WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

Vol. 8
No. 1
Spring
2013

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

Contents

Ion IVAN
PhD, University Professor, Department of Economic Informatics
Faculty of Cybernetics, Statistics and Economic Informatics
University of Economics, Bucharest, Romania

E-mail: ionivan@ase.ro ; **Web Page:** http://www.ionivan.ro

Claudiu HERTELIU
PhD, Assistant Professor, Department of Statistics and Econometrics,
University of Economics, Bucharest, Romania

E-mail: claudiu.herteliu@gmail.com, **Web page:** http://www.hertz.ase.ro

Gheorghe NOSCA
PhD, Association for Development through Science and Education, Bucharest, Romania

E-mail: r_g-nosca@yahoo.com

We are happy to announce the winners of JAQM 2012 Awards.

After deliberations, the winners are:

1st Category

For his very important contribution in promoting and developing Quantitative Methods in Academia, as a recognized scholar, we grant to:

Alexandru ISAIC-MANIU

from **University of Economics, Bucharest, Romania**

the ***2012 JAQM Medal of Honor***

2nd Category

For the most valuable, Quantitative Methods related, paper published in JAQM:
"Credit Accelerator, CDS Rate and Long Term Yields: Empirical Evidences from the CEE Economies", JAQM Fall Issue, 2012, pp. 1-12
we granted to:

Bogdan-Gabriel MOINESCU
from **University of Economics, Bucharest, Romania**

the **2012 JAQM Best Paper Award**

3rd Category

For the most promising young researcher in Quantitative Methods area, we grant to:

Michael A. LEWIS

from **Hunter College School of Social Work, New York, USA**

the ***2012 JAQM Distinction***

EXTERNAL ANALYSIS IN PLS-PATH MODELING FOR THE EVALUATION OF THE PASSANGER SATISFACTION

Anna CRISCI
Second University of Naples

E-mail: crisci.anna@virgilio.it

Antonello D'Ambra
Second University of Naples

E-mail: antonello.dambra@unina2.it

ABSTRACT:

In recent years the need to verify the degree of user satisfaction and the quality of services provided has become a priority for transportation companies. The requirement is not only to move towards higher quality of services provided, but also to strengthen confidence in the companies that provide the service.

This paper shows a study to evaluate the overall passenger satisfaction of public transport service. To explore the collected data, PLS-Path Modeling approach has been adopted, and later the analysis of the data has been further supported by means of the joint use of the PLS-PM and the External Analysis Method, catching the advantages of both.

In particular, the additional element that the present contribution wants to highlight is the use of so-called external information. In fact, passenger satisfaction may be influenced by external factors such as sex, age, educational level, profession, etc.

Therefore, the joint use of the PLS-PM and the External Analysis Method could help the researcher interpret the results objectively.

Key words: *External Analysis, Structural Equation Modeling, PLS- Path Modeling, Passenger Satisfaction.*

1. INTRODUCTION

The activities of Passenger Satisfaction (PS) have a more and more strategic role for the achievement of the objectives of the leader company in the supply of a services [3].

The "total quality" of the services and the full satisfaction of all the users are often unattainable with limited resources. Therefore, it is important to set a goal of possible quality, and an interesting perspective could be "***The satisfaction of users concerning their most important expectations***". The main problem of this approach is that the expectations modify over the time, often while the service is carried out. The PS must be managed along with the service cycle; moreover, in order to estimate the quality of the services supplied, the *so-called "user feedback"* is important [24]. The intangible part in the performance service cannot be subject to a testing or check as in the case of tangible goods. In other words, to measure an effective performance we have to refer to the people involved in it, that is to users who are the only ones that actually test the service, with reflections both on the efficiency of the system and on the quality of the performance.
It is, therefore, essential that the activities of PS:

- are aligned with the more general needs of the company;
- meet the user needs;
- supply elements that allow to improve systematically and competitively organizational structure, directing it to the PS.

In this paper a study of PS evaluation in the transport service in Benevento (Italy) is shown.

In 2011 to estimate the PS a survey was carried out and, a questionnaire, agreed with the responsible of the service, composed by analytical sections of interest (***Quality Factors***) was arranged. The method of extraction of the data is the sample random with a sampling bias and a confidence level, 4.5% and 95%, respectively. The interviewed sample of 400 units is determined by a population of 9,000[1] users daily.

The method of extraction of the data is the sample random with a sampling bias and a confidence level, 4.5% and 95%, respectively. The interviewed sample of 400 units is determined by a population of 9,000[1] users daily.

The five dimensions of the quality are:

- accessibility to the service (A.S);
- condition of the bus (C.B);
- security on board (S.B);
- reliability of the service (R.S);
- overall passenger satisfaction (O.P.S);

To explore the collected data, PLS-Path Modeling approach [4,5,8,21,22] has been adopted, and later the analysis of the data has been further supported by means of the joint use of the PLS-PM and the External Analysis Method [15], catching the advantages of both.
The assumed model is composed of five latent variables and seventeen manifest variables:

[1] The population of 9,000 users was determined by a survey concerning the adoption of the service that the public service company had carried out previously.

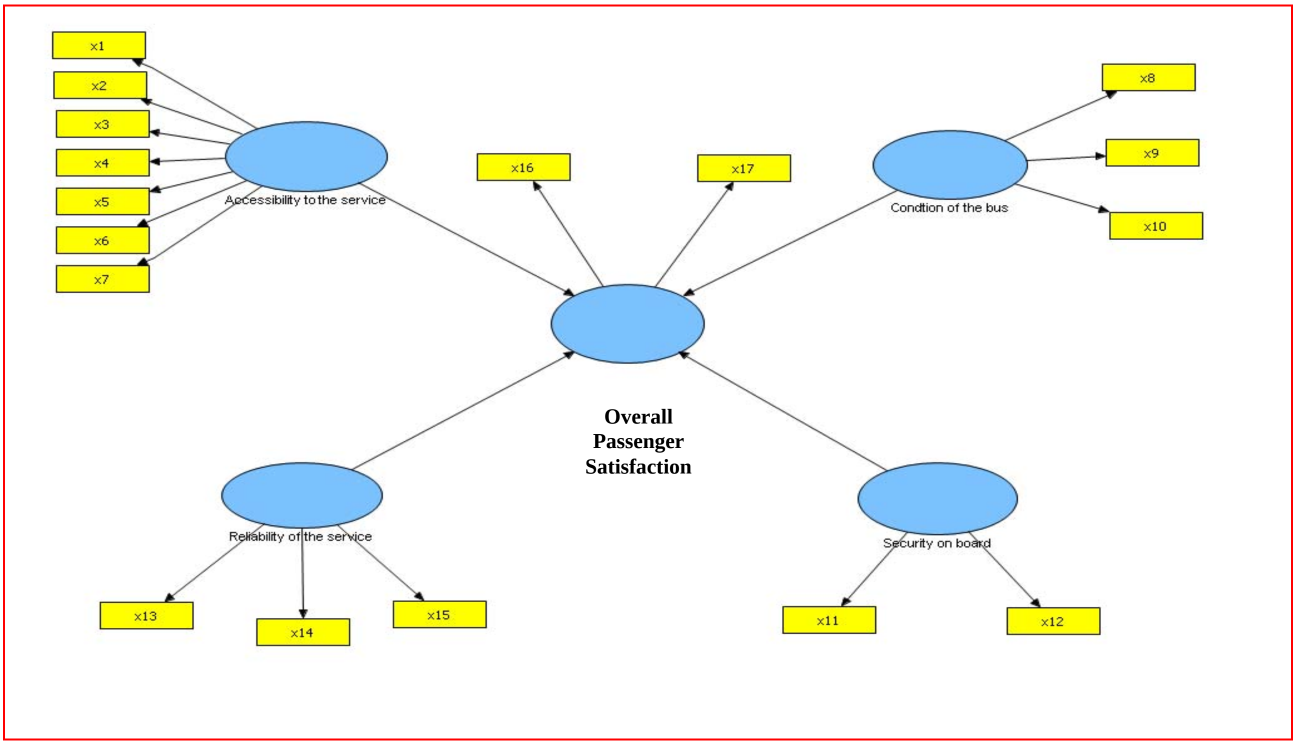

Figure 1: The Theoretical Model for the evaluation of the Passenger Satisfaction of the Public Transport Service

The PLS Path Modeling is a statistical method which was developed for the Analysis Structural Models with latent variables. One of the difficulties for a researcher in the economic-social sciences in the specification of a statistical model describing the casual-effect relationships between the variables derives from the fact that the variables which are object of the analysis are not directly observable (*i.e. latent*), for example, the performance, the passenger satisfaction, the social status etc. Although such latent variables (*LVs or latent construct*) cannot be directly observable, the use of proper indicators (*i.e. manifest variables, MVs*) can make the measurement of such constructs easy.

As opposed to the covariance-based approach or LISREL [2,10,11,12,13,14], the aim of the PLS is to obtain the scores of the latent variables for predicted purposes without using the model to explain the covariation of all the indicators [9,23]. For example, a researcher may be interested in what dimensions of the service quality can more influence the PS.

The joint analysis of the original data with external information (e.g. socio-demographic features of the subjects: sex, age, education, job, etc.) can lead the researcher to a more objective interpretation of the obtained results. In other words, the role of the external information may be that of establishing, for example, the existence of a relationship between the assumed links and the demographic information on the subjects.

The paper is organized as follows: in sections 2 and 3, the PLS-PM approach and the External Analysis are shown, respectively. In section 4, the estimations of the parameters, the general results of the Path-model, and the results of the joint analysis of the PLS-PM and External Analysis are shown.

2. The PLS-Path Modeling Approach

PLS –Path Modeling aims to estimate the relationships among *J* blocks of variables, which are expression of unobservable constructs. Specifically, PLS-PM estimates the network of relations among the manifest variables and their own latent variables, and the latent variables inside the model through a system of interdependent equations based on simple and multiple regression. Formally, let us usually assume *K* variables observed on *N* units. The resulting data x_{nkj} are collected in a partitioned table of standardized data $\boldsymbol{X}$:

$$X = [X_1, \dots, X_j, \dots, X_J],$$

where X_j is the generic *j*-th block.

The path models in the PLS involve three sets of relations:

1. ***Inner Model or Structural model,*** which refers to the structural model and specifies the relationships between the latent variables LVs. A latent variable can play both predictand role and predictor one; a latent variable which is never predicted is called an exogenous variable, otherwise, it is called endogenous variable.

The structural model can be expressed as:

$$\xi_j = \sum_i \beta_{ji} \xi_i + \zeta_j \qquad (1)$$

where β_{ji} is called the path coefficient (representing the path from *i*-th latent variable to the *j*-th latent variable) and we indicate with B the matrix of all the path coefficients in the model. It is a square matrix of 0/1, and its dimensions are equal to the number of LVs. This matrix indicates the structural relationship between LVs. ζ_j is the inner residual term, and the diagonal variance/covariance matrix among inner terms is indicated with Ψ .

2. ***Outer Model or Measurement Model***, which refers to the measurement model and specifies the relationships between the constructs and the associated indicators MVs. Two ways to establish these links can be distinguished as follows [6]:

- ***Reflective way***: in which the indicators (manifest variables) are regarded to be reflections or manifestations of their latent variables: a variation of the construct yields a variation in the measures. As a result, the direction of causality is from the construct to the indicator.

Each manifest variables represents the corresponding latent variable, which is linked to the latent variable by means of a simple regression model.

The reflective indicators of the latent construct should be internally consistent, and, as it is assumed that all the measures are indicators equally valid of a latent construct, they are interchangeable.

The reflective measures are at the basis of the theory of the classical tests, of the reliability estimation, of and factorial analysis, each of them considers the manifest variable x_{jk} being a linear combination of its latent variable ξ_j:

$$x_{jk} = \lambda_{jk} \xi_j + \varepsilon_{jk} \qquad (2)$$

where $\lambda_{jk}$is the generic loading coefficient associated to the *k*-th manifest variable in the *j* block, and we indicate with Λ the matrix containing all the loading coefficients in the block.

ε_{jk} represents the generic outer residual term associated to the generic manifest variable and the corresponding diagonal variance/ covariance matrix is indicated with Θ^{ε} .

Predictor specification is adopted,

$$E(x_{jk}|\xi_j) = \lambda_{jk}\xi_j \tag{3}$$

which implies that

$$E(\varepsilon_{jk}) = E(\xi_j \varepsilon_{jk}) = 0 \tag{4}$$

the residuals have zero mean and are uncorrelated with MVs.

- ***Formative way***: in which the indicators are regarded as causes of their latent constructs: a variation of the measures yields a variation in the construct. As a result, the direction of causality is from the indicator to the construct. The elimination of items that have low correlations compared with the overall indicators will compromise the construct validation, narrowing the domain.

This is one of the reasons by which the reliability measures of the internal consistency should not be used to estimate the fitting of the formative models. Moreover, the multi-collinearity between the indicators may be a serious problem for the parameter estimations of the measurement model when the indicators are formative, but it is a good point when the indicators are reflective.

