ORIENTATION OF PUBLIC POLICY BY HIGHLIGHTING THE RELATIONSHIP BETWEEN ECONOMIC DEVELOPMENT, EDUCATION AND ECO-INNOVATION IN EU COUNTRIES

POWER AND SIZE ANALYSIS OF CO-INTEGRATION TESTS IN CONDITIONAL HETEROSKEDASCITY: A MONTE CARLO SIMULATION

HIERARCHIES OF ASSOCIATIVE DYNAMICS, STARTING FROM ROMANIA'S MACRO-ECONOMIC IMBALANCES IN THE EU-28. WHAT DOES ROMANIA'S ECONOMIC EVOLUTION IN THE EU-28 LOOK LIKE?

PRESERVING LOGICAL RELATIONS WHILE ESTIMATING MISSING VALUES

SIMULATION PROCESS, THEORETICAL PARADIGM AND OPERATIONAL-STRATEGIC REALITY AS TOOL FOR MANAGERIAL DECISION MAKING

IMPACT OF DIVIDEND TAXATION CHANGING ON DIVIDEND POLICY OF ROMANIAN COMPANIES LISTED ON STOCK MARKET

THEORETICAL ASPECTS REGARDING THE OPTIMAL TAXATION OF EFFORT WITH MORE CONDITIONS

3/2017

www.revistadestatistica.ro
CONTENTS 3/2017

ORIENTATION OF PUBLIC POLICY BY HIGHLIGHTING THE RELATIONSHIP BETWEEN ECONOMIC DEVELOPMENT, EDUCATION AND ECO-INNOVATION IN EU COUNTRIES 3
Florin VĂDUVA
Titu Maiorescu University, Bucharest, Romania
Rodica GHERGHINA
Bucharest University of Economic Studies, Romania
Ioana DUCA
Titu Maiorescu University, Bucharest, Romania

POWER AND SIZE ANALYSIS OF CO-INTEGRATION TESTS IN CONDITIONAL HETEROSKEDASCITY: A MONTE CARLO SIMULATION 17
Osabuohien-Irabor OSARUMWENSE
Ambrose Alli University (A.A.U), Ekpoma, Edo, Nigeria
Julian I. MBEGBU
University of Benin, Benin City, Nigeria

Prof. PhD. habil. Gheorghe SĂVOIU
University of Piteşti, Romania
Senior lecturer PhD. Emilia GOGU
Bucharest University of Economic Studies, Romania
Lecturer PhD. Marian ȚAICU
University of Pitești, Romania

PRESERVING LOGICAL RELATIONS WHILE ESTIMATING MISSING VALUES 47
Ton DE WAAL
Statistics Netherlands, & Tilburg University
Wieger COUTINHO
Statistics Netherlands

Romanian Statistical Review nr. 3 / 2017
SIMULATION PROCESS, THEORETICAL PARADIGM AND OPERATIONAL-STRATEGIC REALY AS TOOL FOR MANAGERIAL DECISION MAKING
PhD Professor Monica Viorica NEDELCU
Bucharest University of Economic Studies, Romania
PhD Professor Cosmin Octavian DOBRIN
Bucharest University of Economic Studies, Romania
PhD Associate Professor Marian Sorin IONESCU
Bucharest University of Economic Studies & University “POLITEHNICA” Bucharest, Romania

IMPACT OF DIVIDEND TAXATION CHANGING ON DIVIDEND POLICY OF ROMANIAN COMPANIES LISTED ON STOCK MARKET
Stela JAKOVA
Bucharest University of Economic Studies, Romania

THEORETICAL ASPECTS REGARDING THE OPTIMAL TAXATION OF EFFORT WITH MORE CONDITIONS
Prof. Dumitru MARIN PhD
Bucharest University of Economic Studies, Romania
Prof. Ion PARTACHI PhD
Academy of Economic Studies of Moldova
Assoc. prof. Madalina ANGHEL PhD
Artifex University, Bucharest, Romania
Orientation of Public Policy by Highlighting the Relationship between Economic Development, Education and Eco-Innovation in EU Countries

Florin VĂDUVA (florin_vaduva@yahoo.com)
Titu Maiorescu University, Bucharest, Romania

Rodica GHERGHINA (rodica.gherghina@fi n.ase.ro)
Bucharest University of Economic Studies, Romania

Ioana DUCA (duciaioana@yahoo.com)
Titu Maiorescu University, Bucharest, Romania

ABSTRACT
The authors’ scientific endeavour takes into account the fact that one of the most important drivers of a nation’s economic growth is the population’s level of education and skills. Nowadays, when growth is slowed down and highly volatile, many countries are looking for policies that will stimulate growth. The objective of this research is to investigate some of the aspects concerning the relationship between economic development and human capital through the role of education and eco-innovation activities in EU countries. For the empirical results, some multivariate regression models are computed in R software. The outputs include parameter estimates and standard errors, as well the residual standard error and multiple R-squared. The research results obtained emphasised the fact that “the school expectancy”, “training” and “eco-innovation” indicators have a significant impact on the GDP per capita.

Keywords: public policies, economic growth, education, training, eco-innovation
JEL Classification: I, I2, I25, I28, O1, O15, M21

INTRODUCTION
It is well known that one of the most important determinants of a nation’s economic progress is the population’s level of education and skills. In all countries, designing and implementing public policies to stimulate economic growth is a constant concern.
The main objective of this research endeavour is to investigate some of the aspects concerning the relationship between economic development and human capital through the role of education and eco-innovation activities in EU countries.

Using two multiple linear regression models, we will study the relationship between the GDP per capita and the variables School expectancy, Training and Eco-innovation, for a total of 28 European countries.

To ensure data consistency, we will use the Eurostat database.

The paper is structured in such a way, presenting the literature review, the research methodology, the data and software used and the main empirical results followed by the main conclusions.

**LITERATURE REVIEW**

Although the general theme of economic and social development was intensively approached and investigated in the literature, there is a relative few studies that examines education and eco-innovation in EU countries in relationship with GDP. It is a well-known that education is a dynamic development tool that requires constant investment and innovation and contributes to increasing the educational potential. Investments in education bring benefits to both individuals and society, but their effects are materialised over a longer period of time.

In this context, the economic development of society must be based on investment in education, because “only science that is filtered through the minds and hearts of people, creating systems of attitudes, skills and spiritual capabilities becomes an active and beneficial driver of the economic and socio-cultural development”, Vaideanu (1998).

Over the years, many authors have tried to assess the contribution of education to economic growth. The most famous studies were those by Schultz and Denison according to Suciu, (2000). The two authors, using two different analysis methods, noticed that, theoretically, they yield relatively similar results.

Other authors Plant and Welch (1989) highlighted the impact of education on labour productivity, considering that “the phenomenon most directly linked to economic growth is increased productivity as a result of investment in education“. The two authors have used the method of opportunity costs related to education.

Through their empirical research, Mc.Mahon (2004) confirms the “positive effect of education on economic growth and highlights the positive relationship between education and productivity.”

The authors Mingat and Tau (1996), in the studies they performed, showed that marginal macroeconomic impact of the different levels of education varies widely, depending on each country’s development level.

Investment in education and its economic effects are analysed by Becker (1997), in his famous book “Human Capital”. Starting from the estimated costs and benefits of education, the author shows “correlations between the level of education, wages or getting a job”. The performed analysis indicates that “the wages of educated and trained persons increase more rapidly than those of less educated and
trained persons”, which leads to the concept of age-wages or age-wealth curves, and emphasises that their shape is determined by investment in the individual’s education.

In the paper *Landmarks of Tomorrow*, Druker (1999) emphasises the fact that, “in a constantly changing society, where the labour market requirements are more and more refined, it is necessary that individuals continue their lifelong vocational training”.

In their studies, Wilson and Briscoe (2004) show that “education and training are key contributors to the development of skills and knowledge”. And increasing investments in education leads to higher productivity and wages for each individual, as well as for society.

In the literature, authors Lucas (1988), Romer (1990), Aghion and Howitt (1998) emphasise that “education may trigger an increase in the innovative capacity of the economy, the development of new knowledge by means of the new technologies”, thus being an important factor of economic growth.

Otherwise, “the introduction of innovative teaching methods and the use of interactive software and online materials require the existence of modern ITC equipment in schools, as prerequisite. Yet, integrating ITC in school education is a complex process and it is affected by several different factors”, Balanskat, Blamire and Kefala (2006).

At the same time, the rapid development of information technology and communications in recent years has had “a major impact on the global society and economy, with fundamental changes in production and distribution models, employment and daily life”, Dumitrescu, (2006). On the other hand, the new software as well as R are able to perform graphical visualization, “prompt and effective response, given the rapid changes in data sources”, Alexandru and Caragea (2016).

**RESEARCH METHODOLOGY**

In order to analyze the relationship between economic development, education and eco-innovation in EU countries, certain regression models were used. The output variable is mainly Gross Domestic Product computed for all 28 countries across the EU. The explanatory variables are the potential factors that could have contribution to GDP growth of countries, as well as the education and eco-innovation. Two indicators were used in order to capture the impact of education on the GDP: training and School expectancy.

The indicators were described by reference to the Eurostat database, [http://ec.europa.eu/eurostat/data/metadata](http://ec.europa.eu/eurostat/data/metadata).

**Description of the variables**

**Dependent variable (Y)** - as the variable of interest in this study - is the *Gross domestic product* (GDP per capita Euro).

Independent variables (Xi) are considered the followings:

**X1 – School expectancy** - stands for “the expected years of education in a lifetime and has been calculated by adding the single-year enrolment rates for all ages.
This type of estimate will only be accurate if current enrolment patterns continue in the future. Headcount data is the basis for such estimates. To illustrate what school expectancy means, we will take the following example: school expectancy would be one year for the age of 8 if all 8-year-old students (in the year the data was collected) were enrolled. But, if, let’s say, only 50% of 8-year-olds were enrolled, then school expectancy would be half a year for the age of 8”, http://ec.europa.eu/eurostat/data/metadata.

\[\text{X2 – Training} \] is the rate of Participation in education and training. This variable is provided by the Adult Education Survey (AES) which encompasses the participation of adults in education and training (whether formal, non-formal or informal learning); it is one of the main sources of data for statistics regarding EU lifelong learning. The AES concentrates on people aged 25-64 who live in private households. The reference period for their participation in education and training is the interval of twelve months preceding the interview.

Learning activities: these are any activities an individual performs with the aim to improve one’s knowledge, abilities, and competences. The definition of intentional learning (unlike random learning) is that of deliberate quest for knowledge, abilities or competences. Organized learning is to be understood as learning mapped out in a pattern or sequence with explicit or implicit goals. The various types of learning activities are synchronised with a classification of learning activities (CLA), namely:

- **Formal education and training** means “education as it is provided within the system of schools, colleges, universities as well as other formal education institutions that usually represents a continuous “ladder” of full-time education for children and young people, which generally begins at the age of 5 to 7 and continues to up to the age of 20 or 25”, http://ec.europa.eu/eurostat/data/metadata;

- **Non-formal education and training** means “any learning activities which are organized and sustained and which do not correspond entirely to the definition of formal education mentioned above. Thus, non-formal education may take place both inside and outside education institutions and addresses people of all ages”, http://ec.europa.eu/eurostat/data/metadata.

According to national contexts, it may include educational programs which disseminate literacy, life-skills, work-related skills, and general culture among adult. There are four types of activities associated with non-formal learning that can be emphasised (details of those categories are not available in the online tables):
- **Courses**;
- **Workshops or seminars**;
- **Instructor-led on-the-job training** (scheduled periods of time allocated to education, instruction and/or training directly at the workplace, which the employer organizes with the support of an instructor);
- **Lessons**.

Informal learning (only referenced in 2007 data in the domain trng_aes_007h) means “learning which is intentional but less organized and less structured compared to the previous types. It may include a variety of learning activities (or
events) for an every person, for example, some that occur in the family, some at work, and in their daily life, on a self-directed, family-directed or socially-directed basis”, http://ec.europa.eu/eurostat/data/metadata.

The **education and training participation rate** refers to participation in both formal and non-formal education and training. Lifelong learning becomes possible through participation in education and training.

**X3 - Eco-Innovation Index (%)** - This index is based on 16 indicators from eight contributors in five areas: eco-innovation contributions, eco-innovation activities, eco-innovation products, environmental outcomes and socio-economic outcomes. The unweighted mean of the 16 sub-indicators is used in calculating an EU Member State’s overall score. It shows the performance level achieved by individual Member States in eco-innovation as compared to the EU average, which is equated with 100. For 2010-2012, the average used for indexing to 100 is the average of EU-27. From 2013 onwards, the average used is calculated from the data for 28 EU Member States. The relevant target in the Roadmap is for an increase in the funding for research that contributes to the environmental knowledge base. Such increases will improve a Member State’s positioning according to the index. Although the index is published annually, its sub-indicators are often not, so the index is a collation of the most recent data available each year. As its units are relative it cannot indicate progress in absolute terms.

**Description of regression models**

Regression analysis is used for explaining or modelling the relationship between a single variable Y, called the response or dependent variable, and one or more predictor, independent or explanatory variables (Xi). “We generally are interested in finding a weighted combination of some set of variables that reproduces or predicts as well as possible the values that we have observed on the response or outcome variable”, Babyak (2004). “If this aim is achieved, the model we develop will predict well not only in the sample data set at hand but also in new data sets”, Babyak (2004).

In this research study Multiple Linear Regression model Baltagi (2008), Freedman (2005) were applied to observe the influence of certain socio-economic factors on the economic development level. Our goal is to find and explain the relationship between the interest variable (GDP per capita in EURO) and particular independent variables. We chose several explanatory variables for testing.

In theory, the model takes the following form, given n observations:

\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i, i = 1,2,\ldots,n \]

with the equation having the following form:

\[ \hat{y} = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_p x_{ip}, i = 1,2,\ldots,n \]

b0, called intercept parameter, shows the value of the response variable considering the predictor variables as null. It is more useful in calculation since it does not have a definite significance.

The rest of bi parameters show how the response variable (y) modifies when a predictor variable (xi) changes with one unit, and all other variables remain constant.
“The sign of the parameters shows the type of relationship between the dependent and independent variables; when the parameter is below 0 they are related in a negative linear sense, when the sign is positive there is a positive linear sense”, Popa (2016).

The OLS - ordinary least squares method for estimating the unknown parameters in a linear regression model - is the most commonly used method, finding regression parameters that gives the best fit of dependent variable. Least squares method minimizes the residual sum of squares where the residuals \((e_i)\) are given by the differences between observed and expected values of response. In other words, we must find the theoretic values of the response as “close” as possible to the observed values. Residuals:

\[ e_i = y_i - \hat{y}_i \]

Residual sum of squares (Sum of Squares for Errors):

\[ SSE = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \beta_0 - \beta_1 x_i)^2 \]

The confidence interval gives an estimated range of values for the estimated parameters, with a selected probability.

**Data source and software used**

In order to measure the dimensions of economic development, we established a set of indicators that have been selected on the basis of literature review and the available data in the from Eurostat database, http://ec.europa.eu/eurostat/data/database.

The comparability of data between countries is assured by the application of common definitions. The data refers to 28 countries (EU countries) in the reference years 2011 respectively 2014. The reason for choosing these reference years was data availability and to test the time evolution of the impact of independent variables (Eco-Innovation Index, School_Expectancy and Training) on the independent variable (GDP). To assure the comparability between countries for GDP we used GDP per capita. The data used in models are presented in Table 1.

To compute the multiple linear regression model the lm function in R was used Dusa et al, (2015). The goodness of fit of the model was established using the ANOVA function in R, and the confidence interval for the estimators was established using the confint function in R. In R, the lm function computes the coefficients. The output includes a table with parameter estimates and standard errors, residual standard error and multiple R-squared.
DATA USED IN MODELS

Table 1

<table>
<thead>
<tr>
<th>Country</th>
<th>GDP_per_capita_2011_Euro</th>
<th>School_Expectancy_2011 (years)</th>
<th>Training_2011 (%)</th>
<th>Eco_Innov_2011 (% EU average=100)</th>
<th>GDP_per_capita_2014_Euro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>34500</td>
<td>19.6</td>
<td>37.7</td>
<td>115</td>
<td>35900</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>5600</td>
<td>16.4</td>
<td>26.0</td>
<td>67</td>
<td>5900</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>15600</td>
<td>18.1</td>
<td>37.1</td>
<td>91</td>
<td>14700</td>
</tr>
<tr>
<td>Denmark</td>
<td>44200</td>
<td>19.6</td>
<td>58.5</td>
<td>138</td>
<td>46200</td>
</tr>
<tr>
<td>Germany</td>
<td>33700</td>
<td>18.1</td>
<td>50.2</td>
<td>123</td>
<td>37100</td>
</tr>
<tr>
<td>Estonia</td>
<td>12500</td>
<td>18.2</td>
<td>49.9</td>
<td>74</td>
<td>15200</td>
</tr>
<tr>
<td>Ireland</td>
<td>38000</td>
<td>17.3</td>
<td>24.4</td>
<td>118</td>
<td>41000</td>
</tr>
<tr>
<td>Greece</td>
<td>18600</td>
<td>18.1</td>
<td>11.7</td>
<td>59</td>
<td>16200</td>
</tr>
<tr>
<td>Spain</td>
<td>22900</td>
<td>17.7</td>
<td>37.7</td>
<td>128</td>
<td>22400</td>
</tr>
<tr>
<td>France</td>
<td>31500</td>
<td>16.5</td>
<td>50.5</td>
<td>99</td>
<td>32200</td>
</tr>
<tr>
<td>Croatia</td>
<td>10400</td>
<td>16.2</td>
<td>21.2</td>
<td>47</td>
<td>10200</td>
</tr>
<tr>
<td>Italy</td>
<td>27300</td>
<td>17.3</td>
<td>35.6</td>
<td>90</td>
<td>26500</td>
</tr>
<tr>
<td>Cyprus</td>
<td>23000</td>
<td>14.9</td>
<td>42.3</td>
<td>71</td>
<td>20400</td>
</tr>
<tr>
<td>Latvia</td>
<td>9800</td>
<td>18.1</td>
<td>32.3</td>
<td>77</td>
<td>11800</td>
</tr>
<tr>
<td>Lithuania</td>
<td>10300</td>
<td>19.1</td>
<td>28.5</td>
<td>52</td>
<td>12400</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>81300</td>
<td>15.0</td>
<td>70.1</td>
<td>130</td>
<td>87600</td>
</tr>
<tr>
<td>Hungary</td>
<td>10100</td>
<td>17.7</td>
<td>41.1</td>
<td>83</td>
<td>10600</td>
</tr>
<tr>
<td>Malta</td>
<td>16500</td>
<td>15.0</td>
<td>35.9</td>
<td>82</td>
<td>19000</td>
</tr>
<tr>
<td>Netherlands</td>
<td>38500</td>
<td>19.0</td>
<td>59.3</td>
<td>109</td>
<td>40000</td>
</tr>
<tr>
<td>Austria</td>
<td>36800</td>
<td>17.1</td>
<td>48.2</td>
<td>125</td>
<td>38500</td>
</tr>
<tr>
<td>Poland</td>
<td>9900</td>
<td>18.1</td>
<td>24.2</td>
<td>50</td>
<td>10700</td>
</tr>
<tr>
<td>Portugal</td>
<td>16700</td>
<td>18.4</td>
<td>44.4</td>
<td>81</td>
<td>16700</td>
</tr>
<tr>
<td>Romania</td>
<td>66000</td>
<td>17.5</td>
<td>8.0</td>
<td>67</td>
<td>7500</td>
</tr>
<tr>
<td>Slovenia</td>
<td>18000</td>
<td>18.4</td>
<td>36.2</td>
<td>109</td>
<td>18100</td>
</tr>
<tr>
<td>Slovakia</td>
<td>13000</td>
<td>16.5</td>
<td>41.6</td>
<td>52</td>
<td>13900</td>
</tr>
<tr>
<td>Finland</td>
<td>36500</td>
<td>20.6</td>
<td>55.7</td>
<td>149</td>
<td>37600</td>
</tr>
<tr>
<td>Sweden</td>
<td>42900</td>
<td>19.8</td>
<td>71.8</td>
<td>142</td>
<td>44400</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>29600</td>
<td>16.6</td>
<td>35.8</td>
<td>105</td>
<td>34900</td>
</tr>
</tbody>
</table>


RESEARCH RESULTS

Model 1

The research authors aimed at identifying the relationship between the GDP per capita and the two indicators, namely School_Expectancy, respectively Education and Training, and a composite indicator in the field of eco-innovation. When selecting these indicators, we considered the following arguments:

- In time, numerous studies have emphasised the contribution of education to economic growth - Schultz and Denison in Suciu (2000), Beker (1997).
The originality of our research is the fact that we identify and emphasise those indicators in the domain of education and training which have a significant impact on economic growth. The purpose of the study is to bring scientific arguments to substantiate the future public policies with long-term significant impact on GDP growth;

• The School_Expectancy indicator captures also school drop-out. In our opinion, medium and long-term school drop-out may have a substantial effect on the potential for economic growth. Moreover, the results obtained by means of the econometric models employed confirmed this assumption;

• The indicator Participation rate in education and training captures lifelong learning. In the modern economy, where changes occur at very high pace, it is necessary to update the knowledge and skills specific to the human capital permanently, so that the latter contributes to economic growth Vaideanu (1998), Druker (1999);

• Eco_Innovation Index is a composite indicator which captures the research field.

The model tested in R software is:

\[ \text{GDP}_{\text{per capita} 2011} \text{ _Euro} = \beta_0 + \beta_1 \text{School_Expectancy}_{2011} + \beta_2 \text{Training}_{2011} + \beta_3 \text{Eco_Innov}_{2011} \]

By running the multiple linear regression model in R software, the following equation is obtained, according with the results presented in Table 2:

\[ \text{GDP}_{\text{per capita} 2011} \text{ _Euro} = 34034.17 -3109.69 \text{School_Expectancy}_{2011} + 329.62 \text{Training}_{2011} + 346.57 \text{Eco_Innov}_{2011} \]

**THE RESULTS ON TESTING THE MODEL 1 WITH R SOFTWARE**

|                          | Estimate  | Std. Error | t value | Pr(>|t|) |
|--------------------------|-----------|------------|---------|----------|
| Intercept                | 34034.17  | 22074.28   | 1.542   | 0.136205 |
| School_Expectancy 2011   | -3109.69  | 1315.99    | -2.363  | 0.026571 |
| Training 2011            | 329.62    | 158.40     | 2.081   | 0.048286 |
| Eco_Innov 2011           | 346.57    | 84.25      | 4.145   | 0.000395 |

Residual standard error: 9241 on 24 degrees of freedom
Multiple R-squared: 0.7141
Adjusted R-squared: 0.6783
F-statistic: 19.98 on 3 and 24 DF
p-value: 1.031e-06

Parameters:
\[ \beta_1 = -3109.69 \] - estimated difference in the GDP per capita with one point increase of School Expectancy. We can see that School Expectancy has the greatest impact on GDP per capita among the independent variables analyzed in model. As
we expected it is a negative relationship between this variable and GDP because the school dropout decrease the capacity of human capital to generate economic increase.

$\beta_2 = 329.62$ - estimated the increase in the GDP per capita with one percentage point increase of Participation rate in education and training (Training) variable.

$\beta_3 = 346.57$ - estimated the increase in the GDP per capita with one percentage point increase of Eco-Innovation Index.

The estimated parameters have statistical significance at 95% level. The independent variables explain 71.41% of response variation (GDP per capita).

In Figure 1 we can see that there are correlations between the dependent variable and each independent variable.

THE CORRELATION BETWEEN GDP PER CAPITA IN 2011 AND INDEPENDENT VARIABLES

Model 2
The next model we kept the same variable independent but there is changed the dependent variable. We replaced the corresponding data of the indicator GDP per capita of 2011 corresponding to those in 2014. The goal is to see if the independent variables over time increasing their impact on the dependent variable.

The model tested in R software is:

$$GDP_{\text{per capita}}_{2014\_\text{Euro}} = \beta_0 + \beta_1 \times \text{School\_Expectancy\_2011} + \beta_2 \times \text{Training\_2011} + \beta_3 \times \text{Eco\_Innov\_2011}$$

By running the multiple linear regression model in R software, the following equation is obtained, according with the results presented in Table 3 and Figure 2:

$$GDP_{\text{per capita}}_{2014\_\text{Euro}} = 37452.64 - 3402.32 \times \text{School\_Expectancy\_2011} + 356.86 \times \text{Training\_2011} + 366.32 \times \text{Eco\_Innov\_2011}$$
The results on testing the model 2 with R software

Table 3

|             | Estimate  | Std. Error | t value | Pr(>|t|) |
|-------------|-----------|------------|---------|----------|
| Intercept   | 37452.64  | 23741.14   | 1.578   | 0.127762 |
| School Expectancy 2011 | -3402.32  | 1415.37    | -2.404  | 0.024305 |
| Training 2011 | 356.86    | 170.36     | 2.095   | 0.046928 |
| Eco Innov 2011 | 366.32    | 90.61      | 4.043   | 0.000473 |

Residual standard error: 9939 on 24 degrees of freedom
Multiple R-squared: 0.71
Adjusted R-squared: 0.6737
F-statistic: 19.59 on 3 and 24 DF
p-value: 1.22e-06

The correlation between GDP per capita in 2014 and independent variables

The estimated parameters have statistical significance at 95% level. The explanatory variables explain 71.0% of response variation (GDP per capita). In this situation, it was estimated that the increase of GDP per capita in 2014 across the EU countries are influenced by education patterns (measured by school expectancy and formal and non-formal training of population) and eco-innovation level registered in 2011 (three years lag).

Testing the multicolinearity of the models, it appears a medium intensity level of correlation between explanatory variables Training and Eco-Innovation (figure 3). For example, the correlation between the result and an explanatory variable might be changed when another predictor is introduced into the model.
Taking into account the definitions and the methodologies to produce data for the indicators under observation, isn’t a correlation based on calculation methodology but it obviously that the correlation could be explained: more developed EU member states are characterized by higher levels of eco-innovation outcomes (environmental protection activities have a higher impact on economic and social development). In the same time, all these countries have lifelong learning more developed then less developed EU countries. In this case, multivariable analysis is a useful instrument, yet only if its users have thorough/ full understanding of the premises and drawbacks of these methods.

CONCLUSIONS

During our research, we tested the same model for GDP per capita for 2011 and 2014. In the analysis of the independent variables coefficient considered, we notice an increase over time of their impact on the dependent variable. For example, the impact of the training factor for the year 2011 on the GDP increases from 329.62, if we start from the first model with GDP per capita in 2011, from 356.86, and compared to the second model with GDP per capita in 2014. Consequently, the public policies aimed at stimulating lifelong learning and decreasing school drop-out, respectively at developing eco-innovation, may have a significant impact on the evolution of the GDP. Simultaneously, the public policies in the field of education have a multiplying effect over time. In terms of the investment in education, this is an obvious aspect, as a better trained individual is much more active on the labour market and triggers a positive effect on economic growth.
At the same time, the results of the eco-innovation process may have a certain delay in capturing the impact on the GDP. These empirical assumptions were emphasized by means of econometric tools, in the two multiple linear regression models. In their future research, the authors aim at testing the models again (based on the new statistical data available) because the models may exhibit certain errors caused by the change in the manner of calculating the Eco-innovation Index (%) indicator in 2013 as compared to 2011. Yet, in the analysis of the two models, we notice that emphasizing such an error, this cannot be significant and it may affect only the importance of the cumulated impact of this indicator on the GDP per capita over time.

