Theoretical and Applied Economics
Volume XXI (2014), No. 4(593), pp. 53-66

Using the regression model for the portfolios analysis and management

Constantin ANGHELACHE
Bucharest University of Economic Studies, Romania
"Artifex" University of Bucharest, Romania
actincon@yahoo.com
Mădălina Gabriela ANGHEL
"Artifex" University of Bucharest, Romania
madalinagabriela_anghel@yahoo.com

Abstract. *In the frame of this article we submitted the practical possibilities to analyse the stock market activity in Romania, by means of the linear regression model. Thus, the study is focusing on the existing correlations between the yield of a portfolio formed by ten financial assets issued by companies quoted on the main market at the Bucharest Stock Exchange and the overall evolution of the capital market in Romania, expressed with the help of the Bucharest Exchange Trading index. A linear regression model implies that the methods used for estimating the two parameters, the methods applied for testing the properties of the regression model estimators as well as the main aspects concerning the utilization of the regression model in making predictions are well known.*

Keywords: linear regression, financial instruments portfolio, BET index, yields, correlation coefficient.

JEL Classification: C23, E44, G11.
REL Classification: 11B.

1. Introduction

The evolution of today's economy, characterized by the increase of the complexity and intensity of all its specific processes, made it more and more frequently needed the setting up of specific applications meant to secure an as much as true reflection of the various situations which may occur in the daily activity, as well as the favourable or unfavourable influences which those ones may induce on the entire analysed system. This statement grounded the researchers' concern as to reflect as exactly as possible, the relations existing between the various evolutions of the human life – in general terms – and its economic side – in particular terms-. Such econometric models can be largely utilized in the analysis performed on the financial activities and, implicitly, on the capital market.

The first attempts to formalize a coherent structure for conceiving an econometric model is dated back the mid-last century when these kind of attempts could have been found out, for instance, in the works signed by Walters, *An introduction to econometrics* or Frank, *Statistics and econometrics.* These have been taken over studied thoroughly at the beginning of the years 2000 by Gujarati, the one who, in his work *Basic econometry*, Fourth edition (2005) is formulating a set of six stages meant to conceive and implement an econometric model.

The drawing of econometric models applicable directly to activities specific to the financial markets and, mainly, to the capital market, took place during the second half of the last century when, based on the existing theoretical models, the first models strictly applicable to the financial and capital investment domain, have been created (Milles, *The econometric modelling of financial time series*, 1993; Diebild, *Elements of forecasting*, 2002).

2. The simple regression model

The simple regression model is representing a correlation where the dependent out coming variable, *Y,* is linearly dependent on the explicative, factorial variable, *X,* but also influenced by the residual ε, as follows:

$Y_i = b + aX_i + \varepsilon_i$, for all values „*i*”.

The two parameters are estimated through the data series set up for the two variables. The estimators for the two parameters are defined as $\hat{b}$ and $\hat{a}$. The model parameters are established in a stochastic manner, on the basis of the estimators.

The residual variable ε is normally distributed, having the mean 0 and a constant dispersion and being included in the model because of the fact that in the economy a functional linear dependence between two variables is not always to be met but merely one of probabilistic kind, as the data series are jeopardized by measurement errors with influence on the estimates for the two parameters while the data series are set up through observations on a number of samples.

When defining the linear regression, there are a number of hypotheses to be considered. The estimated value of the out coming variable, of the estimators of the model parameters and their properties depend on characteristics of the independent variable and residual variable properties, so that the above mentioned hypotheses are referring to the variables which define the regression mode as well as to the residual variable.

1) The data series are not affected by recoding errors. The hypothesis is postulating the characteristics of the series of values which are utilized for estimating the parameters. The estimation of the parameters is achieved on the basis of a values sample $(x_i, y_i)_{i=1,n}$ which represents values for the two variables. In the case of the classical model of regression, it is considered that the values of the factorial characteristic are determinist (fix values) while those of the out coming characteristic are stochastic. The values of the factorial characteristic are non-stochastic if for each value of this characteristic there is a corresponding family of values of the out coming characteristic. For each value x_i of the factorial characteristic, there is an average of the family of the out coming characteristic which is calculated and the series of values $E[Y|X = x_i], i = 1, n$ is set up.