The latent variable ξ_j is assumed to be a combination linear of its manifest variables x_{jk}:

$$\xi_j = \sum_k \pi_{jk} x_{jk} + \delta_j \tag{5}$$

assuming predictor specification:

$$E(\xi_j|x_{jk}) = \sum_k \pi_{jk} x_{jk} \tag{6}$$

which means that residuals have zero mean and are uncorrelated with the MVs

$$E(\delta_j) = E(\xi_j \delta_j) = 0 \tag{7}$$

3. ***Weight relations.*** The specification of the relations between LVs and their set of indicators is carried out at a conceptual level. In other words, the outer relations refer to the indicator and the "true" LV, which is unknown. As a result, the weight relations must be defined for completeness. The estimation of LVs are defined as follows:

$$\hat{\xi}_j = \sum_k \tilde{w}_{jk}\, x_{jk} \tag{8}$$

where $\tilde{w}_{jk}$ are the weights used to estimate the LVs as linear combinations of their observed MVs.

2.1 The PLS Algorithm

By means of this algorithm both the estimation of the latent variables and the estimation of the parameters is obtained. The PLS-PM implies some steps.

The first step consists in an iterative procedure made up of simple and/or multiple regressions by taking into account the relation of the internal model, of the external model, and of weight relations. The result is the estimation of a set of weights which are used to calculate the scores of the latent variables as linear combinations of their manifest variables. Once the estimations are obtained the following steps imply the non-iterative estimation of the structural model and measurement model coefficients.

2.1.1 *PLS Algorithm: stage 1*

Stage 1.1: outer approximation and inner approximation

In order to estimate the parameter, two double approximations for LVs are considered by PLS algorithm [17,20]:

- ***the outer approximation or external estimation***, called Y_j, is used for the measurement model. In this stage we find an initial proxy of each LV, ξ_j, as a linear combination of its MVs x_{jk}. The external estimation is obtained as the product of the block of MVs and the outer weights w_{jk};
- ***the inner approximation or internal estimation***, called Z_j, is used for the structural model. The connections among LVs are taken into account in order to get a proxy of each LV worked out as weighted aggregate of its adjacent LVs. The internal estimation is obtained as the product of the external estimation Y_i (of ξ_i) and the so-called inner weights, e_{ji}.

There are three ways to calculate the internal weights:

- **centroid scheme (Wold)**: Il centroid scheme is the scheme of the original algorithm by Wold. This scheme considers only the direction of the sign among the latent variables $sign\{cor(Y_j, Y_i)\}$, without considering neither the direction nor the power of the paths in the structural model.

- **factorial scheme (Lohmöller)**: this scheme uses the correlation coefficients $cor(Y_j, Y_i)$ as internal weights instead of using only the correlation sign and, therefore, it considers not only the direction of the sign but also the power of link of the paths in the structural model.

- **path weighting scheme**: in this case the latent variables are divided into predictors and followers according to the cause- effect relations between the two latent variables. A variable can be either a follower (if it is yielded by another a latent variable), or a predictor (if it is the cause of another latent variable. If ξ_i is follower of ξ_j, then, the internal weight is equal to the correlation between Y_j e Y_i. On the other hand, for the predictors ξ_i of ξ_j, the internal weights are the regression coefficients of the Y_i in the multiple regression of Y_j on Y_i associated with the predictors ξ_j.

The path weighting scheme has the advantage to consider both the direction and the power of the paths in the structural model.

Even though the path weighting scheme seems the most coherent with direction of the structural relations between latent variables, the centroid scheme is very often used as it adapts well to cases where the manifest variables in a block are strongly correlated to each other. The factorial scheme, instead, is better suited to cases where such correlations are weaker. In spite of the different common practices, it is strongly recommended to use the path weighting scheme. Indeed, this is the only estimation scheme that explicitly considers the direction of relationships as specified in the predictive path model [8].

Stage 1.2: Updating outer weights

The external estimation is conceived as a step in which the information contained in the internal relations are incorporated in estimation process of the latent variables.

Once the internal approximation is carried out, the internal estimation Z_i must be considered with respect to their indicator.

There are two ways to calculate the external weights:

1. Mode A: is preferred when the indicators are linked to their latent variables by means of the reflective way, in which each weight w_{jk} is the coefficient regression of Z_j in the simple regression of x_{jk} on Z_j, that is the simple regression $x_{jk} = w_{jk}Z_j$ in which:

$$w_{jk} = (Z_j^t Z_j)^{-1} Z_j^t x_{jk} = cor(x_{jk}, Z_j) \quad (9)$$

2. ***Mode B***: : is preferred when the indicators are linked to their latent variables by means of the formative way, in which Z_j is regressed on the block of indicators linked to the latent construct ξ_j, and the w_j of weight w_{jk} is the regression coefficients in the multiple regression

$$Z_j = \sum_k w_{jk} x_{jk} \quad (10)$$

and it is defined by means of:

$$w_j = (X_j^t X_j)^{-1} X_j^t Z_j \quad (11)$$

in which X_j is the matrix with columns of manifest variables x_{jk}.

Stage 1.3: Checking convergence

The algorithm is repeated as far as the weight convergence. In each iterative step, for example, S=1,2,3.., the convergence is verified comparing the external weights of the S step (new weights) with the weights of the S-1 step (old weights). For example, Wold (1982) suggested $|w_{jk}^{S-1} - w_{jk}^{S}| < 10^{-5}$ as a criterion of convergence.

2.1.2 algorithm PLS: stage 2

The second step of the algorithm consists in the estimation of the structural coefficients (path coefficients) and of the parameters of the measurement model (loading coefficients).

For the structural model the path coefficients are estimated by ordinary last squares in the multiple regression of Y_j on the Y_i.

For the measurement model the loading coefficients are estimated depending on the corresponding way. In particular, in the reflective way, the loading coefficients are the regression coefficients of the simple linear regression of x_{jk} on Y_j. By contrast, in the formative way, the weight coefficients π's coincide with outer weights as we perform the multiple linear regression Y_j on x_{jk}.

2.2 **Summary**

The procedure works on centered (or standardized) data, and it starts by choosing arbitrary weights (e.g *1,0..0*)[19]. Chin (1999) suggested starting with equal weights for all indicators (the loadings are set to 1) to get a first approximation of the LVs as a simple sum of its indicators. The inner relations among LVs are considered to estimate the internal approximation by choosing three options: centroid, factoring and path scheme. After obtaining the internal approximation, the algorithm turns around the external relations with the estimate of outer weights obtained by means of mode A (*reflective*) or by mode B (*formative*). The procedure is repeated until convergence of the weights is obtained. Once convergence of the weights is obtained and LVs are estimated, the parameters of the structural and measurement models are calculated by means of the ordinary least squares (OLS).

3. THE EXTERNAL ANALYSIS

A way to add external information to the model on the subjects (e.g, age, sex, education, job, etc..) and/or on the variables is the External Analysis (Takane e Shibayama, 1991). The addition of external information on the subjects and/or variables could help the researcher interpret the results objectively.

This method decomposes the original data into several components (External Analysis), those that can be explained, and those that cannot be explained, by the external information.

We denote a matrix of the data composed of *N*-subject and *n*-variables with *X*. The data may consist in *N* subjects' judgment on *n*-dimensions of the supplied service public quality, or any other multivariate observations. The data may be raw or pre-processed, for example, by standardizations or other transformations.

We denote a matrix of the information on the users with G, and the matrix of the information on the variables with H. For example, G may be an *N*-component vector of ones, a matrix of dummy variables, or a matrix of continuous variables characterizing the subjects.

Similarly, H may be an *N*-component vector of ones, or any other matrix of explanatory variables characterizing the relationships among columns of *X*. Both G and H can be put in any analysis by means of orthogonal projector operators, that is, orthogonal projection operators onto spaces spanned by the column vectors of G and H with $P_G = G(G'G)^- G'$ e $P_H = H(H'H)^- H'$, respectively. It is well known that both P_G and P_H are unique even if $(G'G)^-$ and $(H'H)^-$ are unique. Moreover, $Q_G = I - P_G$, and $Q_H = I - P_H$ are both orthogonal projector operators that are orthogonal to P_G and P_H, respectively.

The matrix of the measurements *X* can be decomposed according to the following relationship: $P_G X P_H + Q_G X P_H + P_G X Q_H + Q_G X Q_H$. Each term has a precise statistic meaning in it; in particular: $P_G X P_H$ indicates the effect of row and column information; $Q_G X P_H$ is the effect of column information without the effect of the row one; $P_G X Q_H$ is the effect of row information without the column ones and $Q_G X Q_H$ is the part which disregards external information.

Once the matrix *X* is decomposed, it may be interesting to carry out the scheme of analysis introduced a prior on each term to catch (to clean) the influence of the external information on the evaluation of the supplied service quality. The external information has different roles, for example, eliminating the different reference systems of the users/citizens as well as catching the effects of the information both on the components and on the dimensions.

4. RESULTS OF THE ANALYSIS

The objective of the research is to analyse the passenger satisfaction of the public transport service by means of the PLS-PM approach and, later, of the combined use of the PLS approach and External Analysis Method. For this reason, five latent variables (LVs) have been identified, each of them measured by proper indicators (i.e. MVs). The LVs and the corresponding MVs identified are:

Table 1: Latent Variable and Manifest Variables

Latent Variables	**Manifest Variables**
Accessibility to the service (Exogenous)	X1: availability of the company time- tables; X2: information on the time tables and the runs; X3: purchase of the tickets and season-tickets; X4: access to the buses; X5: services for disabled people; X6: bus-shelter condition; X7: behaviour of the company staff (kindness, propriety, presentability)
Condition of the buses (Exogenous)	X8: motor vehicle cleanness; X9: journey comfort; X10: reduction of the environmental impact;
Security on board (Exogenous)	X11: journey Security; X12: safety for the users as regards stealing and attacks;
Reliability of the service (Exogenous)	X13: punctuality, i.e observance of the times. X14: regularity, i.e ability to carry out the predicted runs; X15: company time-tables and their connections with other means of transport.
Overall Passenger Satisfaction (Endogenous)	X16: judgment on the company as a whole ; X17: degree of improvement of the service reached in the latest year.

The LVs- ***accessibility to the service, condition of the buses***, ***security on board***, ***reliability of the service-*** are exogenous LVs, i.e. they are variables which are never predicted and behave just only as predictors, while the ***overall passenger satisfaction*** latent dimension is an endogenous LV (i.e. dependent).

Any analysis of the structural equation modeling implies an explicit specification of the path-model which, in its turn, is composed of the measurement model (Outer Model) and of the structural model (Inner Model).

4.1 The Outer Model and the Inner Model

● **Outer Model**

The outer model establishes the relationship between the block of the manifest variables and their corresponding latent variables. As regards the LVs **accessibility to the service**, **condition of the buses**, **security on board**, **reliability of the service** and **overall passenger/user satisfaction** the MVs are linked to them in a "reflective" way. In other words, the MVs are considered reflections or manifestations of the LVs. For these LVs, in correspondence with their respective MVs, we read the std. loadings, that is standardized regression coefficients (it is about a simple linear regression).

Test of the unidimensionality of the reflective MVs block.

A way to check the quality of the measurement model is the verification of the unidimensionality of the reflective MVs block. The reflective indicators of the LV should be coherent internally, and, as it is supposed, that all the measures are indicators equally valid of a LV, they are interchangeable.

- ***Dillon –Goldestein's ρ (o Jöreskog):*** a block is unidimensional if this index is >0.7.

$$\rho = \frac{(\sum_{k=1}^{p} \lambda_{jk})^2 \operatorname{var}(\xi_j)}{\sum_{k=1}^{p} \operatorname{var}(\xi_j) + \sum_{k=1}^{p} \operatorname{var}(\varepsilon_{jk})} \tag{12}$$

As, in practice, we do not know the real values of λ_{jk} and ξ_j, an estimate of Dillon Goldstein's ρ is needed. The approximation of the latent variable is achieved by using the first principal component t_{j1} of the *j-th* block of indicators; the approximation of the loading coefficient is taken as the correlation between t_j and the observed variable x_k, cor(t_1,x_k); the term var (ε_k) is approximated by 1-cor^2(t_1,x_k). Then, estimate of Dillon-Goldstein's is given by

$$\hat{\rho} = \frac{[\sum_{k=1}^{p} cor(x_{jk}, t_{j1})]^2}{\sum_{k=1}^{p} [cor(x_{jk}, t_{j1})]^2 + \sum_{k=1}^{p} [1 - cor^2(x_{jk}, t_{j1})]} \tag{13}$$

The value of the index is > 0.7 for all the observed VMs blocks (Table 2) .

