

ECONOMETRIC ANALYSIS REGARDING THE DEPENDENCE OF THE FUTURES PRICE, THE UNDERLYING ASSET AND THE ROBOR

Assist. Mircea Gabriel Ciolpan PhD Student
University of Financial Banking
Faculty of Financial Management
Bucharest, Romania
Prof. Jenica Popescu Ph. D
University of Craiova
Faculty of Economics and Business Administration
Craiova, Romania

Abstract: : The derivatives market is an important barometer for the investments return and the national economy trends. To check the efficiency of the Romanian derivatives market represents a milestone for the professional investor. How information contributed to the futures price represented a guarantee for the investors that the derivatives follow the reality of the economy. As well, we considered that the banking interest rate represented a decisive aspect for making an investment decision. By marking the level of profit for the riskless investment an investor could determine the risk and the necessary return for any other market. The multifactorial regression is an econometrical model suitable for the analysis of the links and dependencies of the above mentioned aspects.

JEL classification: C01, E44, G13, G24

Key words: critical; derivatives, futures price, market efficiency, interest rate, multifactorial regression

1. INTRODUCTION

Understanding the fluctuations of the course of the futures contracts is a very important aspect in reaching the goal of the actions carried out on the derivatives financial markets. The importance of **underlying assets** is decisive, as a financial derivative cannot exist without the support which has determined it. The support assets also have a very important role in mathematical calculus in the assessment of the price of futures contracts. Also, on maturity the price of futures is aligned to the spot price for the transacted asset. The importance of the correlation with the basis asset is very important in the context of derivative financial products, since it checks the way information are sent at the level of the Romanian capital market as well as level o effectiveness of the latter. The effectiveness of the market of derivative financial products is indicated by the following aspects: the capacity to anticipate the price, the capacity to offer possibilities of risk de management and the capacity to efficiently reallocate the resources.

Through this multiple regression we intended to establish the effectiveness of the futures market in Romania and for that purpose we used the most transacted derivate financial product of the Sibiu stock exchange, the futures contract on the Dow Jones index.

The first variable inserted in the mathematical model was that of the basic asset, the course of the Dow Jones index on the American market. The attractiveness of these contract futures was very high and the large volume of operations allowed us to make some econometric calculi.

The second variable, the level of the interbank interest rate ROBOR was taken into account as any investment is based on profit. We have considered that it is necessary to correlate the placements in derivative financial instruments with the level of the interest rate. We have used the ROBOR because the level of the passive interest for the funds attracted by the bank from the non/banking clients is based on this interest. The fluctuation of ROBOR is transposed to a time interval in the changes of the deposit interest rate.

2. THE METHODOLOGY USED. THE REGRESSION AND CORRELATION METHOD

For a deeper analysis of the connection and interdependence of economic-social phenomena, elementary statistic methods are often insufficient. Therefore, the analysis of the connections between the price of the futures contract, the price of the basic assets and the interest rate can be exemplified by means of the *correlation and regression method*.

By determining the most influential factors in the development of the analyzed phenomenon, we can design and take practical measures in order to provide for optimal conditions for its development, intensification of the influence of positive factors, eliminate or alleviate the influence of unfavorable factors. In case of complex connections, where together with the dependent variable there appear several independent variables, the calculation of correlation cannot however be limited only to variables in pairs, it must also include other independent variables, with significant influence on the analyzed dependent variable. Also, within the multiple connections, factorial variables have different influences on the resulting variable, some exert an important influence on the effect phenomenon and must be taken into account in the regression and correlation calculations, while others have a less important influence and may be neglected. The effects of the correlation methods are the simplification of calculations and conclusions, as it is very difficult to quantify the aggregate of all the causal factors acting on a phenomenon of economic-social process.

Therefore, in the second step of the statistic approach the main dependences between the above-named variables are characterized quantitatively and qualitatively. Thus, for a more complex analysis of the connection and interdependence of the economic-social phenomena, elementary statistical methods are often insufficient. At the same time, **multi-factor regression** allows the analysis of indicators which characterize the investment activity due to several factors and, implicitly, the comparison of typical values in order to establish whether there are any significant differences between them.

