Operations Research; Statistical Decision Theory

An Analysis of Two Types of Regressions for the same Dataset

Cătălin Angelo Ioan[1], Gina Ioan[2]

Abstract: The paper analyzes two regression methods for a set of data relative to the absolute values, respectively their variation indices. A number of conclusions are drawn regarding the closest forecast to reality.

Keywords: regression; forecast; index

JEL Code: E17; E27

1. Introduction

In the forecasting activity, the used methods are of particular importance. Thus, for the same set of data, the absolute data and the growth indexes can be used.

The problem studied here is which prognosis is closer to the real situation.

In the article, we will consider a set of data relative to a certain period, performing both types of regressions after which we compare the results with the real ones obtained in the next period.

Also, the predicted value is often based on a final value of the estimated data set due to insufficient reporting. The question is to what extent interim data influence the forecast in both cases.

[1] Associate Professor, PhD, Danubius University of Galati, Department of Economics, Romania, Address: 3 Galati Blvd., Galati 800654, Romania, Tel.: +40372361102, Corresponding author: catalin_angelo_ioan@univ-danubius.ro.
[2] Senior Lecturer, PhD, Danubius University of Galati, Department of Economics, Romania, Address: 3 Galati Blvd., Galati 800654, Romania, Tel.: +40372361102, E-mail: ginaioan@univ-danubius.ro.

2. The Analysis

Considering a set of indicators $(t_k, x_k)_{k=\overline{1,n}}$, the regression equation corresponding to this set obtained by the least squares method is x=at+b, where:

$$a=\frac{n\sum_{k=1}^{n}t_kx_k-\sum_{k=1}^{n}t_k\sum_{k=1}^{n}x_k}{n\sum_{k=1}^{n}t_k^2-\left(\sum_{k=1}^{n}t_k\right)^2},\ b=\frac{\sum_{k=1}^{n}t_k^2\sum_{k=1}^{n}x_k-\sum_{k=1}^{n}t_k\sum_{k=1}^{n}t_kx_k}{n\sum_{k=1}^{n}t_k^2-\left(\sum_{k=1}^{n}t_k\right)^2}$$

Considering a set of temporal indicators: $(x_k)_{k=\overline{1,T}}$, from the above formulas, $t_k=k$, n=T (where T is the time period of the analysis), we obtain:

$$a_1=\frac{6}{T(T+1)(T-1)}\left(2\sum_{k=1}^{T}kx_k-(T+1)\sum_{k=1}^{T}x_k\right),\ b_1=\frac{2}{T(T-1)}\left((2T+1)\sum_{k=1}^{T}x_k-3\sum_{k=1}^{T}kx_k\right)$$

Let us now consider the growth indices corresponding to the values $(x_k)_{k=\overline{1,T}}$: $(I_k)_{k=\overline{2,T}}$, $I_k=\frac{x_k}{x_{k-1}}$, k≥2.

From above, we obtain:

$$a_2=6\frac{2\sum_{k=2}^{T}kI_k-(T+2)\sum_{k=2}^{T}I_k}{T(T-1)(T-2)}=6\frac{2\sum_{k=2}^{T}k\frac{x_k}{x_{k-1}}-(T+2)\sum_{k=2}^{T}\frac{x_k}{x_{k-1}}}{T(T-1)(T-2)},$$

$$b_2=2\frac{(2T^2+5T+6)\sum_{k=2}^{T}I_k-3(T+2)\sum_{k=2}^{T}kI_k}{T(T-1)(T-2)}=$$

$$2\frac{(2T^2+5T+6)\sum_{k=2}^{T}\frac{x_k}{x_{k-1}}-3(T+2)\sum_{k=2}^{T}k\frac{x_k}{x_{k-1}}}{T(T-1)(T-2)}$$

Let us consider the forecast at the time T+1 through the first relationship:

$$\tilde{x}=a_1(T+1)+b_1=\frac{2}{T(T-1)}\left(3\sum_{k=1}^{T}kx_k-(T+2)\sum_{k=1}^{T}x_k\right)$$

By the second relationship, we obtain:

$$\tilde{I} = a_2(T+1)+b_2 = \frac{2}{(T-1)(T-2)}\left(3\sum_{k=2}^{T} kI_k - (T+4)\sum_{k=2}^{T} I_k\right)$$

The predicted value will therefore be:

$$\tilde{\tilde{x}} = \tilde{I}\, x_T = \frac{2x_T}{(T-1)(T-2)}\left(3\sum_{k=2}^{T} k\frac{x_k}{x_{k-1}} - (T+4)\sum_{k=2}^{T} \frac{x_k}{x_{k-1}}\right)$$