REFERENCES

Power and Size analysis of Co-integration tests in Conditional Heteroskedascity: A Monte Carlo Simulation

Osabuohien-Irabor Osarumwense (osabuohienosa@auekpoma.edu.ng)
Ambrose Alli University (A.A.U), Ekpoma, Edo, Nigeria

Julian I. Mbegbu
University of Benin, Benin City, Nigeria

ABSTRACT
This paper investigates the finite sample performance of power and size properties of several major co-integration tests using simulation analysis. These tests include; the co-integration Regression Durbin-Watson test (CRDW), Eagle-Granger test, Dicky Fuller unit root test with $(T(\hat{\rho} - 1))$ statistics, Johansen likelihood ratio tests, and Phillips-Ouliaris test. Comparisons of tests are evaluated based on the proportion of rejects of the hypothesis of a no co-integration. This study answers the question of which co-integration test is better, particularly between the Eagle-Granger two-step test and the Johansen’s tests for co-integration, when the sets of parameters in models are persistence and spiky. The bivariate Generalized Autoregressive Conditional Heteroskedasticity (GARCH(1,1)) model with Gaussian innovations, is used in the data generating process (DGP). Our simulation results reveal that there is size distortion in the different co-integration test considered. The Eagle-Granger two-step test shows good robustness with respect to heteroskedasticity for the different sample sizes applied. However, the Johansen’s test for co-integration still proves to be powerful in capturing co-integration relationship, particularly for large sample when the co-integration innovations are Gaussian.

Key words: Co-integration test, Size, Monte-Carlos Simulation, Power, Heteroskedasticity

JEL Classifications: C12, C15 C32

INTRODUCTION
The idea of co-integration relationship between two or more time series was originally proposed by Granger (1983). It implies that $y_t$ and $x_t$ variables shares similar stochastic trend. If their difference ($e_t$) are similar, they will never diverge too far from each other. The test for co-integration is effectively a test of the stationarity of the residuals. If the residuals are stationary, then $y_t$ and $x_t$ are said to
be co-integrated. But if residual are non-stationary, then and are not cointegrated and any apparent regression relationship between them, is said to be spurious (Hill, et al (2011)). Brooks (2008) also stated that, if two variables that are $I(1)$, are linearly combined, then the combination will also be $I(1)$ i.e if variables with differing order of integration are combined, the combination will have an order of integration equal to the largest. Therefore, if $X_{i,t} \sim I(d_i)$ for $i = 1, 2, 3, ..., k$ so that there are $k$ variables each integrated of order $d_i$ and letting

$$z_t = \sum_{i=1}^{k} \alpha_i X_{i,t}$$

(1)

Then $z_t \sim I(\max d_i)$. $z_t$ in this context is simply a linear combination of the $k$ variables $X_t$. Rearranging (1),

$$X_{i,t} = \sum_{i=2}^{k} \beta_i X_{i,t} + z'_i$$

(2)

Where

$$\beta_i = -\frac{\alpha_i}{\alpha_1}, \quad z'_i = \frac{z_i}{\alpha_1}, \quad i = 2, ..., k$$

$\alpha, \beta$ are the coefficients of $X_1$ and $X_2$ respectively, and $z'_i$ is the disturbance, which is non stationary and autocorrelated if all $X_i$ are $I(1)$.

Over the years, homoskedasticity is one of the basic conditions underlying the applicability of many financial and econometric analyses, and in essence, the co-integration test with a view to avoiding spurious results affected by heteroskedasticity. To capture these conditional variances in the underlying return time series, Engle (1982) first introduced the ARCH model (AutoRegressive Conditional Heteroskedasticity) which was later developed independently into the GARCH model (Generalized AutoRegressive Conditional Heteroskedasticity) by Bollerslev (1986) and Taylor (1986). Heteroskedasticity assumes that the random errors in model have normal conditional distribution. The GARCH model is today the most popular model among the Non-linear models, which captures the basic stylized facts such as persistence, volatility clustering, leverage effects etc in financial time series data. And the GARCH (1,1) has been proven to be sufficient in capturing these stylized fact. The co-integration errors are fitted by a GARCH model with Gaussian or non-normal random error.

The concept and application of co-integration in financial time series and econometric analysis has received much attention in recent years, particularly, from practitioners and academics who test and apply the stationarity of econometric and financial time series variables. Hence, a number of co-integration tests have been developed and used to examine the behavior of these co-integration errors and the relationship among time series. However, the need to evaluate the performance and applicability of these different co-integration tests with regard to its powers and sizes cannot be over-emphasized.

There are avalanche of literature on the applicability of some co-integration tests, and some recently notable works includes; Cho and Ramirez (2016), estimates
the demand for real money in Korea over the 1973Q3 to 2014Q4 period via unit root and co-integration methods. They utilizing the Johansen co-integration methodology and the Pantula principle, and established that a long-term relationship exists among the included variables. It finds that the broader definition of money, M2, serves as a relatively better measure of the money aggregate than M1 when evaluating the stability of the real demand for money. The long-term interest (LR) rate also seems to provide better results than the short-term rate (SR), which is consistent with economic theory given that it refers to a long-run equilibrium relationship. Caruso and Domizio (2016), study the interdependence of military spending between US and a panel of European countries in the period 1988-2013. They based their empirical estimation on a: (i) unit root tests and a co-integration analysis and (ii) fully modified ordinary least squares and dynamic ordinary least squares estimations. Their results highlight that military spending of European countries is: (i) positively associated with US military spending and (ii) negatively associated with average military spending of other European countries. Ferreira and Oliveira (2014), also investigates the existence of long-run relations between the Portuguese and other markets under stress. Interestingly, they found that the only market that did not follow this trend was Spain. Their results found six co-integration vectors: two within the group of European emerging markets (Portugal, Italy and Ireland) and the other four between the Portuguese market and the mature markets (France, United Kingdom, Germany and United States).

However, some notable literature on the properties of co-integration tests also includes; Kosapattarapim et al (2013), test several co-integration tests; such as Dicky Fuller, co-integration Regression Durbin-Watson test, Wild Bootstrap test and the Johansen tests. Using a simulation design of co-integration system for power analysis and non-cointegrated system for size test, they compare the Johanson’s tests to other co-integration test for $n=100$, and $1000$ when the co-integration error distribution is non-normal. Their analysis reveals that the Johanson’s tests are still more powerful than the alternative tests. Krauss and Herrmann (2017), examined the power and size of several co-integration tests with high frequency data. The AR(1), AR(1)-GARCH(1,1) and the multiple regime STAR(1)-GARCH(1,1) model were used in capturing the co-integration relationship. Their analysis revealed that under high frequency data, the selected co-integration tests still exhibit high power especially Phillip-Parron and Pantula, Gonzalez-Farias and Fuller tests which persons at with limited size distortion. Cavaliere et al, (2007), analysed the vector autoregression with non stationary volatility of general form. Their analysis revealed that the conventional rank statistics of Johansen are potentially unreliable. They also showed that large sample distribution depends on the integrated covariation of the underlying multivariate volatility process which impacts on both the size and power of the associated co-integration tests. Bernstein et al, (2014) considers co-integration tests when the co-integration rank is deficient. The asymptotic theory test for co-integration ranks and co-integrating vectors are derived, and the limiting distribution revealed. Their study also apply to the U.S. treasury.

Gerolimetto and Procidano (2003), investigated and compared the Wild Bootstrap test with the DF ($t^2$) test, CRDW test and the Johansen’s test when co-integration errors follows GARCH(1,1) model with normal random error distribution.
They also showed that the Wild Bootstrap test is robust to heteroskedasticity. Mallory and Lence (2012), explores performance of the Johansen co-integration statistics on data containing negative moving average (NMA) errors. A Monte Carlo experiment is used to demonstrate that the asymptotic distributions of the statistics are sensitive to NMA parameters, and that using the standard 5% asymptotic critical values results in severe underestimation of the actual test sizes. They also demonstrate that problems associated with NMA errors do not decrease as sample size increases; instead, they become more severe. Further, they examine the evidence that many U.S. commodity prices are characterized by NMA errors. Pretesting data is recommended before using standard asymptotic critical values for Johansen’s co-integration tests. Moral-Benito and serven (2013), Proposes a new test of weak exogeneity in panel co-integration models. The test has a limiting Gumbel distribution that is obtained by first letting $T \to \infty$ and then letting $N\to \infty$. They evaluated the accuracy of the asymptotic approximation in finite samples via simulation experiments and an empirical illustration. Weak exogeneity of disposable income and wealth in aggregate consumption was tested. Boswijk et al (1999) consider the performance of several contemporary tests based on Gaussian quasi-likelihoods using either the kernel estimation or semi-nonparametric density approximation. They observed that, in small samples, the overall performance of the semi-nonparametric approach appears best in terms of size and power. But in large samples and for heavily skewed or multi-modal distributions, the kernel based adaptive method dominates.

Müller, and Watson (2013), consider low-frequency tests about co-integrating vectors under a range of restrictions on the common stochastic trends. They quantify how much power can potentially be gained by exploiting correct restrictions, as well as the magnitude of size distortions if such restrictions are imposed erroneously. A simple test motivated by the analysis in Wright (2000) is developed and shown to be approximately optimal for inference about a single co-integrating vector in the unrestricted stochastic trend model. Other includes; Maki (2012), Demetrescu et al (2014), Tursoy and Faisal (2017) etc

However, this paper is an extension of Kosapattarapim et al (2013) with an objective to evaluate the performances of several major and contemporary co-integration tests such as the Co-integration Regression Durbin-Watson test (CRDW) (1983), the Eagle-Granger two-step (1987) test, the Dicky Fuller (1979) unit root test with ($T(\hat{ρ} - 1)$) statistics, the Johansen’s test (1988), and the Phillips-Ouliaris (1990) test. The power and size of these co-integration tests for finite sample performance are investigated, when the set parameters in the models are persistence and spiky with the co-integration errors being Gaussian innovation distributions. This study also ask the question, on whether the Eagle-Granger two-step approach and the Phillips-Ouliaris test for co-integration analysis can replace, compete or be used as close substitute to the powerful Johansen’s test for co-integration. This rest of this paper is organized as follows; Section 2 deals with the methodology of the study using simulation-based data generating process (DGP). In section 3, the analysis of results of our simulation experiment with regards to power and size of co-integration test are discussed, while section 4 discusses the conclusion and finding from our simulated experiment.
METHODOLOGY

This paper applied simulation-based approach to assess the power and size properties of co-integration tests when innovation is Gaussian GARCH (1,1) distributed. These tests include; the Co-integration Regression Durbin-Watson test (CRDW), Eagle-Granger two-step test, Dicky Fuller unit root test with \( T(\hat{\rho} - 1) \) statistics, Johansen’s test, and the Phillips-Ouliaris test. The comparison of these several co-integration test is done with the proportion of rejects of the null hypothesis of a no co-integration.

Design of the Monte-Carlo simulation-based experiment

The Monte-Carlo Simulation is carried out by generating artificial data series of ‘daily’ time series \( y_t \) and \( x_t \). The data generating process (DGP) will be repeated for \( dNt = 1, 2, 3, 4 \) and the first value of \( d = 100 \) observations are discarded. This helps in removing the initial value effect. An independent simulation of \( N = 3000 \) replications are carried out for a variety of sample sizes for \( T = 100, 500, 750, 1500, \) and \( 2000 \) respectively from the designed system and apply to the six co-integration tests (CRDW, Eagle-Granger two-step test, Dicky Fuller unit root test \( T(\hat{\rho} - 1) \), Johansen (trace test \( \lambda_{trace} \) and the maximum eigenvalue test \( \lambda_{max} \), Phillips-Ouliaris). The comparison of the different co-integration tests are considered based on the frequency of rejection of the null hypothesis of no co-integration.

The Size test

The size of a test is the probability that the null hypothesis is rejected when it is true. It is evaluated by the frequency of the null hypothesis stating that truly non co-integrated series is not co-integrated in \( 3000 \) trials. To investigate the size properties of co-integration test, a simulated sample sizes from bivariate non-co-integrated system with GARCH (1,1) Gaussian innovations are generated. Where; is the variance from the GARCH (1,1) model.

Let \( Y = (Y_1, Y_2, ..., Y_n) \) be an \( NXN \) vector of cointegration series.

The bivariate system is given as;

\[
\begin{align*}
\nabla Y_{1t} &= e_{1t} \\
\nabla Y_{2t} &= e_{2t} 
\end{align*}
\]

In other words, this implies that the data are generated by a random walk given as;

\[
Y_t = Y_{t-1} + e_{1t} \quad \text{and} \quad Y_{2t} = Y_{2t-1} + e_{2t}, \quad e_t \sim iid(\sigma, \sigma^2), \quad t = 1, 2, ...
\]

The errors \( e_{1t} \) and \( e_{2t} \) follows GARCH (1,1) model

\[
\begin{align*}
\eta_t &= \eta \sqrt{h_t} \\
h_t &= \omega_1 + \alpha_1 e_{u-1}^2 + \beta_1 h_{u-1}
\end{align*}
\]
Where \( h_{it} \) is the variance conditional on information available up to time \( t-1 \) and \( \eta_{it} \) are iid with \( E(\eta_{it}) = 0 \) and \( \text{var}(\eta_{it}) = 1 \).

### The Power test

The power of a test is the probability that the null hypothesis is rejected when the null hypothesis is false. The frequencies of rejection of no co-integration from a co-integrated system are counted. To examine the power of co-integration test, a co-integrated system is simulated. In this paper, the Phillips’ (1991) triangular representation equations as simulated and applied by Zivot (2000) and Fassati (2013), are used in simulating a bivariate co-integration system. It’s of the form;

\[
Y_{1,t} = \beta_1 Y_{2,t} + e_{1t} \quad \text{(7)}
\]

\[
Y_{2,t} = Y_{2,t-1} + e_{2t} \quad \text{(8)}
\]

Where \( e_{1t} \) and \( e_{2t} \) follows GARCH (1,1) Gaussian innovation model

Therefore,

\[
Y_{1,t} = Y_{1,t-1} + u_t \quad \text{(9)}
\]

\[
Y_{2,t} = Y_{2,t-1} + e_{2t} \quad \text{(10)}
\]

\[
u_t = 0.7u_{t-1} + e_t \quad \text{(11)}
\]

Where \( u_t = Y_{1,t} - Y_{2,t} \) are co-integrated residual with AR(1) of \( \phi = 0.75 \)

The hypotheses to be tested are;

\[
H_0 : e_u = Y_u \sim I(1) \quad \text{(No co-integration)}
\]

\[
H_1 : e_u = Y_u \sim I(0) \quad \text{(Co-integration)}
\]

### The Co-integration Regression Durbin-Watson test (CRDW)

The Co-integration Regression Durbin-Watson test (CRDW) test statistics can be used as a quick test of co-integration. The estimate assumes a co-integrating equation;

\[
x_{1,t} = \beta_1 + \beta_2 x_{2,t} + \cdots + \beta_p x_{p,t} + u_t \quad \text{(12)}
\]

The test statistics for first order autocorrelation is calculated. Under the null hypothesis that \( x_{1,t} \) is a random walk and that \( \beta_2 = \cdots = \beta_p = 0 \), so there is no co-integration and \( u_t \) become a random walk with theoretical first order autocorrelation equal to unity. Under the null of no co-integration, the CRDW value will not be significantly different from zero. Therefore, a co-integrating Regression Durbin-Watson (CRDW) test statistics different from zero implies co-integration.

### The Engle and Granger Two-step procedure

The intuition behind this test motivates its role as the first co-integration test to be learned. It starts by estimating the so called co-integrating regression (the first step);

\[
x_{1,t} = \beta_1 + \beta_2 x_{2,t} + \cdots + \beta_p x_{p,t} + u_t \quad \text{(12')}
\]

Where \( p \) is the number of variables in the equation, \( u_t \) is the co-integrating error and \( \beta_i \) is the coefficient of the variable \( x_i \). We assume that all the variables are \( I(1) \) and co-integrate to form stationary relationship and thus a stationary residual term;

\[
u_t = x_{1,t} - \beta_1 - \beta_2 x_{2,t} - \cdots - \beta_p x_{p,t}
\]

\[
u_t = x_{1,t} - \beta_1 - \beta_2 x_{2,t} - \cdots - \beta_p x_{p,t}
\]
If variables are co-integrated, they will share a common trend and form a stationary relationship in the long run. The second step in the Engle and Granger two-step procedure, is to test for a unit root in the residual process of the co-integrating regression above; therefore,

$$\Delta \hat{u}_t = \alpha + \sum_{i=1}^{k} \gamma_i \Delta \hat{u}_{t-i} + \epsilon_t$$  \hspace{1cm} (14)$$

The constant term $\alpha$ (in most cases) can be left out to improve the efficiency of the estimate. Under the null of no co-integration, the estimated residual is $I(1)$ because $x_{1,t}$ is $I(1)$ and all parameters are zero in the long run. Thus finding a significant $\pi$ implies co-integration. The alternative hypothesis is that the equation is a co-integrating equation, meaning that the integrated variables $x_{1,t}$ co-integrate at least with one of the variables on the right hand side.

The test statistics is:

$H_0: \pi = 0$ No co-integration
$H_a: \pi < 0$ Co-integration

**Johansen Test of co-integration**

The Johansen’s tests are called the maximum eigen value test and the trace test. The Johansen tests are likelihood-ratio tests. There are two tests; the maximum eigen value test and the the trace test. For both test statistics, the initial Johansen test is a test of the null hypothesis of no co-integration against the alternative of co-integration. The test differs in term of the alternative hypothesis.

**The maximum Eigen value Test:** The maximum eigen value test examines whether the largest value is zero relative to the alternative that the next largest eigen value is zero. The first is a test whether the rank of the matrix is zero. The null hypothesis is that rank ($\Pi$) = 0 and the alternative hypothesis is that rank ($\Pi$) = 1 for further test, the null is that rank ($\Pi$) = 2,3,... This is a test using the largest eigen value. If the rank of the matrix is zero, the largest eigen value is zero, then there is no co-integration. The test of the maximum eigen value is a likelihood ratio test. The test statistics is:

$$LR(r_0, r_{0+1}) = -Tln(1 - \lambda_{r_{0+1}})$$  \hspace{1cm} (15)$$

Where $LR(r_0, r_{0+1})$ is the likelihood ratio test statistics for testing whether rank ($\Pi$) = $r_0$ versus the alternative hypothesis that rank ($\Pi$) = $r_{0+1}$. The likelihood ratio statistics does not have the usual asymptotic $\chi^2$ distribution.

**Trace Test:** The trace test is a test whether the rank of the matrix ($\Pi$) is $r_0$. The alternative hypothesis is that $r_0 < \text{rank}(\Pi) \leq n$. Where $n$ is the maximum number of possible co-integrating vectors. If the null hypothesis is rejected, the next null hypothesis is that the rank ($\Pi$) is $r_{0+1}$ and the alternative hypothesis is that $r_{0+1} < \text{rank}(\Pi) \leq n$.

$$LR(r_0, n) = -T \sum_{i=r_{0+1}}^{n} \ln(1 - \hat{\lambda}_i)$$  \hspace{1cm} (16)$$
Where $LR(r_0, n)$ is the likelihood ratio test statistics for testing whether rank $(\Pi) = r_0'$ versus the alternative hypothesis that rank $(\Pi) = \leq n$.

**Dicky Fuller Test**

The basic idea of a DF test is to test if $\gamma = 1$ in the AR equation;

$$y_t = u + \gamma y_{t-1} + u_t$$  \hspace{1cm} (17)

$$y'_t - y_{t-1} = u + \gamma y'_{t-1} - y_{t-1} + u_t$$  \hspace{1cm} (17')

$$\Delta y_t = u + \gamma y_{t-1} + u_t$$  \hspace{1cm} (17'')

Where;

$\gamma' = \gamma - 1$. The hypotheses are;

$H_0: \gamma' = 0$ (a unit root (No co-integration))

$H_0: \gamma' < 0$ (a unit root (co-integration))

**Phillips-Ouliaris Test**

The Phillip-Ouliaris introduced two residual based test, namely the variance ratio test and the multivariate trace statistics (Phillips-Ouliaris, 1990). These tests are used in the same way as the unit root test but data are residuals from co-integrating regressions. They are implemented on matrice or multivariate series and are both based on residual of the first order vector autoregression equation.

$$z_t = \Pi z_{t-1} + e_t$$  \hspace{1cm} (18)

Where $z_t$ is partitioned as $z_t = (y_t', x_t')$ with a dimension of $x_t$ equal to $(m = n + 1)$, $\Pi$ is a regression coefficient and $n$ is equal to the number of variable under study.

**RESULTS AND DISCUSSION**

The results shown in tables and clearly report the two different GARCH (1,1) parameter models used in the analysis of the power and size properties of the several co-integration test considered in this paper. The co-integrating error follows two types of Gaussian innovations ($e_1$ and $e_2$) in GARCH (1,1) model. The two sets of parameter $(\omega, \alpha, \beta)_i$, $i = 1, 2$ considered for our simulation analysis are;

$(\omega_1 = 0.01, \alpha_1 = 0.85, \beta_1 = 0.05)$

and $(\omega_1 = 0.10, \alpha_1 = 0.15, \beta_1 = 0.65)$

The $\alpha_i$ and $\beta_i$ coefficients capture the market information. A high value of $\alpha_i$ indicates that the volatility is sharp, spiky and also responds well to market movements, where as a low value shows that the coefficient responds poorly to market movements. High value of $\beta_i$ has a large persistent influence (Dowd, 2005). In other words, the $\beta_i$ focuses on the degree of persistent of the market news. Our simulation results and analysis reveals that when the GARCH parameter in the model is spiky, there is size distortion in all the co-integration tests across the different sample sizes – see Table 1. The Dickey Fuller test (1979) has the largest values of size distortion when co-integrated innovations are Gaussian distributed.
However, in the power test, the Eagle-Granger two-step test and the Johansen’s test are much higher and consistence when compare to the other tests. In small sample sizes, the Eagle-Granger two-step co-integration test shows to be more powerful than the Johansen test. Results are reported in Table 2.

### Proportion of Rejects for Size of the Test

<table>
<thead>
<tr>
<th>TEST</th>
<th>T=100</th>
<th>T=500</th>
<th>T=750</th>
<th>T=1500</th>
<th>T=2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRDW</td>
<td>0.094</td>
<td>0.097</td>
<td>0.003</td>
<td>0.081</td>
<td>0.091</td>
</tr>
<tr>
<td>Engle-Granger</td>
<td>0.186</td>
<td>0.002</td>
<td>0.016</td>
<td>0.091</td>
<td>0.272</td>
</tr>
<tr>
<td>$DF \hat{T}(\hat{\rho} - 1)$</td>
<td>0.272</td>
<td>0.091</td>
<td>0.284</td>
<td>0.364</td>
<td>0.182</td>
</tr>
<tr>
<td>$\lambda_{\text{Trace}}$</td>
<td>0.351</td>
<td>0.023</td>
<td>0.014</td>
<td>0.025</td>
<td>0.454</td>
</tr>
<tr>
<td>$\lambda_{\text{Max}}$</td>
<td>0.091</td>
<td>0.004</td>
<td>0.029</td>
<td>0.007</td>
<td>0.364</td>
</tr>
<tr>
<td>Phillips-Ouliaris</td>
<td>0.175</td>
<td>0.179</td>
<td>0.011</td>
<td>0.161</td>
<td>0.187</td>
</tr>
</tbody>
</table>

The proportions of rejects are at 5% critical value for set parameter $\omega = 0.01, \alpha = 0.85, \beta = 0.05$, $\omega = 0.05, \alpha = 0.90, \beta = 0.05$ with Garch(1,1) Gaussian innovation distributed. The co-integration tests are applied in 3000 replication from the designed experiment for size test analysis.

### Proportion of Rejects for Power of the Test

<table>
<thead>
<tr>
<th>TEST</th>
<th>T=100</th>
<th>T=500</th>
<th>T=750</th>
<th>T=1500</th>
<th>T=2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRDW</td>
<td>0.636</td>
<td>0.182</td>
<td>0.636</td>
<td>0.545</td>
<td>0.364</td>
</tr>
<tr>
<td>Engle-Granger</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.636</td>
</tr>
<tr>
<td>$DF \hat{T}(\hat{\rho} - 1)$</td>
<td>0.454</td>
<td>0.091</td>
<td>1.180</td>
<td>0.545</td>
<td>0.273</td>
</tr>
<tr>
<td>$\lambda_{\text{Trace}}$</td>
<td>0.636</td>
<td>1.000</td>
<td>0.983</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\lambda_{\text{Max}}$</td>
<td>0.272</td>
<td>0.919</td>
<td>0.954</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Phillips-Ouliaris</td>
<td>0.185</td>
<td>0.096</td>
<td>0.018</td>
<td>0.166</td>
<td>0.273</td>
</tr>
</tbody>
</table>

The proportions of rejects are at 5% critical value for set parameter $\omega = 0.01, \alpha = 0.85, \beta = 0.05$, $\omega = 0.05, \alpha = 0.90, \beta = 0.05$ with Garch(1,1) Gaussian innovation distributed. The co-integration tests are applied in 3000 replication from the designed experiment for power test analysis.

Just like when the parameter is spiky in the GARCH model, Table 3 reveals that there is irregularity and distortion in size as the sample size increases, when the model is persistent. Table 3 also reveals, that the CRDW and the Johansen’s Max co-integration test have smaller size distortion when compare to the other co-integration tests. The simulation analysis proves that the Eagle-Granger two-step test is as powerful...
as the Johansen’s trace test, but slightly more powerful than the Johansen’s Max test. Besides, the Phillips-Ouliaris test shows to be poor co-integration test when co-integration innovations are distributed Gaussian − see Table 4. Figs 1-6, shows some selected series, variance series, and GARCH series for both the cointegrated and non-cointegrated series for the different sample sizes. Trials of 3000 replications in five different sample sizes for \( T = 100, 500, 750, 150 \) and 2000 are voluminous, hence the author shown selected copies. kindly email the corresponding author for request of full analysis with graphical details.