3. MAIN RESULTS. DEFINING THE MODEL OF MULTI-FACTOR REGRESSION

In order to render the model of multiple linear regression we have used data concerning Romania for the rate of futures on the Dow Jones index, transacted at Sibex, the rate of the base asset, represented by the values of the Dow Jones index as published by the Financial Times, and the quotation of the inter-banking interest rate ROBOR as

published by the BNR, for the period 22.03.2011-29.02.2012, thus implementing the extracted data in Eviews.

The specification of an econometric model also supposed the selection of a mathematical function ($f(x)$) which can be used to describe the connection between the variables.

The form of the model of multiple linear regression is:

$$curs_futures_t = \alpha + \beta_1 * curs_activ_baza_t + \beta_2 * Robor_t + \varepsilon_t; \quad t=1,2,\dots,n.$$

After assessing the parameters in Eviews, the following equation was obtained:

$$LS \text{ Curs_futures} = C(1) + C(2) * curs_activ_baza + C(3) * Robor$$

The results obtained are synthesized in the following table:

Table no. 1: Results of the Application of the multi-factor model

LS CURS_FUTURES=C(1)+C(2)*CURS_ACTIV_BAZA+C(3)*ROBOR				
EstimationEquation:				
CURS_FUTURES =-550.96+85.137* CURS_ACTIV_BAZA +44.639* ROBOR				
CURS_FUTURES=C(1)+C(2)*CURS_ACTIV_BAZA+C(3)*ROBOR				
SubstitutedCoefficients:				
Dependent Variable: CURS_FUTURES				
Method: LeastSquares				
Sample: 224				
Included observations: 224				
CURS_FUTURES=C(1)+C(2)*CURS_ACTIV_BAZA+C(3)*ROBOR				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-550.96	2327.8	-0.23668	0.8156
C(2)	85.137	33.444	2.5457	0.0203
C(3)	44.639	117.70	0.37927	0.7089
R-squared	0.28364	Mean dependent var		4757.2
AdjustedR-squared	0.20405	S.D. dependent var		2434.1
S.E. of regression	2171.6	Akaikeinfocriterion		18.336
Sumsquaredresid	84889076	Schwarz criterion		18.485
Log likelihood	-189.52	Durbin-Watson stat		1.8463

From the Output table, as generated by Eviews, we note the estimates of the coefficients, their standard errors, the value of the Student test, as well as the corresponding p value.

Testing the parameters of the Multi-factor Regression Model

Done by means of the Student Test.

➤ The Student Test

We have the hypothesis:

- Null hypothesis, $H_0 : \alpha = 0$ or $\beta_t = 0, t = 1,2$
- Alternative hypothesis, $H_1 : \alpha \neq 0$ or $\beta_t \neq 0, t = 1,2$

Thus the coefficient of the rate of the basic asset in the regression model is $\hat{\beta}_1 = 85.13$, standard error $SE(\hat{\beta}_1) = 33.4$, and the Student test $\hat{t}_1 = 2.54$, thus calculated:

$\hat{t}_1 = \frac{\hat{\beta}_1}{SE(\hat{\beta})_1} = \frac{\text{Coefficient}}{\text{Std.Error}}$; p value = 0.02, which shows us that the rate of the basic asset is also an important factor of influence on the rate of futures.

The coefficient of the ROBOR quotation is $\hat{\beta}_2 = 44.63$, standard error $SE(\hat{\beta}_2) = 117.69$, and the test $\hat{t}_2 = 0.37$. The probability value here is 0.70, so the Robor quotation does not represent a significant component for the rate of futures in the assessed regression model.

The coefficient of the free term in the regression model is $\hat{\alpha} = -550.96$, the standard error $SE(\hat{\alpha}) = 2327.83$, the t test expressed by $\hat{t}_\alpha = -0.23$, with the value of the probability p of 0.81. Thus neither is the free term significant for the chosen regression model.