The difference between the two forecasts is therefore:

$$\tilde{\tilde{x}} - \tilde{x} =$$

$$\frac{2}{T(T-1)}\left(3\sum_{k=1}^{T} kx_k - (T+2)\sum_{k=1}^{T} x_k\right) - \frac{2x_T}{(T-1)(T-2)}\left(3\sum_{k=2}^{T} k\frac{x_k}{x_{k-1}} - (T+4)\sum_{k=2}^{T} \frac{x_k}{x_{k-1}}\right)$$

Let us consider that x_T is provisionally calculated. Let therefore the function:

$$f(x_T) = \frac{2}{T(T-1)}\left(3\sum_{k=1}^{T} kx_k - (T+2)\sum_{k=1}^{T} x_k\right) - \frac{2x_T}{(T-1)(T-2)}\left(3\sum_{k=2}^{T} k\frac{x_k}{x_{k-1}} - (T+4)\sum_{k=2}^{T} \frac{x_k}{x_{k-1}}\right)$$

Noting:

$$\alpha = 2\frac{3\sum_{k=1}^{T-1} kx_k - (T+2)\sum_{k=1}^{T-1} x_k}{T(T-1)}, \quad \beta = 2\frac{3\sum_{k=2}^{T-1} k\frac{x_k}{x_{k-1}} - (T+4)\sum_{k=2}^{T-1} \frac{x_k}{x_{k-1}}}{(T-1)(T-2)}$$

we have:

$$f(x_T) = -\frac{4}{(T-1)x_{T-1}}x_T^2 + \left(\frac{4}{T} - \beta\right)x_T + \alpha$$

The second grade polynomial has $\Delta = \left(\frac{4}{T} - \beta\right)^2 + \frac{16\alpha}{(T-1)x_{T-1}}$. If $\Delta \geq 0$ then:

$$f(x_T) \geq 0 \Leftrightarrow$$

$$x_T \in \left[\frac{\left(\frac{4}{T}-\beta\right) - \sqrt{\left(\frac{4}{T}-\beta\right)^2 + \frac{16\alpha}{(T-1)x_{T-1}}}}{\frac{8}{(T-1)x_{T-1}}}, \frac{\left(\frac{4}{T}-\beta\right) + \sqrt{\left(\frac{4}{T}-\beta\right)^2 + \frac{16\alpha}{(T-1)x_{T-1}}}}{\frac{8}{(T-1)x_{T-1}}}\right]$$

Therefore, if x_T belongs to the above range, then the index-based forecast will provide a higher value than the one based on absolute data.

3. Case Study

Consider the absolute values of Romania's GDP (adjusted quarterly) between Q1-2015 - Q1-2017.

Table 1

Quarter	Absolute GDP	GDP Index
QIV 2014	34009.2	-
QI 2015	34459.7	1.013
QII 2015	34439.8	0.999
QIII 2015	35093	1.019
QIV 2015	35443.6	1.010
QI 2016	35853.2	1.012
QII 2016	36399.3	1.015
QIII 2016	36650.7	1.007
QIV 2016	37233.1	1.016
QI 2017	37897.2	1.018
QII 2017	**38540.1**	**1.017**

Source: INSSE

where the data in the last line is provisional but useful (in our case) to see the quality of the forecast for the data in the analyzed period. We also mention that QIV 2014 data was only considered for the calculation of the growth index.