### Proportion of Rejects for Size of the Test

**Table 3**

<table>
<thead>
<tr>
<th>TEST</th>
<th>( T = 100 )</th>
<th>( T = 500 )</th>
<th>( T = 750 )</th>
<th>( T = 1500 )</th>
<th>( T = 2000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRDW</td>
<td>0.162</td>
<td>0.027</td>
<td>0.060</td>
<td>0.004</td>
<td>0.013</td>
</tr>
<tr>
<td>Engle-Granger</td>
<td>0.195</td>
<td>0.250</td>
<td>0.137</td>
<td>0.182</td>
<td>0.100</td>
</tr>
<tr>
<td>( DF \ T (\hat{\rho} - 1) )</td>
<td>0.182</td>
<td>0.167</td>
<td>0.014</td>
<td>0.091</td>
<td>0.102</td>
</tr>
<tr>
<td>( \lambda_{\text{Trace}} )</td>
<td>0.092</td>
<td>0.177</td>
<td>0.011</td>
<td>0.182</td>
<td>0.013</td>
</tr>
<tr>
<td>( \lambda_{\text{Max}} )</td>
<td>0.087</td>
<td>0.018</td>
<td>0.063</td>
<td>0.010</td>
<td>0.044</td>
</tr>
<tr>
<td>Phillips-Ouliaris</td>
<td>0.182</td>
<td>0.250</td>
<td>0.007</td>
<td>0.318</td>
<td>0.100</td>
</tr>
</tbody>
</table>

The proportions of rejects are at 5% critical value for set parameter \( (\omega_i = 0.10, \alpha_i = 0.15, \beta_i = 0.65) \), \( (\omega_i = 0.05, \alpha_i = 0.20, \beta_i = 0.75) \) with Garch(1,1) Gaussian innovation distributed. The co-integration tests are applied in 3000 replication from the designed experiment for size test analysis.

### Proportion of Rejects for Power of the Test

**Table 4**

<table>
<thead>
<tr>
<th>TEST</th>
<th>( T = 100 )</th>
<th>( T = 500 )</th>
<th>( T = 750 )</th>
<th>( T = 1500 )</th>
<th>( T = 2000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRDW</td>
<td>0.727</td>
<td>0.636</td>
<td>0.300</td>
<td>0.727</td>
<td>0.918</td>
</tr>
<tr>
<td>Engle-Granger</td>
<td>0.888</td>
<td>0.988</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>( DF \ T (\hat{\rho} - 1) )</td>
<td>0.182</td>
<td>0.182</td>
<td>0.300</td>
<td>0.019</td>
<td>0.020</td>
</tr>
<tr>
<td>( \lambda_{\text{Trace}} )</td>
<td>0.909</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>( \lambda_{\text{Max}} )</td>
<td>0.818</td>
<td>0.901</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Phillips-Ouliaris</td>
<td>0.061</td>
<td>0.273</td>
<td>0.178</td>
<td>0.242</td>
<td>0.113</td>
</tr>
</tbody>
</table>

The proportions of rejects are at 5% critical value for set parameter \( (\omega_i = 0.10, \alpha_i = 0.15, \beta_i = 0.65) \), \( (\omega_i = 0.05, \alpha_i = 0.20, \beta_i = 0.75) \) with Garch(1,1) Gaussian innovation distributed. The co-integration tests are applied in 3000 replication from the designed experiment for power test analysis.
CONCLUSIONS

A Monte Carlo simulation experiment was performed to assess the power and size properties of six different co-integration tests in five different sample sizes, with Gaussian innovation. Two different sets of GARCH parameter models were experimented. We can summarize that there is size distortion in the several tests analyzed. And the power of the Eagle-Granger two-step test is quiet robust to heteroskedasticity and as powerful as the Johansen’s trace test. The Phillip-Ouliaris co-integration test proves not to be a good co-integration test.

In addition to comparing the different classical econometric approaches in solving co-integration tests, as well as analyzing its size and power properties for finite samples, statistical properties of co-integration test using the Bayesian framework could be compared to these classical tests as part of future research.

ACKNOWLEDGEMENT:

Our sincere thanks go to the Journal Editor and the anonymous reviewer(s) for their painstaking effort in reading this manuscript. Their useful criticisms, suggestions and corrections have greatly improved the standard of this paper.

BIBLIOGRAPHY

APPENDIX I

Graphs showing some selected co-integrated and non co-integrated series with different sample sizes for the parameter sets \( \{\alpha_0 = 0.01, \alpha_1 = 0.85, \beta_1 = 0.05\} \)

Figure 1
Graphs showing some selected simulated variance series with different sample sizes for parameter set \( (\omega_1 = 0.01, \alpha = 0.85, \beta = 0.05) \)

*Figure 2*
Graphs showing some selected simulated GARCH series with different sample sizes for the parameter set \((\omega_i = 0.01, \alpha_i = 0.85, \beta_i = 0.05)\)

*Figure 3*
Graphs showing some selected co-integrated and non co-integrated series for different sample sizes for parameter set \((\omega = 0.10, \alpha = 0.15, \beta = 0.65)\)

*Figure 4*
Graphs showing some selected simulated variance series for different sample sizes for parameter set $(\omega_i = 0.10, \alpha_i = 0.15, \beta_i = 0.65)$

Figure 5
Graphs showing some selected simulated GARCH series with different sample sizes for the parameter set \( (\omega_t = 0.10, \alpha_t = 0.15, \beta_t = 0.65) \)

*Figure 6*
Hierarchies of Asociative Dynamics, Starting From Romania’s Macro-Economic Imbalances in the EU-28. What Does Romania’s Economic Evolution in the EU-28 Look Like?

Prof. PhD. habil. Gheorghe SĂVOIU
University of Pitești

Senior lecturer PhD. Emilia GOGU
Bucharest University of Economic Studies

Lecturer PhD. Marian ȚAICU
University of Pitești

ABSTRACT

The authors’ answer to the second part of the title question is a threefold offer. First of all, they propose to improve the classic statistical ranking methods by capitalizing on the dynamic support of data series that are essentialized by the correlation or association ratio, as a structuring variable. Secondly, they provide an original method of ranking or hierarchizing a set of associative dynamics, or correlative evolutions. Lastly, they offer a disaggregation into partial equilibria turned into the analysis criteria of the general equilibrium theory through Nicholas Kaldor’s magical square, which became the magic rectangle of Lionel Stoléru’s strategy. All these three contributions are made in the standardized format of the sections of a paper of economic, statistical and econometric research.

Key words: macroeconomic equilibrium, magic quadrant or square, associative dynamical hierarchy, correlation matrix, econometric model, determinant coefficient.

Jel codes: B22, C46, C62, D58, F41, R13

INTRODUCTION

A simple question concerning the association of Romania’s economic evolution with any other dynamics of an economy in the EU-28 invariably triggers
a standard answer, designating the Bulgarian economy. The authors started from the dominant response or the preconceived solution, or else the hypothesis of self-induced validation, which gradually become false or outdated, as they attempt to give a chance to the alternative or alternatives appropriate to the new realities related to new similarities of national economy and other economic areas. At the same time, the authors exploit an econometric model as a tool to rank the territorial association or correlation, hoping that progress indicators that generate the main macroeconomic balances in Romania will place its economy on short-term horizons and influences of sustainable development, with other dynamics and trends at the end of more than two decades of evolutionary analysis.

An innovative approach, based on the Keynesian acceptance of the similarities of the “magic” balances belonging to Nicholas Kaldor (Kaldor, 1939; 1957; 1967), which became “strategic”, in the vision of Lionel Stoléru (Stoléru, 1967; 1972; 1974), generated in this article an original method of associative or correlative ranking dynamics, both econometric by integrating major modeling balances of growth, prices, labour market and external, and statistical, by evaluating the ranks in keeping with the intensity of associating to the macroeconomic construct the consolidated or stabilized balances according to the coefficient of determination in the final model, described briefly by the value of $R^2$ (Rsquared).

In summary, the present paper begins with a very brief introduction, brings together a presentation section concerning economic equilibrium theories, choosing the magical square solution, synthesizes the specificity of the new dynamic hierarchy-making (statistics of a both classical and econometric type, through the $R^2$ model of which it is being evaluated), then it details a section of results and discussion, identifying the hierarchies of associative dynamics, starting from Romania’s macroeconomic imbalances in pre-accession and post-accession to the EU-28. Some remarks concerning the limits of creativity in classical statistical hierarchies and the utility of the new method are the end of the article.

**A BRIEF REVIEW OF THE LITERATURE DEDICATED TO ECONOMIC EQUILIBRIUM THEORIES**

General economic equilibrium describes a situation of interdependent markets characterized by the absence of excess demand or supply, or in the view of mathematical inference, $n$ interdependent markets are in equilibrium if in $n - 1$ markets excess demand or supply is equal to zero (Genereux, 2001). In 1803, Jean-Baptiste-Say took the first step towards a more elaborate analysis of the imbalance between global supply and global demand, by stating the law of outlets. The first demonstration and the first structured statement of economic equilibrium theory are attributed to Léon Walras (1874), in his work *Elements of Pure Economics*, one of the first rigorous mathematical analyses of general equilibrium, centered on a specific exploration or tentative checking that still bears his name. In Walrasian theory, perfect price flexibility appeared as a condition for achieving and stability of the overall equilibrium or balance.
In 1911 there appeared a variant of the economic equilibrium theory drafted by Irving Fisher, which equalizes the product of the monetary mass and the circulation speed of the currency to the product of the general price level and the volume of transactions, and in 1917 Arthur Pigou originally nuances the economic equilibrium or balance, capitalizing on real income and a constant derived from a fraction of the real income that companies want to keep in the form of cash (the Cambridge solution). In 1936, Maynard Keynes developed the general theory of employment, interest and money, challenging from the outset the existence of a Walrasian exploration mechanism. Economic equilibrium becomes the object of a more careful study of the negotiations, and the Nash equilibrium represents a remarkable example in this respect, which transformed a much-discussed economic term into a central concept of the mathematical theory of games (Morgenstern and von Neumann, 1947; Nash, 1950). Kenneth Arrow and Gerard Debreu identified, as early as the 1950s, the main conditions for a Walrasian general balance, ranging from rationality, to pure and perfect competition, from complete markets to survival endowment, from the convexity of indifference curves to scale yields or absence of fixed costs. The Arrow-Debreu theorem disputes the anticipated effects of the exploration and imposes this very broad set of conditioning or restrictions. However, in 1973-1974, the Sonnenschein–Mantel–Debreu theorem stresses that, even if all conditions were met, nothing can guarantee that essential market mechanisms will achieve and maintain the general balance. Even when all the Arrow-Debreu hypotheses capable of generating a general balance were respected, the form of the functions of net demand for various goods remains undetermined, or more specifically, even if prices are perfectly flexible, nothing guarantees that price fluctuations ensure simultaneous convergence of all markets towards equilibrium (Genereux, 2001).

The relativity of general economic equilibrium is more and more frequently invoked as the depth of analysis grows, or attempts to focus on an economic and social optimum. Kenneth Arrow and Amartya Sen’s impossibility theorems (1970) found that simply defining an optimum becomes impossible, just like trying to imagine some democratic rules of public decision-making or policy capable of respecting freedom of expression and the principle of unanimity, or avoiding ineffective choices. If, over the course of two centuries, theorists tried to prove that the general balance exists, and real economic competition and price flexibility can contribute to the stability of a general equilibrium, some pragmatic economists, such as Nicholas Kaldor and Lionel Stoléru, managed to decompose a general economic balance in partial balances, reaggregating them into a complex concept, which transforms Kaldor’s magical square centred on the rate of economic growth, inflation, unemployment, and the trade balance in GDP (expressed as a percentage) into a quadrilateral of the main strategies for balancing an economy according to the Stoléru model. Linking the four major indicators, the four nodes or points on each axis of the square, defining four objectives of the economic policy of a state, describes a stable geometric figure with trends of continuity in development and diminishing the trade deficit, and the larger the surface of the square, the healthier the real network of an economy (obviously, with the restrictions of a sustainable development, with limiting inflation and unemployment thresholds).
Kaldor’s Magic quadrant, which became le carré magique with Stoléru, has a polar diagram as its visual expression, which synthesizes four objectives of macroeconomic policy: a) economic growth; b) the full employment of the labour force; c) price stability; s) external balance. The magic quadrant reveals, and even accentuates the conflictuality of some goals (Bezbakh, Gherardi, 2000, p. 36), especially those related to the hopefully full employment of labour force and price stability (theorized in the Phillips curve). In Figure 1 are presented the dimensions of the magical quadrant and the concrete ways of calculating the indicators on the quadrant axes. The quadrant was called “magic” because, according to Kaldor, it is impossible to attain the four goals at the same time.

Nicholas Kaldor’s magical quadrant, which became the quadrilateral of economic strategy, in the vision of Lionel Stoléru

![Figure 1](http://les-yeux-du-monde.fr/ressources/14053-quest-ce-le-carre-magique-de-kaldor)

In order to achieve the associated hierarchies, we made use of the classical methods of ranks and the relative distance, innovatively made use of by means of the correlation rate, matricially deduced with the Eviews software package, from the Eurostat databases ([http://ec.europa.eu/eurostat/data/database](http://ec.europa.eu/eurostat/data/database)).

The correlation matrix concerning economic growth initially had 44 or 45 series, from which the authors selected Romania, followed by two comparable EU-28
The new method proposed by the authors starts from the four equilibria or axes of the magic quadrant, transformed into endogenous variables (real GDP rate, or economic growth) and exogenous (inflation rate, unemployment rate and trade balance as a percentage of GDP) of an econometric model of multifactorial regression, applied to all 30 series of data for a longer time interval.

The method responds to a modern question and solves the problem of dynamic or evolutive hierarchies, over longer periods of time, in relation to one or more econometric modelling hypotheses, which gives it originality. The final ranking or hierarchy was made by the value $R^2$ or Rsquared, i.e. the coefficient of determination in relation to the level it reached in the model of the four constitutive equilibria of the magic quadrant: equilibrium of economic growth, price equilibrium, labour market equilibrium and external or macroeconomic equilibrium:

$$\text{Rate of PIB}_i = \alpha + \beta \times \text{Inflation rate}_i + \beta \times \text{Unemployment rate}_i + \gamma \times \text{Net balance of trade in GDP}_i + \varepsilon_i$$ (1)

The econometric model, which reflects the four previously commented balances, is built for both Romania and the other 27 European states, including two aggregations in the form of the EU-28 and the distinct Euro zone (19 countries), the authors opting for statistical gaps determined as discriminatory deviations calculated in the module (a simple difference between the determinations of the econometric models describing intensities, close or not, within the general balance or imbalance specific to each economy):

$$|R^2_i - R^2_{\text{Romania}}|$$ (2)

The originality of the method proceeds from solving a question related to the integration of econometric models derived from economic laws or theories, validated in dynamic hierarchies or rankings (associative or correlative hierarchies), which classical statistical methods cannot make in practice, both for reasons of static applicability, and for reasons related to the volume of data involved (120 separate series of data, for the period 1996-2016, were used in the correlation ratio calculations and in the econometric modeling).

**RESULTS AND DISCUSSION**

The first result, or the first step in the complex approach of the proposed method, after the realization of a correlation matrix, is obtaining an accurate database (which explains the option for six decimal figures), focused on the values of the correlation ratio between similar generators or quantifiers of the equilibria in the magical quadrant, both in Romania and in the other 27 EU-28 states (alphabetically
ordered, by their official names in English, but also in comparison with the EU-28 average and the Euro zone average (19 countries). The values of the correlation ratio \((R)\) for each economy or aggregate area, following the quantification of the intensity of the link or the evolutive association over the period 1996-2016, are presented in Table 1:

The values of the correlation ratio between the signalling indicators of equilibria specific to the magic quadrant, compared to the Romanian economy (1996-2016)

<table>
<thead>
<tr>
<th>Areas or economy correlated with the Romanian economy</th>
<th>Real GDP Ratio (RPR) or economic growth rate</th>
<th>Rate of inflation (RI)</th>
<th>Unemployment rate (RS)</th>
<th>Net balance of trade in GDP (SBCP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union (28 countries)</td>
<td>0.745384</td>
<td>0.833512</td>
<td>0.372838</td>
<td>0.795822</td>
</tr>
<tr>
<td>Euro area (19 countries)</td>
<td>0.734858</td>
<td>0.756100</td>
<td>-0.013902</td>
<td>0.745533</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.623701</td>
<td>0.555229</td>
<td>0.082971</td>
<td>-0.341178</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>0.889865</td>
<td>0.831962</td>
<td>0.788160</td>
<td>0.580668</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.769932</td>
<td>0.621872</td>
<td>0.828067</td>
<td>0.404197</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.651018</td>
<td>0.829612</td>
<td>-0.001994</td>
<td>0.729305</td>
</tr>
<tr>
<td>Germany</td>
<td>0.506316</td>
<td>0.657877</td>
<td>0.470365</td>
<td>0.136952</td>
</tr>
<tr>
<td>Estonia</td>
<td>0.592918</td>
<td>0.688922</td>
<td>0.459508</td>
<td>0.626486</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.280610</td>
<td>0.342164</td>
<td>-0.289093</td>
<td>0.816060</td>
</tr>
<tr>
<td>Greece</td>
<td>0.713799</td>
<td>0.842596</td>
<td>-0.209807</td>
<td>0.829388</td>
</tr>
<tr>
<td>Spain</td>
<td>0.778304</td>
<td>0.799258</td>
<td>-0.240992</td>
<td>0.901820</td>
</tr>
<tr>
<td>France</td>
<td>0.677817</td>
<td>0.730256</td>
<td>-0.154851</td>
<td>-0.171086</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.882371</td>
<td>0.814656</td>
<td>0.497383</td>
<td>0.710690</td>
</tr>
<tr>
<td>Italy</td>
<td>0.674100</td>
<td>0.756496</td>
<td>-0.040208</td>
<td>0.493712</td>
</tr>
<tr>
<td>Cyprus</td>
<td>0.625202</td>
<td>0.842015</td>
<td>-0.291483</td>
<td>0.500477</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.712356</td>
<td>0.630028</td>
<td>0.274686</td>
<td>0.847320</td>
</tr>
<tr>
<td>Lithuania</td>
<td>0.780126</td>
<td>0.751085</td>
<td>0.500731</td>
<td>0.810198</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.499522</td>
<td>0.766910</td>
<td>-0.548910</td>
<td>0.263037</td>
</tr>
<tr>
<td>Hungary</td>
<td>0.760474</td>
<td>0.764813</td>
<td>-0.227101</td>
<td>0.620520</td>
</tr>
<tr>
<td>Malta</td>
<td>0.284758</td>
<td>0.655127</td>
<td>0.715725</td>
<td>0.611976</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.757312</td>
<td>0.580973</td>
<td>-0.134210</td>
<td>0.373462</td>
</tr>
<tr>
<td>Austria</td>
<td>0.738401</td>
<td>0.516921</td>
<td>-0.223400</td>
<td>-0.045005</td>
</tr>
<tr>
<td>Poland</td>
<td>0.363192</td>
<td>0.801348</td>
<td>0.849096</td>
<td>0.605644</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.610539</td>
<td>0.516516</td>
<td>-0.391356</td>
<td>0.707721</td>
</tr>
<tr>
<td>Romania</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
<tr>
<td>Slovenia</td>
<td>0.876634</td>
<td>0.793488</td>
<td>0.053315</td>
<td>0.702531</td>
</tr>
<tr>
<td>Slovakia</td>
<td>0.789251</td>
<td>0.732481</td>
<td>0.902304</td>
<td>0.489558</td>
</tr>
<tr>
<td>Finland</td>
<td>0.756108</td>
<td>0.692679</td>
<td>0.596629</td>
<td>-0.433893</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.521128</td>
<td>0.733623</td>
<td>-0.147979</td>
<td>-0.813690</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.634270</td>
<td>0.821490</td>
<td>-0.166594</td>
<td>0.213048</td>
</tr>
</tbody>
</table>

Software used: Eviews

What can be noticed at this stage of intermediate processing, is a stronger similarity per indicator for Romania’s economy as compared to the EU-28 average (in relation to the EURO Zone), as well as a more pronounced similarity of the evolution of associations with certain economies, of which Bulgaria and Greece present values in the first three, two indicators generating specific balances of the magic quadrant.
By applying the classical rank-establishing method (according to the value of the correlation ratio with the data from the Romanian economy, as values expressed in the mathematical module), the following results are obtained (Table 2)

The final associative hierarchy, obtained by the rank method, improved according to the correlation ratio with the dynamics of Romania’s similar indicators

<table>
<thead>
<tr>
<th>Countries</th>
<th>Real GDP rate (RGR)</th>
<th>Rate of inflation (RI)</th>
<th>Rate of unemployment (RU)</th>
<th>Net trade balance in GDP (NTBG)</th>
<th>RGR</th>
<th>RI</th>
<th>RU</th>
<th>NTBG</th>
<th>Final rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgaria</td>
<td>0.889650, 0.831962</td>
<td>0.788160</td>
<td>0.589066</td>
<td></td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.882371, 0.814656</td>
<td>0.497383</td>
<td>0.710690</td>
<td></td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>Spain</td>
<td>0.778304, 0.799259</td>
<td>-0.249992, 0.901820</td>
<td></td>
<td></td>
<td>6</td>
<td>8</td>
<td>16</td>
<td>1</td>
<td>31</td>
</tr>
<tr>
<td>Lithuania</td>
<td>0.780126, 0.751085</td>
<td>0.500731</td>
<td>0.810198</td>
<td></td>
<td>5</td>
<td>13</td>
<td>8</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>Greece</td>
<td>0.713789, 0.842596</td>
<td>-0.209807, 0.829388</td>
<td></td>
<td></td>
<td>12</td>
<td>1</td>
<td>19</td>
<td>3</td>
<td>35</td>
</tr>
<tr>
<td>Slovakia</td>
<td>0.782925, 0.732481</td>
<td>0.902304</td>
<td>0.489558</td>
<td></td>
<td>4</td>
<td>15</td>
<td>1</td>
<td>18</td>
<td>38</td>
</tr>
<tr>
<td>Slovenia</td>
<td>0.876634, 0.793488</td>
<td>0.633315</td>
<td>0.702531</td>
<td></td>
<td>3</td>
<td>9</td>
<td>25</td>
<td>10</td>
<td>47</td>
</tr>
<tr>
<td>Hungary</td>
<td>0.760474, 0.764813</td>
<td>-0.227101, 0.630520</td>
<td></td>
<td></td>
<td>8</td>
<td>11</td>
<td>17</td>
<td>12</td>
<td>48</td>
</tr>
<tr>
<td>Poland</td>
<td>0.365192, 0.801348</td>
<td>0.849696</td>
<td></td>
<td></td>
<td>25</td>
<td>7</td>
<td>2</td>
<td>14</td>
<td>48</td>
</tr>
<tr>
<td>Cyprus</td>
<td>0.625202, 0.341018</td>
<td>-0.294583, 0.503477</td>
<td></td>
<td></td>
<td>15</td>
<td>2</td>
<td>13</td>
<td>16</td>
<td>49</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.712560, 0.630028</td>
<td>0.274686</td>
<td></td>
<td>0.847320</td>
<td>13</td>
<td>21</td>
<td>15</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.769932, 0.621872</td>
<td>0.828067</td>
<td></td>
<td>0.404197</td>
<td>7</td>
<td>22</td>
<td>3</td>
<td>20</td>
<td>52</td>
</tr>
<tr>
<td>Finland</td>
<td>0.756108, 0.692679</td>
<td>0.596629</td>
<td>-0.433893</td>
<td></td>
<td>10</td>
<td>17</td>
<td>6</td>
<td>19</td>
<td>52</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.651018, 0.829612</td>
<td>-0.001994, 0.729305</td>
<td></td>
<td></td>
<td>16</td>
<td>4</td>
<td>27</td>
<td>7</td>
<td>54</td>
</tr>
<tr>
<td>Estonia</td>
<td>0.592918, 0.688922</td>
<td>0.459508</td>
<td>0.626486</td>
<td></td>
<td>21</td>
<td>18</td>
<td>11</td>
<td>11</td>
<td>61</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.521128, 0.736323</td>
<td>-0.147979, -0.813690</td>
<td></td>
<td></td>
<td>22</td>
<td>14</td>
<td>22</td>
<td>5</td>
<td>63</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.495522, 0.760910</td>
<td>-0.548910, 0.265037</td>
<td></td>
<td></td>
<td>24</td>
<td>10</td>
<td>7</td>
<td>23</td>
<td>64</td>
</tr>
<tr>
<td>Malta</td>
<td>0.284758, 0.655127</td>
<td>0.715725</td>
<td>0.611976</td>
<td></td>
<td>26</td>
<td>20</td>
<td>5</td>
<td>13</td>
<td>64</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.634270, 0.821400</td>
<td>-0.166594, 0.213048</td>
<td></td>
<td></td>
<td>17</td>
<td>5</td>
<td>20</td>
<td>24</td>
<td>66</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.610519, 0.516516</td>
<td>-0.391136, 0.707721</td>
<td></td>
<td></td>
<td>20</td>
<td>26</td>
<td>12</td>
<td>9</td>
<td>67</td>
</tr>
<tr>
<td>Italy</td>
<td>0.674100, 0.756496</td>
<td>-0.049028, 0.497672</td>
<td></td>
<td></td>
<td>15</td>
<td>12</td>
<td>26</td>
<td>17</td>
<td>70</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.280619, 0.342164</td>
<td>-0.208993, 0.816000</td>
<td></td>
<td></td>
<td>27</td>
<td>27</td>
<td>14</td>
<td>4</td>
<td>72</td>
</tr>
<tr>
<td>France</td>
<td>0.677817, 0.730256</td>
<td>-0.14851, -0.171086</td>
<td></td>
<td></td>
<td>14</td>
<td>16</td>
<td>21</td>
<td>25</td>
<td>76</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.757312, 0.580973</td>
<td>-0.134210, 0.373462</td>
<td></td>
<td></td>
<td>9</td>
<td>23</td>
<td>21</td>
<td>21</td>
<td>76</td>
</tr>
<tr>
<td>Germany</td>
<td>0.506316, 0.657877</td>
<td>0.470365</td>
<td>0.136952</td>
<td></td>
<td>23</td>
<td>19</td>
<td>10</td>
<td>26</td>
<td>78</td>
</tr>
<tr>
<td>Austria</td>
<td>0.738401, 0.516921</td>
<td>-0.224260, -0.045005</td>
<td></td>
<td></td>
<td>11</td>
<td>25</td>
<td>18</td>
<td>27</td>
<td>81</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.623701, 0.555229</td>
<td>0.082971</td>
<td>-0.341718</td>
<td></td>
<td>19</td>
<td>24</td>
<td>24</td>
<td>22</td>
<td>89</td>
</tr>
</tbody>
</table>

Source: The data in Table 2 were processed by the authors and reordered according to the final ranks

Only seven of the top ten countries, according to the final rankings, remain in the first third of ranking in the confrontation of the two improved classical methods, to be transformed from static to dynamic methods (Table 3), and only Bulgaria retains its rank, and implicitly the position, which confirms the highest evolutionary similarity with the Romanian economy.
The final associative hierarchy, obtained by the relative distance method, improved according to the correlation ratio with the dynamics of Romania’s similar indicators