The Determination Ratio (R^2) shows the percentage which explains the influence of the significant factors. It is calculated as: $R^2 = \frac{SSR}{SST} = 1 - \frac{\sum \varepsilon_i^2}{SST}$ It is used for the assessment of the quality of the model. This can only have values situated within the interval [0,1]. The closer the values are to the value 1, the better the model is. The value it takes here is 0.283 and thus we can say that the regression model is not that good. Approximately 28.3% of the variation of the futures rate is explained by means of the chosen multiple linear regression model.

$$\text{curs_futures}_t = \hat{\alpha} + \hat{\beta}_1 * \text{curs_activ_baza} + \hat{\beta}_2 * \text{cot atia_Robor}$$

$$\text{curs_futures}_t = -550.96 + 85.13 * \text{curs_activ_baza}_t + 44.63 * \text{cot atia_Robor}_t$$

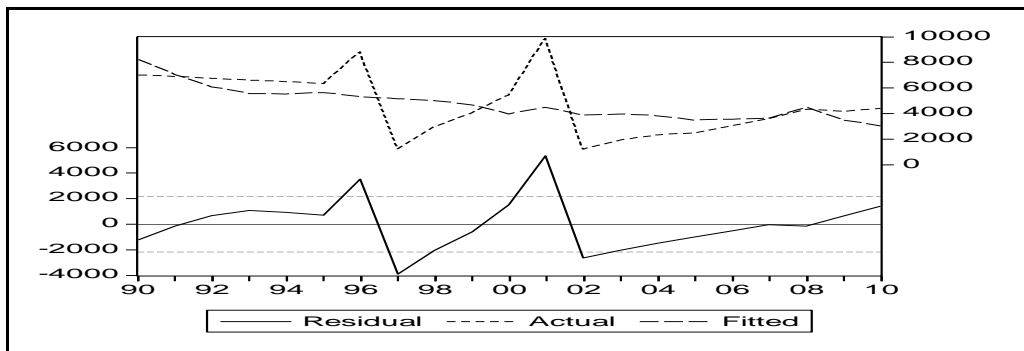


Figure no. 2
Graph of Values of Residual Variable,
Value of Dependent and Fitted Variable

The graphs explaining the values of the residues were also extracted from Eviews as follows: the first line of the graph shows the values taken by the calculated residues, taking as interval (-2000; 2000), and the second also shows the graphs according to which the residues were calculated, that is both the graph of the rate of futures in the source table, represented by the Actual line as well as the graph of the rate of futures taken as fitted value, represented by the Fitted line. The line of residues is represented precisely by the difference between the other two values previously presented.

Verification of the hypothesis of the multi-factor regression model

Test F is used to test the validity or likelihood of the model as a whole.

The value of this test is established as a ratio between variation explained on the basis of regression and variation not explained by regression, each of these being on their turn divided by their degrees of freedom. The calculus formula looks like this:

$$F = \frac{\sum (\hat{y}_i - \bar{y})^2 / k}{\sum (y_i - \hat{y}_i)^2 / (n - k - 1)}$$

where k = number of variables for the model, here 3, and T = number of observations, i.e. 224.

When analyzing the data in our model we note that we have $F = 3.56$ and a likelihood of 0.04. Therefore, we can accept that on the whole the studied multiple linear regression model is good.

➤ **Testing multi-collinearity: The Klein Test**

We calculate the Pearson correlation coefficients between any two independent variables, $r^2_{x_i, x_j}$, and we have the hypothesis:

$$H_0 : \exists r^2_{x_i, x_j} > R^2 \text{ phenomenon of multi-collinearity;}$$

$$H_1 : r^2_{x_i, x_j} < R^2 \text{ the phenomenon of multi-collinearity does not appear.}$$

From Eviews we have the following results:

Table no. 2 Testing of multi-collinearity

	CURS_FUTURES	CURS_ACTIV_BAZA	ROBOR
CURS_FUTURES	1.000000	0.527179	-0.160424
CURS_ACTIV_BAZA	0.527179	1.000000	-0.433633
ROBOR	-0.160424	-0.433633	1.000000

The value for R^2 is 0.283 and we see that it is not larger than all the Pearson coefficients so the phenomenon of multi-collinearity is present in this model of multiple regression.

➤ **Farrar-Glauber Test**

We calculate the correlation matrix of the exogenous variables of the model of multiple regression.

Table no. 3 Correlation Matrix

	C(1)	C(2)	C(3)
C(1)	5418822.	-74630.19	-163023.0
C(2)	-74630.19	1118.489	1706.893
C(3)	-163023.0	1706.893	13852.78

The hypotheses are defined:

H_0 : the module of the determinant of the matrix of correlation coefficients is equal to 1, there is no phenomenon of colinearity

H_1 : the module of the determinant of the matrix of correlation coefficients is smaller than 1, there is a phenomenon of colinearity.