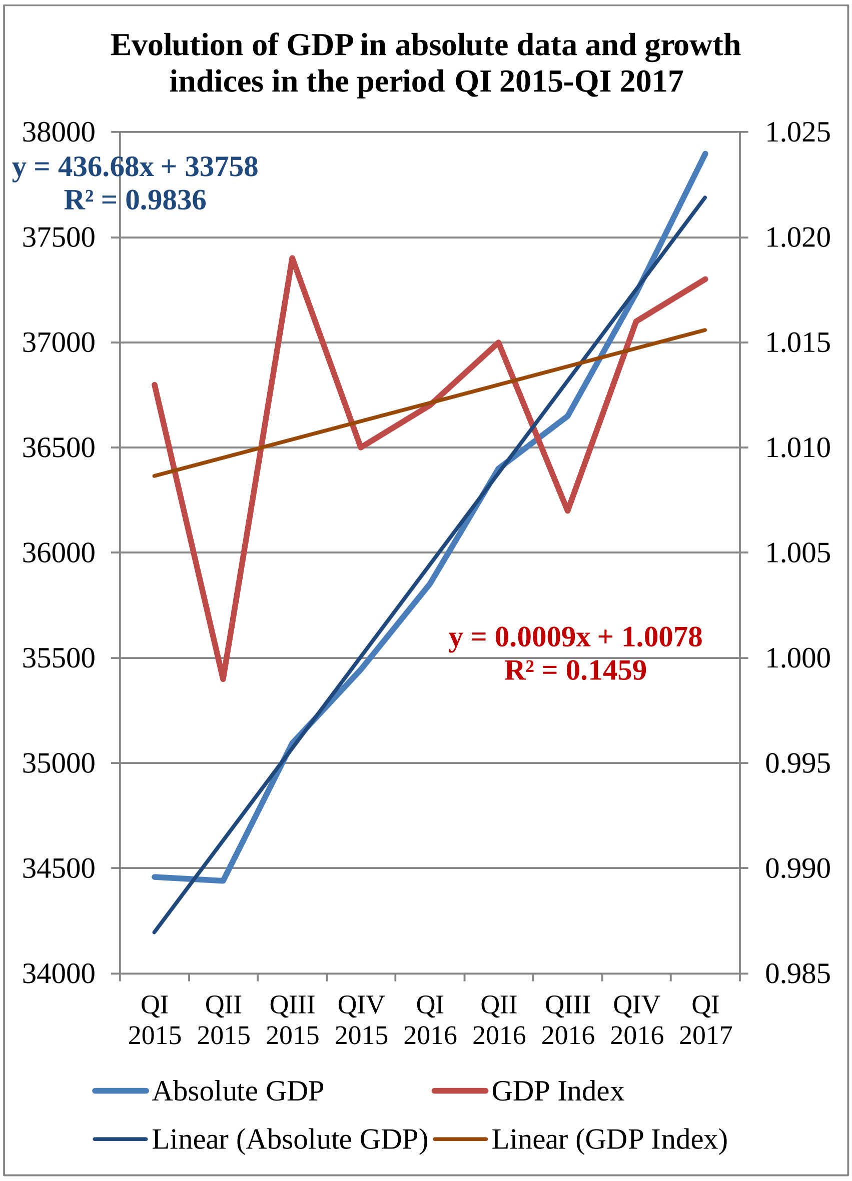
Evolution of GDP in absolute data and growth indices in the period QI 2015-QI 2017
38000
37500
37000
36500
36000
35500
35000
34500
34000
1.025
1.020
1.015
1.010
1.005
1.000
0.995
0.990
0.985
y = 436.68x + 33758
R² = 0.9836
y = 0.0009x + 1.0078
R² = 0.1459
QI 2015
QII 2015
QIII 2015
QIV 2015
QI 2016
QII 2016
QIII 2016
QIV 2016
QI 2017
Absolute GDP
GDP Index
Linear (Absolute GDP)
Linear (GDP Index)

Figure 1.

Regression analysis for absolute data reveals:

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.991778					
R Square	0.983624					
Adjusted R Square	0.981285					
Standard Error	164.9608					
Observations	9					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	11441540	11441540	420.4582	1.65E-07	
Residual	7	190484.5	27212.07			
Total	8	11632025				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	33320.97	139.1072	239.5345	5.83E-15	32992.03	33649.9
X Variable 1	436.6833	21.29635	20.50508	1.65E-07	386.3255	487.0412

therefore the regression is: GDP=436,6833·Quarter+33320,97.

The forecast for QII 2-17 is therefore: 38124.5 with a relative error: -1.08%.

Regression analysis for index data reveals:

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.3819676					
R Square	0.1458993					
Adjusted R Square	0.0238849					
Standard Error	0.0061391					
Observations	9					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	4.50667E-05	4.50667E-05	1.19575472	0.31036807	
Residual	7	0.000263822	3.76889E-05			
Total	8	0.000308889				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.0069111	0.005176967	194.498266	2.50677E-14	0.994669529	1.0191527
X Variable 1	0.0008667	0.000792558	1.0935057	0.31036807	-0.0010074	0.0027407

therefore the regression is: GDP%=0.0008667·Quarter+1.0069111.

The forecast for QII 2-17 is therefore: 1.016 or for absolute data: 38503.5 with a relative error: -0.09%.

The interval for a higher forecast in the case of indices is: [-51219.3, 37480,3].

4. Conclusions

As a result of the above analysis, we find that for the case of Romania, the regression based on growth indices provides conclusions closer to reality than relative absolute values.

5. References

Ioan, C.A. (2011). *Elements of econometrics*. Galați: Zigotto Publishers.

Ioan, C.A. & Ioan G. (2011). *n- Microeconomics*. Galați: Zigotto Publishers.

Ioan, C.A. & Ioan, G. (2013). The LSM adjustment method of time series. *Euroeconomica*, nr. 3(32), pp. 150-155.