<table>
<thead>
<tr>
<th>Țări</th>
<th>Real GDP rate (RGR)</th>
<th>Rate of inflation (RI)</th>
<th>Rate of unemployment (RU)</th>
<th>Net trade balance in GDP (NTBG)</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>Average D</th>
<th>Final Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgaria</td>
<td>0.889650</td>
<td>0.788160</td>
<td>0.580668</td>
<td>100</td>
<td>98.7395</td>
<td>67.34972</td>
<td>64.3846</td>
<td>86.33</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Slovakia</td>
<td>0.782516</td>
<td>0.902304</td>
<td>0.489558</td>
<td>98.69354</td>
<td>86.93146</td>
<td>100</td>
<td>54.2856</td>
<td>80.43</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Croatia</td>
<td>0.882711</td>
<td>0.497383</td>
<td>0.710690</td>
<td>90.15785</td>
<td>96.68406</td>
<td>55.12366</td>
<td>78.8061</td>
<td>80.33</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td>0.780126</td>
<td>0.500731</td>
<td>0.810198</td>
<td>87.6679</td>
<td>89.1394</td>
<td>55.49471</td>
<td>89.8402</td>
<td>79.01</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.769352</td>
<td>0.828067</td>
<td>0.404197</td>
<td>86.52234</td>
<td>73.40829</td>
<td>91.77251</td>
<td>94.8204</td>
<td>71.59</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>0.363192</td>
<td>0.849669</td>
<td>0.605644</td>
<td>40.84126</td>
<td>95.10654</td>
<td>94.16959</td>
<td>76.1579</td>
<td>70.39</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>0.751089</td>
<td>0.596269</td>
<td>0.433893</td>
<td>94.89809</td>
<td>82.2072</td>
<td>86.12268</td>
<td>76.0224</td>
<td>68.66</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>0.778304</td>
<td>0.240992</td>
<td>0.901820</td>
<td>97.86154</td>
<td>94.85661</td>
<td>26.70852</td>
<td>100</td>
<td>68.61</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Estonia</td>
<td>0.592918</td>
<td>0.459508</td>
<td>0.626486</td>
<td>66.63011</td>
<td>81.76184</td>
<td>80.92697</td>
<td>96.4907</td>
<td>66.26</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td>0.713789</td>
<td>0.209807</td>
<td>0.829388</td>
<td>60.21388</td>
<td>100</td>
<td>23.25236</td>
<td>91.9824</td>
<td>64.36</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td>0.712356</td>
<td>0.274686</td>
<td>0.847320</td>
<td>68.61026</td>
<td>91.30055</td>
<td>43.37297</td>
<td>78.4707</td>
<td>61.51</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>0.760474</td>
<td>0.227101</td>
<td>0.620520</td>
<td>85.49489</td>
<td>90.76865</td>
<td>25.16901</td>
<td>68.80752</td>
<td>60.54</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Malta</td>
<td>0.284758</td>
<td>0.715725</td>
<td>0.611976</td>
<td>32.00013</td>
<td>77.75102</td>
<td>79.21941</td>
<td>67.86011</td>
<td>60.49</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>0.625202</td>
<td>0.291483</td>
<td>0.500477</td>
<td>70.25087</td>
<td>99.93105</td>
<td>32.3043</td>
<td>55.4633</td>
<td>59.56</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.499522</td>
<td>0.548910</td>
<td>0.263037</td>
<td>56.13458</td>
<td>91.01752</td>
<td>60.83426</td>
<td>29.16735</td>
<td>54.87</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>0.521128</td>
<td>0.147979</td>
<td>0.813698</td>
<td>58.56259</td>
<td>87.06699</td>
<td>16.40013</td>
<td>90.22754</td>
<td>52.41</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td>0.876834</td>
<td>0.033535</td>
<td>0.702531</td>
<td>98.31513</td>
<td>94.17182</td>
<td>89.20876</td>
<td>77.9014</td>
<td>45.46</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>0.206010</td>
<td>0.299993</td>
<td>0.816060</td>
<td>31.534</td>
<td>40.60031</td>
<td>32.03942</td>
<td>90.4934</td>
<td>43.90</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.573132</td>
<td>0.134210</td>
<td>0.373462</td>
<td>85.10418</td>
<td>98.95036</td>
<td>84.47144</td>
<td>41.41203</td>
<td>43.60</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>0.506316</td>
<td>0.470365</td>
<td>0.136952</td>
<td>56.89067</td>
<td>78.07739</td>
<td>52.12935</td>
<td>55.18018</td>
<td>43.31</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.634700</td>
<td>0.166594</td>
<td>0.213048</td>
<td>71.2771</td>
<td>97.49151</td>
<td>18.46131</td>
<td>23.6243</td>
<td>41.73</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>0.677917</td>
<td>0.154851</td>
<td>0.171086</td>
<td>76.1707</td>
<td>86.66739</td>
<td>17.16735</td>
<td>18.9713</td>
<td>38.29</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>0.674100</td>
<td>0.040208</td>
<td>0.493712</td>
<td>75.75308</td>
<td>80.78158</td>
<td>45.54168</td>
<td>42.47418</td>
<td>35.89</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>0.627011</td>
<td>0.082971</td>
<td>0.341178</td>
<td>70.0894</td>
<td>65.89004</td>
<td>91.95456</td>
<td>37.83216</td>
<td>35.60</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>0.738401</td>
<td>0.222460</td>
<td>0.045005</td>
<td>82.97896</td>
<td>91.34862</td>
<td>24.64648</td>
<td>49.90406</td>
<td>28.13</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>0.651081</td>
<td>0.282961</td>
<td>0.021994</td>
<td>73.79053</td>
<td>73.15918</td>
<td>98.45055</td>
<td>22.00908</td>
<td>80.87035</td>
<td>18.94</td>
<td>26</td>
</tr>
</tbody>
</table>

Source: The data in Table 2 were processed by the authors and reordered according to the final ranks.
The econometric model of the four partial equilibria that make up Romania’s magical quadrant (1996-2016)

Table 4

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-8.310266</td>
<td>8.889692</td>
<td>-0.934820</td>
<td>0.3638</td>
</tr>
<tr>
<td>SER55 (RI)</td>
<td>-0.054916</td>
<td>0.023616</td>
<td>-2.325345</td>
<td>0.0335</td>
</tr>
<tr>
<td>SER85 (RS)</td>
<td>1.412508</td>
<td>1.232565</td>
<td>1.145990</td>
<td>0.2687</td>
</tr>
<tr>
<td>SER115 (SBCP)</td>
<td>-0.393538</td>
<td>0.209670</td>
<td>-1.876943</td>
<td>0.0789</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.393376</td>
<td>Mean dependent var 2.830000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.279634</td>
<td>S.D. dependent var 4.281183</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>3.633627</td>
<td>Akaike info criterion 5.595196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>211.2519</td>
<td>Schwarz criterion 5.794343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-51.95196</td>
<td>F-statistic 3.458494</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>1.511674</td>
<td>Prob(F-statistic) 0.041475</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The determinant coefficient R², or Rsquared, subsequently hierarchizes in relation to the effectively achieved level of 0.393376 in the model of the four constitutive balances of the magic quadrant, starting from the deviation from the value in the econometric model specific to Romania for the 1996-2016 period (Table 5 and Table 6). Comparing the final gaps of R² as to the two EU-28 final aggregates and the EURO zone, by means of the new method, it can be noticed that Romania is already closer, in point of a correlative intensity, to this monetary area made up of 19 countries, more economically developed as a matter of principle.

The final correlative hierarchy of EU aggregates, obtained by the new method centred on econometric modeling, structured according to the determinant coefficient R² or Rsquared

Table no. 5

<table>
<thead>
<tr>
<th>Aggregated areas</th>
<th>The ranking method centered on R², or Rsquared, of the reunited equilibrium models Rhythm GDP = Inflation rate + Unemployment rate + Trade balance in GDP</th>
<th>Deviation (gap) R² - R² Romania</th>
<th>Considerations relating to evolutive similarities according to the comparison of the intensity of the models</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union (28 countries)</td>
<td>0.163045</td>
<td>0.230331</td>
<td>Romania is closer to EU-28 as intensity of the correlation</td>
</tr>
<tr>
<td>Euro area (19 countries)</td>
<td>0.232663</td>
<td>0.160713</td>
<td>Romania is closer to Euro area as intensity of the correlation Zona EURO</td>
</tr>
</tbody>
</table>

Software used: EViews
Comparing the final rankings of the new method, it is obvious that only five countries are among the top ten in the face of the improved classical rank method, and as many out of the comparison with the improved method of relative distance, and the hierarchy of the first two places is completely changed, this time Bulgaria appearing only on the third position. Hungary tends to be the closest to Romania’s economy, in point of intensity of the model of the equilibria in the magic quadrant (intrinsic determination), and surprisingly Ireland appears on the second place (some values in equilibrium evolution confirm contradictory tendencies similar to the Romanian economy, although they seem to be lack level similarities). The new method centred on the econometric modeling and assigning ranks according to the determination coefficient $R^2$ or Rsquared reached in relation to a well-established economic theory (in this case, the general equilibrium of the magic quadrant, disaggregated in four specific equilibria)  hierarchizes according to the intensity of the correlation of the general equilibrium, or in other words, the method is simultaneously statistical, econometric and economic by the theoretical arguments of the final ordering of gaps (the deviations of specific determinations).

The final correlative hierarchy of the other EU-28 countries, obtained by the new method focused on econometric modelling and ranking by the $R^2$ or Rsquared determination coefficient reached in relation to an established economic theory

<table>
<thead>
<tr>
<th>Countries</th>
<th>Ranking method centered on R2 or R squared of the reunited equilibrium models Rhythm GDP = Inflation rate + Unemployment rate + Trade balance in GDP</th>
<th>Gap (deviation)</th>
<th>Final rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hungary</td>
<td>0.378879</td>
<td>0.014497</td>
<td>1</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.378705</td>
<td>0.014671</td>
<td>2</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>0.430023</td>
<td>0.034646</td>
<td>3</td>
</tr>
<tr>
<td>Greece</td>
<td>0.337695</td>
<td>0.055681</td>
<td>4</td>
</tr>
<tr>
<td>Finland</td>
<td>0.336478</td>
<td>0.056898</td>
<td>5</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.452832</td>
<td>0.059456</td>
<td>6</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.324212</td>
<td>0.069164</td>
<td>7</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.303679</td>
<td>0.089697</td>
<td>8</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.491723</td>
<td>0.098347</td>
<td>9</td>
</tr>
<tr>
<td>Lithuania</td>
<td>0.286133</td>
<td>0.107243</td>
<td>10</td>
</tr>
<tr>
<td>Estonia</td>
<td>0.327350</td>
<td>0.121006</td>
<td>11</td>
</tr>
<tr>
<td>Cyprus</td>
<td>0.514462</td>
<td>0.121086</td>
<td>12</td>
</tr>
<tr>
<td>France</td>
<td>0.271329</td>
<td>0.122047</td>
<td>13</td>
</tr>
<tr>
<td>Poland</td>
<td>0.250791</td>
<td>0.142585</td>
<td>14</td>
</tr>
<tr>
<td>Malta</td>
<td>0.557546</td>
<td>0.164170</td>
<td>15</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.217856</td>
<td>0.175520</td>
<td>16</td>
</tr>
<tr>
<td>Spain</td>
<td>0.509992</td>
<td>0.197616</td>
<td>17</td>
</tr>
<tr>
<td>Slovenia</td>
<td>0.186531</td>
<td>0.200843</td>
<td>18</td>
</tr>
<tr>
<td>Austria</td>
<td>0.177602</td>
<td>0.215774</td>
<td>19</td>
</tr>
<tr>
<td>Germany</td>
<td>0.149991</td>
<td>0.243355</td>
<td>20</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.139586</td>
<td>0.253790</td>
<td>21</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.672651</td>
<td>0.279275</td>
<td>22</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.108965</td>
<td>0.284411</td>
<td>23</td>
</tr>
<tr>
<td>Slovakia</td>
<td>0.070646</td>
<td>0.322730</td>
<td>24</td>
</tr>
<tr>
<td>Italy</td>
<td>0.061261</td>
<td>0.332115</td>
<td>25</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.055903</td>
<td>0.337473</td>
<td>26</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.051573</td>
<td>0.341603</td>
<td>27</td>
</tr>
</tbody>
</table>

Software used: EViews
We should perceive the contribution of econometric thinking not as one of modeling validation, nor as a strict assessment of a number of tested hypotheses, but rather as one of ordering models that are relatively valid over time, and ultimately offering a more general criterion in a context of a well-established economic theory.

**CONCLUSIONS**

Of the three methods presented, with their different results in proportions ranging from 50% to 66%, the authors give greater credibility to the method of correlation-based ranking, or the method centred on associative dynamics, which performs an evolutive statistical analysis, economically discriminates from a criterion point of view, and relativizes stationary data, conferring authority on dynamics by integrating a modelling economic theory. This theory is clearly different in comparison to its application to various territorial hierarchies of specific domains, for example in the case of foreign direct investment analyses, one can opt for John H. Dunning’s eclectic theory, which allows the step-planning and selection of hierarchy factors with much rigor.

The originality of the new complex dynamic hierarchy-building method for economies following the associated analysis of major macroeconomic equilibria in a well-established manner, applied in relation to Romania, for four international statistical indicators as endogenous (economic growth) and exogenous variables (inflation rate, unemployment rate, net trade balance as a percentage of GDP), represents, in the opinion of the authors, a necessary and pragmatic stratification activity, and is the result of a compromise dictated by the existence of a limited number of statistically comparable series, although the authors worked on 120 series of data collected from Eurostat.

The method of correlative ranking, or the method centred on associative dynamics, provides continuity to classic methods in the stationary stalemate, or even static information from a temporal point of view. The pragmatism and originality of the method provide another answer to the question in the second title of the paper, as its quality depends on the validity and perishability of the economic theory that constitutes the source of the discriminated and hierarchical model, in this case the theory of the general economic equilibrium materialized in the magic quadrant or quadrilateral.

**Bibliography**

Preserving Logical Relations while Estimating Missing Values

Ton de Waal (t.dewaal@cbs.nl)
Statistics Netherlands, & Tilburg University

Wieger Coutinho
Statistics Netherlands

ABSTRACT

Item-nonresponse is often treated by means of an imputation technique. In some cases, the data have to satisfy certain constraints, which are frequently referred to as edits. An example of an edit for numerical data is that the profit of an enterprise equals its turnover minus its costs. Edits place restrictions on the imputations that are allowed and hence complicate the imputation process. In this paper we explore an adjustment approach. This adjustment approach consists of three steps. In the first step, the imputation step, nearest neighbour hot deck imputation is used to find several pre-imputed values. In a second step, the adjustment step, these pre-imputed values are adjusted so the resulting records satisfy all edits. In a third step, the best donor record is selected. The adjusted record corresponding to that donor record is the final imputed record. In principle, a potential donor that is not the closest to the record to be imputed may still give the best results after adjustment. In this paper we therefore focus on the number of potential donor records that are considered in the imputation step.

Keywords: Nearest-neighbour Imputation, Edit restrictions, Linear programming, Data adjustment

JEL classification: C13, C14, C61, C81, C83

INTRODUCTION

Item-nonresponse is a frequently occurring problem in survey data. An often used approach to treat missing data due to item-nonresponse is imputation, where individual values are estimated and filled into the data set.

In some cases, the data have to satisfy certain constraints, which are often referred to as edit rules or edits for short. Examples of such edits for numerical data are that the profit of an enterprise equals its turnover minus its costs, and that the turnover of an enterprise should be at least zero. These edits place restrictions on the imputations that are allowed and hence complicate the imputation process.

In this paper we describe an imputation approach that satisfies the edits. We will refer to this approach as adjusted nearest-neighbour imputation.

The remainder of this paper is organized as follows. Section 2 gives a literature review. Section 3 describes the edits we consider in this paper and the proposed imputation approach. Section 4 gives the results of several variations of the imputa-
tion approach and discusses these results. Section 5 ends the paper by drawing some conclusions and identifying some possible topics for future research.

LITERATURE REVIEW

Many imputation methods have been developed for many different situations and kinds of data sets, and are discussed in a large number of articles and books, such as Andridge and Little (2010), Kalton and Kasprzyk (1986), Rubin (1987), Schafer (1997), Little and Rubin (2002), De Waal, Pannekoek and Scholtus (2011) and Van Buuren (2012).

In particular several imputation methods satisfying edits have been developed, see Geweke (1991), Raghunathan, Solenberger and Van Hoewyk, (2002), Tempelman (2007), Coutinho, De Waal and Remmerswaal (2011), Coutinho and De Waal (2012), Pannekoek, Shlomo and De Waal (2013), Kim et al. (2014) and De Waal, Coutinho and Shlomo (forthcoming). These methods are generally quite complicated, and are based either on sequentially imputing the variables with missing values or on somehow truncating a statistical distribution, such as the multivariate normal, to the region defined by the edits. We note that, besides satisfying edits, the methods proposed by Pannekoek, Shlomo and De Waal (2013 and De Waal, Coutinho and Shlomo (forthcoming) also preserve known or previously estimated totals.

Pannekoek and Zhang (2015) proposed a much simpler adjustment approach for satisfying edits, consisting of two steps. In the first step, the imputation step, nearest neighbour hot deck imputation is used to find a donor record that is closest to the record to be imputed. Missing values in the record to be imputed are pre-imputed with values from that donor record. In a second step, the adjustment step, these pre-imputed values are adjusted so the resulting record satisfies all edits.

In the current paper we explore this approach in more detail. In particular, we will focus on the number of potential donor records that are considered in the imputation step. In principle, a potential donor that is not the closest to the record to be imputed may still give the best results after adjustment. Whereas Pannekoek and Zhang (2015) considered only the closest donor record to the record to be imputed in the imputation step, we will consider the closest records, where \( k = 1, 2, 5 \) or 10. We will also provide more evaluation results than Pannekoek and Zhang (2015).

METHODOLOGY

Before we describe the proposed methodology, we first need to describe the kind of edits that we consider as this clarifies the problem we are trying to solve. This is the topic of subsection 3.1 below. Our adjusted nearest-neighbour method itself is described in subsection 3.2.

Edits

In this paper we consider edits for numerical data. Such edits are generally either linear equations or linear inequalities. In other words, an edit can be written as:
We start by describing the idea of our adjusted nearest-neighbour imputation method. For each record to be imputed our method consists of 3 steps. In the first step we use observed values from potential donor records to construct a pre-imputed record for each potential donor record. These pre-imputed records may, and often will, violate some of the edits. In a second step we therefore adjust the imputed values in these pre-imputed records so these adjusted records satisfy all edits. In a third step we select the donor record for which the distance of the donor record to the record to be imputed plus the distance of the pre-imputed record to the adjusted record is minimal. The adjusted record corresponding to that donor record is our final imputed record.

Since using all potential donor records may be too time-consuming, in practice we restrict ourselves to the \( k \) nearest-neighbour donor records instead of all potential donor records.

We will now describe the above idea in mathematical terms. Let us suppose that we want to impute a certain record \( x_{r_0} \), the recipient record. Let \( x_{\text{hot}}(x_{r_0}, x_r) \) denote the record that would be obtained if the missing values in record \( x_{r_0} \) were imputed with the corresponding (observed) values in record \( x_r \). Also, let \( x_{\text{adj}}(x_{r_0}, x_r) \) denote the record after adjusting the record \( x_{\text{hot}}(x_{r_0}, x_r) \) so the adjusted record satisfies the edit restrictions.

We aim to find the donor record that we will use to impute the recipient by minimizing the distance of the donor record \( x_r \) to the recipient record \( x_{r_0} \), plus the distance of \( x_{\text{hot}}(x_{r_0}, x_r) \) to \( x_{\text{adj}}(x_{r_0}, x_r) \). That is, we solve the following problem

\[
\min_{x_r \in D(x_{r_0})} \left[ d_{\text{obs}}(x_{r_0}, x_r) + d_{\text{mis}}(x_{r_0}, x_{\text{hot}}(x_{r_0}, x_r), x_{\text{adj}}(x_{r_0}, x_r)) \right]
\]

where \( D(x_{r_0}) \) is the set of potential donor records for \( x_{r_0} \), (the so-called donor pool), \( d_{\text{obs}}(x_{r_0}) \) the distance function used to measure the distance between records restricted to the variables observed in \( x_{r_0} \), and \( d_{\text{mis}}(x_{r_0}) \) the distance function used restricted to the variables with missing values in \( x_{r_0} \).
The adjusted record is then given by

\[ x_{\text{adj}}(x_{r_0}, x_r) \equiv x_{\text{hot}}(x_{r_0}, x_{\text{opt}}(x_{r_0}, x_r)), \]

In words, the record \( x_{\text{adj}}(x_{r_0}, x_r) \) is obtained by adjusting the donor values from record \( x_r \) to values from a synthetic record \( x_{\text{opt}}(x_{r_0}, x_r) \) so that the adjusted donor values together with the observed values in \( x_{r_0} \) satisfy all edits.

As we already mentioned at the beginning of this subsection, solving (3) to optimality is in many cases quite time-consuming since all potential donor records have to be considered. Instead of solving the above problem to optimality we will instead apply a less time-consuming approach. This approach is likely to find the optimal solution for most records in any case. In our approach we first select the neighbors of \( x_{r_0} \) according to the distance function \( d \), say \( x_{i}^{*} (i = 1, \ldots, k) \). For each \( x_{i}^{*} \) we then solve

\[ x_{\text{opt}}(x_{r_0}, x_{i}^{*}) = \arg \min_{x} \left[ d_{\text{mis}}(x_{r_0}) (x_{\text{hot}}(x_{r_0}, x_{i}^{*}), x_{\text{hot}}(x_{r_0}, x_{a})) \right], \quad x_{\text{hot}}(x_{r_0}, x_{d}) \in A \]  

Next, we calculate

\[ d_{\text{obs}}(x_{r_0}, x_{i}^{*}) + d_{\text{mis}}(x_{r_0}) (x_{\text{hot}}(x_{r_0}, x_{i}^{*}), x_{\text{adj}}(x_{r_0}, x_{i}^{*})) \]  

and select the record \( x_{i}^{*} (i = 1, \ldots, k) \) for which (5) is the smallest. Our imputed record is then given by \( x_{\text{adj}}(x_{r_0}, x_{i}^{*}) \). This heuristic returns the optimal solution to (3), unless the optimal donor record is not among the \( k \) nearest-neighbours of \( x_{r_0} \).

In this paper we will use the sum of absolute differences to measure the distance between two records \( x_1 \) and \( x_2 \), i.e. we will use

\[ d_s(x_1, x_2) = \sum_{j \in S} u_j |x_{1j} - x_{2j}| \]  

with \( S \) the set of variables that are used to calculate the distance function, and \( u_j \) a weight for the \( j \)-th variable indicating the importance of a change in that variable.

Substituting (6) into (4), we see that in step 2 of our approach, i.e. the adjustment step, we need, for each \( x_{i}^{*} (i = 1, \ldots, k) \), to solve a linear programming problem with objective function given by

\[ \sum_{j \in \text{mis}(x_{r_0})} u_j |x_{i,j}^{*} - x_{a,j}| \]  

subject to the constraint that record \( x_{\text{adj}}(x_{r_0}, x_{i}^{*}) \) satisfies all edits. This can be formulated as a linear programming problem and can, for instance, be solved by the well-known simplex algorithm.
A technical problem is that many implementations of the simplex algorithm cannot minimize a sum of absolute distances directly. We overcome this technical problem by introducing variables $\lambda_j (j \in \text{mis}(x_{r_n}))$ that have to satisfy

$$\lambda_j \geq x^*_i - x_{a_j} \text{ for } j \in \text{mis}(x_{r_n})$$

(8)

$$\lambda_j \geq x_{a_j} - x^*_i \text{ for } j \in \text{mis}(x_{r_n})$$

(9)

Our minimization problem is now given by:

$$\text{Minimize } \sum_{j \in \text{mis}(x_{r_n})} u_j \lambda_j$$

(10)

subject to (8), (9) and the constraint that $x_{\text{tot}}(x_{r_n}, x_a)$ satisfies all edits. Since in an optimal solution $\lambda_j = x^*_i - x_{a_j}$ or $\lambda_j = x_{a_j} - x^*_i \text{ (for } j \in \text{mis}(x_{r_n}))$, we have that in an optimal solution $\lambda_j = |x^*_i - x_{a_j}| \text{ (for } j \in \text{mis}(x_{r_n}))$. Hence, minimizing (10) subject to (8), (9) and the constraint that $x_{\text{tot}}(x_{r_n}, x_a)$ satisfies all edits leads to the same solution for $x_a$ as minimizing (7) subject to the constraint that $x_{\text{tot}}(x_{r_n}, x_a)$ satisfies all edits.

We illustrate our imputation method by means of the example below.

**Example**

Suppose there are four variables, $T$ (turnover), $P$ (profit), $C$ (costs), and $N$ (number of employees in fulltime equivalents), and that the edits are given by

$$T = P + C$$

(11)

$$P \leq 0.5T$$

(12)

$$0.1T < P$$

(13)

$$T \leq 550 N$$

(14)

$$T \geq 0$$

(15)

$$N \geq 0$$

(16)

$$C \geq 0$$

(17)

Let us suppose that the weight of variable $N$ in (6) equals 500 and of the other three variables 1. Suppose furthermore that in a certain record $N = 5$, $T = 2000$ and the values of $P$ and $C$ are missing.

Now suppose that in one of the $k$ nearest (potential) donor records the following values are observed: $N^* = 6$, $T^* = 2200$, $P^* = 900$ and $C^* = 1300$. We then find adjusted values $\hat{P}$ and $\hat{C}$ such that $\hat{P}$ and $\hat{C}$ together with the observed values $N = 5$ and $T = 2000$ satisfy (11) to (17) and

$$|\hat{P} - 900| + |\hat{C} - 1300|$$

is minimized. The problem can easily be formulated as a linear programming problem, which can, for instance, be solved by means of the simplex algorithm. A solution is $\hat{P} = 900$ and $\hat{C} = 1100$.

The total distance (6) is then given by

$$500|5 - 6| + |2000 - 2000| + |900 - 900| + |1300 - 1100| = 700$$
Likewise we calculate this distance for the other \((k - 1)\) nearest (potential) donor records. From all \(k\) imputed records we select the record with the smallest value for (5).

**RESULTS AND DISCUSSION**

**Evaluation approach**

In our evaluation study we have used two data sets. The true values for the data in the two data sets are known. In the completely observed data sets values were deleted, using a Missing At Random (MAR) mechanism. When the missing data mechanism is MAR, there is a relation between the missing data pattern and the values of the observed data, but not between the missing data pattern and the values of the missing data.

For each of our evaluation data sets we thus have two versions available: a version with missing values and a version with complete records. The former version is imputed, without making any use of the complete records. The resulting data set is then compared to the version with complete records.

**Methods evaluated**

In our evaluation study we have used a weighted and an unweighted version of our adjusted nearest-neighbour approach. In the weighted version the weight \(u_j\) for the \(j\)-th variable in (6) is set to the reciprocal of the observed mean for this variable in the data set with missing data. In the unweighted version the weight \(u_j\) is set to 1 for all variables. We will denote our imputation methods by W 1, W 2, W 5, W 10, NW 1, NW 2, NW 5 and NW 10, where “W” indicates a weighted version and “NW” an unweighted version, and the number indicates the number of nearest-neighbours considered. For instance, NW 5 denotes the imputation method where all weights \(u_j\) in (6) have been set to 1, and 5 nearest-neighbours are considered.

