52,35056203	59,81610464
X Variable 8	55,96527778	1,732368441	32,3056	3E-12	52,15236055	59,778195
X Variable 9	56,84722222	1,770854377	32,1016	3,2E-12	52,94959802	60,74484642
X Variable 10	53,22916667	1,811280502	29,3876	8,3E-12	49,24256516	57,21576817
X Variable 11	53,61111111	1,853519872	28,9239	9,9E-12	49,53154138	57,69068084
X Variable 12	53,99305556	1,897451394	28,4556	1,2E-11	49,8167932	58,16931791
X Variable 13	50,875	1,942960292	26,1843	2,9E-11	46,59857323	55,15142677

Fig. 3 Regression summary output

The dependant variable becomes:

$$Y_t = 0{,}118 \cdot t + 51{,}173 \cdot S_1 + 52{,}555 \cdot S_2 + \cdots + 50{,}875 \cdot S_{12}\,.$$

=$M$50+$M$51*N27+$M$52*O27+$M$53*P27+$M$54*Q27+$M$55*R27

L	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
	t	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	**Expected**
November	23	0	0	0	0	0	0	0	0	0	0	1	0	
December	24	0	0	0	0	0	0	0	0	0	0	0	1	
January	25	1	0	0	0	0	0	0	0	0	0	0	0	**54,125**
February	26	0	1	0	0	0	0	0	0	0	0	0	0	**55,625**
March	27	0	0	1	0	0	0	0	0	0	0	0	0	**55,625**
April	28	0	0	0	1	0	0	0	0	0	0	0	0	**57,125**

Fig. 4 Regression output

3. Results

Results are shown in figure 5 and table

2.

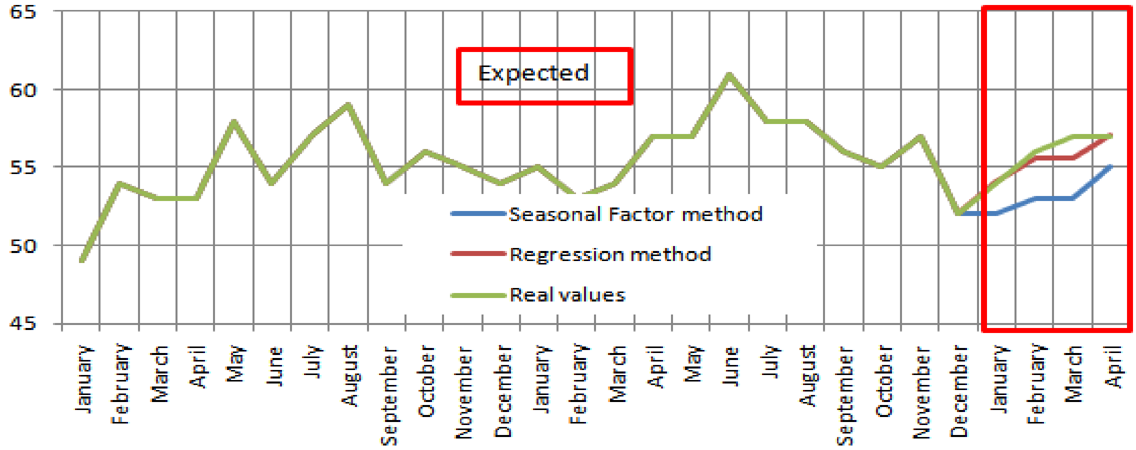

Fig. 5 Results Graph

Tabelul 2. **Results and relative errors**

	Real values	Seasonal Factor method	Error [%]	Regression method	Error [%]
January	54	52	3,70	54	0,00
February	56	54	3,57	56	0,00
March	57	54	5,26	56	1,75
April	57	55	3,51	57	0,00
		Auverage	**4,01**		**0,44**

4. **Conclusions**

An analysis of the results shows that the multiple regression method predicts the values very close the actual values for the period. The relative error in this case is 0.44%.

The seasonality factor method is less precise in this case, with a relative error of 4.01%.

The multiple regression method is recommended if:

– Adjusted R Square>0,8;
– F>>>Fsignificance
– T_{stat}>>>p_{value} for all α_i coefficients.

References

Robert Nau, "Statistical forecasting: notes on regression and time series analysis" available on-line at www. *people.duke.edu/~rnau/411home.htm;*

Rob J Hyndman, George Athanasopoulos, "Forecasting: principles and practice" on-line at https://www.otexts.org/books 2015.