For each fixed value of the factorial characteristic, the residual variable is of mean zero, namely: $E[\varepsilon_i|X = x_i] = 0$, for each i

Based on this statement it is resulting that he other un-recorded factors, excepting the factorial characteristic, have no systematic influence on the mean of the out coming characteristic. If the hypothesis is met by the linear regression model, it is written: $E[Y|X = x_i] = b + ax_i$

2) The homoscedasticity hypothesis – the residual dispersion is constant. This property shows that the conditional distributions ($y_i/x=x_i$) have the same dispersion, represented by the following equality: $\mathrm{var}[\varepsilon_i|X = x_i] = \sigma_\varepsilon^2$, is constant for any i.

If the residual variables do not meet this property, the regression model is heteroscedastic and the residual variables have different variances:

$$\mathrm{var}[\varepsilon_i|X = x_i] = \sigma_i^2$$

3) The lack of the residuals correlation. This property is expressing the fact that between the residual terms the covariance phenomenon is not showing up, it being written in form of:

$\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$, for any $i \neq j$. If the residual variable meets the hypotheses „2" and „3", the relation is resulting:

$$\mathrm{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, i \neq j \\ \sigma_\varepsilon^2, i = j \end{cases}$$

A quite different situation happens when the residual variable is submitting an auto-correlation of first degree, namely:

$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, where u_t is white noise.

4) Non-correlation of the residual variable with the independent variable. When this hypothesis is met, it is written:

$\mathrm{cov}(X, \varepsilon_j) = 0$, for any j, which means that an increase of the values of the factorial variable is not leading automatically to an increase of the residual variable values. The residual values are distributed upon a normal repartition of mean 0 and dispersion σ_x^2 . Then, for the residual variable, we shall write that $\varepsilon_i \in N(0, \sigma^2)$.

Analysis and interpretation of the residual variable

In the frame of the linear regression model, ε_i is representing the residual variable while e_i or $\hat{\varepsilon}_i$ is measuring the gap between the real value y_i and the adjusted value through the regression model. We define $e_i = y_i - \hat{y}_i$.

In order to set up an estimate for the two parameters of the regression line, an estimate for the residual variable dispersion is established, with the following properties:

1. For the series of gaps, the sum of its terms equals to zero. In this case, the following equality is written $\sum_i \varepsilon_i = 0$. The calculation formula for the gap, the estimator formula for the free term and the fact that the sum of the gaps of the series terms as against the mean is zero, are factors being taken into consideration, where off it is resulting:

$$\sum_i \varepsilon_i = \sum_i \left(y_i - \hat{b} - \hat{a}x_i\right) = \sum_i \left(y_i - \bar{y}\right) - \hat{a}\sum_i \left(x_i - \bar{x}\right) = 0$$

The property is not valid for the series of the residual variables but only for the situation when the hypothesis $E(\varepsilon_i)=0$ is met for all indices *i*. We shall express the gap of a value as against the adjusted value in connection with the residual variable, the following equalities resulting:

$$e_i = y_i - \hat{y}_i = (b-\hat{b}) - (a-\hat{a})x_i + \varepsilon_i$$

Considering the relations $\hat{b} = \overline{y} - \hat{a}\overline{x}$, $b = \overline{y} - a\overline{x} - \varepsilon$ and $a - \hat{a} = \sum_i w_i\varepsilon_i$, we get:

$$e_i = \varepsilon_i - \overline{\varepsilon} - (x_i - \overline{x})\sum_i w_i\varepsilon_i$$

2. The dispersion of the residual variable for the classical regression model (the parameters are estimated through the least squares method) is estimated through the relation:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-2}\sum_i e_i^2$$