Table 2: Block Unidimensionality

LVs	Type Measure	MVs	DG.rho
A.S	Reflective	7	0,8373
C.B	Reflective	3	0,8025
S.B	Reflective	2	0,7258
R.S	Reflective	3	0,8706
O.P.S	Reflective	2	0,8180

● **Inner Model**

The inner model considers only the LVs, which are assumed to be linearly interconnected according to a casual-effect relationship model. The present study aims at verifying, from an explorative and non confirmative view point, the existence of meaningful relations between the following LVs:

1. accessibility to the service and overall passenger satisfaction
2. condition of the buses and overall passenger satisfaction
3. security on board and overall passenger satisfaction
4. reliability of the service and overall passenger satisfaction

Like the measurement model, the structural model also requires to be validated. In particular, in correspondence with the endogenous LV, we read the coefficient of determination R^2. For each regression in the structural model we have an R^2 that is interpreted similarly as any multiple regression analysis. R^2 indicates the amount of variance in the endogenous latent variable explained by its independent latent variables. A satisfying R^2 (0,57) is obtained for the latent variable Overall Passenger Satisfaction (Table 3). Moreover, if we consider the ***Average Redundance*** (note that the redundancy index represents the power of the set of independent latent variables to explain the variation in the dependent latent variable), a satisfying prediction endogenous LV -**Overall Passenger Satisfaction** (0.38)- from LVs exogenous LVs is obtained.

It is important to note in table 3 the values of AVE (Average Variance Extracted) that tries to measure the amount of variance that a LV captures from its indicators in relation to the amount of variance due to measurement error. AVE is, in most cases, > 0.50, which means that 50% or more variance of the indicators should be accounted for.

Table 3: Summary Inner Model

	LV.Type	**Measure**	**MVs**	**R-square**	**Av.Redun**	**AVE**
A.S	Exogen	Rflct	7	0	0	0.43
C.B	Exogen	Rflct	3	0	0	0.58
S.B	Exogen	Rflct	2	0	0	0.53
R.S	Exogen	Rflct	3	0	0	0.69
O.P.S	Endogen	Rflct	2	0,57	0,38	0.68

GoF (Goodness of fit):

The Goodness of Fit (GoF) is a global criterion developed by Amato, S., Esposito Vinzi, V., and Tenenhaus, M. [1], and it is represents a compromise between the quality of the measurement model and the quality of the structural model. A satisfying GoF is obtained both for the outer model (0,98) and for the inner model (0,62).

4.2 Parameter estimation and validation by re-sampling methods

To estimate the parameter of the model, we have used the module R-package. To calculate the inner estimates of the latent variables, we have used the path weighting scheme.

The non-parametric bootstrap [7,16] procedure can be used in PLS-PM to provide confidence intervals for all parameter estimation, building the basis for statistical inference. In general, the bootstrap technique provides an estimation the shape, spread, and bias of the sampling distribution of a specific statistic.

Bootstrapping treats the observed sample as if it represented the population. Bootstrap samples are created by randomly drawing cases with replacement from the original sample.

In particular, bootstrap has been based on 100 samples, and 95% confidence intervals have been asked for. The confidence intervals indicate the regression coefficients which are significant [18].

If a confidence interval for an estimated path coefficient does not include zero, the hypothesis that the parameter equal zero is rejected.

As shown in the table 4, all the parameters of the measurement model (loading coefficients) are significant (the confidence intervals never include zero).

Moreover, the p-value associated to the t- statistic bootstrap, in this specific case, is, in most cases, > 0.05 and this implies the acceptation of the hypothesis according to which the distribution sampling is centered at the true value of the parameter.

Moreover, the MVs that best reflect their corresponding LVs are:

● ***accessibility to the service***: behaviour of the company staff (loading coefficient: 0,72), availability of the company time- tables (loading coefficient: 0,67), access to the buses (loading coefficient: 0,67).

● ***condition of the bus***: motor vehicle cleanness (loading coefficient: 0.86); journey comfort (loading coefficient: 0,85);

● ***security of board***: journey security (loading coefficient: 0,98) ;

● ***reliability of the service:*** regularity, i.e ability to carry out the predicted runs (loading coefficient: 0,84), company time-tables and their connections with other means of transport (loading coefficient:0,83), punctuality, i.e observance of the times (loading coefficient:0,82);

● ***overall passenger satisfaction:*** judgment on the company as a whole (loading coefficient:0,92);

Table 4: Bootstrap validation for loading coefficients (** significant 5%)

	Original (O)	Mean Boot.	Std.Err	Bootstrap t-statistic	p-value	Conf. Intervals (95%)
X1	0,6296	0,6285	0,0420	-0,2628	0,7932	0,5459;0,7111
X2	0,6711	0,6684	0,0384	-0,6949	0,4888	0,593;0,7439
X3	0,5983	0,6004	0,0497	0,4342	0,6651	0,5026;0,6982
X4	0,6753	0,6717	0,0409	-0,8917	0,3747	0,5913;0,7520
X5	0,6331	0,6359	0,0378	0,7294	0,4675	0,5616;0,7102
X6	0,6153	0,6167	0,0513	0,2722	0,7860	0,5159;0,7175
X7	0,7261	0,7269	0,0306	0,2532	0,8006	0,6667;0,7871
X8	0,8665	0,8687	0,0198	1,1066	0,2712	0,8298;0,9075
X9	0,8544	0,8546	0,0213	0,1125	0,9106	0,8128;0,8965
X10	0,5116	0,5082	0,06	-0,5752	0,5665	0,3902;0,6261
X11	0,9830	0,978	0,0207	-2,4172	0,0175	0,9374;1,0187
X12	0,3185	0,3272	0,1207	0,7171	0,4750	0,09;0,5643
X13	0,8226	0,8229	0,0258	-0,2140	0,8310	0,7713;0,8728
X14	0,8438	0,8434	0,0246	-0,1678	0,8671	0,7951;0,8917
X15	0,8271	0,8286	0,0235	0,6284	0,5312	0,7825;0,8747
X16	0,9267	0,9246	0,0163	-1,2653	0,2052	0,8926;0,9566
X17	0,7029	0,7089	0,0457	1,3232	0,1888	0,6191;0,7987

If we consider the analysis of the path- coefficient we can infer a significance of all the links.

In particular, the exogenous LVs that mostly affect the endogenous LV (overall passenger satisfaction) are: **accessibility to the service** (path coefficient: 0,41) and **reliability of the service** (path coefficient: 0,22).

The value of the link between the **accessibility to the service** and the **overall passenger satisfaction** indicates that when the accessibility to the service rises by one unit (in particular, the behavior of the company staff, the access to the buses, the information on time table and runs) the ***overall users satisfaction*** rises by 0,41. Similarly, the value of the link between the **reliability of the service** and the **overall passenger satisfaction** indicates that when the **reliability of the service** rises by one unit (in particular, regularity, company time- table and their connection with other means of transport, punctuality) the ***overall users satisfaction*** rises by 0.22.

Table 5: Bootstrap Validation for path coefficients (**significant 5%)

	Original (O)	Mean Boot	Std.Err	Bootstrap t-statistic	p-value	Conf. Intervals (95%)
A.S→ O.P.S	0,4105	0,4121	0,0612	0,2578	0,7971	0,2918;0,5324
C.B→ O.P.S	0,1364	0,1337	0,0537	-0,5038	0,6158	0,0281;0,2392
S.B→ O.P.S	0,1423	0,1445	0,0464	0,4640	0,6437	0,0532;0,2358
R.S→ O.P.S	0,2228	0,2227	0,0473	-0,0186	0,9852	0,1297;0,3158

The full specification of the path diagram (Structural model and Measurement model), along with the indication of the loading coefficients and path coefficients, is the following:

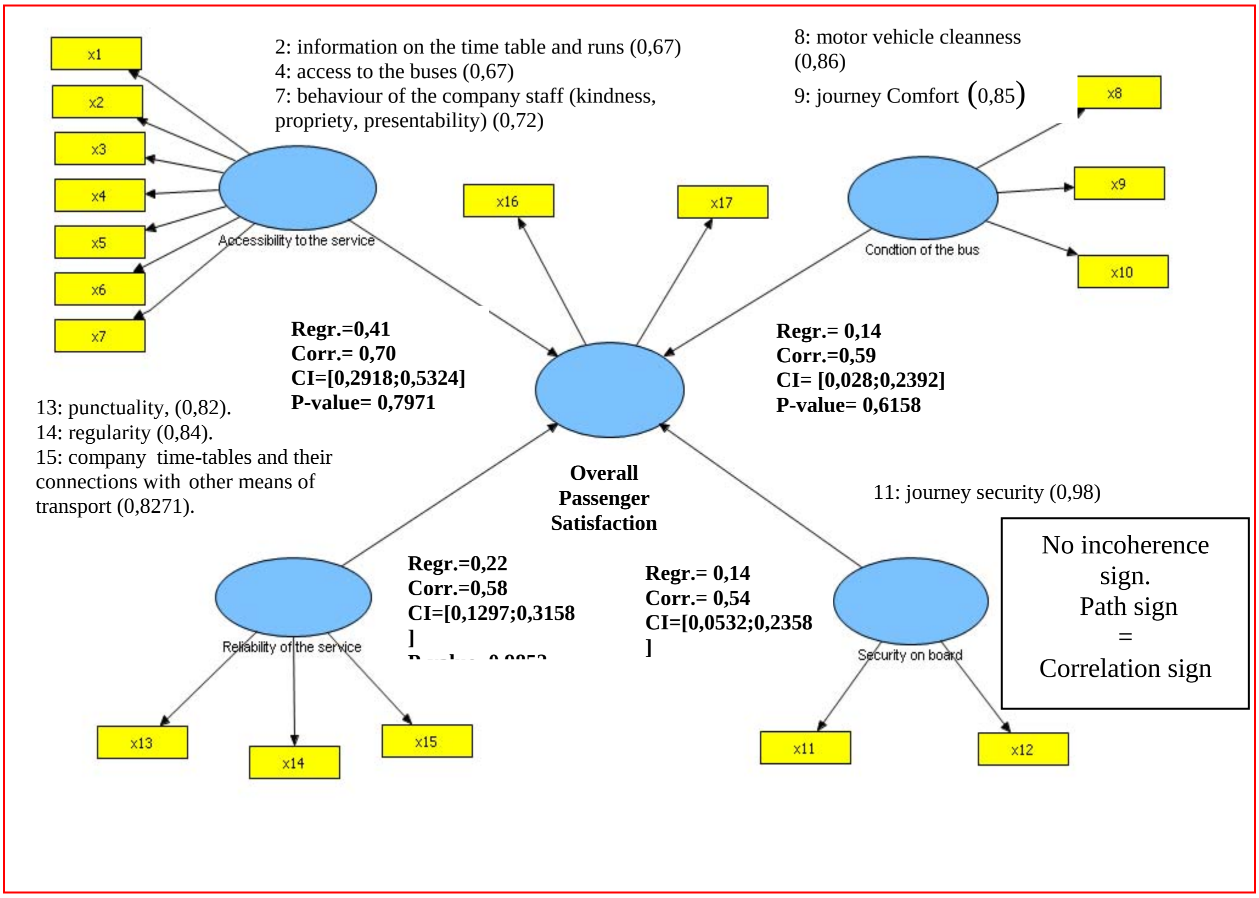

Figure 2: The specification of the path diagram (Structural Model and Measurement Model): path coefficients, correlations, loading coefficients and bootstrap results.

It is important to observe that the figure 2 also shows the correlations between LVs as there is no incoherence between the path coefficient sign and the correlations sign; for this reason, none of the original relationships between LVs has been suppressed.

4.3 PLS-PM with external information on subjects

The analysis of the data has been further supported by means of the joint use of two methodologies: the PLS-PM and the External Analysis, catching the advantages of both.

For this purpose, the external information on subjects that has been taken into account is the *job of the interviewed subjects*: 1. Student, 2. Employee, 3. Housewife, 4. Retired, 5. Other".

The choice of working only with the external information on the subjects has been determined by the interest in testing how these preference data are related to the subjects' demographic information, that is their job.

As previously shown, the aim of the External Information is the decomposition of the matrix of the original data (X) in different components, those which can be explained by the external information: "job", and those that cannot be explained by external information.

This external information can be put in the analysis by means of orthogonal projector operators; in this specific case, the orthogonal projector is:

$$P_G = G_{(N,B)}(G'_{(B,N)}G_{(N,B)})^{-}G'_{(B,N)} \tag{14}$$

As it is known, the decomposition of the matrix *X* is the following:

$$X = P_G X P_H + Q_G X P_H + P_G X Q_H + Q_G X Q_H \tag{15}$$

The PLS-PM has been applied to the first and the last matrix *X* component, and both the predictive power of the model and the differences in terms of path coefficients between the two PLS- models (i.e: PLS with information and PLS without information) have been evaluated in order to verify whether a relationship between the information taken into account and the theoretical links specified in the path diagram (figure 2) exists.

The addition of the "job" external information improves the quality of the structural model. The coefficient of determination R^2 changes from 0.56 (for the PLS-model without information) into 0.90 (for the PLS-model with information).

Table 6: Determination Coefficient (R^2)

PLS-Model with information	**PLS-Model without information**
0.90	0.56

The redundancy index also improves by changing from 0,39 into 0,65.

Table 7: Redundance Index

PLS-Model with information	**PLS-Model without information**
0.65	0.39

The values of AVE (Average Variance Extracted) for PLS-Model with information improved in most cases.