We have also examined taking the average over all 8 imputations methods. We consider this as a ninth method, and will be denoted by Mixed. By taking the average over the above-mentioned 8 imputation methods, we have constructed a kind of implicit fractional imputation method (see e.g. Kim and Fuller 2004 and Kim 2011). In fractional imputation, an imputed value is actually a weighted sum of several imputed values, for instance obtained from several donor records. In fractional imputation explicit weights are used. In our case, the weighting is implicit, and depends on how often the same donor value is used to impute a record.

To evaluate the results of our imputation methods we have compared them to nearest-neighbour hot deck imputation, using the sum of the absolute differences (6) as distance function, and random hot deck imputation. We will refer to nearest-neighbour hot deck imputation as NN HD and to random hot deck imputation as Random HD.

**Evaluation data**

The main characteristics of the data sets are presented in Table 1.
The characteristics of the evaluation data sets

<table>
<thead>
<tr>
<th></th>
<th>data set 1</th>
<th>data set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of records</td>
<td>3096</td>
<td>500</td>
</tr>
<tr>
<td>Number of records with missing values</td>
<td>287</td>
<td>250</td>
</tr>
<tr>
<td>Total number of variables</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Total number of edits</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>Number of balance edits</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total number of inequality edits</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>Number of non-negativity edits</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Tables 2 and 3 give the numbers of missing values and (unweighted) means of the variables of our data sets. In brackets the percentages of records with a missing value for the corresponding variable out of the total number of 3,096 records for data set 1 and 500 records for data set 2 is given. The means are taken over all observations in the complete versions of the data sets. Variable $R_8$ in data set 1 does not contain any missing values. This variable is only used as auxiliary variable.

The numbers of missing values and the means in data set 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of missing values</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_1$</td>
<td>0 (0.0 %)</td>
<td>37.4</td>
</tr>
<tr>
<td>$R_2$</td>
<td>79 (2.2 %)</td>
<td>777.6</td>
</tr>
<tr>
<td>$R_3$</td>
<td>76 (4.2 %)</td>
<td>11574.8</td>
</tr>
<tr>
<td>$R_4$</td>
<td>73 (4.8 %)</td>
<td>209.9</td>
</tr>
<tr>
<td>$R_5$</td>
<td>67 (2.6 %)</td>
<td>169.2</td>
</tr>
</tbody>
</table>

The numbers of missing values and the means in data set 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of missing values</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>61 (12.2 %)</td>
<td>97.8</td>
</tr>
<tr>
<td>$S_2$</td>
<td>86 (17.2 %)</td>
<td>175018.3</td>
</tr>
<tr>
<td>$S_3$</td>
<td>131 (26.2 %)</td>
<td>731.0</td>
</tr>
<tr>
<td>$S_4$</td>
<td>61 (12.2 %)</td>
<td>175749.3</td>
</tr>
<tr>
<td>$S_5$</td>
<td>109 (21.8 %)</td>
<td>154286.5</td>
</tr>
<tr>
<td>$S_6$</td>
<td>91 (18.2 %)</td>
<td>7522.3</td>
</tr>
</tbody>
</table>

When method $W_1$, $W_2$, $W_5$, $W_10$, $NW_1$, $NW_2$, $NW_5$ or $NW_{10}$ is used, sometimes values from a donor record are used directly whereas in other cases these values are adjusted as a result of solving a linear programming problem. Table 4 below reports how many times donor values were used directly (see the columns “Donor”) and how many times donor values were adjusted (see the columns “Adjusted”). As could be expected, the number of times donor values were used directly increases as the number of nearest-neighbour donors considered increases.
Number of times donor respectively simplex is used

Table 4

<table>
<thead>
<tr>
<th></th>
<th>Data set 1</th>
<th>Data set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Donor</td>
<td>Simplex</td>
</tr>
<tr>
<td>W 1</td>
<td>266</td>
<td>21</td>
</tr>
<tr>
<td>W 2</td>
<td>273</td>
<td>14</td>
</tr>
<tr>
<td>W 5</td>
<td>280</td>
<td>7</td>
</tr>
<tr>
<td>W 10</td>
<td>281</td>
<td>6</td>
</tr>
<tr>
<td>NW 1</td>
<td>266</td>
<td>21</td>
</tr>
<tr>
<td>NW 2</td>
<td>276</td>
<td>11</td>
</tr>
<tr>
<td>NW 5</td>
<td>279</td>
<td>8</td>
</tr>
<tr>
<td>NW 10</td>
<td>282</td>
<td>5</td>
</tr>
</tbody>
</table>

Evaluation measures

In our evaluation study we focus on three measures proposed by Chambers (2003). Each of these measures examines a different aspect, namely the preservation of individual values, the preservation of totals and means, and the preservation of univariate statistical distributions.

The preservation of individual values is measured by the $d_{L1}$ measure. The $d_{L1}$ measure for variable $x_j$ is defined as

$$d_{L1} = \frac{\sum_{i \in M(j)} w_i |\hat{x}_{ij} - x_{ij}^{true}|}{\sum_{i \in M(j)} w_i}$$

where $\hat{x}_{ij}$ is the imputed value in record $i$ of the variable $x_j$ under consideration, $x_{ij}^{true}$ the corresponding true value, $M(j)$ the set of records for which the value on variable $x_j$ is missing, and $w_i$ is the survey weight of record $i$. This measure calculates the average distance between the imputed and true values.

The preservation of totals and means is measured by the $m_1$ measure, which is defined as

$$m_1 = \frac{\sum_{i \in M(j)} w_i (\hat{x}_{ij} - x_{ij}^{true})}{\sum_{i \in M(j)} w_i}$$

The $m_1$ measure calculates the preservation of the first moment of the empirical distribution of the true values.

The preservation of univariate distributions is measured by the $KS$ Kolmogorov-Smirnov distance. For weighted data, the empirical distribution of the true values is defined as

$$F_{x_j}(t) = \sum_{i \in M(j)} I(w_i x_{ij} \leq t) / |M(j)|$$

with $|M(j)|$ the number of records with missing values for the variable $x_j$ under consideration and $I$ the indicator function. Similarly, we define $F_{\hat{x}_j}(t)$. The $KS$ distance is defined as

$$KS = \max_k |F_{x_j}(t_k) - F_{\hat{x}_j}(t_k)|$$

where the $t_k$ values are the $2|M(j)|$ jointly ordered true and imputed values. The $KS$ compares the empirical distribution of the original values to the empirical distribution of the imputed values.
Smaller absolute values of the evaluation measures indicate better imputation performance.

We have also compared the correlations in the (partly) imputed data to the correlations in the true data in order to evaluate to what extent the relationships between different variables are preserved. In particular, we have calculated the average absolute difference between the correlations in the true complete data and in the imputed data, where we have taken the average over all 10 pairs of variables for data set 1 and all 15 pairs for data set 2. We have also calculated the average of the absolute percentage differences, where the percentage is calculated with respect to the correlations in the complete data over all pairs of variables in each of the data sets.

**Evaluation results**

The evaluation results for data set 1 are presented in Tables 5 to 7. As variable $R_1$ has no missing values it is not included in Tables 4 to 6. “Average” is the average of the absolute results over all 4 variables mentioned in these tables.

**Results for $d_{L1}$ for data set 1**

Table 5

<table>
<thead>
<tr>
<th></th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
<th>$R_5$</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>W 1</td>
<td>689</td>
<td>5803</td>
<td>72</td>
<td>20</td>
<td>1646</td>
</tr>
<tr>
<td>W 2</td>
<td>689</td>
<td>5793</td>
<td>72</td>
<td>20</td>
<td>1644</td>
</tr>
<tr>
<td>W 5</td>
<td>694</td>
<td>3967</td>
<td>315</td>
<td>19</td>
<td>1249</td>
</tr>
<tr>
<td>W 10</td>
<td>694</td>
<td>3.967</td>
<td>315</td>
<td>19</td>
<td>1249</td>
</tr>
<tr>
<td>NW 1</td>
<td>767</td>
<td>3843</td>
<td>315</td>
<td>19</td>
<td>1236</td>
</tr>
<tr>
<td>NW 2</td>
<td>702</td>
<td>3967</td>
<td>315</td>
<td>19</td>
<td>1250</td>
</tr>
<tr>
<td>NW 5</td>
<td>694</td>
<td>3967</td>
<td>315</td>
<td>19</td>
<td>1249</td>
</tr>
<tr>
<td>NW 10</td>
<td>694</td>
<td>3967</td>
<td>315</td>
<td>19</td>
<td>1249</td>
</tr>
<tr>
<td>Mixed</td>
<td>625</td>
<td>4532</td>
<td>185</td>
<td>20</td>
<td>1340</td>
</tr>
<tr>
<td>NN HD</td>
<td>690</td>
<td>5803</td>
<td>79</td>
<td>28</td>
<td>1650</td>
</tr>
<tr>
<td>Random HD</td>
<td>1140</td>
<td>8080</td>
<td>81</td>
<td>26</td>
<td>2331</td>
</tr>
</tbody>
</table>

**Results for $m_{4}$ for data set 1**

Table 6

<table>
<thead>
<tr>
<th></th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
<th>$R_5$</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>W 1</td>
<td>93</td>
<td>865</td>
<td>32</td>
<td>18</td>
<td>253</td>
</tr>
<tr>
<td>W 2</td>
<td>98</td>
<td>936</td>
<td>33</td>
<td>18</td>
<td>271</td>
</tr>
<tr>
<td>W 5</td>
<td>96</td>
<td>857</td>
<td>259</td>
<td>19</td>
<td>308</td>
</tr>
<tr>
<td>W 10</td>
<td>96</td>
<td>857</td>
<td>259</td>
<td>19</td>
<td>308</td>
</tr>
<tr>
<td>NW 1</td>
<td>38</td>
<td>1126</td>
<td>258</td>
<td>17</td>
<td>360</td>
</tr>
<tr>
<td>NW 2</td>
<td>89</td>
<td>857</td>
<td>259</td>
<td>19</td>
<td>306</td>
</tr>
<tr>
<td>NW 5</td>
<td>96</td>
<td>857</td>
<td>259</td>
<td>19</td>
<td>308</td>
</tr>
<tr>
<td>NW 10</td>
<td>96</td>
<td>857</td>
<td>259</td>
<td>19</td>
<td>308</td>
</tr>
<tr>
<td>Mixed</td>
<td>8</td>
<td>3</td>
<td>153</td>
<td>18</td>
<td>46</td>
</tr>
<tr>
<td>NN HD</td>
<td>93</td>
<td>865</td>
<td>26</td>
<td>10</td>
<td>249</td>
</tr>
<tr>
<td>Random HD</td>
<td>182</td>
<td>432</td>
<td>9</td>
<td>12</td>
<td>159</td>
</tr>
</tbody>
</table>
Results for KS for data set 1

<table>
<thead>
<tr>
<th></th>
<th>R_1</th>
<th>R_2</th>
<th>R_3</th>
<th>R_4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>W 1</td>
<td>0.10</td>
<td>0.20</td>
<td>0.19</td>
<td>0.92</td>
<td>0.35</td>
</tr>
<tr>
<td>W 2</td>
<td>0.10</td>
<td>0.20</td>
<td>0.19</td>
<td>0.98</td>
<td>0.36</td>
</tr>
<tr>
<td>W 5</td>
<td>0.18</td>
<td>0.10</td>
<td>0.46</td>
<td>0.97</td>
<td>0.43</td>
</tr>
<tr>
<td>W 10</td>
<td>0.18</td>
<td>0.10</td>
<td>0.46</td>
<td>0.97</td>
<td>0.43</td>
</tr>
<tr>
<td>NW 1</td>
<td>0.16</td>
<td>0.08</td>
<td>0.46</td>
<td>0.67</td>
<td>0.35</td>
</tr>
<tr>
<td>NW 2</td>
<td>0.18</td>
<td>0.10</td>
<td>0.46</td>
<td>0.97</td>
<td>0.43</td>
</tr>
<tr>
<td>NW 5</td>
<td>0.18</td>
<td>0.10</td>
<td>0.46</td>
<td>0.97</td>
<td>0.43</td>
</tr>
<tr>
<td>NW 10</td>
<td>0.18</td>
<td>0.10</td>
<td>0.46</td>
<td>0.97</td>
<td>0.43</td>
</tr>
<tr>
<td>Mixed</td>
<td>0.08</td>
<td>0.15</td>
<td>0.35</td>
<td>0.81</td>
<td>0.35</td>
</tr>
<tr>
<td>Random HD</td>
<td>0.32</td>
<td>0.51</td>
<td>0.47</td>
<td>0.92</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 7

The evaluation results for data set 2 are presented in Tables 8 to 10. “Average” is the average of the absolute results over all 6 variables mentioned in these tables.

Results for $d_{L1}$ for data set 2

<table>
<thead>
<tr>
<th></th>
<th>S_1</th>
<th>S_2</th>
<th>S_3</th>
<th>S_4</th>
<th>S_5</th>
<th>S_6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>W 1</td>
<td>28</td>
<td>13549</td>
<td>537</td>
<td>19470</td>
<td>39322</td>
<td>2132</td>
<td>12507</td>
</tr>
<tr>
<td>W 2</td>
<td>32</td>
<td>13584</td>
<td>477</td>
<td>19455</td>
<td>42757</td>
<td>2092</td>
<td>13056</td>
</tr>
<tr>
<td>W 5</td>
<td>32</td>
<td>13855</td>
<td>471</td>
<td>19791</td>
<td>44222</td>
<td>2092</td>
<td>13410</td>
</tr>
<tr>
<td>W 10</td>
<td>32</td>
<td>9975</td>
<td>504</td>
<td>14347</td>
<td>43081</td>
<td>2403</td>
<td>11724</td>
</tr>
<tr>
<td>NW 1</td>
<td>31</td>
<td>11329</td>
<td>400</td>
<td>16133</td>
<td>17668</td>
<td>2617</td>
<td>8030</td>
</tr>
<tr>
<td>NW 2</td>
<td>31</td>
<td>7579</td>
<td>442</td>
<td>10968</td>
<td>14943</td>
<td>2703</td>
<td>6111</td>
</tr>
<tr>
<td>NW 5</td>
<td>32</td>
<td>7814</td>
<td>450</td>
<td>11229</td>
<td>16718</td>
<td>2703</td>
<td>6491</td>
</tr>
<tr>
<td>NW 10</td>
<td>28</td>
<td>11718</td>
<td>430</td>
<td>16707</td>
<td>20294</td>
<td>2692</td>
<td>8645</td>
</tr>
<tr>
<td>Mixed</td>
<td>23</td>
<td>8988</td>
<td>342</td>
<td>12775</td>
<td>22272</td>
<td>1909</td>
<td>7718</td>
</tr>
<tr>
<td>NN HD</td>
<td>31</td>
<td>52319</td>
<td>553</td>
<td>62618</td>
<td>57207</td>
<td>2316</td>
<td>29174</td>
</tr>
<tr>
<td>Random HD</td>
<td>36</td>
<td>118355</td>
<td>548</td>
<td>110335</td>
<td>102786</td>
<td>3338</td>
<td>55899</td>
</tr>
</tbody>
</table>

Table 8

Results for $m_1$ for data set 2

<table>
<thead>
<tr>
<th></th>
<th>S_1</th>
<th>S_2</th>
<th>S_3</th>
<th>S_4</th>
<th>S_5</th>
<th>S_6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>W 1</td>
<td>4</td>
<td>8049</td>
<td>153</td>
<td>11677</td>
<td>9134</td>
<td>1135</td>
<td>5025</td>
</tr>
<tr>
<td>W 2</td>
<td>4</td>
<td>3107</td>
<td>150</td>
<td>4704</td>
<td>19932</td>
<td>831</td>
<td>4788</td>
</tr>
<tr>
<td>W 5</td>
<td>4</td>
<td>10000</td>
<td>100</td>
<td>14312</td>
<td>15037</td>
<td>850</td>
<td>6717</td>
</tr>
<tr>
<td>W 10</td>
<td>4</td>
<td>7468</td>
<td>202</td>
<td>10961</td>
<td>19045</td>
<td>1269</td>
<td>6491</td>
</tr>
<tr>
<td>NW 1</td>
<td>4</td>
<td>1790</td>
<td>24</td>
<td>2576</td>
<td>600</td>
<td>1526</td>
<td>1087</td>
</tr>
<tr>
<td>NW 2</td>
<td>1</td>
<td>1515</td>
<td>134</td>
<td>2425</td>
<td>75</td>
<td>1490</td>
<td>940</td>
</tr>
<tr>
<td>NW 5</td>
<td>3</td>
<td>4899</td>
<td>24</td>
<td>6959</td>
<td>4144</td>
<td>1652</td>
<td>2947</td>
</tr>
<tr>
<td>NW 10</td>
<td>2</td>
<td>122</td>
<td>110</td>
<td>407</td>
<td>2460</td>
<td>1652</td>
<td>792</td>
</tr>
<tr>
<td>Mixed</td>
<td>1</td>
<td>4619</td>
<td>112</td>
<td>6753</td>
<td>7593</td>
<td>1270</td>
<td>3391</td>
</tr>
<tr>
<td>NN HD</td>
<td>7</td>
<td>2381</td>
<td>58</td>
<td>17754</td>
<td>8751</td>
<td>951</td>
<td>4983</td>
</tr>
<tr>
<td>Random HD</td>
<td>8</td>
<td>156</td>
<td>32</td>
<td>15369</td>
<td>6044</td>
<td>495</td>
<td>3684</td>
</tr>
</tbody>
</table>

Table 9
Examine the results for methods W 1, W 2, W 5, W 10, NW 1, NW 2, NW 5 and NW 10, we see that they are rather erratic. For instance, W 1 performs best of these methods on the $m_1$ measure for data set 1, but rather bad for data set 2. For data set 2, W 1 is outperformed by the standard methods NN HD and Random HD. NW 1 performs worst on that measure for data set 1, but best for data set 2.

A general result is that our imputation methods perform better on the $d_{L_1}$ measure than NN HD and Random HD.

The results for Mixed are much more stable than for our individual imputation methods. Mixed is among the best performing approaches for all evaluation measures and both data sets. Mixed also performs better than NN HD and Random HD in all examined cases.

Table 11 gives the average (over all pairs of variables) of the absolute deviation of the correlations in the imputed data and the correlations in the true, complete data. Between brackets the average (over all pairs of variables) of the absolute percentage differences is given.

### Average absolute deviation from true correlations

<table>
<thead>
<tr>
<th></th>
<th>Data set 1</th>
<th>Data set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>W 1</td>
<td>0.0016 (0.50%)</td>
<td>0.0328 (13.00%)</td>
</tr>
<tr>
<td>W 2</td>
<td>0.0016 (0.46%)</td>
<td>0.0387 (18.48%)</td>
</tr>
<tr>
<td>W 5</td>
<td>0.0033 (0.98%)</td>
<td>0.0438 (20.23%)</td>
</tr>
<tr>
<td>W 10</td>
<td>0.0033 (0.98%)</td>
<td>0.0343 (14.90%)</td>
</tr>
<tr>
<td>NW 1</td>
<td>0.0034 (0.98%)</td>
<td>0.0381 (20.33%)</td>
</tr>
<tr>
<td>NW 2</td>
<td>0.0033 (0.98%)</td>
<td>0.0433 (24.14%)</td>
</tr>
<tr>
<td>NW 5</td>
<td>0.0033 (0.98%)</td>
<td>0.0417 (21.39%)</td>
</tr>
<tr>
<td>NW 10</td>
<td>0.0033 (0.98%)</td>
<td>0.0293 (14.86%)</td>
</tr>
<tr>
<td>Mixed</td>
<td>0.0021 (0.76%)</td>
<td>0.0360 (16.00%)</td>
</tr>
<tr>
<td>NN HD</td>
<td>0.0016 (0.49%)</td>
<td>0.0371 (10.72%)</td>
</tr>
<tr>
<td>Random HD</td>
<td>0.0035 (0.90%)</td>
<td>0.1457 (37.53%)</td>
</tr>
</tbody>
</table>

Mixed is again performing quite well, although it is slightly outperformed by NN HD for data set 1 on this measure.

By design our imputation methods do not violate any edits in any of the imputed records, including Mixed since an average of values satisfying linear edits
again satisfies these edits. The standard methods NN HD and Random HD do lead to violated edits and violated records, i.e. records in which one or more edits are violated. The number of edit violations for these methods are reported in Table 12 below.

<table>
<thead>
<tr>
<th></th>
<th>Data set 1</th>
<th></th>
<th>Data set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Failed edits</td>
<td>Failed records</td>
<td>Failed edits</td>
</tr>
<tr>
<td>NN HD</td>
<td>21</td>
<td>21</td>
<td>258</td>
</tr>
<tr>
<td>Random HD</td>
<td>55</td>
<td>55</td>
<td>250</td>
</tr>
</tbody>
</table>

Of the 258 failed edits for NN HD for data set 2, 189 times the balance edits was failed and 69 times an inequality edit. Of the 250 failed edits for Random HD for data set 2, 193 times the balance edits was failed and 57 times an inequality edit.

If edit violations are to be prevented, the use of NN HD or Random HD should therefore be avoided.

**CONCLUSIONS**

Considering the results on the evaluation measures, Mixed would be our recommended method of all methods examined in the current paper. The Mixed method does not lead to any violated edits, it leads to good results for all evaluation measures examined in this paper, and it performs better than NN HD and Random HD in almost all cases. Apparently, the use of edit restrictions pushes the imputations in the right direction for this method, in any case for the data sets considered in our evaluation study.

The imputations produced by Mixed are averages over 8 other imputation approaches. These 8 imputation methods show some rather erratic evaluation results. A method that performs quite well on a certain measure for one data set can perform rather badly on the same measure for the other data set. Apparently, taking the average reduces the erratic behavior of the results, and brings out the best of the 8 imputation methods.

Another advantage of Mixed is that it is a relatively simple method to develop and implement. The most complicated part is using the simplex algorithm, but for that several software packages, e.g. in R, are nowadays freely available. Mixed is considerably simpler than earlier imputation methods that preserve edits that were mentioned in the Introduction.

In the current paper we have not made an attempt to optimize the Mixed approach, i.e. we simply took the average over all other adjusted nearest-neighbour imputation methods we proposed and evaluated. An interesting topic for future research could be finding an optimal mix for the Mixed approach: should we combine only unweighted adjusted nearest-neighbour imputation methods, only weighted adjusted nearest-neighbour imputation methods, or a combination of both? In the latter two cases: what should these weights be? In this paper we took the reciprocal...
of the average of the observed values for a variable as the weight for this variable, but other weights might possibly lead to better results. Finally, also the number of adjusted nearest-neighbour imputation methods should be decided upon as well as the optimal number of nearest neighbours for each of these methods.

A possible extension of Mixed would be to include the preservation of known or previously estimated totals in the imputation process. We leave such an extension to possible future research.

References
Simulation Process, Theoretical Paradigm And Operational-Strategic Realy As Tool For Managerial Decision Making

Monica Viorica NEDELCU, Professor, Ph.D., Bucharest University of Economic Studies, Bucharest, ROMANIA
Cosmin Octavian DOBRIN, Professor, Ph.D., Academy of Economic Studies, Bucharest, ROMANIA
Marian Sorin IONESCU, Ph.D., Associate Professor, Academy of Economic Studies, Bucharest, and, University “POLITEHNICA” Bucharest

ABSTRACT

The present study is concerned with the concept of simulation and the development of simulation models, transposed into mathematical formalism from engineering sciences into economic context, as a powerful and effective tool for managerial decision making. The applications under consideration involve deterministic systems with continuous time and states, as well as with discrete ones.

There presented simulations are of type G/G/1 activity, considered to be representative for models of modern business-type of problems, as well as paradigms, concepts and the mathematical formalism from engineering sciences, which have been successfully applied to economic organization-type problems, such as the Pol-laczek – Khinchin formula, Lindley’s equation, the Wiener-Hopf equation.

Finally, two classical, representative methods for simulations are briefly and synthetically discussed, the Monte-Carlo method and the Metropolis method, together with the methodology of implementation via the specialized software Microsoft Excel, and with the convergence of the simulation processes.

Key Words: Simulation, modeling, business model, mathematical formalism.
JEL Classification: C63, L25

INTRODUCTION

The complexity of the managerial decision, aimed at consolidating the market position and subsequently the development of a modern economic organization, requires increasing the profit rates, which demands establishing appropriate business models, innovative, cross-disciplinary concepts, adapted from other areas of fundamental and applied research.

We propose here a methodology based on the concept of simulations, which has been extensively used in fundamental research in the engineering sciences.
The explosive growth of digital resources provides this paradigm new, ample dimensions, as the modern simulations models are nothing else but virtual information structures (Bratley et al., 1983). Simulations, together with the development of models, play a great role in the process of design, implementation, and exploitation of complex technological systems. In the case of an economic organization system, these are described through the multitude of its components, the strategic-operational strategic relations among them, and the multiple variables that define the state of the system at any point in time, the simulation process is used to reproduces a given process, via some model elaborated based on the original process, at some point in time when the original process is not present.

The situations when simulations are recommended can be identified for the following types of economic processes: activities with significant financial allocations and with large numbers of operational alternatives for distribution chains, optimal decisions, high risk economic-industrial processes – air transportation, industrial-economic processes involving dangerous substances, chemical industry, slow agricultural processes, and rapid industrial, chemical, or physical processes. It was stated that “the simulation of a given economic process (operational-strategically identifiable) is designed based on a second economic process, distinct from the one under consideration, which however follows the same rules encapsulated in analogous equations that are applicable for both models.” (Heche et al, 2003)

Rather than deterministic systems (for which the current states are uniquely and continuously determined by the previous states), via differential equations, it is discrete, stochastic systems that contributed to the development of the simulation paradigm.

Summarizing, simulation amounts to reproducing, with some degree of accuracy, an economic phenomenon over some well defined time horizon.

A system is said to be of “continuous time” if the time variable takes real numerical values within some interval (not necessarily bounded or open).

A system is said to be of “discrete time” if the set of all possible values of the time variable forms an ordered, countable sequence, possibly finite.

The set of time values can be thought of a spatial structure on the real axis, obtained by sampling the continuous time axis at regular time intervals.

Another possible definition, which can be useful in processes based on iterative algorithms, is that of a discrete, zero measure set, that is a subset of some larger set (the real axis), similarly to the classification of the time variable, the set of state of the system can be classified as continuous or discrete.

The system itself can be considered as deterministic (when the evolution in the future is precisely determined from its present state by some evolution law), or stochastic or random (when the evolution on the future cannot be precisely derived from its present state but only probabilistically).
THE CONCEPT OF SIMULATION AND MODEL DEVELOPMENT

The quality of the chosen model

The validity of the simulation of the economic process depends on the accuracy of the chosen model, a high quality model yields simulations that provide precise, clear, quantifiable results that are close to reality.