We shall take into account the hypotheses grounding the classical regression model $E(\varepsilon_i)=0$, $\mathrm{var}(\varepsilon_i)=\sigma_\varepsilon^2$, $\mathrm{cov}(\varepsilon_i,\varepsilon_j)=0, \forall i \neq j$ and $\mathrm{cov}(X,\varepsilon_i)=0$ for all indices *i*. Calculating e_i^2 and applying the mean operator, the following equality is resulting for each index *i* :

$$E(e_i^2) = \sigma_\varepsilon^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_\varepsilon^2(x_i-\overline{x})^2\sum_k w_k^2 - 2E\left[\left(\varepsilon_i - \overline{\varepsilon}\right)\cdot(x_i-\overline{x})\sum_k w_k\varepsilon_k\right]$$

$$= \sigma_\varepsilon^2 + \frac{\sigma_\varepsilon^2}{n} + \sigma_\varepsilon^2(x_i-\overline{x})^2\frac{1}{\sum_k\left(x_k-\overline{x}\right)^2} -$$

$$-2\sigma_\varepsilon^2\left[\left(x_i-\overline{x}\right)w_i - \frac{1}{n}(x_i-\overline{x})\sum_k w_k + (x_i-\overline{x})w_i\right].$$

Using the addition operator and considering the properties of the se

series $(w_i)_{i=\overline{1,n}}$, we get:

$$E(\sum_i e_i) = (n-2)\sigma_\varepsilon^2$$

We establish the estimator of the residual variation, which is to compare with the estimator $\hat{a}^2 = \frac{\sum_i e_i^2}{n}$, as non-shifted estimator.

For the linear regression model, the size of the dispersion of the gaps series $(e_i)_{i=\overline{1,n}}$ is higher to the extent the series of the out coming characteristic values is larger but smaller if the dependence between the two characteristics is stronger.

Between the dispersion of the gaps series, of the endogenous characteristic values and the linear correlation coefficient the equality: $\sigma_e^2 = (1-r^2)\sigma_y^2$ is confirmed.

We shall demonstrate the last relation taking into consideration the fact that, in case there is a linear dependence between the two variables, $r^2 = \frac{SPE}{SPT}$, where $SPT = SPR + SPE$. Since $1-r^2 = \frac{SPR}{SPT}$ and out of the formula of calculation for the residual variable dispersion, we get:

$$\left(1-r^2\right)\sigma_y^2 = \frac{SPR}{SPT}\frac{SPT}{n} = \sigma_\varepsilon^2$$

The gaps dispersion has been calculated through the formula:

$$\sigma_\varepsilon^2 = \frac{1}{n}\sum_i e_i^2$$

Out of the residual normality hypothesis, it is resulting:

$$\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}(n-2) \to x_{n-2}^2$$

A confidence interval for the residual variable dispersion can be set up if a significance threshold α, is fixed, the confidence interval being:

$$(n-2)\frac{\hat{\sigma}_\varepsilon^2}{\lambda_2^2} < \sigma_\varepsilon^2 < (n-2)\frac{\hat{\sigma}_\varepsilon^2}{\lambda_1^2}$$

λ_1^2 and λ_2^2 are values provided by the repartition λ^2 for *n-2* freedom degrees and a set up significance threshold α.

By graphical plotting of the coordinates points $(y_i, e_i)_{i=\overline{1,n}}$, we can verify empirically whether the homoscedasticity hypothesis is met. By graphical plotting of the data series there are two situations occurring: if the point are defining the *point cloud,* then the homoscedasticity hypothesis is not met; while when the points are disposed in the form of a horizontal strip, the hypothesis is valid for the data series $(x_i, y_i)_{i=\overline{1,n}}$.

3. The application of the regression model on the capital market in Romania

In order to take advantage of the submitted theoretical aspects, we used the regression econometric model meant to evaluate the relation between the yield of a portfolio of financial instruments and the overall evolution of the capital market.

A significant element which has to be known by any portfolio manager consists of the selection of the component financial instruments. Therefore, we submitted to the fundamental analysis a number of 20 companies quoted at the Bucharest Stock Exchange (see table no.1), which equities are presently transacted on the main REGS market, both in the category I of financial instruments and the II one. The analysis was based on information out of the synthesis accounting documents, respectively the accounting balance sheet and the profit and losses account. The selected companies as subject of the analysis are part of a large range from the pont of view of their field of activity as a portfolio diversified on more economic sectors will be less risky as comparatively with a portfolio including assets from a single branch of activity.