Table 8: Average Variance Extracted

LVs	**AVE PLS-Model with information**	**AVE PLS-Model without information**
A.C	0.43	0.42
C.B	0.78	0.58
S.B	0.93	0.53
R.B	0.51	0.69
O.P.S	0.70	0.67

Finally, the Gof (Goodness of fit) for the Inner Model also improves, (from 0,61 for the PLS-model without information to 0,66 for the PLS-model with information) .

By comparing path coefficients of the PLS-PM with information and the PLS-PM without information it follows that, if the external information is absent (e.g: equal jobs), all the links are positive and significant. The addition of the external information further strengthens some of these links, in particular, the relationship between the access to the service and the overall passenger satisfaction, and between the bus condition and the overall passenger satisfaction. Probably, there are specific categories of subjects whose judgment of preference may be actually linked to the external information taken in consideration, for example, students and employees, workers whose use of the means of the transport should be more frequent. As a result, the company, in order to better the quality of the service, should direct its efforts to improve this aspect of the service.

Table 9: Comparison between the path coefficients of the PLS-PM with external information and of the PLS-PM without external information

	PLS-PM with information	**PLS-PM without information**	**abs. diff**
A.S->O.P.S	**0,5943**	**0,3965**	**0,1978**
C.B->O.P.S	**0,3638**	**0,1416**	**0,2222**
S.B->O.P.S	0,0913	0,1243	0,0330
R.S->O.P.S	0,2860	0,2478	0,0382

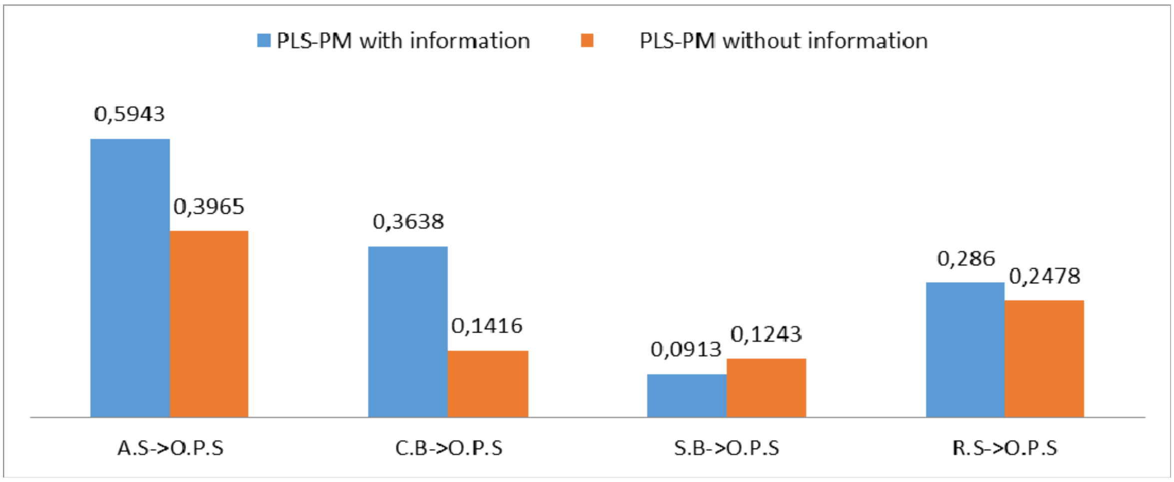

Figure 3: Comparison between path coefficients of the PLS-PM with information and the PLS-PM without information

5. CONCLUSION

The objective of the research has been to analyse the passenger satisfaction of the public transport service by means of the PLS-PM approach and, later, of the combined use of the PLS approach and External Analysis Method. In particular, the External Analysis Method represents the additional element that this contribution has sought to highlight.

The use of the PLS Path modeling has shown that the exogenous LVs that mostly affect the endogenous LV are the **accessibility to the service** and **reliability of the service**. A satisfying GoF is obtained both for the outer model and for the inner model.

Later, the analysis of the data has been further supported by means of the joint use of two methodologies: the PLS-PM and the External Analysis. The external information on subjects that has been taken into account is the Job of the interviewed subjects. The addition of the *job external information* improves the quality of the structural model, of both R^2 and GoF.

In particular, the relationship between the access to the service and the overall passenger satisfaction, and between the bus condition and the overall passenger satisfaction further strengthens.

REFERENCES

1. Amato, S., Esposito Vinzi, V., and Tenenhaus, M., **A global Goodness-of-Fit for PLS structural equation modelling**. Oral communication to PLS Club, HEC School of Management, 2004.
2. Bagozzi, R.P, **"Structural Equation Models in Experimental Research"**, Journal of Marketing Resarch,14, 1977, pp 209-226.
3. Busacca, B., **Le risorse di fiducia dell'impresa. Soddisfazione del cliente, creazione di valore, strategie di accrescimento**, UTET, Libreria, 1994.
4. Chin, W.W. and Newsted, P.R., **Structural Equation Modeling Analysis with Small Samples using Partial Least Squares**. In: *Statistical Strategies for Small Sample Research*, 1999, 307-341. Hoyle R. (Ed). London: Sage.
5. Chin, W.W, **Partial Least Squares for research: An overview and presentation of recent advance using the PLS approach**, 2000, p.68.
6. Edwards, Jeffrey R. and Bagozzi, P.R, **"On the Nature and Direction of Relationships between Construct an Measures"**, *Psycological Method,5*(2), 2000, 155-74.
7. Efron,B, **The jackknife, the bootstrap and other resampling plans**, SIAM,USA: Philadelphia, 1982.
8. Esposito Vinzi, V., Chin, W.W., Henseler,J., and Wang. H, **Handbook of Partial Least Squares, Concepts, Methods and Application**, Springer Handbooks of Computational Statistics, New York, 2010.
9. Fornell, C., and Bookstein, F.L., **Two Structural equation model: LISREL and PLS applied to consumer exit-voice theory**. Journal of Marketing Research,19, 1982, 440-452.
10. Jöreskog, K.G., **Factor Analysis and Its Extensions. Factor Analysis at 100: Historical Developments and Future Decisions**. Department of Psychology, University of North Carolina, 2004, May 13-15.
11. Jöreskog, K.G., and Wold, H., **The ML and PLS Techniques for Modeling with Latent Variables: Historical and Comparative Aspects.** In: Systems under indirect observation: Causality, structure, prediction. Part I, 263-270. K.G. Jöreskog and H. Wold (Eds). Amsterdam: North Holland. 149, 1982.
12. Jöreskog, K.G., **"Testing Structural Equation Models"**. In: *Testing Structural Equation Models.* (Ed) Bollen, K.A and Lang, J.S., Beverly Hills, Sage, 1993, pp. 294-316.
13. Jöreskog, K.G and Sorbon, D., **LISREL 7: A Guide to the Program and Application***,* Chicago, IL, SPSS, Inc, 1989.
14. Jöreskog, K.G and Sorbon, D., ***LISREL 7: User's Reference Guide***. In: Scientific Software, Inc., Mooresville, 1989.

15. Takane, Y. and Shibayama T., **Principal Component Analysis with externalInformation on both subject and variables**, Psicometrica-vol.56, NO. 1, 1991, 97-120.
16. Temme, D., Kreis, H., and Hildebrandt, L., **PLS Path Modeling: A Software Review**. SFB 649 Discussion Paper 2006-084, Economic Risk.
17. Tenenhaus, M. (1999), **L'Approche PLS**, *Statistquèe Applique* XLVII(2), 1999, 5-40.
18. Tenenhaus, M. (2008) Structural Equation Modeling for small samples. HEC school of management (GRECHEC).
19. Tenenhaus, M., Esposito Vinzi, V., Chatelin,Y.M, and Lauro,C., **PLS Path Modeling**, Computational Statistics & Data Analysis 48, 2005, pp.159-205.
20. Wold, H., **Soft modeling: The Basic Design and Some Extensions.** In: *Systems under indirect observation: Causality, structure, prediction*. Part II, 1-54. K.G. Jöreskog and H. Wold (Eds). Amsterdam: North Holland, 1982.
21. Wold, H., **Non Linear estimation by iterative least squares procedures**. In: Researche Paper Statistcs, Festchrft for J. Neyman, 411-444. F.David (Ed).New York: Wiley, 1966.
22. Wold, H., **Estimation of Principal Components an Related Models by Iterative Least Squares**, In: Multivariate Analysis, Proceedings o fan International Symposium, 391-430, P.R. Krishmaiah (Ed). New York: Academic Press, 1966.
23. Sanchez, G., **Partial Least Squares**, Phd thesis, 2009
24. Zeithaml V.A., Parasuraman A., Berry L.L., **Servire la qualità**, McGraw-Hill, 1991.

TREND OF FOREIGN EXCHANGE EARNINGS BY THE VISITORS IN BANGLADESH

M. Taj UDDIN
Professor, Department of Statistics,
Shahjalal University of Science and Technology, Sylhet, Bangladesh

M. Anamul HUQ
MSc students, Department of Statistics,
Shahjalal University of Science and Technology, Sylhet, Bangladesh

M. Nazrul ISLAM
PhD, Associate Professor, Department of Statistics,
Shahjalal University of Science and Technology, Sylhet, Bangladesh

M. Ohid ULLAH
hD, Associate professor, Department of Statistics,
Shahjalal University of Science and Technology, Sylhet, Bangladesh.

E-mail: ohidullah@gmail.com

ABSTRACT

This study was an attempt to know the trend of foreign exchange earnings by the visitors in Bangladesh. To estimate this trend, secondary data was used for the period 1986-2008. Different statistical techniques and trend models were used to perform the analysis. An upward trend of exchange earning has been observed over the study period. The semi-log trend model was found to be the best fitted model for the foreign exchange earnings. The results showed significantly upward trend with a growth rate 18.05 percent per year. Thus it is gratifying that the tourism industry of the country is expanding day by day and to increase the growth rate government and non-government organization need to take proper action for better economy.

Key words: *Visitors, Trend, Foreign exchange, Tourism industry, Economy*

INTRODUCTION

Tourism has become a very important sector in the developing countries. Its growth affects not only linked to the activities of the tourism industry but also other sectors. Bangladesh is a developing country in Asia, holding high potentiality for tourism. Bangladesh Parjatan Corporation (BPC) plays an important role for the development of this sector. For a long time, Bangladesh has been an attractive destination for tourists. But at present, its position is not significant in terms of the international tourism market. In the last few decades, in particular, tourism has been developed by their integrated tourism planning (Buhalis, 1999; Butler, 2002; Vanhove, 2005).

In many countries, tourism is an industry for earnings revenues and foreign exchange (Hossain, 2007). Many businesses that grow concurrently with the development of tourism include airlines, shipping, hotels and restaurants, finance companies, tour operators, travel agents, car rental firms, caterers and retail establishments and together, they contribute significantly to the overall development of a country's economy and to its cultural diversification and adaptation (Islam, 2009). Tourism in its modern sense is a relatively recent phenomenon, and it has been begun in the present Bangladesh area only during the 1960s. Tourists from abroad came to see and enjoy the beaches, the scenic beauty of the landscape covered with lavish greens and the web of rivers, tribal culture, religious rituals, historical places, forests, wild life and hill resorts (Hasan, 2006). The private sector involvement in tourism in Bangladesh is still not adequate. However, the Bangladesh government has been taken remedial measures to encourage the private sector to play a positive role in the development and diversification of tourist facilities to promote domestic and international tourism in the country (Rahman, 2004). Tour operators have a significant role to play in tourism business. Over 40 private tour operators have already been engaged in tourism marketing in Bangladesh. Some of them conduct only domestic (inbound) tours while others offer both domestic and outbound tours (Bangladesh Monitor, 2007). Thirty two such private tour operators are members of an association named "Tours Operator Association of Bangladesh (TOAB)", formed to carry out their activities more efficiently, to lobby the BPC and the government for the realization of justified rights (Siddiqi, 2006), and to promote the country's tourism together. The natural beauty and panoramic views of Bangladesh is recognized worldwide to attract tourists. But due to lack of effective initiatives, proper management plan, quick and sincere effort of government this sector is not expanding up to the mark (Akther and Shelina, 2001). To take a proper decision for the betterment of this sector the level and nature of existing growth of foreign visitors should be addressed rigorously. So far our knowledge no study has been performed to measure the trend of foreign exchange earnings by the visitors in Bangladesh. Therefore, we aim to study the trend of foreign exchange earnings by the visitors in Bangladesh.

MATERIALS AND METHODS

Time series data on foreign exchange earning for the period (1986 – 2008) were collected from the publication of Bangladesh Parjatan Corporartion (BPC). To carry out the objective of the study different statistical measures and trend models were used to analyze the data.

METHODS OF DETERMINING TREND

To search a suitable model of foreign exchange earnings the following trend models have been considered.

(a) Straight-line trend: $Y_t = \beta_0 + \beta_1 t$

(b) Semi-log Trend/growth trend: $\log_e Y_t = \beta_0 + \beta_1 t$

(c) Exponential trend: $Y_t = \beta_0 . e^{\beta_1 t}$

(d) Semi-log parabolic Trend: $\log_e Y_t = \beta_0 + \beta_1 t + \beta_2 t^2$

(e) Logarithmic trend: $Y_t = \beta_0 + \beta_t \ln t$

(f) Parabolic Trend/quadratic trend: $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2$

Where, Y_t is the amount of foreign exchange earning by visitor in Bangladesh at time t (1986-2008).