The procedure (Heche et al, 2003) of identifying the most appropriate model can be schematically presented as follows:

\[
\begin{align*}
\text{Primary} & \quad \rightarrow \quad \text{Original} \quad \rightarrow \quad \text{Experiment} \quad \rightarrow \quad \text{Primary Quantifiable rule} \quad \rightarrow \quad \text{process} \quad \rightarrow \quad \text{monitoring} \quad \rightarrow \quad \text{results} \\
\text{Comparative Procedures} & \quad \rightarrow \quad \text{Secondary} \quad \rightarrow \quad \text{Chosen} \quad \rightarrow \quad \text{Process} \quad \rightarrow \quad \text{Secondary Quantifiable rule} \quad \rightarrow \quad \text{model} \quad \rightarrow \quad \text{simulation} \quad \rightarrow \quad \text{results}
\end{align*}
\]

In the process of choosing the most adequate model for simulations, we identify the following concepts, and situations:

- both the rules that make the initial (original) process operational and the rules that drive the model are known, this is possibly an ideal situation, when the rules can be compared, by interchanging the variables or other entities, possibly leading to high complexity optimization problems.
- only one set of rules is known, in which case one has to make assumptions on the other set of rules and on the validity of the experimental results obtained by observing the economic phenomena, leading to non-linear optimization problems.
- both sets of rules are unknown, however it is possible to quantify and compare the outcomes of the original process with the results of the simulations of the chosen model.

In this case, an essential role is played by statistical tests, which are used to decide whether the chosen model needs to be rejected in the light of the data. Hypothesis testing are a powerful instrument in such approaches.

It is possible to obtain results from simulations of the chosen model, however data on the original process is not available. In this case accepting the results of the simulations, without a comparison with the original process, assumes major risks in the subsequent decision making.

The simulation of processes with continuous times and states, these processes are subordinated to a mathematical formalism (Heche et al, 2003) based on differential equations, for which we adopt the following notation convention:
The law that describes the dynamics of the economic system is given by the formula:

\[
\begin{align*}
x' &= Ax + Bu \\
y &= Cx
\end{align*}
\]

(1)

If a feed-back loop is present, with the matrix \( R \) representing the transformation of a state of the system into a command, with \( u = Rx \), we have the following differential equations:

\[
\begin{align*}
x' &= (A + BR)x \\
y &= Cx
\end{align*}
\]

(2)

The system is described as follows:

\[
-u \rightarrow (B) \rightarrow (A) \rightarrow x \rightarrow (C) \rightarrow y \rightarrow
\]

Figure 1

The symbol \( A \) in the above diagram corresponds to the linear system \( x' = Ax + w \), obtained under the assumption that \( w \) represents an input of the system.

As scientifically validated and operationally applicable software solutions, (Cormen et al, 2009) it is recommended to use the computer algebra systems **Maple, Matlab, Mathematica**, which offer general modules for system integration, concept adapted from the engineering sciences, these software packages present exceptional facilities, due to their friendly graphic interface, and the ability to take in an and executing algebraic operations with an arbitrary number of vectors, each vector consisting of \( n \) components that are time dependent.

Starting with an initial condition represented by a vector \( x(0) \) of \( n \) components, the summation or integration operator computes the sum or integral, from \( 0 \) to \( t \), of the previous vectors. We identify a module that takes in as input vectors of the form \( x(t) = (x_1(t), \ldots, x_n(t)) \), all whose components are functions of time, and yields as outputs functions of time obtained by multiplying the input vector \( x \) by the matrix \( A: n \times n \).

The general scheme is to translate concepts from advanced technology (from the world of electronics) to economic processes (managerial processes, capital price dynamics), this simulations are based on linear differential systems with constant coefficients.
The time dependent vectors represent the interaction between the system and the environment, and are generated via some simple circuits, which makes this type of simulation models adequate for economic problems from a wide spectrum of business structures.

To analyze the process following this scheme, if we put \( u \equiv 0 \), and suppressing the feedback loop, that is, for \( R \equiv 0 \) we obtain the homogeneous system \( x' = Ax \).

Independently of initial condition \( x(0) \) the system is stable if all roots of the characteristic polynomial \( \det[A - \lambda I] \) have negative real parts; these roots are the eigenvalues of the system.

Considering the feedback loop described by \( R \), one obtains stability for the system from figure graphics 1, independently on the initial conditions, in the case when all roots of the characteristic polynomial \( \det[A + BR - \lambda I] \) have negative real part.

We point out that the simulation of non-linear system of the type \( y' = \phi(y) + w \), present a higher degree of complexity than in the linear case, as the qualitative analysis of the solution is much more intricate.

**THE SIMULATION OF DISCRETE SYSTEMS AND EVENTS: METHODOLOGY AND MODELS**

In simulation of economic processes of stochastic type, the most common paradigms that of modeling via discrete events, via markovian processes (Cormen et al., 2009) in continuous time, where the state space and the transition law are explicit, it is important to point out that the simulation of discrete events has generated the creation of some specialized computer programming languages, of expert type, which are adapted to the requirements of the scientific models as well as to the business-operational type paradigms, such as **SIMULA, GPSS, SIMSCRIPT**, in subsequent development, object oriented programming languages concepts, and discrete events simulation models have been developed, which influenced in a significant way the development of modern client systems.

In what follows, we present a simulation model based on a sequence of type G/G/1, specific to economic phenomena in the areas of logistics and distribution of goods.

**ARRIVALS** | **WAITING** | **SERVER** | **DEPARTURES**
--- | --- | --- | ---
Arrival processes | Service disciplines | Service processes | Departure processes

We introduce the following notation that is utilized in probability laws:

- "M", is associated with exponential law, "markovian", "memoryless",
- the arrival process is of type **POISSON**,
- the time intervals between two successive arrivals of clients are random variables.

- "D", is associated with a "Degenerate law",
- the arrival of clients occurs at regular time intervals.

- "E_k", process associate to a process when the time intervals between two successive arrivals is a random process following **ERLANG** law of order "k".
In waiting-time theory (Filipowicz and Kwiecien, 2008), which is part of probability theory, we denote by $G/G/1$ a sequence (queue) of waiting times, in a single-server system, when the arrival and the services are distributed within some arbitrary interval, as described below:

**G** – general type of distribution of arrival times, no hypothesis

$G$ – general distribution of waiting times, $S^{-1} = 1/\mu$

1 – single service, load $\rho = \lambda S^{-1}$, for a stable sequence, $\rho < 1$

The model the waiting sequence, $G/G/1$ is a complex system for which utilizes general types of approximations, one of the simplest being described by the **Pollaczek – Khinchin** formula

$$W_{G/G/1}^{-} = \left(\frac{C_1 + C_2}{2}\right)W_{M/M/1}^{-}$$  \hspace{1cm} (3)

In waiting-time theory, the previous formula describes the relation between the length of the queue and the distribution of service times via the [LAPLACE transform](#), and it also relates the average length of the waiting queue and the average service time, in this expression, $C_A^2$ is the square of the variation of the random variable that describes the time intervals between two successive arrivals of clients, distribution following a probabilistic law under the assumption that the events are identically distributed and independent, and $C_S^2$ is the random variable that models the service times which are also assumed to be identically distributed and independent, the value $W_{M/M/1}^{-}$ corresponds to the average waiting time in a queue $M/M/1$, with the same rate of arrival and service as in the system under consideration.

**Lindley Equation**

We propose the simulation of the behavior of such a queue with a model of the type FIFO, which studies the waiting times via Lindey equation, we consider a sequence of arrivals of clients indexed by $n$, using the formalism from below:

| $C_n$ | the $n$ - th client arriving to the system |
| $a_n$ | the time of the arrival of $C_n$ |
| $d_n$ | the time of departure of $C_n$ |
| $t_n$ | time interval between $C_{n-1}$ and $C_n$ |
| $s_n$ | service time of $C_n$ |
| $w_n$ | waiting times for $C_n$ |

The process of interest is defined by the succession \( \{ w_n \mid n = 0, 1, \ldots \} \) of waiting times, before establishing the fact that the process is markovian, the time of departure of $C_{n+1}$ is
The first restriction corresponds describes the situation that \( C_{n+1} \) needs to wait for the departure of \( C_n \) in order to access the server, the second restriction describes the situation when \( C_{n+1} \) arrives to an empty system, by definition, the waiting time of \( C_{n+1} \) is given by

\[
w_{n+1} = d_{n+1} - a_{n+1} - s_{n+1}
\]

subtracting \((a_{n+1} + s_{n+1})\) from (4), it follows

\[
w_{n+1} = \begin{cases} 
  d_n - a_{n+1}, & \text{if } d_n - a_{n+1} \geq 0 \\
  0, & \text{if } d_n - a_{n+1} \leq 0
\end{cases}
\]

Since \( w_n = d_n - a_n - s_n \), it follows:

\[
w_{n+1} = \begin{cases} 
  w_n + s_n - t_{n+1}, & \text{if } w_n + s_n - t_{n+1} \geq 0 \\
  0, & \text{if } w_n + s_n - t_{n+1} \leq 0
\end{cases}
\]

Introducing the variable \( u_n = s_n - t_{n+1} \), we obtain the fundamental relation:

\[
w_{n+1} = \begin{cases} 
  w_n + u_n, & \text{if } w_n + u_n \geq 0 \\
  0, & \text{if } w_n + u_n \leq 0
\end{cases}
\]

The sequences \( \{t_n, n \geq 0\}, \{s_n, n \geq 0\} \) are mutually independent, and the form independent random variable, the sequence \( \{u_n, n \geq 0\} \) is also form by mutually independent random variables, and, due to (5), the sequence \( \{w_n, n \geq 0\} \), defines a markovian process, analyzing the distribution function \( U_n(z) \) and \( u_n \) in terms of the time intervals between successive arrivals of the clients, denoted by \( A(t) \), and service times denoted by \( B(s) \), for \( z \in R \),

\[
U_n(z) = P[u_n = s_n - t_{n+1} \leq z],
\]

\[
u_n(z) = P[u_n = s_n - t_{n+1} \leq z] = \int_{0}^{\infty} \int_{0}^{w_n} g(z + t) \, dt \, d\mathbf{P}(t)
\]

The previous expression shows that the distribution function \( U_n(z) \) is independent of \( n \), hence it can be written as \( U(z) \), thus, the waiting queue is stable and the single variable \( u_n \) has negative expectation if

\[
E[u_n] = E[s_n - t_{n+1}] = E[s_n] - E[t_{n+1}] = E[S] - E[A] = e[A](\rho - 1)
\]

where \( E[S], E[A] \) represent the expectation of the service time and of the time intervals between successive arrivals of clients, the expectation \( u_n \) is negative if and only if \( \rho = \frac{E[S]}{E[A]} < 1 \), denoting \( W_n(y) \) the distribution function of the waiting times \( w_n \) for \( C_n \) we obtain for \( y \geq 0 \) according to (6),

\[
\text{Romanian Statistical Review nr. 3 / 2017}
\]

67
The waiting time is assumed to be stable, and the processes are of ergodic type, a process is said to be “ergodic” if its statistical properties can be derived from a randomly chosen sample, and any collection of random samples extracted from the process exhibit the typical statistical properties of the process in its entirety, non ergodic processes have a chaotic rate of change, the process under consideration admits an invariant distribution $W(y)$, which verifies the functional equation:

$$W_{n+1}(y) = P[w_{n+1} = y | w_n = w]dW_n(w) = \int_0^\infty U(y-w)dW_n(w)$$

(7)

The waiting time is assumed to be stable, and the processes $\{w_n, n \geq 0\}$ are of ergodic type, a process is said to be “ergodic” if its statistical properties can be derived from a randomly chosen sample, and any collection of random samples extracted from the process exhibit the typical statistical properties of the process in its entirety, non ergodic processes have a chaotic rate of change, the process under consideration admits an invariant distribution $W(y)$, which verifies the functional equation:

$$W(y) = \int_0^\infty U(y-w)dW(w), y \geq 0$$

(8)

The waiting times have non-negative dimension, hence $W(y) = 0$, for $y < 0$, for a queue of the type $G/G/1$ that is stable, (Kleinrock, 1975) the stationary distribution of the waiting times satisfies Lindley’s integral equation:

$$W(y) = \left\{ \begin{array}{ll}
\int_0^\infty U(y-w)dW(w), & y \geq 0 \\
0, & y < 0
\end{array} \right.$$
events that take place within a certain time horizon in the future, or modifies the evolution of posterior events that were already scheduled to take place in the future, these events only occur at the discrete times, as suggested by the name of the model. On the modeling aspect, we remark that for a process structured on specific activities, for example the successive operations on an assembly line, decisional tree specific to production management, we are not concerned with the specific activities that mark the beginning and the final of the process, but only the resulting final effect.

As long as the activity takes place, there is no visible effect on the system besides that locking of the resources, so the state of the system does not change for this type of activity throughout its evolution, this is in contrast with models based on continuous time, described previously, for which the state of the system is continuously updated, in applications to economic operations, for case studies with specific requirements, sometimes one uses “mixed” simulations, which comprise both a continuous part and a discrete part.

Simulation elements, discrete events
We identify the following main elements in the process of discrete event simulation, relevant for our discussion:

- descriptor of status elements of the systems that underlie the demand for the determination of simulation running;
- indicators (counters, quantifiers) of positions (places) where the results obtained are stored;
- a chronology for future events, together with upgraded algorithms and procedures, allowing the management of an event insertion in the correct position, searching the next event and its exclusion from this chronology;
- various types of events, each with its own description, those inducing actions of events on the system status, conditional triggers, the chronology position of the future events, upgrading and updating the various recordkeeping and statistics indicators.

OPERATIONALIZATION OF SIMULATION: ALGORITHMS AND IMPLEMENTATIONS

The operationalization of simulation activity (Bratley et al., 1983) presupposes the elaboration of three fundamental phases, (Heche et al., 2003), (Cormen et al, 2009) as follows:

**START**: At the beginning, the system is positioned in its initial status, a first event, **START** initiating the process onset time, from the point \( t = 0 \), the record of subsequent events in the time spreadsheet and forecasts especially the event **STOP** marking the end of simulation;

**PROCESS**: Successive events are processed in their chronological order of occurrence within the time agenda: the movement takes place from the initial moment of the first temporary event, this event being then removed from the time schedule (spreadsheet), after it is executed (operationalized), the requested status
change takes place, we record the subsequent event in the appropriate position of the
time spreadsheet, we upgrade the corresponding indicators and statistics, the time the
next event produces being mentioned;

**STOP**: Reaching the time when the significant event takes place **STOP**
marks the end of simulation, this can be triggered by other events, or it is a temporary
scheduled event from the beginning of the simulation, to indicate that the time granted
to this process type expired, the algorithms and procedures cease their operationality
and the results are presented.

For the transposition in operationality for the simulation of managerial
decision, high-level simulation software solutions are accessed, but which request a
high number of parameters to induce the particular simulation model.

Software facilities specialised and dedicated to simulations offer various
graphical interfaces leading to a better understanding of phenomena, graphical
animation representing a communication platform between the actors involved in
the modelled and simulated systems, advanced-level simulation digital solutions are
suitable for the rapid elaboration of simplified prototypes, the systems that must be
simulated in managerial operationality have a higher and higher complexity degree,
with the evolution and development of business models and the quantity of information
that must be identified, selected and submitted to the simulation process.

Compared to much more rigorous and relatively more stable engineering
processes, economic processes have a higher dynamism and a much more complex
uncertainty degree, due to the involvement of human decision-maker, as an uncertainty
and perturbation factor.

Generic simulation of discrete events is advantaged by the increasing
capacity of the computing power of the new computer generations, but also by the
data stocks under the form of databases and even data deposits.

**Application, simulation of a waiting line, \( G/G/1 \)**

We have briefly presented (Kleinrock, 1976) the waiting line of the
type \( G/G/1 \): the total customers arriving into a system of this type are taken
over under the responsibility of the one serving, if it is free, it is composed of an
infinite capacity queue, where each person waits that the precedent person be
served in his/her arriving order, before its turn, this representing the operation rule
**FIFO** (an acronym for "First In First Out", **method for row management, of the waiting queues**), (Filipowicz and Kwecien, 2008).

The time intervals between two successive customer arrivals form a row
\( \{\alpha_i\} \) of independent random variables, submitted in the same distribution functions
\( F_\alpha(t) \) with \( E[\alpha] = \frac{1}{\lambda} \).

Within an analogous approach, the serving times for successive customers
form a row \( \{\beta_i\} \) of independent random variables, all of them distributed according to
a distribution function \( F_\beta(t) \) with \( E[\beta] = \frac{1}{\mu} \), the row of serving times is composed
of independent entities of the customer arriving processes.
The system is stable and reaches its stationary regime if and only if \( \rho = \frac{\lambda}{\mu} < 1 \), as we assume in the continuation of the developed reasoning.

The final target of simulation is the estimation of an average number \( N^- \) of present customers, so that the average standing time \( T^- \) of a customer within the system when it evolves in stationary regime.

In specialty literature, it has an increased difficulty degree for lines with the structure \( G/G/1 \) and it is much easier for particular cases such as waiting lines \( M/M/1 \). This particularities offer good opportunities of testing the simulation software, comparing the generated results with the values resulted by mathematical calculations, we define the variable \( n \) indicating at each moment \( t \) the value of the function \( n_t \), the number of customers that are present in the system, and the one in progress of being served.

We note that if \( n \) defines the status of a Markov process in the case of a waiting line of the type \( M/M/1 \), this is not also valid for the line \( G/G/1 \) being the argument for which this more general waiting line has a higher complexity degree to be analytically treated as Markov type and, therefore (Kleinrock, 1976).

It is preferably to proceed to simulations to obtain quantitative indications on the behaviour of this type of systems, the value \( \tau \) represents the total number of customers \( X \) the time units being served by the system in the timeframe \([0, t]\).

We calculate the average number of customers present in the system at some moment in the interval \([0, t]\): \( \hat{N}^- = \frac{\tau}{t} \).

The simulation process (Heche et al, 2003) imposes the definition of four structural events, \textit{START}, \textit{STOP}.

\textit{ARRIVAL, DEPARTURE} being thus four algorithm structures changing the system status, triggering in their turn events and managing statistical scheduling.

\textbf{THE ALGORITHM OF SIMULATED EVENTS}

\textit{START} := Simulation initialization and launching

The number of customers existing in the system, \( n := 0 \).

The cumulation of system existence times, \( n_{\text{cum}} = 0 \).

Current times \( t := 0 \).

Total duration of the simulation process \( D_{\text{tot}} \).

Forecasting the occurrence of the STOP event, at the moment \( t_{\text{fin}} := D_{\text{tot}} \).

The operationalization of the law \( F_a(t) \), the arrival of the first customer, achievement of \( a \).

Forecasting a first event of ARRIVAL type, at the moment \( t_{\text{arr}} := a \), within the scheduling.
ARRIVAL: the process of arrival of a new customer takes place at the moment $t_{arr}$.

The event is removed from the time spreadsheet.

Be $\Delta := t_{arr} - t$, the update of the clock for the times for the occurrence of the current ARRIVAL event, after which the time $\Delta$ is calculated between this one and the previous one.

We generate a random variable $\alpha$, which is submitted to the distribution function $F_{\alpha}$, up to a new ARRIVAL.

We schedule a new ARRIVAL event for the time moment $t_{arr} = t + \alpha$ within the temporary forecast.

If $n = 0$, the line is free, then

We generate a random variable $\beta$, for the distribution function $F_{\beta}(\cdot)$, representing the serving time for the new customer.

We forecast the DEPARTURE event at the time moment $t_{dep} = t + \beta$.

Be $n_{\text{cum}} := n_{\text{cum}} + n \times \Delta$, the statistics update.

Be $n := n + 1$, the system status update.

There follows the search of the next event within the temporary forecast.

DEPARTURE: A customer leaves the system at a moment $t_{dep}$.

The event is removed from the calendar schedule.

Be $\Delta := t_{dep} - t$, the clock update at the time $t_{dep}$, for the current DEPARTURE event, after the calculation of the time $\Delta$, between an eveniment and the previous one.

If $n > 1$, at least one customer is waiting within the line, then

We generate a random variable $\beta$, according to the law $F_{\beta}(\cdot)$.

We forecast (schedule) the DEPARTURE event, at the moment $t + \beta$.

Be $n_{\text{cum}} := n_{\text{cum}} + n \times \Delta$.

Be $n := n - 1$, the update of the number of customers in the system.

There follows the search of the next event in the forecast.
STOP: we present the results and the simulation end process at the moment $t_{fin}$.

Be $N^* = \frac{n_{\text{sum}}}{t_{fin}}$, the average number of customers present in the system.

Be $T^* = \frac{1}{k}$, the average waiting times for the customers in the system, that can be calculated Little.

STOP

This presentation of the simulation algorithm, (Heche et al, 2003) offers the possibility that two or more simultaneous events be distanced, for the simulation of the previous algorithm, the occurrence of simultaneous events is null, being thus required to elaborate a way of tie breaking of simultaneous events and the coherence of the process must be preserved.

Table, scheme of discrete event simulation running, row G/G/1, we highlight four events, START, ARRIVAL, DEPARTURE, STOP

<table>
<thead>
<tr>
<th>FORECAST – CALENDAR</th>
<th>EVENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Times</td>
<td>Events</td>
</tr>
<tr>
<td>0</td>
<td>START →</td>
</tr>
<tr>
<td>$t_1$</td>
<td>ARRIVALS &lt; −</td>
</tr>
<tr>
<td>...</td>
<td>→</td>
</tr>
<tr>
<td>$t_m$</td>
<td>ARRIVAL &lt; −</td>
</tr>
<tr>
<td>$t_{m+1}$</td>
<td>DEPARTURE &lt; −</td>
</tr>
<tr>
<td>...</td>
<td>→</td>
</tr>
<tr>
<td>$t_q$</td>
<td>ARRIVAL</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>$t_s$</td>
<td>DEPARTURE</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>$t_{STOP}$</td>
<td>STOP</td>
</tr>
</tbody>
</table>

DISCUSSION OF MOST REMARKABLE METHODS USED IN THE SIMULATION PROCESSES

MONTE CARLO method is at the origin of simulation of stochastic processes (Fishman, 1996), the hazard within these processes being minimized, and the method is recommended for the resolution of various problems, involving a low computing effort compared to the difficulty of the problem.

The simulation of managerial decision-making problems (Bratley et al., 1983) involving economic organizations is applied to all classes of challenges that include operational rules, algorithms and procedures, involving the adaptation of decisions, their control, price policies, the resolution of problems that are specific to economic organizations, by simulation technologies, impose the use of some...
interactive algorithms and the existence of some rigorous phases, aiming at reaching a final presupposed objective, using entry data randomly generated.

The **Monte Carlo method** is successfully applicable in the multiple integral calculation (Rubinstein, 1981), within the resolution of differential equations with partial derivatives and combination optimization, modelling of an economic system presupposes the modelling of time variables for the treatment of an item within the operationalized processes and the two arrivals by probability distributions.

Within the theory of waiting lines and those of modelling/simulation of processes that are specific to economic organizations, the arrivals admit a dual-type perception, either by random numbers of items throughout a reference period (defined time units), or by random time intervals separating two successive arrivals within the system.

Two distributions are usually used in the random arrival processes of the customers in an activity specific to business models:

- **Poisson distribution** corresponds to the distribution of a number of items that occur within a space-time frame given (for example, spatial framework - the access to a retail point, organizational sale, temporary framework - time frame mentioned, event - the arrival of a customer). The **Poisson probability distribution** of parameter $\lambda$ is $P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$; this distribution corresponds, dually, to a time frame $T$ separating two successive events following the exponential law of the parameter $\lambda = \frac{1}{T}$, under the formalism $P(T > t)$.

$$\int_{t}^{\infty} \lambda \cdot e^{-\lambda t} = e^{-\lambda t}$$

This distribution formulation being preferred to the **Poisson** of the modelling, simulation of economic processes, the probability that an event occur in the time frame $t$, $t + \epsilon, \epsilon \to 0$ is the same, regardless the time run $\omega$ since the previous event, this conditioned probability corresponds to what we call **hazard function**, specialised software EXCEL offers the possibility of quantification with the spreadsheet **Arrival Poisson Introduction Simulation.xls**, the **Poisson** -type process being a process with a **without memory**.

- The second distribution used for modelling the random entries of items in an economic system is the **Erlang distribution**, this being characterized by an Erlang distribution is characterized by a parameter of the form $\alpha$, positive whole number and a parameter of the scale $\lambda$, as follows

$$P(T > t) = \frac{1}{\lambda} \cdot e^{-\lambda t} \cdot \frac{\lambda^t}{t!}$$

with $\alpha = 1$, we find the **exponential distribution** for superior values, the probability density function being unimodal and approximated with a **normal law**, for values of $\alpha$ exceeding the positive whole number 5.6, $d^\alpha$Erlang law is defined as exponential, the probability of occurrence of an event in the time frame $t, t + \epsilon, \epsilon \to 0$ depends on the time since the occurrence of the previous event and on the value of the parameters $\alpha$ and $\lambda$, the approach within the specialised software EXCEL offers the possibility of quantification with the spreadsheet **Erlang distribution**.
Another method specific to the simulation processes is \textbf{METROPOLIS method}, applied on a large scale of combination optimization processes, within this method being generated a row \( \{X_1, X_2, \ldots\} \) of random variables that can be interpreted as a trajectory of \textbf{Markov chain}, with \( n \) statuses with a stationary distribution \( \pi = (\pi_1 \ldots \pi_n) \), with \( \pi_i \geq 0 \), whatever \( i \) and \( \sum_i \pi_i = 1 \).

\textbf{METROPOLIS method} (Heche et al., 2003) is specific to the applications of fundamental sciences (physics), engineering sciences, physical systems in thermodynamic balance, to the acquisition of random points in high complexity degree areas, characteristic which imposed its extrapolation to economic organizations, this last application representing the continuous case of the method, and, integrated with \textbf{MONTE CARLO method} (Fishman, 1996) offers significant results where other simulations fail, in statistics, the simulation method presented is known under the acronym \textit{Markov Chain Monte Carlo}, having a high applicability degree.

\section*{THE ANALYSIS AND DISCUSSION OF SIMULATED PROCESS CONVERGENCE}

The purpose of simulating the stochastic behaviour of economic system is to assess its real behaviour and performances, (Aarts and Korst, 1989), (Bratley et al, 1983) generated by an added value and profit margins, managerial performance outputs, a number of unknown parameters, susceptible to provide the searched information, must be estimated, eventually internally, within an optimization loop, and the time allotted to the performance of simulation is also an essential element for this type of approach.

One must determine „the number of replications” of simulation experiences, sufficient for the identification of simulated trajectory, after a possible elimination of the initial transitory phase, which could influence the results, and the precision degree must be high enough.

It is possible to estimate the precision of a simulation process starting from the results generated by it, namely sampling, planned experiments, probabilistic estimate of the confidence intervals, described in statistical surveys, without the certainty of a clear answer, only for certain punctual, particular situations.

The simulation results are formally represented by a row (vector) of values \( X_1, X_2, \ldots, X_i, \ldots, X_{n-1}, X_n \) \((9)\) they can be assimilated to the reproduction of a certain trend or specific trajectory of a stochastic process, from which we want the estimation of some parameters, (Heche et al., 2003).