Table 1. *Companies submitted to the fundamental analysis*

No	Company's name	Symbol	Category	Field of activity
1.	Aerostar	ARS	II	Manufacturing of aircrafts and space craft
2.	ALRO	ALR	I	Aluminium metallurgy
3.	Antibiotice	ATB	I	Manufacturing of basic pharmaceutics
4.	Banca Comercială Carpatica	BCC	I	Monetary intermediation activities
5.	Banca Transilvania	TLV	I	Monetary intermediation activities
6.	Biofarm	BIO	I	Manufacturing of pharmaceuticals
7.	BRD-Groupe Societe Generale	BRD	I	Monetary intermediation activities
8.	BERMAS	BRM	II	Beer manufacturing
9.	Boromir Prod Buzau (Spicul)	SPCU	II	Manufacturing of bread: cakes and fresh pastry
10.	Calipso Oradea	CAOR	II	Bars and other activities meant to serving drinks
11.	Electromagnetica Bucureşti	ELMA	I	Manufacturing of instruments and devices for measuring, checking, control, navigation
12.	Farmaceutica Remedia	RMAH	II	Wholesales of pharmaceuticals
13.	OMV Petrom	SNP	I	Extraction of crude oil
14.	Prodplast	PPL	II	Manufacturing certain products of plastics
15.	SIF Muntenia	SIF 4	I	Other financial intermediations n.c.a.
16	C.N.T.E.E. Transelectrica	TEL	I	Transport of electric energy
17.	S.N.T.G.N. Transgaz	TGN	I	Transports by pipelines
18.	Turism Felix Băile Felix	TUFE	II	Hotels and other similar accommodation facilities
19.	SIF Banat Crişana	SIF 1	I	Other financial intermediations n.c.a.
20.	SIF Oltenia	SIF 5	I	Other financial intermediations n.c.a.

Source: self-systematization.

Out of the 20 analysed companies, we have selected a number of 10 companies to be included in the portfolio, namely: five commercial companies, two national companies, two financial investments companies and a banking company, for which the daily evolution of the prices has been analysed over the period 01.01.2012–31.12.2012. The companies have been selected according to rigorous criteria by granting them marks which have been turned into scores afterwards, whereof the final classification:

Table 2. *Selected companies which equities are considered to enter into the portfolio*

No	Company	Symbol
1.	Banca Transilvania	TLV
2.	Bermas	BRM
3.	Biofarm Bio	BIO
4.	Electromagnetica	ELMA
5.	OMV Petrom	SNP
6.	SIF Oltenia	SIF5
7.	SIF Muntenia	SIF4
8.	Transgaz	TGN
9.	Turism Felix	TUFE
10.	Transelectrica	TEL

Source: self-calculations.

Further on, we have analysed the data concerning the daily evolution of the prices for the ten financial assets quoted on the main market of the Bucharest Stock Exchange in the year 2012 (366 calendar days/ 250 actual transacting sessions – the data being used are daily surveys over a time horizon between January 1^{st} 2012 and December 31^{st} 2012, respectively 250 surveys, excepting week-end days and legal holidays, the data being processed in form of yields), as well as those directly connected with the BET index evolution (the reference index of the capital market in our country) during the considered period.

Based on this information, we have set up the yield of the portfolio which includes in equal proportions the ten types of equities previously mentioned (equally weighted portfolio) as well as the yield of the BET index for the one year period submitted to the (see Table 2).

When utilizing this model, the data processed by the help of the Microsoft Excel data processor have been imported into a newly created application with the support of the Eviews informatics programme. We have applied statistical tests for each and every of the two series of data emphasized by the graph of the two indicators, submitted bellow:

- For the data series concerning the portfolio yield

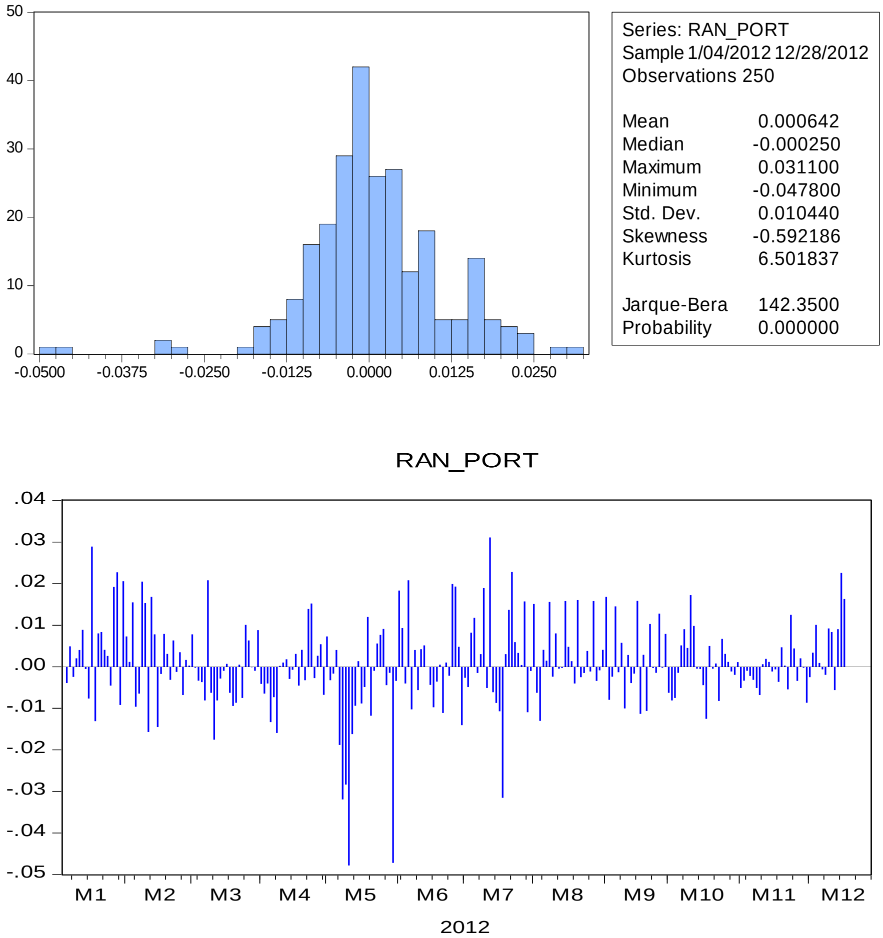

Figure 1. *Statistical tests for portfolio performance*

- For the data series concerning the BET index variation

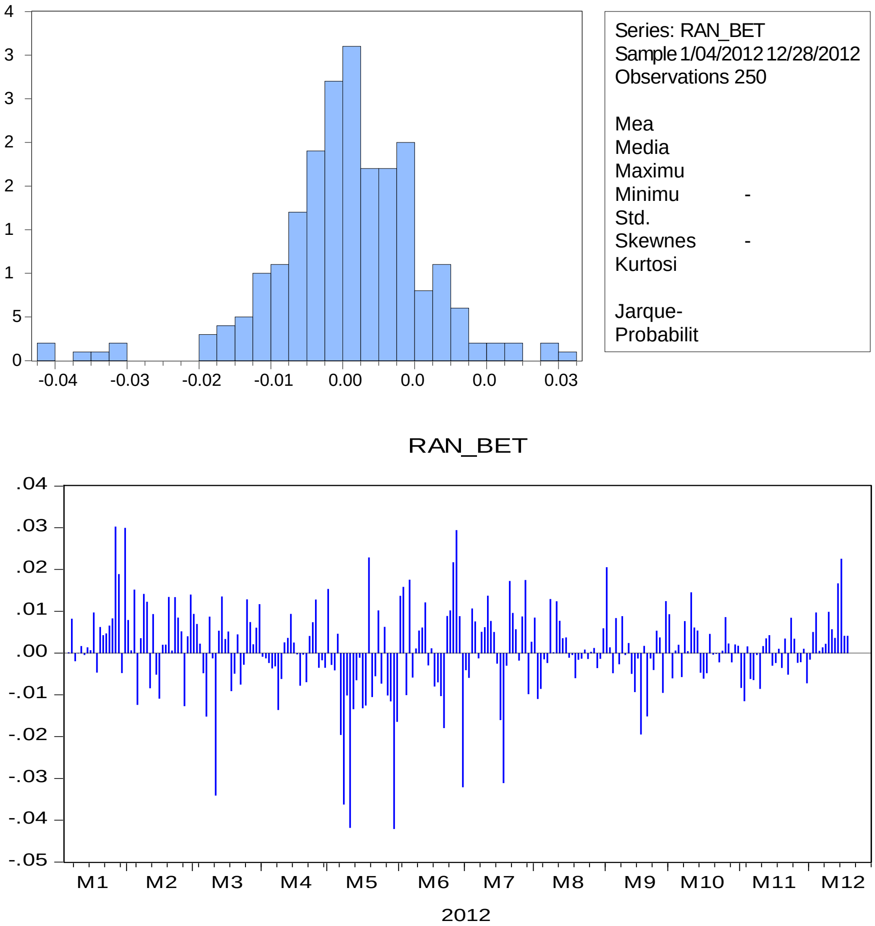

Figure 2. *Statistical tests for BET index*

Out of the graphs above it can be noticed that the repartition of the two data series is a very similar one which means that there is a pronounced dependence between the two sizes. This idea can be sustained also by the argument of the graph representation of the evolution of the two analysed indicators.

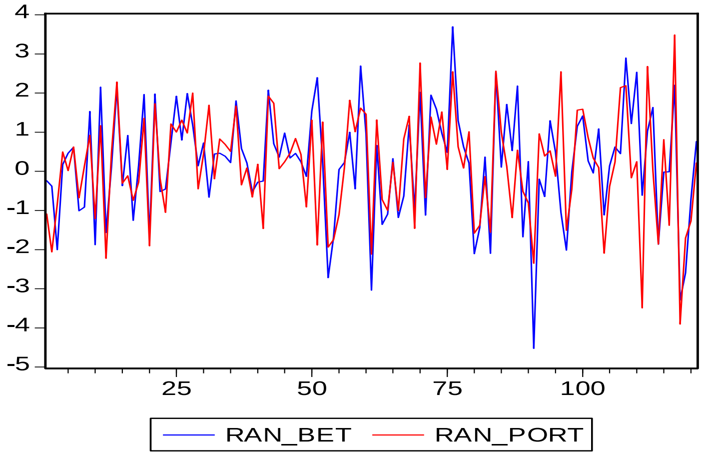

Figure 3. *Evolution of BET index and portfolio performance*

From the above graph it is obvious that the evolution of the two data series is more than similar which leads to the conclusion that there is a pronounced dependence between the two analysed indicators.