These models are fitted separately to the annual data using regression analysis methodology. From these fitted models the most appropriate one is selected according to the following rule:

R^2 - Criteria: It is known that one of the measures of goodness of fit of a regression model is R^2, which, is defined as:

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

R^2 thus defined, of necessity lies between 0 and 1. The closer it is to 1, the better is the fit. But there are problems with R^2. First, it measures in-sample goodness of fit in the sense of how close an estimated Y value is to its actual value in the given sample. There is no guarantee that it will forecast well out of sample observations. Second, in comparing two or more R^2, the dependent variable, must be the same. Third, the more importantly, an R^2 can be increased when more explanatory variables are added to the model. Therefore, there is very temptation to play the game of "maximizing the R^2" by simply adding more variables to the model. Of course, adding more variables to the model may increase R^2 but it may also increase the variance of forecast error.

Adjusted R^2 Criteria: Each additional regressor variable added to the model increases R^2. Thus, since R^2 can be made larger simply by adding more predictor variables to the model, a modification of R^2 has been proposed. This adjusted R^2 does not automatically increase when new predictor variables are added to the model. In fact, the adjusted R^2 may actually decrease, because the decrease in ESS may be more than offset by the corresponding decrease in the error degrees of freedom.

$$\overline{R}^2 = 1 - \frac{ESS/(n-k)}{TSS/(n-1)}$$

Here, $\overline{R}^2 \leq R^2$, showing how the adjusted R^2 penalizes for adding more regressors. For comparative purposes, therefore, $\overline{R}^2$ is a better measure than R^2.

RATE OF GROWTH OF EXCHANGE EARNINGS:

An exponential model (Jeromi and Ramanathan,1993) was used to estimate growth rates for different segments of continuous time-series based on prior differentiation of rule-periods. The rate of growth implicit in the semi-log trend is derived by the solving the following equation.

$$\log_e(1+r) = \beta_1$$

Where, β_1 is the slope of the semi-log trend equation and r is the annual rate of growth.

RESULTS AND DISCUSSION

Trend in Foreign exchange earning by visitors

The trend models for the exchange earning in Bangladesh during the Year 1986-2008 was shown in Table-1.

Table 1: Trend models for the exchange earnings by the visitors

Types of Models	Estimated Coefficients			R^2	$\overline{R}^2$	F- Value
	β_0	β_1	β_2			
Linear	-1006.67 (251.757)	267.728* (18.361)		0.910	0.906	231.72*
Semi-log/growth Trend	5.173* (0.150)	0.166* (0.011)		0.917	0.913	231.03*
Logarithmic Trend	-1902.97 (726.85)	1831.31* (304.769)		0.632	0.615	36.11*
Semi-Log Parabolic Trend	4.965* (0.237)	0.216* (0.045)	-0.002 (0.002)	0.922	0.914	117.64*
Parabolic Trend (Quadratic Trend)	157.063 (235.417)	-11.569 (45.192)	11.637* (1.828)	0.970	0.967	326.59*

*Significant at 1% level.

Results showed that the value of $\overline{R}^2$ is the highest for parabolic trend model but the estimated co-efficient β_1 is not significant at 5% level. The second highest $\overline{R}^2$ value is 0.914 for the semi log parabolic trend model but β_2 is not significant at 5% level. Fulfilling all the criteria, the semi-log trend model appeared to be the most appropriate for the trend pattern of the series. Therefore the semi-trend model for the exchange earnings from the foreign tourist arrived in Bangladesh is

$$\log_e Y_t = 5.173 + 0.166t$$

Where, Y_t is the amount of exchange earnings (in million taka) in Bangladesh and t is time (in year).

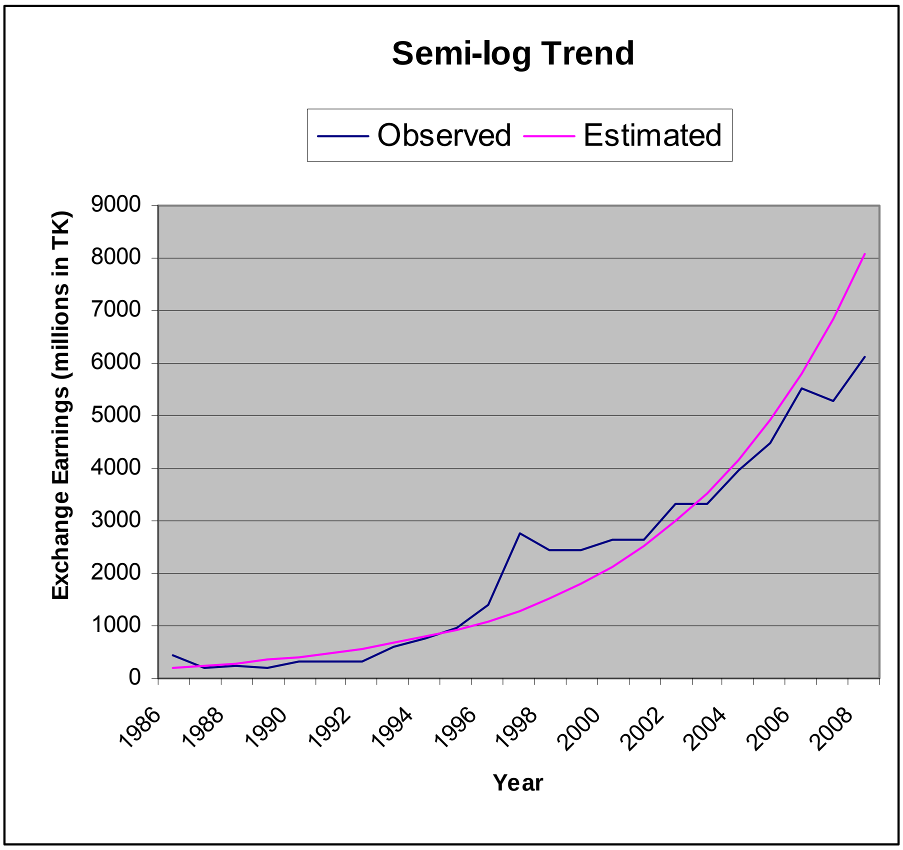

Figure 1: Semi log trend model of the exchange earnings by the visitors

From the above graph it is observed that the foreign exchange earning shows an upward trend during the study period. It is also observed that at the year 1986 the trend value was only 208.38 (millions TK) but after 23 years at the year 2008 the trend value is remarkable high and it is 8097.54 (millions TK). The above results of regression analysis clearly revealed that over the past 23 years the amount of exchange earnings in Bangladesh has trended up significantly.

Growth Rate of foreign exchange earnings

A study has been conducted to analyze the tourist profile and tourist expenditure pattern in Bangladesh between the period 1972-1990 by United Nations on 1993 (United Nations, 1993). Applying the descriptive analytical tools they found that international tourist arrivals in Bangladesh increased in absolute terms from 34,580 in 1972 to 115,369 in 1990 and the trend has fluctuated widely, recording even negative growth rates in some years. Another study has also been done to know the trend of foreign visitors in Bangladesh during the period 1976-2008 (Taj Uddin, 2012). They found that semi-log trend has been performed better with growth rate 5.97 percent per year. However, none study has been conducted to estimate the growth rate of foreign exchange earnings based on trend equations or models. Therefore, we tried to estimate the growth rate based on trend equations. We found that the rate of growth based on the semi-log trend (appropriate trend in this study) is 0.1805, indicated that the growth rate of the amount of exchange earnings from the foreign tourist arriving in Bangladesh between 1986 and 2008 had been 18.05 percent per year.

CONCLUSION

This study has investigated using the time series data to achieve its objective. The several trend models were used and found that the semi-log trend model was more suitable for foreign exchange earnings of Bangladesh. For the time period 1986-2008 it is found that the foreign exchange earning is significantly upward trend with a growth rate of 18.05 percent per year. Therefore, it can be recommended that the government, semi government and NGO's should take sufficient initiative to maintain this growth and further improvement of this sector for better economic development of the country. It is hopeful that the findings of this research may be considered by the government and private policy makers while formulating both short and long term policies for the development of tourism industry as well as the economy of Bangladesh.

ACKNOWLEDGEMENTS:

We would like to acknowledge Bangladesh Parjatan Corporation (BPC) to provide us data.

Conflict of interest: The authors declared that there is no conflict of interest for this study.

REFERENCES

1. Akhter and Shelina**, Tourism in Bangladesh: An Evaluation**. *Journal of Asiatic Society Bangladesh*, December, 2001.

2. Bangladesh Monitor, **Tourism: Sustainable tourism in developing countries**. *A travel trade journal on aviation, Tourism, Cargo-Shipping, Hotel-Restaurant and bank-commerce*, XVI, 17, 2007.

3. Buhalis, D., **Limits of tourism development in peripheral destinations: problems and challenges**. Tourism Management, University of Surrey, 1999, Accessed on 25th August

4. Butler, R. W., **The Development of Tourism in Frontier Regions: Issues and Approaches**. *Tourism in Frontier Areas*, 2002, 3-20

5. Hasan, S. R., **Problems and Prospect of Bangladesh Tourism Industry**. *Research Book- Bureau of Business Research*, University of Dhaka, Bangladesh, 2006.

6. Hossain, M. A. **Strategic Promotion Approaches to Developing Tourism in Bangladesh: An Empirical Study of Some Selected Tour Operators**. *Research Book-Bureau of Business Research*, University of Dhaka, Bangladesh, 2007

7. Islam, S., **Tourism potential in Bangladesh**, *The Daily Star*, Dhaka, Bangladesh, 2009.

8. Jeromi, P.D. and Ramanathan, A., **World paper market in India: An analysis of growth and instability**, *Indian Journal of Agriculture Economics*, 48(1), 1993, 88-97.

9. Rahman, M. M., **Prospects of Tourism Industry in Bangladesh**. *The Financial Express,* Dahak, Bangladesh, 2004.

10. Siddiqi, R., **Tourism in Bangladesh suffers: Policy makers lack understanding, government lacks strong intention**. *The daily Independent*, Dhaka, Bangladesh, 2006.

11. Taj Uddin, M., Anamul Huq, M and Baker Chowdhury, A., **Trend and variability of visitors in Bnagladesh**. *Journal of Statistics & Management Systems*, 15(6), 2012, 685-698.

12. United Nations, **The economic impact of tourism in Bangladesh**. *ST/ESCAP/1118,* 1993,
URL: http://www.unescap.org/ttdw/Publications/TPTS_pubs/pub_1181_tor.pdf

13. Vanhove, N., **Tourism and employment**, *International Journal of Tourism Management*, 2(3), 2005, 162-175

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

STYLIZED FACTS OF THE DAILY AND MONTHLY RETURNS FOR THE EUROPEAN STOCK INDICES DURING 2007-2012

Iulian PANAIT
PhD candidate, Department of International Business and Economics
The Bucharest Academy of Economic, Bucharest, Romania

E-mail: iulian.panait@gmail.com

Alexandru CONSTANTINESCU
PhD candidate, Department of International Business and Economics
The Bucharest Academy of Economic Studies, Bucharest, Romania

E-mail: alex.constantinescus@gmail.com

Abstract: *This study is intended to investigate the connection between the complexity of a capital market and the occurrence of dramatic decreases in transaction prices. The work hypothesis is that such episodes, characterized by sudden and dramatic decreases in transaction prices mostly occur in period of market inefficiency, when the level of complexity reaches a local minimum. In this regard, we introduce a complexity estimator, through differential entropy. The connection between the market complexity level and the appearance of extreme returns is illustrated in a logistic regression model.*

Key words: *differential entropy; stock market crash; logistic regression*

1. INTRODUCTION

This paper is dedicated to the study of the particularities in daily and monthly stock index returns for European markets during April 2007 and March 2012. Our comparative approach is based on three different dimensions. First we try to identify if the statistical behavior of the stock indices` returns is different between mature, emerging and frontier markets. Second, we look for particularities of monthly returns that are different from ones of the daily returns. Third, we document the specific behavior of the European stock indices` returns during the 2007-2009 stock market crisis in comparison with the evolution of the same indices after the March 2009 mid-term stock market bottom.

Over time, many investment managers and researchers studied the statistical characteristics of various stock market indices. Also, from obvious practical reason, the reactions of the stock markets on many types of previous financial and economic crisis were examined in detail by the research and academic community. More recently, starting with 2008, many authors showed interest to study the behavior of stock market returns during the 2007-2009 financial crisis.

Especially for investment managers, knowing the statistical characteristics of assets returns (such as mean, variance, skewness, kurtosis, the form of the distribution, the evolution in time of the correlation coefficients, the presence of autocorrelation in returns and squared returns etc.) represents an important step forward towards creating optimal portfolios.