The parameters can be of hope time \( E[\psi(X_i)] \), whatever \( i \), or \( \psi(\cdot) \) is a given function, a correct choice of this function allows the estimation of the dimensions of hope functions \( X, \psi(X) = X \); hope function of \( X^2, \psi(X) = X^2 \), which in its turn
allows the estimation of the variation of $X$, as being $\text{Var}(X) = E[X^2] - E[X]^2$, we insert the distribution function $P[X \leq x]$ of $X$, we also use the indicator function $\psi(x) = 1_{\{x \leq x\}}$

If the differences $X_i$ have all the same distribution, then it is obvious that $E[\psi(X_i)] = \gamma$ for any $i$, the type of basic stochastic process is the decision-making factor for the chosen simulation method, we extremely identify two cases:

- the row (9) represents especially the achievement of independent random variables, identically distributed;
- the row (9) represents a fraction of the trajectory of non-stationary stochastic processes;

The first one represents the classical analysis of the sampling, one can identify $n$ independent replications for the same experiment performed more times in identical conditions, the size of the sampling must be great enough to allow the application of "great number law" and "central limit theorem" allowing the estimation of a confidence interval for the parameter $\mu = E[X_i]$ the second case is identifiable as being non-stationary, the processes evolving in time, generating a behaviour trend of the average value $E[X_i] = \mu_i$.

The processes can have an evolution with a certain variability degree, variant or covariant between the variables shifted by a same time interval, and there is the possibility that the two effects overlap, when there is no information on the process non-stationary nature, the estimation of the parameters $\mu_i$ is possibly intuitive with a number $K$ of independent replications of all experiments, obtaining the result rows (vectors), presented in a matrix structure, picture, as follows

\[
\begin{pmatrix}
X_{i1}^{(1)}, \ldots, X_{i1}^{(K)} \\
\ldots \ldots \ldots \ldots \ldots \ldots \ldots \\
X_{i1}^{(1)}, \ldots, X_{i1}^{(K)} \\
\ldots \ldots \ldots \ldots \ldots \ldots \ldots \\
X_{iK}^{(1)}, \ldots, X_{iK}^{(K)}
\end{pmatrix}
\]  

(10)

For each value of the index $i$, the “transversal” row of values within the previously defined matrix $X_{i1}^{(1)}, \ldots, X_{i1}^{(K)}$ for the estimation of the unknown parameter, $E[X_i] = \mu_i$ the situation is limited to the previous case of independent observation sampling, this simple modelling being achieved with high costs and it does not use complementary information being available in the basic structures.

The analysis of the series occurred in chronological order and the elaboration of correct estimates within the non-stationary processes are major importance
challenges for economic organizations and applications of superior added value generation (Heche et al., 2003).

One can identify, in addition to the row of independent random and identically distributed variables, the non-stationary stochastic processes, “stationary stochastic processes”, and we create the hypothesis that „the statistical balance” of these processes lead to an analysis that is similar to independent sampling; in the stationary case, the simulation results are identified within a row of random variables identically distributed but independent one from another, this dependency remaining invariant in relation to the time shifts (Bratley et al., 1983).

Considering these hypotheses, we identify an ergodic process, (a stochastic process is ergodic if the statistical properties are deducted from a single random sample, representative for enough long temporary sequences), confidence estimates are generated for \( \mu = E[X_j] \), starting from an unique row.

The temporary average of the values of a particular process achievement tends to the hope \( X_j \), therefore existing for a particular moment \( j \), the probability 1, it results:

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = E[X_j] = \mu
\] (11)

If the simulations are replications of that in non-stationary case, the occurrence average for any replication tends to the same limit as any average, transversally with the probability 1, it results

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i^{(k)} = \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} X_i^{(k)} = E[X_j] = \mu, \text{ a.s.} \] (12)

the phrase is practically a representation of the great number law, the annotation a.s. comes from the theory of probabilities meaning almost sure.

Following the developed reasoning, we can conclude the convergence of the simulation processes, even if the economic problems aim at disjunctive phenomena and the elaboration methods for the simulation processes are different as vision and approach.

CONCLUSIONS

The concepts induced by the simulation process within engineering sciences sees new dimensions and development trends for the economic processes, with a special focus on the processes involving managerial decision.

Within the latter, the development is structured on two hubs:

1. The simulation obliges the elaboration of structured models, described by mathematical formalisms with a high rigour degree, the analyses of economic processes being ample and profound;

2. The simulation is that process implicitly offering an easy and “friendly” communication relationship between humans and “machines”, allowing the use of data storage power and the computing one, offered by modern IT solutions, to elaborate the optimal answers, or a row of answers tending asymptotically towards the optimal operational strategic decision, for
complex and dynamic challenges generated from the modern business environment;

Within the scientific debate developed, the simulation of economic events, to be modelled by continuous and discrete events, is also taken into account.

Two modern modelling methods have been briefly presented and discussed, namely the Monte Carlo method and the Metropolis method, translated from the applications that are specific to technical, engineering sciences to economic phenomenology, (Heche et al., 2003).

We can notice the explosive development of operational strategic software products that amplify a lot the complex managerial process of decision elaboration, both operational and strategic, in temporary well-delimited horizons.

The vulnerability identified in simulation processes is accentuated simplification, economic problems having in general, in addition to the certain, quantifiable factors, a multitude of other uncertain and unpredictable factors, coming both from the interior of economic organization, decision-making organisational entities or task-forces, as well as from external ones, the business environment, legislative, environmental, competitive factors (Knuth, 1981).

With a greatly diminished mathematical formalism, following only the strictly necessary things in the comprehension and construction of models suitable for the representation of modern economic problems, the article offers an introduction to the principles and utilities offered by modelling, in the search for the uniquely optimal managerial decision, or represented by a sequence of intermediary decisions tending asymptotically towards it.

It is recommended that the results generated by the simulation processes be perceived and analysed with a certain suspicion dose, the decision-making manager having in them a considerable theoretical and practical support, but he/she has to manifest a critical, innovative and creative spirit, sometimes being close to the spirit of art creators, based on the data generated scientifically and rigorously spiritually.

REFERENCES

4. Mario G.C.A. CIMINO and Gigliola VAGLINI., Interval-Valued Approach to Business Process Simulation Based on Genetic Algorithms and the BPMN, Department of Information Engineering, University of Pisa, Largo Lucio Lazzarino 1, Pisa 56122, Italy.
Impact of Dividend Taxation Changing on Dividend Policy of Romanian Companies Listed on Stock Market

Stela JAKOVA (stelajakova@gmail.com)
Bucharest University of Economic Studies

ABSTRACT

The aim of this paper is to analyze factors (with an influence on value of transactions on stock exchange) which induce companies listed on Bucharest Stock Exchange to pay dividends for period 2008 – 2015. First we will estimate five regression models, taking into account some financial and macroeconomics variables. Second we will rebuild these five regression models including a dummy variable which indicates the change of dividend tax rate. Our analyses emphasized the fact that net income, liquidity index, BET return and total assets have a significant effect on dividend payment policy which keeps their significance also after introducing the dummy variable, except the BET rate which decreased its significance. Also only the change of dividend tax rate and net income have a positively significant impact over dividend payment policy of Romanian companies. The findings from this study are useful to be taken in account by the board of managers of companies when they decide the dividend policy for the company. In the same time our study extends the existing literature on this topic, by analyzing the impact of the most recent change in dividend tax rate from 2015, and also by extending the factors which affect the dividend policy of the companies listed on stock exchange.

Keywords: dividend payout, dividend tax, stock market, stock exchange

JEL Classification: G34, G35, G38

INTRODUCTION

Sometime is very difficult to consider the factors which indicate an increase of willingness to pay dividend by companies listed on Stock Exchange. In many researches is analyzed the influence of change in taxation as an important factor over the dividend payment of companies’ behavior. In Romania, starting from last decade until now the Fiscal Code has been changed three times regarding the dividend tax rate. Until the end of the year 2005 according to Art. 67 lit. 1 of the Fiscal Code dividends were imposed with a share of 10% of their value and starting from 2006 the dividend tax rate has been modified to 16%. This means that the dividends from 2004 has been distributed in 2005, and companies has profited from this low taxation in order to pay their dividends until the end of 2005 because in 2006 was expected a rise due to some rumors regarding the market expectation (Dragotă et al., 2009).
Knowing that from 2015, according the new law number 227/2015 regarding the Fiscal Code, the dividend tax was also reduced to 5%, this paper wants to know if the companies’ behavior to pay more dividends in 2016 has been modified or there are more other financial variables from the companies’ balance sheet which indicate that there exist other influences determining the dividends payout.

The paper is organized as follows: section 2 reviews the literature regarding to the factors which impact the decisions of companies regarding to the dividend policy; in section 3 we present the methodology used; section 4 is reporting the main descriptive statistics and the correlation analyze of used data, while main results are presented in section 5. Finally, in the sixth section, we present the conclusions of our study.

LITERATURE REVIEW

A significant number of past studies have concentrated their attention to the tax changes with an impact over the companies’ behavior. Due to different period of time of business cycle that a company is does not look to exist a strict reaction of companies when a decrease of dividend tax rate is occurred. Some ideas are saying that if companies record profit this does not mean that they are going to pay necessarily dividends. Some of them chose to allocate the profits for investments reasons, as Djankov et al., (2010) mention in their paper that a decrease of tax rates, in most cases, induces the companies to make higher investments. On the other hand, subsequently some of them, chose to pay dividends due to a decrease of tax rates and this according to objectives that each company has. That’s why it can be said that the initiative to pay dividends depends not only by lower dividend tax rate expectations but also by the sustainability of profit recorded by companies, as Lintner, (1956) demonstrates in their model. This because reducing the income tax rates is one of the most important factor with a direct influence in encouraging the investments (Cojocaru and Cocoșilă, 2003).

Only few papers demonstrated the existence of an insignificance relationship between profit and changes of dividend policy (Jensen’s, 1986; Gill et al., 2010; Utami and Inanga, 2011; Hellström and Inagambaev, 2012). A considerable number of studies such is: Fama and Babiak (1968), Kale and Noe, (1990), Charitou (2000), Al-Malkawi (2007), Kowalewski et al. (2007), Anil and Kapoor (2008), Ahmed and Javid (2009), Ramli (2010), Al-Shabibi and Ramesh (2011) and Hashemi and Zadeh (2012), have demonstrated that dividend policy is positively significant influenced by gross and net incomes.

Firstly, Lintner, (1956), is one of the past earliest study which concluded their findings to “Lintner model” as an important basic formula in analyzing the determinants of dividend policy. Then other studies such is Pruitt and Gitman (1991) and Fama and Babiak (1968), support Lintner (1956) conclusion demonstrating for that the dividend policy of USA companies is impacted to a change of current profit and previous dividends.

But knowing that sometimes companies do not allocate their earnings to pay dividends, even if they record profit this means that others factors have intervene, which are determinants in dividend policy decision for the companies. Al-Malkawi
(2007), Kowalewski et al. (2007), Ramli (2010) and Al-Shubiri (2011), showed that if the debts of the companies increase this means a lack of liquidity. Also it can be said that even if companies, with high total debts, have recorded profit they will allocate their profit to investment and financing actions of firms. So companies will not change their decision to pay dividends due to other objectives that they have. The board director of companies must also to consider the profit realized by them, total debts, necessary investments and the opportunity to make good investments and also the size of the shareholders before they decide to pay the dividends (Yusof and Ismail, 2016).

Moreover, analyzing the manner that dividend payment is impacted for 200 countries listed on the Malaysia stock market, Yusof and Ismail, (2016) considered the dividend policy as important and useful to be taken in account by the companies’ director board and managers and also for the companies; shareholders in order to keep the existing investors and creating the new ones. Dragotă (2006) mention in his paper that 50% from companies’ shareholders are minority shareholders. That’s why is very important to extend logically this conclusion and to mention that if companies have larger shareholders and decide to pay them this is an opportunity for the companies to attract new investor and to keep the existing ones knowing that the dividend is a kind of return for the invest that a shareholder has made.

Regarding to market expectation in Romania case have been some rumors about the dividend tax rate which is going to increase next year in 2017 (Dan, 2016; AttoSoft site, 2016), which maybe have had an influence over the decision of dividend policy of companies. Excepting larger rates of dividend taxes in the future is normally to effect the companies’ initiatives to pay the dividends to shareholders in order to maintain their existing investment and maybe to profit from this moment to attract new investors.

Dragotă et.al, (2009), have analyzed the Romanian case for 37-41 companies listed at Bucharest Stock Exchange (BVB) demonstrating a negative impact of corporate tax burden, a positive significant impact of gross income, a negative influence of GDP, a negative impact of Romanian State shareholders, and a negative impact of tax changes on dividend payment which is not similar for all of companies.

Based on many studies which have included more variables in analyzing the manner that dividend policy is impacted or the determinants of the dividend payment, the present study try to extend the existing literature from the stock market of Romanian case using more recent data and extending the number of companies taken under analyze, including all the sectors. In this papers analyze model we have introduced one more variable which is the liquidity index of the firms in compare with the Dragotă et.al, (2009).

**METHODOLOGY – THE DATA**

*Data and descriptive statistics*

This paper research is devoted to analyze in two ways the manner that the dividend payment policy is influenced: first, by some financial and macroeconomic
indicators and other indicators which have a direct influence in stock market; second, by including in the same models also the impact of change in tax dividend from the year 2015. There is one study which treats this topic for Romanian companies, namely Dragota et al., (2009), which found no significant impact between change of cooperate tax and dividend payment but an influence over the dividend payment by companies in 2005, when is recorded a change of dividend tax.

Following Dragota et al., (2009) which taken under analyze 65 listed companies at Bucharest Stock Exchange for the period 1998-2005, we extend and updated the analysis with a database covering the period 2008-2015. The total number of companies listed on BVB for this period is 86 but taking in consideration the available dates which can be taken in analyze we have reduced the number of companies used in regression models to 59.

In table 1, are presented and also described all the variables and estimated variables taken under analyze. Respective some of these variables are downloaded from the links presented in the table 2.

**Variables with impact over the dividend payment of companies’ behavior**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dividend payment (DIV)</td>
<td>Dividends paid by companies only when it is recorded income, according to the Romanian law.</td>
</tr>
<tr>
<td>Gross income (GRE)</td>
<td>Recorded by listed companies</td>
</tr>
<tr>
<td>Net income (NEG)</td>
<td>Recorded by listed companies</td>
</tr>
<tr>
<td>Corporate tax burden (CTB)</td>
<td>Calculated as a difference between GRE and NEG</td>
</tr>
<tr>
<td>Market capitalization (MKCap)</td>
<td>Calculated by multiply the value of closing market price and number of shares</td>
</tr>
<tr>
<td>Market to book ratio (MBR)</td>
<td>Calculated as a ratio between market capitalization and total Shareholders’ Equity</td>
</tr>
<tr>
<td>GDP growth rate (GDP)</td>
<td>Reported Gross Domestic Product growth rate</td>
</tr>
<tr>
<td>Total debts (DBT)</td>
<td>Recorded by listed companies</td>
</tr>
<tr>
<td>Total assets (AST)</td>
<td>Recorded by listed companies</td>
</tr>
<tr>
<td>Market return (BET)</td>
<td>The market index BET returns calculated yearly as a percentage between final price minus initial price reported to the initial price</td>
</tr>
<tr>
<td>Liquidity index (IL)</td>
<td>Estimated as a report between total receivables reported to turnover</td>
</tr>
<tr>
<td>Changes in dividend tax payment in the beginning of the 2015 (TAX)</td>
<td>Dummy variable is equal to 0 for the period 2008-2014 and equal to 1 for the year 2015.</td>
</tr>
</tbody>
</table>

Source: Authors' calculation

Considering ca financial variables reported yearly by companies were not all available for all the period taken under analyze, we have downloaded them by more than two sites.
Data sources for used variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIV, GRE, NEG, AST, DBT, BET</td>
<td><a href="http://www.bvb.ro/FinancialInstruments/Markets/Shares">http://www.bvb.ro/FinancialInstruments/Markets/Shares</a></td>
</tr>
<tr>
<td>Number of Shares/Company</td>
<td><a href="http://www.depozitarulcentral.ro">www.depozitarulcentral.ro</a></td>
</tr>
</tbody>
</table>

Source: Authors calculation

Descriptive statistics for selected variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Median</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln(DIV)</td>
<td>7.8155</td>
<td>0.0000</td>
<td>21.2863</td>
<td>0.0000</td>
<td>8.3594</td>
</tr>
<tr>
<td>ΔLn(GRE)</td>
<td>0.3508</td>
<td>0.2049</td>
<td>9.5325</td>
<td>-4.5849</td>
<td>1.5227</td>
</tr>
<tr>
<td>ΔLn(NEG)</td>
<td>0.3577</td>
<td>0.1628</td>
<td>17.8506</td>
<td>-10.8969</td>
<td>2.1274</td>
</tr>
<tr>
<td>ΔLn(MKCAP)</td>
<td>0.0931</td>
<td>0.1628</td>
<td>17.8506</td>
<td>-10.8969</td>
<td>2.1274</td>
</tr>
<tr>
<td>ΔLn(CTB)</td>
<td>0.4148</td>
<td>0.2344</td>
<td>18.1328</td>
<td>-18.0742</td>
<td>3.3080</td>
</tr>
<tr>
<td>ΔLn(CTB)</td>
<td>0.3864</td>
<td>0.1628</td>
<td>17.8506</td>
<td>-10.8969</td>
<td>2.1274</td>
</tr>
<tr>
<td>ΔLn(DBT)</td>
<td>0.4843</td>
<td>0.0589</td>
<td>7.2185</td>
<td>-5.2316</td>
<td>1.7294</td>
</tr>
<tr>
<td>MBR</td>
<td>0.7274</td>
<td>0.1628</td>
<td>17.8506</td>
<td>-10.8969</td>
<td>2.1274</td>
</tr>
<tr>
<td>IL</td>
<td>0.5084</td>
<td>0.2344</td>
<td>18.1328</td>
<td>-18.0742</td>
<td>3.3080</td>
</tr>
<tr>
<td>BET</td>
<td>0.0445</td>
<td>0.0000</td>
<td>0.6168</td>
<td>-0.7047</td>
<td>0.3596</td>
</tr>
<tr>
<td>TAX</td>
<td>0.1228</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.3285</td>
</tr>
</tbody>
</table>

Source: Authors calculation

Going on, in Table 3 we present the main descriptive statistics for the analyzed indicators. Based on this information, we are able to see that the mean for market return is 4.4%, while the liquidity index record an average of 50%.

Model

Following Dragotă et al., (2009) methodology which has as fundamental the Lintner model (1956), we will estimate 2 types of equations: first type in which we include only financial variables as main factors in dividend payments, and second type of regression models in which we will include the variable dummy which highlights the change in dividend tax rate from 2016, which was applied for paid dividends for 2015.

In order to normalize the series and to be stationary data, we used the logarithm series for: total assets, total debts, corporate tax burden, gross income, net income and market capitalization.

Five regression models which shows how the dividend payment by the listed companies is influenced by other macroeconomics and financial variables, are given by equations (1.1)-(1.5):
Other five regression models which include also a change in dividend tax (dummy variable) besides the macroeconomics and financial variables, are given by equations (2.1)-(2.5):

\[
\begin{align*}
\text{(2.1)} & \quad \ln(\text{DIV})_\text{it} = \beta_1 + \beta_2 \ln(\text{AST})_\text{it} + \beta_3 \ln(\text{CTB})_\text{it} + \beta_4 \ln(\text{MKCAP})_\text{it} + \beta_5 \ln(\text{NEG})_\text{it} + \beta_6 \text{MBR}_\text{it} + \beta_7 \text{IL}_\text{it} + \beta_8 \text{BET}_\text{it} + \epsilon_\text{it} \\
\text{(2.2)} & \quad \ln(\text{DIV})_\text{it} = \beta_1 + \beta_2 \ln(\text{DBT})_\text{it} + \beta_3 \ln(\text{CTB})_\text{it} + \beta_4 \ln(\text{MKCAP})_\text{it} + \beta_5 \ln(\text{NEG})_\text{it} + \beta_6 \text{MBR}_\text{it} + \beta_7 \text{IL}_\text{it} + \beta_8 \text{BET}_\text{it} + \epsilon_\text{it} \\
\text{(2.3)} & \quad \ln(\text{DIV})_\text{it} = \beta_1 + \beta_2 \ln(\text{GRE})_\text{it} + \beta_3 \ln(\text{CTB})_\text{it} + \beta_4 \ln(\text{MKCAP})_\text{it} + \beta_5 \ln(\text{NEG})_\text{it} + \beta_6 \text{MBR}_\text{it} + \beta_7 \text{IL}_\text{it} + \beta_8 \text{BET}_\text{it} + \beta_9 \text{TAX}_\text{it} + \epsilon_\text{it} \\
\text{(2.4)} & \quad \ln(\text{DIV})_\text{it} = \beta_1 + \beta_2 \ln(\text{ASST})_\text{it} + \beta_3 \ln(\text{CTB})_\text{it} + \beta_4 \ln(\text{NEG})_\text{it} + \beta_5 \text{MBR}_\text{it} + \beta_6 \text{IL}_\text{it} + \beta_7 \text{BET}_\text{it} + \beta_8 \text{TAX}_\text{it} + \epsilon_\text{it} \\
\text{(2.5)} & \quad \ln(\text{DIV})_\text{it} = \beta_1 + \beta_2 \ln(\text{ASST})_\text{it} + \beta_3 \ln(\text{CTB})_\text{it} + \beta_4 \ln(\text{NEG})_\text{it} + \beta_5 \text{MBR}_\text{it} + \beta_6 \text{IL}_\text{it} + \beta_7 \text{BET}_\text{it} + \beta_8 \text{TAX}_\text{it} + \epsilon_\text{it}
\end{align*}
\]

Where \( i \) - cross-section observation unit in the sample, stands for company, \( t \) – time period, takes value between 2008 – 2015, \( \beta_1 \) to \( \beta_8 \) - are the parameters of the models that will be estimated, \( \alpha_i \) - is the individual effect, \( \epsilon_\text{it} \) - is error term of above regression models.

Based on the stationarity test Levin, Lin and Chu (2002) presented in the Table 4 it can be seen that all variables being stationary.

**Unit root test for panel data – Levin, Lin & Chu (2002)**

<table>
<thead>
<tr>
<th>Variables</th>
<th>t-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln(\text{DIV}) )</td>
<td>-68.6701***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \Delta \ln(\text{GRE}) )</td>
<td>-9.1673***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \Delta \ln(\text{NEG}) )</td>
<td>-5.5315***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \Delta \ln(\text{MKCAP}) )</td>
<td>-30.6789***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \Delta \ln(\text{CTB}) )</td>
<td>-6.8987***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \Delta \ln(\text{AST}) )</td>
<td>-8.6914***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \Delta \ln(\text{DBT}) )</td>
<td>-8.4579***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \text{MBR} )</td>
<td>-15.5378***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \text{IL} )</td>
<td>-6.5362***</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \text{BET} )</td>
<td>-38.0537***</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

* *, **, *** - Indicates significant at the 0.1 level, 0.05 level and 0.01 level

Before regression estimation based on panel data, we calculated the correlation coefficient between independent variable in order to avoid multicollinearity, the results are presented in Table 5. Based on these the high positive correlation between: the total assets (AST) and total debts (DBT) and gross income (GRE); corporate tax
burden (CTB) and gross and net income (GRE), (NEG) and also between gross income (GRE) and net income (NEG), are taken into consideration when the regressions are built.

The correlation coefficient analysis between the variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>∆Ln(CTB)</th>
<th>∆Ln(DBT)</th>
<th>∆Ln(GRE)</th>
<th>MBR</th>
<th>∆Ln(MKCAP)</th>
<th>∆Ln(NEG)</th>
<th>IL</th>
<th>BET</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆Ln(CTB)</td>
<td>0.1683</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(DBT)</td>
<td>0.5472</td>
<td>0.1843</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(GRE)</td>
<td>0.3940</td>
<td>0.3806</td>
<td>0.4713</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBR</td>
<td>-0.0283</td>
<td>-0.0494</td>
<td>-0.0645</td>
<td>-0.0689</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(MKCAP)</td>
<td>-0.0204</td>
<td>0.0762</td>
<td>-0.0294</td>
<td>-0.0337</td>
<td>0.0878</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(NEG)</td>
<td>0.2499</td>
<td>0.1248</td>
<td>0.3787</td>
<td>0.6903</td>
<td>-0.0543</td>
<td>-0.0328</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>0.0031</td>
<td>0.0419</td>
<td>-0.0191</td>
<td>-0.0023</td>
<td>0.0405</td>
<td>-0.0185</td>
<td>-0.0018</td>
<td>1.0000</td>
</tr>
<tr>
<td>BET</td>
<td>0.1043</td>
<td>0.0548</td>
<td>0.0905</td>
<td>0.0127</td>
<td>-0.0297</td>
<td>0.1545</td>
<td>-0.0080</td>
<td>-0.0349</td>
</tr>
<tr>
<td>TAX</td>
<td>-0.0519</td>
<td>-0.0581</td>
<td>-0.0588</td>
<td>-0.0255</td>
<td>0.0081</td>
<td>-0.0301</td>
<td>-0.0256</td>
<td>0.1347</td>
</tr>
</tbody>
</table>

Source: Authors calculation

RESULTS

The estimation results of regression models based on yearly panel data are presented in Table 6 and Table 7. In these models we have considered as a dependent variable the dividend paid by companies – Ln(DIV). Through these estimated regression models we want to highlight the manner in which financial factors and changing of the legislation regarding the decrease in tax rate for dividends affects the dividend payment policy of analyzed companies.