The statistics of the test will be equal to -530.38. It will be compared to

$$\chi^2_{\alpha; k \frac{(k-1)}{2}} = 3.84.$$

Since the calculated value is smaller than the one in the table, it results that the phenomenon of multi-colinearity can be neglected.

➤ **Verification of Normality**

In order to test whether the errors of the model follow or not a normal distribution we shall use the Jarque-Bera test which shows the following hypotheses:

H_0 : the errors follow a normal distribution: skewness = 0 and kurtosis = 3

H_1 : the errors do not follow a normal distribution.

It is known that, if the errors follow the normal law of average zero and standard deviation $s_{\hat{\varepsilon}}$, then the following relation takes place:

$$P\left(\left|\hat{\varepsilon}_i\right| \leq t_{\alpha} s_{\hat{\varepsilon}}\right) = 1 - \alpha.$$

The verification of the hypothesis of normality of errors will be done by means of the test Jarque-Bera¹, which is also an asymptotic test (valid in the case of a large volume sample), which follows a distribution chi square with a number of the degrees of freedom equal to 2, with the following form:

$$JB = n \left[\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right] \sim \chi_{\alpha;3}$$

where: n = number of observations;

S = the coefficient of skew (skewness), which measures the symmetry of the distribution of errors around their average, which is equal to zero, with the following calculus relation:

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^3}{\sigma^3}$$

K = the smoothing coefficient calculated by Pearson (kurtosis), which measures the vaulting of the distribution (how „pointed” or smooth is the distribution in comparison with normal distribution), with the following calculus relation:

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^4}{\sigma^4}$$

The Jarque-Bera test is based on the hypothesis that normal distribution has a coefficient of skew equal to zero, $S = 0$, and a smoothing coefficient $K = 3$.

If the probability $p(JB)$ corresponding to the calculated value of the test is low enough, then the hypothesis of normality of the errors is rejected, whereas, in the opposite case, for a sufficiently high level of probability the hypothesis of normality of the errors is accepted, or if $JB > \chi_{\alpha;2}^2$, the hypothesis of normality of the errors is rejected.

The value of the JB test is of 1.94

We note that skewness = 1.00, and kurtosis = 3.18, the probability of the test is = 0.37. For this reason we accept the null hypothesis, namely that this regression follows a normal distribution.

Verification of the homoskedasticity

¹EViews, UserGuide, Version 2.0, QMS Quantitative Micro Software, Irvine, California, 1995, p. 140-141

Homoskedasticity concerns that hypothesis of the regression model which affirms that the errors of the model must have the same variance: $Var(\varepsilon_t) = \sigma^2$ for any $t=1, \dots, n$. The presence or not of homoskedasticity can be identified both graphically and with the help of some statistical tests. From the graph of residues we cannot affirm with certainty neither the existence of homoskedasticity nor that of heteroskedasticity. The random (residual) variable is of null average $M(\hat{\varepsilon}) = 0$, and its dispersion $s_{\hat{\varepsilon}}^2$ is constant and independent of X – the hypothesis of homoskedasticity, based on which we can admit that the connection between Y and X is relatively stable.

The verification of the hypothesis of homoskedasticity of errors in the case of this model will be done by means of the White test.

The use of the White test supposes passing through the following steps:

- assessment of the parameters of the initial model and the calculating the estimated values of the residual variable, u;
- construction of an auxiliary regression, based on the supposition of the existence of a relation of dependence between the square of the values of the error, the exogenous variable included in the initial model and the square of her values

$$\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \omega_i$$

and the calculation of the coefficient of determination, R^2 , corresponding to this auxiliary regression;

- verification of the significance of the parameters of the newly-built model, and if one of these is insignificant, then the hypothesis of heteroskedasticity of the errors is accepted.

There are two variants of use for the White test:

- use of the classic Fisher–Snedecor test, based on the hypothesis of parameter nullity, respectively:

$$\mathbf{H}_0: \alpha_0 = \alpha_1 = \alpha_2 = 0$$

If the null hypothesis, according to which the results of the assessment are insignificant ($F_c < F_{\alpha; v_1; v_2}$), is accepted, then the hypothesis of homoskedasticity is verified, the contrary case meaning the presence of heteroskedasticity of errors.