During the last 30 years, the investment community was in particular interested by the less developed markets around the world, in search of larger profits and better portfolio diversification. Our study offers many useful details regarding the statistical behavior of stock market returns in particular for the emerging and frontier markets from Europe, during the recent, very difficult but also very relevant, period of time.

To cast some light on these issues, the rest of the paper is organized as follows: section 2 presents the most relevant Romanian and international related studies; section 3 describes the data that we worked with and the methodology that we have used; section 4 presents the results that we have obtained; finally section 5 summarizes the most important conclusions and proposes further studies in this field.

2. LITERATURE REVIEW

In 1998 Bekaert G., Erb C. B., Harvey C.R. and Vyskanta T.E. identify some clear differences of yield evolution in emerging markets: volatility high, low intensity correlation with mature markets and between emerging markets with each other, long-term high yields, greater predictability than can be achieved in mature markets, more likely to be influenced by external shocks (legislative, policy or exchange rate) [2].

Also, Bekaert G. and Harvey C.R. (1997) analyze the reasons that volatility is different across emerging markets, particularly with respect to the timing of capital market reforms. They argue that capital market liberalizations often increase the correlation between local market returns and the world market, but do not drive up local market volatility [1].

In 2001 Cont R. shows several features of logarithmic returns for a sufficient number of financial assets to believe that they have the character of generality. Account features specified refers to the absence autocorrelations in returns, high probabilities for extreme events (or thick tails of the distribution – "heavy tails"), asymmetry, higher values for standard deviation in comparison with the mathematical simple average, positive autocorrelation in squared returns and variance, leverage, correlation dependence time etc. [5].

Gelos R.G. and Sahay R. (2001) examined financial market co-movements across European transition economies and compared their experience to that of other regions. They found that correlations in monthly indices of exchange market pressures can partly be explained by direct trade linkages, but not by measures of other fundamentals [7].

Forbes K.J. and Rigobon R. (2002) argue that there is a high level of market co-movement during all periods, which he calls "interdependence". Previous research suggested that contagion (defined as a significant increase in market co-movement after a shock to one country) it is often occurring during crises. Forbes and Rigobon's paper is in opposition with that belief and shows that there was virtually no increase in unconditional correlation coeficients (i.e., no contagion) during the 1997 Asia crisis, 1994 Mexican devaluation and 1987 U.S. market crash [6].

Maroney N., Naka A. and Wansi T. explored risk and return relations in six Asian equity markets affected by the 1997 Asian financial crisis and found that after the start of the crisis, national equity betas increased (due to leverage linked to exchange rates) and average returns fell substantially. Subsequently, the authors propose a new probability-based asset pricing model that captures leverage effects using valuation ratios. Their results show the role of leverage in explaining the likelihood of the financial crises [10].

Hartmann P., Straetmans S. and de Vries C.G. (2004) characterize asset return linkages during periods of stress by an extreme dependence measure. Their estimates for the G-5 countries suggest that simultaneous crashes between stock markets are much more likely than between bond markets. Also, their data show that stock-bond contagion is approximately as frequent as flight to quality from stocks into bonds. Also, they found that extreme cross-border linkages are surprisingly similar to national linkages, illustrating a potential downside to international financial integration [9].

Latter, Bekaert G., Harvey C.R. and Ng A. (2005) studies contagion and propose a two-factor model with time-varying betas that accommodates various degrees of market integration. The authors apply this model to stock returns in three different regions: Europe, Southeast Asia, and Latin America. In addition to examining contagion during crisis periods, they document time variation in world and regional market integration and measure the proportion of volatility driven by global, regional, and local factors [2].

Pop C., Curutiu C. and Dumbrava P. (2009) present the Bucharest Stock Exchange evolution before the 2007-2009 crisis started to manifest and try to identify the main factors which influenced its explosive growth. The paper investigates the current financial crisis influences on Bucharest Stock Exchange – with an emphasis over the factors which might have deepen the descendent trend for the Romanian stock exchange market. The authors also present the effects of the current financial crisis on the future development of Bucharest Stock Exchange, taking into consideration the position of the Romanian capital market in Eastern Europe [12].

Harrison B., Lupu R., and Lupu I. (2010) studied the statistical properties of the CEE stock market dynamics using a panel data analysis and found that there is evidence of stationarity for the returns provided by the Romanian stock indices. They have also identified some particular characteristics of returns in these markets such as a great amount of non-linearity and cross correlation [8].

3. DATA AND METHODOLOGY

In our study we used the non-tradable stock market indices computed by the international financial advisory company MSCI Barra. The time series of daily and monthly prices for all the MSCI Barra indices are freely available at the company`s website www.msci.com and we were able to collect such daily and monthly data for the period April 2007 – March 2012.

For our research purpose we have selected 16 European stock markets, 2 international markets and 3 global stock market indices (needed in order to be able to make comparisons of the results). All those 21 indices were grouped in three categories: 6 developed market indices, 6 emerging market indices and 6 frontier market indices.

Because the price time series are not stationary, we preferred to transform all the 21 price time series into returns time series.

Regarding the returns estimation, as Strong (1992, p.353) pointed out "there are both theoretical and empirical reasons for preferring logarithmic returns. Theoretically, logarithmic returns are analytically more tractable when linking together sub-period returns to form returns over long intervals. Empirically, logarithmic returns are more likely to be normally distributed and so conform to the assumptions of the standard statistical techniques."[13]. This is why we decided to use logarithmic returns in our study since one of our objectives was to test of whether the daily returns were normally distributed or, instead, showed signs of asymmetry (skewness). The computation formula of the daily returns is as follows:

$$R_{i,t} = Ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right) \quad (1)$$

where $R_{i,t}$ is the return of asset *i* in period *t*; $P_{i,t}$ is the price of asset *i* in period *t* and $P_{i,t-1}$ is the price of asset *i* in period *t-1*. As already mentioned above, according to this methodology of computing the returns, the prices of the assets must be adjusted for corporate events such as dividends, splits, consolidations and share capital increases (mainly in case of individual stocks because indices are already adjusted).

As a result of this initial data gathering we obtained 21 time series of log-returns, each with 1295 daily observations and 60 monthly observations.

For those 21 time series and two return frequencies we have computed the mean, standard deviation, skewness and kurtosis and also we have applied the Jarque Bera test of the normality of distribution of the daily returns.

For a financial time series the mean represents the simple mathematical average of all the observations within the sample. It is obtained by adding up the series and dividing the result by the number of observations.

The standard deviation of a financial time series is a measure of dispersion or spread in the series. The standard deviation is computed by:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(R_i - \bar{R})^2}{N-1}} \quad (2)$$

Where N is the sample size, R_i represents the individual observations of daily returns, and $\bar{R}$ represents the sample mean computed as above.

Concerning the estimation of skewness, according to most authors a time series of financial asset returns is symmetric around it's mean (noted here with μ) if:

$$\forall k, f(\mu + k) = f(\mu - k) \quad (3)$$

where f is the density function of the returns. If this property is valid then the mean of the returns series coincides with its median.

The skewness of a data population is defined as the third central moment. To be more precise, skewness is computed as the average cubic deviation of the individual observations from the sample mean, divided by the standard deviation raised to the third power. As a consequence of these considerations, we have calculated the sample skewness as follows:

$$\hat{S} = \frac{\frac{1}{N}\sum_{i=1}^{N}(R_i - \bar{R})^3}{\hat{\sigma}^3} \quad (4)$$

where $\hat{S}$ is the sample skewness; N is the total number of individual observations within the sample, R_i is the return of period t, $\bar{R}$ is the sample arithmetic mean and $\hat{\sigma}$ is an estimator for the standard deviation that is based on the biased estimator for variance ($\hat{\sigma} = \sigma\sqrt{(N-1)/N}$)

The skewness of a symmetric distribution, such as the normal distribution, is zero. Positive skewness means that the distribution has a long right tail and negative skewness implies that the distribution has a long left tail.

According to Peiro (1999), under normality hypothesis, the asymptotic distribution of $\hat{S}$ is given by $\hat{S} \rightarrow N(0, \frac{6}{T})$.

Kurtosis is a measure of how outlier-prone a distribution is. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3 [11].

The kurtosis of a distribution is defined as

$$K = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{R_i-\bar{R}}{\hat{\sigma}}\right)^4 \qquad (5)$$

where $\bar{R}$ is the mean of R_i, $\hat{\sigma}$ is the standard deviation of R_i, and N is the sample size. The kurtosis of the normal distribution is 3. If the kurtosis exceeds 3, the distribution is peaked (leptokurtic) relative to the normal. If the kurtosis is less than 3, the distribution is flat (platykurtic) relative to the normal.

The Jarque-Bera test is a two-sided goodness-of-fit test suitable when a fully-specified null distribution is unknown and its parameters must be estimated. The test statistic is

$$JB = \frac{N}{6}\left(s^2 + \frac{(k-3)^2}{4}\right) \qquad (6)$$

where *N* is the sample size, *s* is the sample skewness, and *k* is the sample kurtosis. For large sample sizes, the test statistic has a chi-square distribution with two degrees of freedom. The reported probability (p-value) is the probability that a Jarque Bera statistic exceeds (in absolute value) the observed value under the null hypothesis. A small probability value leads to the rejection of the null hypothesis of a normal distribution.

4. RESULTS AND INTERPRETATIONS

The first step in investigating the properties statistical represents the calculation of the averages, variances, the asymmetry coefficients and the flattening coefficient, according to the methods described in the methodology section. The results obtained for the daily series of returns are presented in the table below. The specific parts for the monthly data will be will be discussed in the second part of this study.

Based on the table below we can already confirm that, for all the European stock markets included in this study, even though we speak about mature markets, emerging or frontier markets, even though the study is done on general indexes or on individual markets, the average of the long-term daily returns tends to zero. Also, included for all the markets in this study, we confirm that the average is statistically significantly close to the value of the median. Apart from observing the effective values from Table 1, these statements have been confirmed also by running the t-statistic test for the hypothesis of an average equal to 0 and respectively by the Sign, Wilcoxon and Van der Wareden tests for the hypothesis of a median equal to 0 in the case of all the 21 assets.

Table 1. Descriptive statistics for the series of daily returns

		Medie	Mediană	Maxim	Minim	Std. Dev.	Skew-ness	Kurtosis	Jarque-Bera	p-v.
Mature markets	Austria	-0.0010	0.0000	0.1277	-0.1119	0.0225	-0.026	6.951	841.994	0
	Franta	-0.0004	0.0000	0.1036	-0.0932	0.0174	0.099	7.899	1,296.085	0
	Germania	-0.0002	0.0000	0.1113	-0.0739	0.0169	0.130	8.245	1,486.650	0
	Italia	-0.0008	0.0000	0.1100	-0.0864	0.0184	0.043	7.269	982.833	0
	UK	-0.0003	0.0004	0.0950	-0.0938	0.0164	-0.106	8.711	1,761.008	0
	SUA	0.0000	0.0007	0.1044	-0.0915	0.0165	-0.151	9.026	1,962.978	0
	DM_Index	-0.0001	0.0007	0.0850	-0.0696	0.0124	-0.220	8.684	1,752.461	0
EM	China	0.0001	0.0002	0.1404	-0.1171	0.0224	0.171	7.665	1,179.811	0
	Cehia	-0.0002	0.0004	0.1675	-0.1568	0.0192	-0.287	16.662	10,081.670	0
	Ungaria	-0.0006	-0.0002	0.1733	-0.1999	0.0270	-0.028	9.208	2,077.982	0
	Polonia	-0.0004	0.0000	0.1125	-0.1124	0.0222	-0.223	6.335	610.344	0
	Rusia	-0.0002	0.0007	0.2376	-0.2334	0.0274	-0.294	18.032	12,201.200	0
	Turcia	0.0000	0.0002	0.1484	-0.1243	0.0246	-0.065	6.736	753.310	0
	EM_Index	0.0001	0.0006	0.1008	-0.0848	0.0151	-0.123	9.051	1,977.405	0
Frontier markets	Bulgaria	-0.0015	0.0000	0.1105	-0.1605	0.0193	-1.456	15.560	8,962.712	0
	Croatia	-0.0005	-0.0002	0.0998	-0.0803	0.0130	-0.199	13.701	6,182.963	0
	Estonia	-0.0005	0.0000	0.1241	-0.0924	0.0192	0.244	7.264	992.929	0
	Romania	-0.0007	0.0000	0.1043	-0.3358	0.0242	-2.179	32.879	49,158.980	0
	Slovenia	-0.0007	0.0000	0.0915	-0.0883	0.0139	-0.404	11.285	3,735.895	0
	Serbia	-0.0013	-0.0008	0.1502	-0.1725	0.0247	-0.352	11.196	2,847.822	0
	FM_Index	-0.0004	0.0004	0.0458	-0.0688	0.0104	-1.187	9.449	2,546.213	0

Source: MSCI Barra, calculations made by the authors

The results of the Sign, Wilcoxon and Van der Wareden tests from above are presented in Table 2 and show that we can not reject the null hypothesis (the mean = 0 and that median = 0) at the maximum permissible error level of 1% for none of the markets included in the study.