In order to established the type of the econometric model to be utilized in order to define the existing relation between the portfolio formed up by the ten assets previously mentioned (a a dependent variable) and the overall evolution of the capital market (shown up by the BET index evolution – as explicative variable) a graph representation of the data series has been achieved, by drawing up the relative regression line.

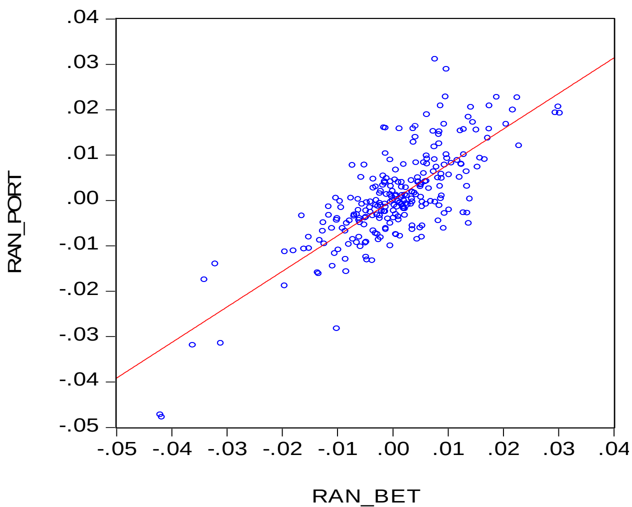

Figure 4. *Relative regression line*

On the basis of the out coming observations, the simple linear regression model has been defined as the form of:

RAN_PORT = α + β x RAN_BET + ε

where:
RAN_PORT = the value of the yield for the considered portfolio of assets (the dependent variable)
RAN_BET = the value of the yield for the BET index (the independent variable)
α, β = the parameters of the linear regression model
ε = the residual value of the regression model.

With the help of the Eviews informatics programme, applying the method of the smallest squares for the estimation, the parameters of the regression model above mentioned have been estimated and specific tests concerning the validity of such a model have been performed.