- the use of the LM test, calculated as product between the number of observations corresponding to the model, n , and the coefficient of determination, R^2 , corresponding to this auxiliary regression. In general, the LM test is asymptotic distributed as a $\chi_{\alpha; v}^2$, for which the number of degrees of freedom is equal to: $v = k$, where k = number of exogenous variables, respectively:

$$LM = n \cdot R^2 \sim \chi_{\alpha; v}^2$$

If $LM > \chi_{\alpha; v}^2$, the errors are heteroskedastic, in the contrary case, they are homoskedastic, respectively the hypothesis of nullity of parameters, $\alpha_0 = \alpha_1 = \alpha_2 = 0$, is accepted.

The best known test is the test of White which verifies the following hypotheses:

- Null Hypothesis $H_0: \sigma_i^2 = \sigma^2$ for all $i = 1, \dots, n$
- Alternative Hypothesis $H_1: \sigma_i^2 \neq \sigma^2$ for at least one index i .

More exactly for the initial regression model the auxiliary regression:

$$\varepsilon_i^2 = \alpha_0 + \alpha_1 * curs_activ_baza + \alpha_2 * cotatia_ROBOR + v_i$$

was built.

The new errors v_i are distributed normally and independently from ε_i .

Under these circumstances I shall have a null hypothesis H_0 : $\alpha_0 = \alpha_1 = \alpha_2 = 0$ with the alternative H_1 : not all α parameters are zero. If we accept the null hypothesis then we accept the homoskedasticity hypothesis, and if there are parameters different from 0 we accept heteroskedasticity.

For this Output table obtained by the new regression model we apply the signification test t for each coefficient separately.

Table no. 4 Test White – Heteroskedasticity

F-statistic	2.792711	Probability	0.062015	
Obs*R-squared	8.633803	Probability	0.070934	
Test Equation:				
Dependent Variable: RESID^2				
Method: LeastSquares				
Date: 01/16/12 Time: 17:28				
Sample: 224				
Included observations: 224				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-27494484	20619033	-1.333452	0.2011
CURS_ACTIV_BAZA	650206.2	620767.9	1.047422	0.3105
CURS_ACTIV_BAZA^2	-3860.775	4505.848	-0.856836	0.4042
ROBOR	2890857.	954088.1	3.029969	0.0080
ROBOR^2	-134724.1	51132.03	-2.634827	0.0180
R-squared	0.411133	Mean dependent var		4042337.
AdjustedR-squared	0.263917	S.D. dependent var		6982388.
S.E. of regression	5990560.	Akaikeinfocriterion		34.25352
Sumsquaredresid	5.74E+14	Schwarz criterion		34.50222
Log likelihood	-354.6620	F-statistic		2.792711
Durbin-Watson stat	2.014792	Prob(F-statistic)		0.062015

Thus the probability for the free term is of 0.20 which exceeds the threshold of 0.05 and is lower than 0.8 i.e., it is situated in the area of incertitude. Most variables coefficients are to be found in this interval, except the Robor coefficient which is significant from a statistical point of view; its probability being of 0.008 and lower than the threshold of 0.05. Also, the probability for the F test is not high enough and is again in the area of incertitude, $p = 0.06$. Considering the value of p we could say that we reject the null hypothesis (the presence of heteroskedasticity) by an error of 94%, consequently, we could accept the null hypothesis (presence of homoskedasticity) with an error of 6% .

Analysis of the autocorrelation of the 1st order

We used the Durbin – Watson test: $cov(\varepsilon_t, \varepsilon_{t-1}) = 0$

For the analyzed regression equation:

$$Future\hat{s}_t = \alpha + \beta_1 * Curs_activ_baza_t + \beta_2 * Robor_t + \varepsilon_t$$

the autocorrelation of 1st order of the errors is expressed by the relation: $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ for $t=2, \dots, n$ where $v_t \sim N(0, \sigma_t^2)$. The DW statistical test uses the pair of hypotheses:

$H_0: \rho = 0$ (null hypothesis); $H_1: \rho \neq 0$ (alternative hypothesis).

The DW statistics is presented in a table, its values depending on a stated level of signification, the number of observations in the sample and the number of influence variables in the regression model. This, for a stated level of signification, has two critical values obtained from the DW tables, d_1 and d_2 .