In this situation, were we can not affirm that for all the investigated assets, the average and the median for the daily returns does not differ significantly from zero, we can also statistically test if the affirmation that states that the averages and the medians for all the 21 temporal series with daily frequency are equal.

In order to test the equality of the medians we used the F test (ANOVA version and the Wech version), and for equal medians we used Chi squared tests, Kruskal-Wallis and Van der Waerden as seen in Table 3. The results for these tests are presented below in table 3 and show indeed that we have additional statistical arguments to state that medians and averages of the series for daily returns in all 21 studied assets are equal and have the value zero.

Table 2. The results for the statistic tests for the null hypothesis for the average=0 and the median=0 for the of daily returns series

		average=0	median=0			
		t-statistic	Sign (exact binomial)	Sign (normal approx.)	Wilcoxon signed rank	van der Waerden (normal scores)
		p-value	p-value	p-value	p-value	p-value
Mature markets	Austria	0.1297	0.627	0.627	0.385	0.2314
	Franta	0.3651	0.9328	0.9328	0.7036	0.4759
	Germania	0.5985	0.4139	0.4139	0.8543	0.8581
	Italia	0.1075	0.9775	0.9775	0.3759	0.1942
	UK	0.5696	0.1688	0.1688	0.8069	0.8561
	SUA	0.9677	0.0231	0.0231	0.2718	0.5542
	DM_Index	0.6927	0.0241	0.0241	0.2735	0.7173
Emergent markets	China	0.8406	0.4512	0.4512	0.7187	0.8293
	Cehia	0.7092	0.3873	0.3873	0.724	0.9393
	Ungaria	0.4278	0.6551	0.6551	0.4414	0.4013
	Polonia	0.4777	0.8011	0.8011	0.9433	0.7087
	Rusia	0.7712	0.3156	0.3156	0.4561	0.7295
	Turcia	0.9551	0.7174	0.7174	0.7222	0.8872
	EM_Index	0.8556	0.0846	0.0847	0.2327	0.4815
Frontier markets	Bulgaria	0.0146	0.1298	0.1298	0.0373	0.0253
	Croatia	0.1527	0.0547	0.0547	0.2115	0.2062
	Estonia	0.3382	0.0652	0.0652	0.1077	0.1431
	Romania	0.3247	0.9554	0.9554	0.7718	0.6253
	Slovenia	0.055	0.1528	0.1528	0.0611	0.0632
	Serbia	0.0925	0.0184	0.0184	0.0106	0.0235
	FM_Index	0.1632	0.1259	0.126	0.5437	0.8401

Source: MSCI Barra, calculations made by the authors

Returning to the other statistical characteristics of daily returns on stock markets, all from Table 1 we observe that for all the 21investigated assets, the standard deviation is higher than the value of the average (which we saw above that we can approximate to zero). This study confirms similar findings of previous research.

Interestingly, the results presented in Table 1 show us that we do not have enough statistical arguments clear to affirm that volatility (risk), measured by standard deviation (and implicitly of the variance) is higher for emerging stock markets. Although, as you can observe, standard deviation values for most of the six mature stock markets included in the study present lower figures compared with those of emerging and frontier markets, however the standard deviation for the daily returns in Austria is 0.0225 which exceeds the values for most emerging and frontier markets. This unusual situation is maintained also when we analyze the global indices.

The statistics presented in Table 1 show that for all the 21 assets studied, the kurtosis value (coefficient of vaulting) is higher than 3 (the specific value of normal distribution). This situation shows that the distributions of stock's daily returns are mostly leptokurtic, sharper than the normal distribution, with many values concentrated around the average values and thicker tails means high probability for extreme values (i.e. higher risks). Within the sample that contained mature markets, the kurtosis value does not exceed 10, showing the lowest levels. The highest levels of kurtosis are found for the frontier stock markets, which according to the figures from above signify a higher risk of investments in undeveloped markets compared to mature markets, findings that confirm results of previous similar studies.

Table 3. The results of statistical tests for the equality of averages and medians for the series of daily returns

	Anova F-test p-value	Welch F-test p-value
Null hypothesis: "all averages are equal"	0.924	0.907

	Med. Chi-squared	Kruskal-Wallis	van der Waerden
Null hypothesis: "all medians are equal"	0.000	0.267	0.617

Source: MSCI Barra, calculations made by the authors

Taking into account that the period analyzed in this study includes both a crisis cycle on the stock markets (with large and persistent declines in the period starting from June 2007 until February 2009) and an accelerated growth phase (between March 2009 - April 2012), offers us interest to study the behavior of standard deviation and maximum amplitude of variation for the two different stages. The result of this investigation is shown below in Table 4. and indicates that for all the 21 assets, the maximum variation amplitude during a trading session was lower during the upward trend compared to the crisis period.
At same the time we observe that for all 21 assets, their corresponding standard deviations had lower values during the upward trend compared with the values during the crisis.

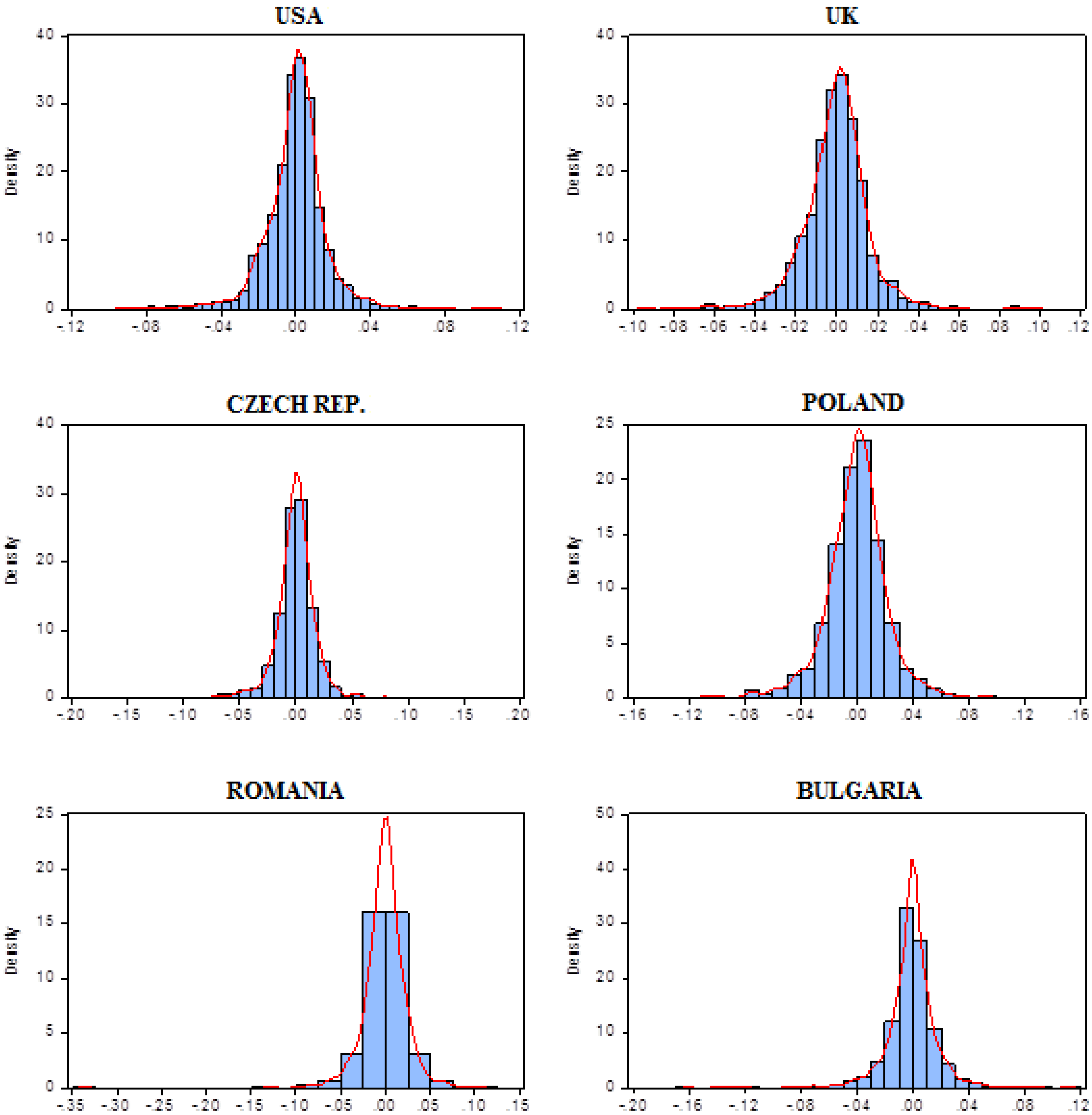

Figure 1 Comparison between actual probability distributions and the normal distribution of series of daily returns
Source: MSCI Barra, calculations made by the authors

Therefore this study confirms previous research findings, according to which the volatility of daily returns is amplified during downturns. The behavior characteristic of the daily stock returns is easily visible in Figure 3. Likewise, we observe that periods of high variance correspond to periods of high amplitude for daily returns.

Table 4. The evolution of volatility and business cycle asymmetry on the types of stock markets

		Standard period		Only the crisis			Only the upward trend		
		Std. Dev.	Skewness	Ampl. max	Std. Dev.	Skewness	Ampl. max	Std. Dev.	Skewness
Mature markets	Austria	0.0225	-0.026	0.1277	0.0269	0.098	0.0962	0.0195	-0.037
	Franta	0.0174	0.099	0.1036	0.0203	0.231	0.0883	0.0154	0.018
	Germania	0.0169	0.130	0.1113	0.0193	0.450	0.0601	0.0152	-0.161
	Italia	0.0184	0.043	0.1100	0.0193	0.411	0.1043	0.0179	-0.207
	UK	0.0164	-0.106	0.0950	0.0211	0.048	0.0642	0.0128	-0.146
	SUA	0.0165	-0.151	0.1044	0.0218	0.000	0.0693	0.0124	-0.188
	DM_INDEX	0.0124	-0.220	0.0850	0.0158	-0.029	0.0522	0.0097	-0.270
Emergent markets	China	0.0224	0.171	0.1404	0.0301	0.221	0.0648	0.0163	0.086
	Cehia	0.0192	-0.287	0.1675	0.0254	-0.257	0.0731	0.0144	0.018
	Ungaria	0.0270	-0.028	0.1999	0.0303	-0.146	0.1478	0.0249	0.172
	Polonia	0.0222	-0.223	0.1125	0.0251	-0.306	0.0985	0.0202	-0.023
	Rusia	0.0274	-0.294	0.2376	0.0356	-0.189	0.1018	0.0210	-0.148
	Turcia	0.0246	-0.065	0.1484	0.0323	0.055	0.0834	0.0186	-0.096
	EM_Index	0.0151	-0.123	0.1008	0.0197	0.014	0.0498	0.0115	-0.093
Frontier markets	Bulgaria	0.0193	-1.456	0.1605	0.0255	-1.536	0.0686	0.0143	0.132
	Croatia	0.0130	-0.199	0.0998	0.0170	0.031	0.0736	0.0098	-0.280
	Estonia	0.0192	0.244	0.0828	0.0193	-0.410	0.1241	0.0190	0.662
	Romania	0.0242	-2.179	0.3358	0.0291	-3.178	0.1285	0.0205	-0.056
	Slovenia	0.0139	-0.404	0.0915	0.0190	-0.282	0.0670	0.0098	-0.231
	Serbia	0.0247	-0.352	0.1725	0.0379	-0.354	0.1233	0.0200	0.450
	FM_Index	0.0104	-1.187	0.0688	0.0127	-1.511	0.0458	0.0088	-0.214

Source: MSCI Barra, calculations made by the authors

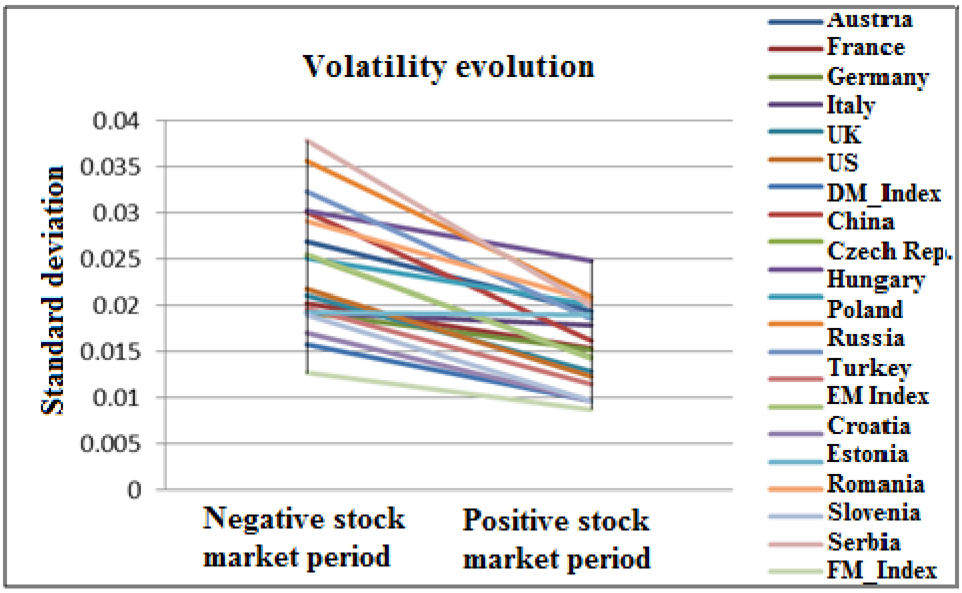

Figure 2.