Dependent Variable: RAN_PORT
Method: Least Squares
Date: 11/20/13 Time: 18:41
Sample (adjusted): 1/04/2012 12/18/2012
Included observations: 250 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.61E-05	0.000415	0.159076	0.8737
RAN_BET	0.783438	0.039942	19.61420	0.0000

R-squared	0.608039	Mean dependent var	0.000642
Adjusted R-squared	0.606459	S.D. dependent var	0.010440
S.E. of regression	0.006549	Akaike info criterion	-7.210973
Sum squared resid	0.010637	Schwarz criterion	-7.182802
Log likelihood	903.3717	Hannan-Quinn criter.	-7.199635
F-statistic	384.7167	Durbin-Watson stat	2.104345
Prob(F-statistic)	0.000000		

Estimation Command:
=========================
LS RAN_PORT C RAN_BET

Estimation Equation:
=========================
RAN_PORT = C(1) + C(2) **x** RAN_BET

Substituted Coefficients:
=========================
RAN_PORT = 6.60561413185e-05 + 0.783437809153 **x** RAN_BET

The regression model describing the linear dependence between the yield of the portfolio formed up of ten financial assets and the BET index evolution can be transcribed as follows:

RAN_PORT = 0,0000660561413 + 0.783438 x RAN_BET + ε

The value recorded by Prob (0.0000) indicates the fact that the variable is significant from statistical point of view. Thus, we can conclude that the influence which the general evolution of the Romanian stock exchange market has upon the considered portfolio of financial instruments is a significant one. The connection between the two variables is a direct one which shows that for an increase of 1% of the BET index yield, the portfolio yield records an increase of 78.34%.

The loose term of the model is presenting a very low value which creates the image of an insignificant influence of the factors not depending on the stock exchange market dynamics on the yield of the analysed portfolio.

The value being recorded by R-squared shows that, at a ratio of 60.80%, the portfolio yield is explained by the BET index yield the difference up to 100% representing the influence of other factors not included in the present model.

4. Conclusions

In order to study the dependence between the yields of an equi-weighted portfolio, formed up by ten equities quoted on the main BVB market and the evolution of the BET index, a number of aspects out coming from the practical activity have been underlined, meant to offer a complete image of the management activity of the financial instruments portfolio. As far as the validity of this econometric model is concerned, we can notice the fact that the tests being automatically done by the Eviews specialized informatics programme have confirmed its sturdiness. In this respect it is enough to mention the values recorded by the test R^2, F-statistic (384.7167) or Prob(F-statistic) (0.00). All these are fully confirming the validity of the considered model and lead to the ideas that it can be successfully utilised in the future for estimating the value of the considered portfolio yield depending on the evolution of the main index of the Bucharest Stock Exchange.

References

Andrei, T., Bourbonnais, R. (2008). *Econometrie*, Editura Economică, Bucureşti

Andrei, T., Stancu, S., Iacob, A.I. (2008). *Introducere în econometrie utilizând Eviews*, Editura Economică, Bucureşti

Anghel, M.G. (2008). "Utilizarea modelului de regresie în analiza pieţei de capital", *Revista Română de Statistică – Supliment*

Anghel, M.G. (2013). "Identification of financial instruments – important step in building portfolios", *Revista Română de Statistică*, no. 9

Anghel, M.G. (2013). *Modele de gestiune şi analiză a portofoliilor*, Editura Economică, Bucureşti

Anghelache C. (2013). *Elemente de econometrie teoretică*, Editura Artifex, Bucureşti

Anghelache C., Anghelache G.V. (2009). *"Utilization of the econometric models in the analysis of the assets rates"*, *Metalurgia international*, vol. XIV, nr. 9 Special Issue, Editura Ştiinţifică, F.M.R.

Anghelache, C., Dinu, M., Barbu, C.M. (2010). "Estimation of the regression functions transformation", *Revista Metalurgia International* vol. XV, no. 1

Dougherty, C. (2008). *Introduction to econometrics.* Fourth edition, Oxford University Press

Guijarati, D. (2005). *Basic Econometrics*, The McGraw-Hill Companies

Wooldrige, J. (2006). *Introductory Econometrics. A modern approach,* Second edition, MIT Press

www.bvb.ro (Bursa de Valori Bucureşti)