The areas of rejection of the null hypothesis are defined as follows:

If $DW \in (d_2, 4 - d_2)$, there is no autocorrelation;

If $DW \in (0, d_1)$ positive autocorrelation of the errors;

If $DW \in (4 - d_1, 4)$ negative autocorrelation of the errors;

If however the value of the DW test is between the remaining intervals (d_1, d_2) or ($4 - d_2, 4 - d_1$) the test is not conclusive.

In the analyzed model, the DW statistics = 1.84. For a threshold of signification of 5%, three influence variables the table values of the statistics are: $d_1 = 1.13$ and $d_2 = 1.54$. The value obtained in the model belongs to the interval ($d_2, 4 - d_2$) so there is no autocorrelation.

The analysis of autocorrelation of a higher level: The Breuch-Godfrey Test

By this test we shall analyze the existence of a autocorrelation of the order k, $k \neq 1$. We suppose that the errors of the regression model are given by the equation:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_k \varepsilon_{t-k} + v_t, \text{ for } t = k, \dots, n, \text{ and } v_t \sim N(0, \sigma_v^2)$$

In order to assess statistically the presence of an autocorrelation of the order k, we shall use the following statistical hypotheses:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0;$$

$$H_1: \rho_1 \neq 0 \text{ or } \rho_2 \neq 0 \text{ or } \dots \rho_s \neq 0$$

We note that the probability of the F statistics is 0.46 (pretty high), so we cannot decide on the existence of the autocorrelation of the 2nd order.

Table no. 5 The Breusch-Godfrey Correlation

F-statistic	0.467976	Probability	0.634575
Obs*R-squared	1.160547	Probability	0.559745
Dependent Variable: RESID			
Method: Least Squares			
Pre sample missing value lagged residuals set to zero.			
Variable	Coefficient	Std. Error	t-Statistic
C(1)	-347.6140	2442.769	-0.142303
C(2)	4.475858	34.94895	0.128068
C(3)	13.79434	122.8738	0.112264
RESID(-1)	0.065832	0.246960	0.266571
RESID(-2)	-0.235215	0.250734	-0.938105
R-squared	0.055264	Mean dependent var	-6.71E-13
AdjustedR-squared	-0.180920	S.D. dependent var	2060.207
S.E. of regression	2238.830	Akaike info criterion	18.46955
Sum squared resid	80197752	Schwarz criterion	18.71825
Log likelihood	-188.9303	Durbin-Watson stat	2.093904

Following the analysis of the data inserted in this model of multiple regression, for better results in relation with the homoskedasticity, autocorrelation of errors, or normality of the model, we can insert more observations to detect the connections between them.

5. CONCLUSIONS

As a result of the use of the econometric model of multi-factor regression we can draw the following **conclusions**:

- The market of futures products in Romania is maturing, but its efficiency cannot be demonstrated very well because the distortions caused by the lack of liquidity and the marked diminishing of activity at the Bucharest Stock Exchange causes abnormal rate changes for an efficient market. Actually, the Sibiu stock exchange has started during the last three years a process of migration towards the transaction of products with international underlying assets. This process proved to be successfully on short term. On long term the relation between SIBEX and Romanian underlying assets could be a difficult matter for the stock exchange board of directors.

- The correlation between the price of the futures contract and the level of the Dow Jones index exists, but is not very strong and therefore, we consider that speculative market operations generate an important component related to the risk premium which the participants are willing to integrate in the price of the transacted product;

- The connection with the interest rate is much lower, because the yields envisaged by the investors on the futures market are much higher than the ones recorded by placing funds with a bank and, therefore, the increase or decrease of the interbank interest rate has no significant repercussions on the appetite for risk. Also, the interest rate has a significant importance for investors with a risk aversion and, because the latter are much less numerous on the market of derivative financial products in Romania the correlation interest rate – futures rate is low.

REFERENCES

1. Chance, D. Analysis of Derivatives for the CFA Program, Association for the Investment Management and Research, USA, 2003
2. Hull, C. J. Options, Futures and other derivatives, 5th edition, Prentice Hall, New Jersey, 2002
3. Wilmott, P. Derivative Inginerie Financiară Teorie & Practică, Editura Economică, București, 2002
4. * * * *EViews, UserGuide*, Version 2.0, QMS Quantitative Micro Software, Irvine, California, 1995, p. 140-141