Source: MSCI Barra, calculations made by the authors

In order to highlight the evolution of the correlation coefficients we used a sample size calculation "rolling" of 130 days (the equivalent of six calendar months of stock trading). The result is shown below in Figure 3 and demonstrates that the value of the correlation coefficient varies over time and their evolution is likely influenced by the stock market situation. For example we observe that during periods of declining stock markets (2007-2008 and then the end of 2010 until mid 2011) the intensity of correlations between all types of markets has been growing, while during periods of an upward trend (2009 and early 2012) the correlation coefficient values are reduced.

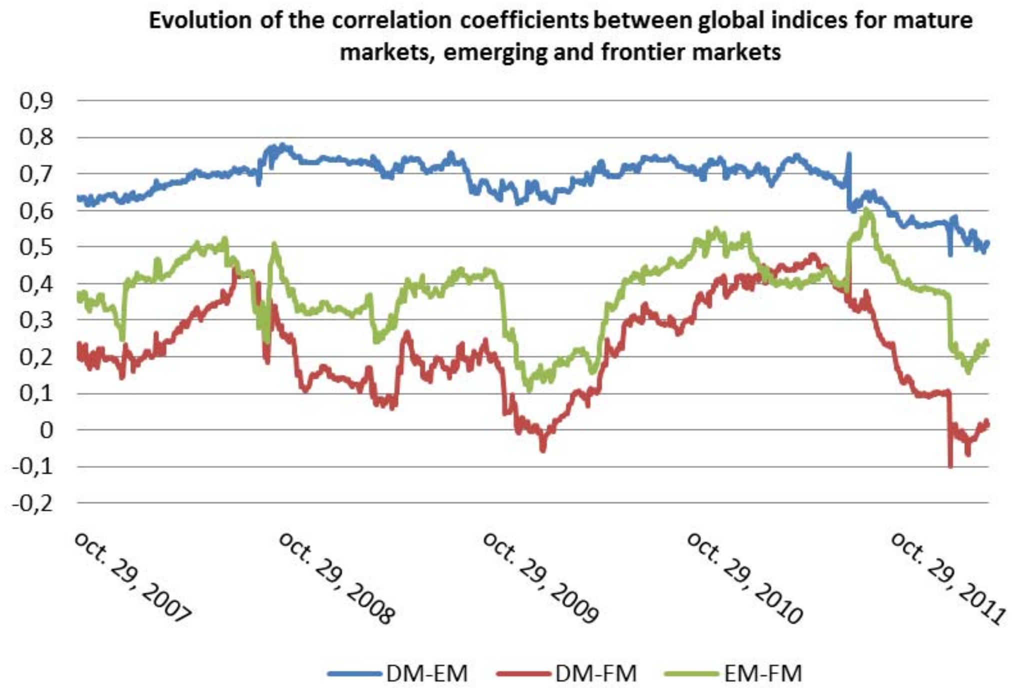

Figure 3. Source: MSCI Barra, calculations made by the authors

We test the presence of the phenomenon of autocorrelation for daily returns by using AC functions ("autocorelation") and CAP ("partial-autocorelation"), for correlations between the current returns and the previous 100 past returns, for all the 21 assets that were analyzed. The values of autocorrelation coefficients and partial autocorrelation respectively show that the phenomenon of autocorrelation in daily returns is not present.

In the second part of the paper we analyze the behavior of low frequency returns (using monthly data) for the 21 assets. Similar to the approach in the first part of this paper, we have calculated the average, variance, the coefficient of asymmetry (skewness) and the flattening coefficient (kurtosis) for the monthly returns. The results are presented in Table 5 below.

As it can also be observed in the case of the monthly returns, we do not have enough statistic arguments to affirm that the average is not equal to zero. This is confirmed both by the values offered in the table and by the t-statistic test applied to each of the 21 monthly time series. At same the time, similar to the situation of the daily returns, the F statistical test (ANOVA and Welch variants) indicate that the value of the averages are equal for all 21 series of monthly returns that were analyzed.

Table 5. Descriptive statistics for the series of monthly returns

		Average	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	p-value
MM	Austria	-0.0197	-0.0150	0.1758	-0.3650	0.0997	-1.07	5.22	23.3857	0.000
	France	-0.0085	-0.0087	0.1197	-0.1517	0.0594	-0.36	2.74	1.4339	0.488
	Germany	-0.0045	0.0057	0.1451	-0.2075	0.0681	-0.68	3.75	5.9795	0.050
	Italy	-0.0164	-0.0233	0.1714	-0.1672	0.0684	0.04	2.97	0.0151	0.992
	UK	-0.0055	0.0010	0.1167	-0.1216	0.0518	-0.32	2.91	1.0090	0.604
	UD	-0.0002	0.0010	0.0922	-0.1029	0.0469	-0.36	2.45	1.9981	0.368
	DM_INDEX	-0.0027	0.0029	0.1055	-0.1090	0.0463	-0.39	2.92	1.5453	0.462
Emerging markets	China	0.0021	0.0116	0.1605	-0.2555	0.0865	-0.76	3.41	6.1297	0.047
	Czech Rep.	-0.0039	-0.0074	0.1685	-0.2465	0.0721	-0.37	4.66	8.1727	0.017
	Hungary	-0.0116	0.0069	0.2259	-0.4660	0.1185	-0.96	5.37	22.7556	0.000
	Poland	-0.0092	-0.0014	0.2356	-0.3110	0.0992	-0.33	4.00	3.5361	0.171
	Russia	-0.0049	0.0117	0.1997	-0.3327	0.1064	-0.65	3.54	4.9252	0.085
	Turkey	-0.0005	0.0033	0.2588	-0.3183	0.1182	-0.23	3.52	1.1591	0.560
	EM_Index	0.0016	0.0068	0.1528	-0.2193	0.0695	-0.69	3.83	6.3317	0.042
Frontier markets	Bulgaria	-0.0326	-0.0108	0.2339	-0.5339	0.1179	-1.60	8.22	92.0749	0.000
	Croatia	-0.0115	-0.0154	0.1835	-0.2659	0.0754	-0.63	5.78	22.8355	0.000
	Estonia	-0.0117	0.0062	0.4322	-0.3796	0.1158	0.34	6.61	33.2325	0.000
	Romania	-0.0146	0.0186	0.2774	-0.5980	0.1423	-1.40	6.71	52.9945	0.000
	Slovenia	-0.0163	-0.0108	0.1645	-0.1712	0.0622	-0.37	4.11	4.3181	0.115
	Serbia	-0.0278	-0.0071	0.3537	-0.6457	0.1719	-1.13	5.99	26.8283	0.000
	FM_Index	-0.0089	-0.0004	0.1001	-0.1981	0.0600	-0.96	4.63	15.6336	0.000

Source: MSCI Barra, calculations made by the authors

The following conclusion is drawn from Table 5 as the monthly data confirms the hypothesis that the value of the standard deviation is significantly higher than the average. At the same time, 19 of the 21 series of monthly returns validate the property of a skewness figure that has negative values, although these values do not have a large dimension. However in terms of flattening coefficient (kurtosis), we observe that unlike the case of the series of daily returns, the monthly returns offer values that are much closer to the value three (the characteristic value of a normal distribution) for 10 of the 21 active investigation. This observation, together with the previous one according to which the skewness values do not have a large dimension, lead us to expect that the form of monthly returns distribution is closer to the normal distribution, which represents an important value for processing and modeling their behavior.

Indeed, the values of the last column in Table 5 show that the Jarque-Bera test results lead to the conclusion that, for 13 of the 21 series of monthly returns analyzed, the hypothesis in which the distribution is described in a Gaussian waveform can not be rejected at an error level of maximum 1%. We obtain statistical arguments to assert that, for more than half of the series of monthly returns analyzed, the shape of the distribution curve does not differ significantly from the normal (theoretical) distribution.

5. CONCLUSIONS

Our paper investigate the statistical characteristics of daily and monthly returns during April 2007 – March 2012 for 16 European national market indices, 2 international indices and 3 global market indices. We compared the results between three categories: developed markets, emerging markets and frontier markets.

(1) The data that we investigate confirmed that the average of returns is not statistically different from zero. This finding is valid both for daily and monthly returns. Also, it is valid for all the three types of markets (developed, emerging and frontier).

(2) Our results also confirm that standard deviation consistently registers higher values comparing with the average, both for daily and monthly returns. We noticed that the developed markets have lower values for standard deviation in comparison with emerging and frontier markets.

(3) For all the types of markets the distribution of daily returns is significantly different from the normal (theoretical) distribution. At the same time, we found evidence that the lower frequency returns (in our case the monthly returns) tend to have empirical distributions close to the normal (theoretical) distribution. For the developed markets the monthly returns are close to the normal distribution, but the monthly returns of the emerging and frontier markets still differ significantly.

(4) We found negative asymmetry for most of the 21 indices investigated, both for the daily and monthly returns.

(5) The daily returns present excess kurtosis for most of the indices and for all types of markets. This conclusion is also valid for monthly returns from emerging and frontier markets. Not surprisingly, the monthly returns of the developed markets (which we found to have empirical distribution close to the normal distribution) have kurtosis values near 3.

(6) For the daily returns we were able to confirm the „leverage" stylized fact described by Cont R. (2001). More specific, we found that during the high volatility periods, the absolute values of effective returns are also higher. We were unable to test this property for the monthly returns because during the period April 2007 – March 2012 we had only 60 empirical observations for each of the 21 time series.

(7) The study that we have conducted confirms that most of the characteristics of the returns change with time. Especially volatility and correlation coefficient tend to register higher values during market crises and lower values during the periods of positive market evolution. This confirms the hypothesis of contagion between markets. Also, we found that mature markets are highly correlated with other mature markets and less correlated with emerging and frontier markets. On the other hand, the frontier markets tend to have lower correlations both with other frontier markets and with emerging and developed markets.

(8) The daily time series show no autocorrelation of simple logarithmic returns, but present autocorrelations of squared returns. On the other hand, we found that the monthly squared returns tend to be less autocorrelated.

ACKNOWLEDGMENT

This paper is the result of the project "Doctorat în economie la standardele Europei cunoaşterii (DOESEC)", POSDRU/88/1.5./S/55287, project co-financed by the European Social Fund Operational Programme Human Resources Development 2007-2013,

coordinated by the Academy of Economic Studies from Bucharest in partnership with the University of West from Timisoara.

REFERENCES

1. Bekaert G., Harvey C.R. **Emerging Equity market volatility**, Journal of Financial Economics, issue 43, 1997, pp. 29-77

2. Bekaert G., Harvey C.R., Ng A. **Market integration and contagion**, Journal of Business, vol. 78, issue 1, 2005, pp. 39-69

3. Bekaert G.E., Erb C. B., Harvey C.R., Vyskanta, T.E. **Distribution Characteristics of Emerging Market Returns and Asset Allocation**, Journal of Portfolio Management, issue 24, 1998, pp.102-116

4. Conover C., Mitchell J.G.R., Johnson R.R. **Emerging Markets: When Are They Worth It?**, Financial analysts Journal, 2002, pp.86-95

5. Cont R., **Empirical properties of asset returns: stylized facts and statistical issues**, Quantitative Finance, vol.1, no.2, 2001, pp. 223-236

6. Forbes K.J., Rigobon R. **No contagion, only interdependence: Measuring stock market co-movements**, Journal of Finance, vol. 57, issue 5, 2002, pp. 2223-2261

7. Gelos R.G., Sahay R. **Financial market spillovers in transition economies**, Economics of Transition, vol. 9, issue 1, 2001, pp. 53-86

8. Harrison B., Lupu R., Lupu I. **Statistical Properties of the CEE Stock Market Dynamics. A Panel Data Analysis**, The Romanian Economic Journal, issue 37, 2010, pp. 41-54

9. Hartmann P., Straetmans S., de Vries C.G. **Asset market linkages in crisis periods**, Review of Economics and Statistics, vol. 86, issue 1, 2004, pp. 313-32611

10. Maroney N., Naka A., Wansi T. **Changing risk, return and leverage: The 1997 Asian financial crisis**, Journal of Financial and Quantitative Analysis, vol 39, issue 1, 2004, pp. 143-166

11. Peiro A. **Skewness in Financial Returns**, Journal of Banking and Finance, issue 23, 1999, pp. 847–862

12. Pop C., Curutiu C., Dumbrava P. **Bucharest Stock Exchange – The effects of the current financial crisis**, Interdisciplinary Management Research V, 2009, pp. 805-818

13. Strong N. **Modeling Abnormal Returns: A Review Article**, Journal of Business Finance and Accounting, vol. 19, issue 4, 1992, pp. 533–553