The results for the first type of estimated regression are presented in Table 6, where are included only the financial variables which indicate the companies’ stance with an impact over the dividend payment.
Model estimation for the determinants of dividend policy for the Romanian listed companies over the period 2008–2015 (excluding tax change dummy)

Table 6

<table>
<thead>
<tr>
<th></th>
<th>Model 1.1</th>
<th>Model 1.2</th>
<th>Model 1.3</th>
<th>Model 1.4</th>
<th>Model 1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>8.1401</td>
<td>8.2575</td>
<td>8.0288</td>
<td>8.1838</td>
<td>-9.4555</td>
</tr>
<tr>
<td></td>
<td>(0.4770)</td>
<td>(0.4854)</td>
<td>(0.4860)</td>
<td>(0.3095)</td>
<td>(2.3490)</td>
</tr>
<tr>
<td>∆Ln(NEG)</td>
<td>0.3893**</td>
<td>0.3809**</td>
<td>0.3677***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1427)</td>
<td>(0.1427)</td>
<td>(0.1422)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(GRE)</td>
<td></td>
<td>0.3044</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(DBT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.1572</td>
<td>(0.1775)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(CTB)</td>
<td>-0.5824**</td>
<td>-0.6728***</td>
<td>-0.6120***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1701)</td>
<td>(0.2153)</td>
<td>(0.2167)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(MKCAP)</td>
<td>0.6569</td>
<td>0.6550</td>
<td>0.7306</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.5378)</td>
<td>(0.5423)</td>
<td>(0.5425)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBR</td>
<td>0.5100</td>
<td>0.4393</td>
<td>0.5664</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4289)</td>
<td>(0.4345)</td>
<td>(0.4365)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>-0.2678</td>
<td>-0.2750**</td>
<td>-0.2670**</td>
<td>-0.2667**</td>
<td>-0.2715**</td>
</tr>
<tr>
<td></td>
<td>(0.1189)</td>
<td>(0.1210)</td>
<td>(0.1205)</td>
<td>(0.1196)</td>
<td>(0.1191)</td>
</tr>
<tr>
<td>BET</td>
<td>-2.7186</td>
<td>-3.1703***</td>
<td>-3.2644***</td>
<td>-2.4700**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.3174)</td>
<td>(1.3306)</td>
<td>(1.3207)</td>
<td>(1.3011)</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.6559</td>
<td>0.6427</td>
<td>0.6415</td>
<td>0.6676</td>
<td>0.6520</td>
</tr>
<tr>
<td>R-squared (adjusted)</td>
<td>0.5669</td>
<td>0.5518</td>
<td>0.6434</td>
<td>0.5610</td>
<td>0.5650</td>
</tr>
<tr>
<td>F-statistic</td>
<td>7.367***</td>
<td>7.0728***</td>
<td>7.2181***</td>
<td>7.4818***</td>
<td>7.4944***</td>
</tr>
</tbody>
</table>

a – (standard errors in parentheses)
*, **, *** - Indicates significant at the 0.1 level, 0.05 level and 0.01 level

We can observe for the first Model 1.1 that variables have all a significant effect over the dividend payment policy. The difference appears regarding the sign of influence. So we are able to see that net income, corporate tax burden, market capitalization and market to book effect positively the dividend payment policy of companies, while total assets, liquidity index and BET effects negatively the companies policy to pay dividends.

In the Model 1.2 has been included another financial variable, total debts which revealed a negative influence over the dividend payment but this is not significant. The result is in accordance with the conclusions of Al-Malkawi (2007), Kowalewski et al. (2007), Ramli (2010) and Al-Shubiri (2011).

Model 1.3 shows the gross income effects positively the dividend payment, but this is not significant. From the other hand the Model 1.4 shows that the net income of companies has a positive significant effect over the changes of dividend payment policy. This is logically explicable by the business cycle point of view because the earning remaining after the discounted taxes (net income) is a direct financial information through which companies can take decision regarding to the dividend payment policy. In this case we can raise two causes: this happen because of a lower demand of these companies listed on BVB as Dragotă et al., (2009) mention in their study, makes companies to pay more dividends; or maybe there is another factor which has a direct influence over the...
positive changes of dividend payment policy.

In the Model 1.5, likewise, is observed a strongly negatively significant effect of total assets of companies such as liquidity index and BET have in dividend payment.

On overall we are able to see that there are net income has a positive and significant effect on dividend payments of companies, at also it is states by the Leitner (1956), because the dividend payment policy decision can be explained by the sustainability of earnings. In the same time, the liquidity, total assets and BET return have all a negative impact on the dividend payment.

For the second type of regression models which calculate the same models but including the variable dummy which signify the change in dividend tax rate from 2016, applied for paid dividends for 2015, the results are presented in the Table 7.

Model estimation for the determinants of dividend policy for the Romanian listed companies over the period 2008–2015 (including tax change dummy)

<table>
<thead>
<tr>
<th></th>
<th>Model 1.1</th>
<th>Model 1.2</th>
<th>Model 1.3</th>
<th>Model 1.4</th>
<th>Model 1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>7.7487**</td>
<td>7.8632***</td>
<td>7.6277***</td>
<td>7.8643***</td>
<td>8.1573***</td>
</tr>
<tr>
<td></td>
<td>(0.5050)</td>
<td>(0.5107)</td>
<td>(0.5124)</td>
<td>(0.3297)</td>
<td>(0.3954)</td>
</tr>
<tr>
<td>∆Ln(NEG)</td>
<td>0.3957***</td>
<td>0.3836***</td>
<td>0.3737***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1417)</td>
<td>(0.1413)</td>
<td>(0.1413)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(GRE)</td>
<td>0.3274*</td>
<td>0.3274*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2002)</td>
<td>(0.2002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(DBT)</td>
<td></td>
<td></td>
<td>-0.1406</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1764)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆Ln(CTB)</td>
<td>-0.5478***</td>
<td></td>
<td></td>
<td></td>
<td>-0.5775***</td>
</tr>
<tr>
<td></td>
<td>(0.2160)</td>
<td></td>
<td></td>
<td></td>
<td>(0.2157)</td>
</tr>
<tr>
<td>∆Ln(MKCAP)</td>
<td>0.6201</td>
<td>0.6159</td>
<td>0.6957</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.5344)</td>
<td>(0.5437)</td>
<td>(0.5387)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBR</td>
<td>0.5305(0.4261)</td>
<td>0.4551(0.4315)</td>
<td>0.5828(0.4332)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td>-0.3069***</td>
<td>-0.3148***</td>
<td>-0.3129***</td>
<td>-0.3107***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1194)</td>
<td>(0.1214)</td>
<td>(0.1209)</td>
<td>(0.1197)</td>
<td>(0.1196)</td>
</tr>
<tr>
<td>BET</td>
<td>-2.0142*</td>
<td>-2.4272*</td>
<td>-2.4899*</td>
<td>-1.7784*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.3457)</td>
<td>(1.3613)</td>
<td>(1.3526)</td>
<td>(1.3284)</td>
<td></td>
</tr>
<tr>
<td>TAX</td>
<td>2.0338**</td>
<td>2.0894**</td>
<td>2.1337**</td>
<td>2.3268**</td>
<td>2.0433**</td>
</tr>
<tr>
<td></td>
<td>(0.9103)</td>
<td>(0.9235)</td>
<td>(0.9193)</td>
<td>(0.8880)</td>
<td>(0.9117)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.6618</td>
<td>0.6490</td>
<td>0.6509</td>
<td>0.6558</td>
<td>0.6579</td>
</tr>
<tr>
<td>R-squared (adjusted)</td>
<td>0.5728</td>
<td>0.5581</td>
<td>0.5609</td>
<td>0.5698</td>
<td>0.5710</td>
</tr>
<tr>
<td>F-statistic</td>
<td>7.4376***</td>
<td>7.1467***</td>
<td>7.3014***</td>
<td>7.6230***</td>
<td>7.5638***</td>
</tr>
</tbody>
</table>

Based in the results of the second type of regression models, we try to analyze if the change of tax rate for dividends (dummy variable) have an impact or not in the moment in which they have been applied. As it is observed no matter of which financial
variables are taken into account in these regression models the dividend tax policy is positively significant effected due to a modification in tax rate. Different from Dragotă et al., (2009) which have concluded that this impact might be different from one to other company our results says that Romanian companies listed on BVB have preferred to profit by this moment of tax change. This difference of our paper result can be explained due to a decrease of the tax rate to 5% in cooperation with the period taken under analyze by Dragotă et al., (2009) where the tax has been reduced to only 10%.

As a result, can be said that the companies’ earnings have influenced positively the dividend tax policy by paying the shareholders. The change in income tax to smaller rate has an impact over the companies’ initiatives to pay the shareholders.

Once again, we are able to see that the same variables which have a significant impact in models estimated in table 6, keeps their significance after introducing the dummy variable (only in BET rate case the significance decreased a little bit).

On the other hand, can be said that gross income influence has become significant positive in the moment that we have introduce the variable dummy which is logically and similar with previous conclusions such is Lintner, (1956), Fama and Babiak, (1968), etc.

CONCLUSIONS

The main purpose of this paper was to analyze the manner that the dividend payment policy undertaken by companies listed on Bucharest Stock Exchange is influenced due to a change of income tax rate, taking into account the period 2008-2015 for a number of 59 of companies. The paper has included in analyze also other financial variables which might be main factors to the change of this policy by Romanian companies.

Our analyses emphasized that a considerable decrease of dividend tax rate is able to strongly statistically influence the decisions of Romanian companies regarding to dividend payment policy, but the opposite relation is valid, where an increase of net income have a strongly influence over the dividend payment.

Romanian companies listed on stock market when registered a sustainable profit they will used to allocate this profit for investment to their market and also to pay dividends. On the other hand, in the moment when a lower dividend tax rate is introduced in 2015 to 5%, according to the law number 227/2015 of the Fiscal Code, seems that Romanian companies have profited by this new rate to pay more dividends. While our conclusions are appropriate with Dragotă et al., (2009), which taken under analyze the period 1998-2005. We can conclude that when this change of income tax rate is given by very lower rates the initiatives of companies to pay dividends is higher.

Also a negative relation is observed in case of financial variables estimated in regression models such are: total assets, liquidity index and BET return which have a significant negative influence over the dividend payment policy. This can be explained starting from the stances of stock market performances. When a market becomes more performant in this phase the companies will choose to concentrate their incomes to increase the investment, to profit from their higher demand in market and maximizing their profits. Otherwise they will choose to allocate their profits paying the shareholders in the periods of time when the market is slowing down and their activity market performance decrees. This will increase the shareholders return and stimulate their
opportunity to reinvest in large companies and lower debts (Yusof and Ismail, 2016). So companies must to control their profits, debts and the lucidity in order to profit by these shareholders, to attract investors and to maintain the existent ones.

Another thing which is very important to mention is that all the variables which indicate a significant impact in dividend payment policy have preserved persistent their influence in all regression models estimated, even we introduced the dummy variable, except of BET return case which significance decreased a little bit. This means that net income, total assets, liquidity index and BET index are stable variables with a big influence over the change of dividend payment policy.

As a conclusion might be said that even if dividend payment policy of Romanian companies is sensitive to a lower dividends tax rates and also to higher net incomes recorded by them, it seems that other factors, such are liquidity index, BET return and total assets, prevent this sensitive reaction because other important objectives which intervene in lodges. Dragotă et al., (2006), Yusof and Ismail, (2016) mention in their studies that some factors may be the size of smaller and larger shareholder that a company has to pay dividends.

It seems that everything keeps their equilibrium in everything, but this depends too much by companies’ policies and is not a strict one. These results add evidence to Dragotă et al., (2009) and Yusof and Ismail, (2016), which are useful to be taken in consideration for the board directors of the companies when they want make decision in dividend policy. Before making the decisions they must take in consideration the principal factors with a significant impact over the dividend payment such is: net income, BET return, liquidity index, change of dividend tax. These conclusions are logic, because in periods of time when the market is growing, the company reinvest its profit to increase the available resources in order to maximize its return from the main activity such that the dividend payment policy is negatively impacted by this decision. Another logically approach to explain this is that dividend payment is considered as a return to shareholders which invest only in large companies with low debts in order to maximize their return. But in the moment that the stock market performance decrease this will disadvantage the investment decision of shareholders. But from the other side when a change of the dividend tax rate is introduced the board of directors must to reconsider all the factors together in order to profit from lower tax rate to pay dividends and also to maintain in control the other variables.

Further studies, can take in consideration other financial variables or extending the period of time and the number of companies taken under analyze to see how dividend tax influence the dividend policy.

References


Theoretical Aspects Regarding the Optimal Taxation of Effort With More Conditions

Prof. Dumitru MARIN PhD (dumitru.marin@csie.ase.ro)
Bucharest University of Economic Studies, Bucharest

Prof. Ion PARTACHI PhD (ionpartachi@yahoo.fr)
Academy of Economic Studies of Moldova

Assoc. prof. Madalina ANGHEL PhD (madalinaanghel08@yahoo.com)
Artifex University, Bucharest

ABSTRACT

In this article, the authors propose to conduct a pertinent analysis of the use of mathematical models in relation to optimal effort taxation. The effort taxation must be a balance between allocative efficiency and distribution. It is a question of determining the optimal tax level for three conditions. The theory is that the shift to a fixed amount, a single share, is not feasible for wage compensation, being important to note that this condition is one that must be considered. The mathematical model proposed by the authors regarding optimal taxation of multi-condition effort is relevant and suggestive for the analysis. It is well known that, in order to increase tax revenue, they must inevitably be based on the unobservable variant, which is the potential gain or, most importantly, on the individual productivity of labor. Solving a problem of this sensibility, such as the optimal taxation of multi-state effort, needs to be analyzed and conditioned by the establishment of a mathematical-econometric model that highlights both the aspects of a minimum expectation threshold or a situation that depends on the economic outcome. The authors consider the existing variants and establish variants based on a mathematical model that is analyzed to best answer the problem of the adverse selection in the balance between allocation efficiency and distribution. We analyze the optimization problem from Lagrange’s function and associated multiplier, highlighting the mathematical functions that can be used. Further, the authors consider the case of adverse selection based on asymmetric information. For the objective function and budget constraint, the authors consider that they need to add incentive restrictions, especially when they are taxed on income. By analyzing this hypothesis in depth, the authors propose and demonstrate a mathematical function that best suits these adjustment needs and solve the adverse selection. The way in which the authors suggest and demonstrate the implementation condition that is equivalent to the assertion that the plurality of admissible solutions is unclear gives a solution perspective applicable in most cases. The authors propose, by synthesizing, conditions (sentences) that demonstrate that the model used is one that must be considered and can be successfully used in optimal taxation of the multi-state
This article aims to solve and perform a scientific analysis of how to balance balance between allocation efficiency and allocation. Income, as inputs or effort, as an output must be analyzed in the two-state version of the most commonly known fixed amount and most often used in practice. However, fixed-rate taxation is limited by the lowest income, as is the case in Romania’s economy, when talking about the minimum wage on the economy. From this point of view, the authors find that a single tax rate applied is not feasible under these circumstances. In the article, the authors inventory a number of analyzes made by other economists, modellers, concerned with establishing a mathematical function with regard to optimal taxation of effort that is characterized by several states. The authors identify and propose a mathematical model to be used at least in two situations. The first is an optimal solution, a context in which the authors identify in their demonstration a series of sentences to be considered by the one who tries this optimization. We will identify from several sentences, and the authors only deal with sentence number one, when symmetric information implies an optimal tax imposed by the government with some characteristics such as budget constraints, marginal utilities, usage levels, Marginal cost of labor, marginal productivity, optimal taxes. All this is the principle from which the authors go. Of course, they demonstrate from Lagrange’s function that the multiplier associated with the participation restriction is in fact a mathematical function that any analyst or user of certain information needs to consider in order to impose management at micro- or macroeconomic level. The authors also deal with the asymmetric information selection based on the fact that the objective function and the budgetary restriction still have to be based on some restrictions in case of income tax. It is demonstrated on the basis of mathematical elements that the computational relationships reached by the authors are those that must mean the way of analysis in order to simplify the optimal optimization of the effort. Appreciating the association of the Kuhn-Tucker multiplier, we analyze the three constraints on which the Lagrange function based on the theorem, „all Kuhn-Tucker multipliers are strictly positive”, and the optimal effort is given by mathematical equality. Finally, the authors consider that the established mathematical relationship is the one and only the one that gives the desired results.

LITERATURE REVIEW

Albanesi and Sleet (2006) are concerned about optimal, dynamic taxation. Aizenman and Frenkel (1985) analyze the correlation between optimal remuneration indexation, markets and monetary policies. Anghelache, Anghelache, Anghel, Niţă and

**RESEARCH METHODOLOGY AND DATA**

A central theme of Adverse Selection is that of the balance between allocative efficiency and distribution. This conflict was highlighted by Mirrlees (1971), winner of the Nobel Prize in Economics.
The issue of optimal taxation of income or input was first analyzed by Mirrlees (1988) in the two-state version of fixed, non-taxable amount of taxation. Fixed charge is limited by the lowest income (single rate is not feasible).

It has been shown that in order to increase tax revenue, it must be based inevitably on the unobservable variable (as potential gains) or on the productivity of the individual.

Maskin and Riky (1989) analyzed how tax levels would be affected if instead of observing „income,“ the government would have noticed the „input“ individually, that is, the equivalent workload, the number of hours worked.

In the paper we determine the optimum level of taxation for three states, both in symmetric information and in asymmetric information. Local and global (upward and downward) inciting locality restrictions are analyzed. Applies when the change in fixed amount (single rate) is not feasible.

Fees must be based on the possible earnings or productivity of the individual.

1. The mathematical model

Assume that an output \( q \) (an income) is obtained with an input (effort) \( e \), according to the productivity function \( q = \theta F(e) \).

The parameter \( \theta \) is one of productivity and can take three values (three tax installments) denoted \( \theta_L, \theta_M, \theta_G \) cu \( \theta_L < \theta_M < \theta_G \) (low, medium and high productivity).

We assume that the proportions in which individuals are L, M or G (subjective or objective probabilities) are \( \Pi_L, \Pi_M, \Pi_G \) cu \( \Pi_L + \Pi_M + \Pi_G = 1 \) and all strictly positive.

Economic agents (individuals) have the same utility function as:

\[ U(q' - t - \psi(e)) \]

where:

\( U'(\cdot) > 0, U''(\cdot) < 0 \), and the variables have the following meanings:

\( q \) = output-ul;

\( t \) = net tax that the individual has to pay or receive from the government (subsidy);

\( \psi(e) \) = the effort function, increasing and convex.

Then the government’s budget cut is:

\[ \Pi_L t_L + \Pi_M t_M + \Pi_G t_G \geq u(p) \]

where:

\( t_L \geq 0, t_M \geq 0, t_G \geq 0 \), and \( u \) a minimum threshold expected by the government.

In the absence of adverse selection (symmetric information), if the government seeks to maximize social utility (the sum of individual utilities weighted with probabilities of realization) then the following problem should be solved:

\[ P(1) \text{ Max } t_L, t_M, t_G, l_L, l_M, l_G \mid \Pi_L \cdot U[\theta_L l_L - t_L - \psi(l_L)] + \Pi_M \cdot U[\theta_M l_M - t_M - \psi(l_M)] + \Pi_G \cdot U[\theta_G l_G - t_G - \psi(l_G)] \]

s.t. \( \Pi_L l_L + \Pi_M l_M + \Pi_G l_G \geq u \)

It is obvious that if the Government, as MAJOR or DECIDENT, can observe \( i \), then it can clearly specify what effort an individual puts forward, when output \( q \) is observable and quantifiable. Under these circumstances, the government can control the individual’s input (effort).
Solving P1 optimization problem
The sentence 1. In the case of symmetric information, the optimal tax imposed by the government has the following characteristics:
i) Budgetary (participation) restriction is saturated;
ii) Marginal utilities are equal between the three types of individuals;
iii) Utility levels are equal for the three types of individuals;
iv) The marginal cost of effort equals marginal productivity for each type;
v) Optimal taxes are also influenced by the proportions of low-, medium- or high-productivity individuals (in the total population).

Demonstration
The function of LAGRANGE, with the \( \lambda \) multiplier associated with the participation restriction, is written:

\[
L(t^L, t^M, t^G, l^L, l^M, l^G; \lambda) = \Pi_L U [\theta_L l^L - t_L - \psi(l^L)] + \Pi_M U [\theta_M l^M - t_M - \psi(l^M)] + \Pi_G U [\theta_G l^G - t_G - \psi(l^G)] + \lambda [\Pi_L l^L + \Pi_M l^M + \Pi_G l^G - u]
\]

By canceling the partial derivatives in relation to taxes, supposedly non-zero, we obtain:

\[
\frac{\partial L}{\partial t^L} = - \Pi_L U' [\theta_L l^L - t_L - \psi(l^L)] + \lambda \Pi_L = 0 \quad (2^o)
\]

\[
\frac{\partial L}{\partial t^M} = - \Pi_M U' [\theta_M l^M - t_M - \psi(l^M)] + \lambda \Pi_M = 0 \quad (3^o)
\]

\[
\frac{\partial L}{\partial t^G} = - \Pi_G U' [\theta_G l^G - t_G - \psi(l^G)] + \lambda \Pi_G = 0 \quad (4^o)
\]

It is noted that Lagrange \( \lambda \) is strictly positive and eliminating it results in equality:

\[
U' [\theta_L l^L - t_L - \psi(l^L)] = U' [\theta_M l^M - t_M - \psi(l^M)] = U' [\theta_G l^G - t_G - \psi(l^G)] \quad (5^o), \text{ie (ii)}.
\]

In addition, the budget restriction is saturated (\( \lambda > 0 \)) so (i) is true.
If the utility function is strictly concave, that is individuals with risk aversion, then the sequence of equality \((5^o)\) becomes:

\[
[\theta_L l^L - t_L - \psi(l^L)] = [\theta_M l^M - t_M - \psi(l^M)] = [\theta_G l^G - t_G - \psi(l^G)] \quad (5^o), \text{which implies identical levels of utility for the three types of individuals, ie (iii)}.
\]

We cancel the partial derivatives in relation to effort levels \( l^L, l^M \) and \( l^G \) and we obtain:

\[
\frac{\partial L}{\partial t^L} = \Pi_L U' [\theta_L l^L - t_L - \psi(l^L)] = \psi'(l^L) \tau = 0 \quad (7^o)
\]

\[
\frac{\partial L}{\partial t^M} = \Pi_M U' [\theta_M l^M - t_M - \psi(l^M)] = \psi'(l^M) \tau = 0 \quad (8^o)
\]

\[
\frac{\partial L}{\partial t^G} = \Pi_G U' [\theta_G l^G - t_G - \psi(l^G)] = \psi'(l^G) \tau = 0 \quad (9^o)
\]

Relationships \((7^o), (8^o)\) and \((9^o)\) lead to the equations:

\[
\psi'(l^L) = \theta_L \quad (10^o)
\]

\[
\psi'(l^M) = \theta_M \quad (11^o)
\]

\[
\psi'(l^G) = \theta_G \quad (12^o), \text{from which conclusion (iv) of sentence 1.}
\]
The solution of the system formed by the equations (10), (11) and (12) leads to the optimal values of the input (level of effort), namely:

\[ \tilde{e}_L = (\psi')^{-1}(\theta_L), \quad \tilde{e}_M = (\psi')^{-1}(\theta_M), \quad \tilde{e}_G = (\psi')^{-1}(\theta_G) \quad (13) \]

We combine the equations (6) and the tight budget constraint and taking into account the optimal effort levels (13) we obtain the system:

\[ \Pi_L t_L + \Pi_M t_M + \Pi_G t_G = y \]

\[-t_L + t_M = \theta_M \tilde{e}_M - \psi(\tilde{e}_M) - (\theta_L \tilde{e}_L - \psi(\tilde{e}_L)) = A \]

\[-t_M + t_G = \theta_G \tilde{e}_G - \psi(\tilde{e}_G) - (\theta_M \tilde{e}_M - \psi(\tilde{e}_M)) = B \]

The system can be resolved by Cramer’s rule, namely:

\[
\Delta = \begin{vmatrix} 
\Pi_L & \Pi_M & \Pi_G \\
-1 & 1 & 0 \\
0 & -1 & 1 
\end{vmatrix} = \Pi_L + \Pi_M + \Pi_G = 1
\]

\[
\Delta_{t_L} = \begin{vmatrix} 
0 & \Pi_M & \Pi_G \\
A & 1 & 0 \\
B & -1 & 1 
\end{vmatrix} = -A(1 - \Pi_L) - \Pi_G B = A\Pi_L - B\Pi_G - A
\]

\[
\Delta_{t_M} = \begin{vmatrix} 
\Pi_L & 0 & \Pi_G \\
-1 & A & 0 \\
0 & B & 1 
\end{vmatrix} = A\Pi_L + \Pi_G (-B) = A\Pi_L - B\Pi_G
\]

\[
\Delta_{t_G} = \begin{vmatrix} 
\Pi_L & \Pi_M & 0 \\
-1 & 1 & A \\
0 & -1 & B 
\end{vmatrix} = \Pi_L (A + B) - \Pi_M
\]

Then the optimal taxes are:

\[ \tilde{e}_L = \Lambda \Pi_L - B \Pi_G - A, \quad \tilde{e}_M = \Lambda \Pi_L - B \Pi_G, \quad \tilde{e}_G = \Pi_G (A + B) - \Pi_M \]

3. Case of ADVERSE SELECTION (asymmetric information)

We also need to add incentive restrictions (when income is taxed) to the objective function and budget constraint.

Restrictions are upward (local and global)

\[ \theta_G l_G - t_G - \psi(l_G) \geq \theta_M l_M - t_M - \psi\left(\frac{\theta_M l_M}{\theta_G}\right) \quad (10) \]

According to this restriction, the G (best) productivity type \( \theta_G \) chooses the contract \((q_G, l_G)\) to the detriment of the contract \((q_M, l_M)\).

It would produce \( q_M \) and would pay \( t_M \) with the effort \( \left(\frac{\theta_M l_M}{\theta_G}\right) \) given the productivity \( \theta_G \).

\[ \theta_M l_M - t_M - \psi(l_M) \geq \theta_M l_M - t_M - \psi\left(\frac{\theta_M l_M}{\theta_G}\right) \quad (20) \]

\[ \theta_G l_G - t_G - \psi(l_G) \geq \theta_G l_G - t_G - \psi\left(\frac{\theta_G l_G}{\theta_G}\right) \quad (30) \]

with an interpretation similar to the one above.
Restrictions (1°) and (2°) are local upward restrictions and (3°) is a global ascending restriction.

The following restrictions are descending (local and global) and ensure that worse-placed agents (with lower productivity) prefer contracts for them than other contracts.

\[
\begin{align*}
\theta_L l_L - t_L - \psi(l_L) & \geq \theta_M l_M - t_M - \psi(l_M) \\
\theta_M l_M - t_M - \psi(l_M) & \geq \theta_G l_G - t_G - \psi(l_G) \\
\theta_L l_L - t_L - \psi(l_L) & \geq \theta_G l_G - t_G - \psi(l_G)
\end{align*}
\]

Restrictions (4°) and (5°) are downstream local restrictions and (6°) is a downward global restriction.

Suppose that \(\theta_M = \theta_G = \theta > 1\) and note \(f(e) = \psi(e) - f(\theta_e)\).

**Proposition 2.** The function \(f(\cdot)\) is negative and decreasing.

**Demonstration.** Obviously \(e \leq \theta_e\) and how \(\psi'(\cdot) > 0\), it follows that \(\psi(e) \leq \psi(\theta_e)\), where from \(f(e) \leq 0\).

By deriving the function \(f(\cdot)\) we obtain:

\[
\Gamma(e) = \psi'(e) - \theta \psi'(\theta_e) < \psi'(\theta_e) < 0 \quad \text{because} \quad \psi'(\cdot) > 0.
\]

**Proposition 3.** The model implementation feasibility is \(\theta_G l_G \geq \theta_M l_M \geq \theta_L l_L\).

**Demonstration.** The implementation condition is equivalent to the assertion that the set of admissible solutions is unclear.

We collect member with inequalities (1°) and (5°) and we obtain:

\[
\begin{align*}
-\psi(l_G) - \psi(l_M) & \geq - \psi(l_G) - \psi(l_M) & \text{or} \\
\psi(l_M) - \psi(l_G) + \psi(l_M) & \leq 0 & \text{The above inequality becomes:} \\
\frac{l_M}{\theta} & \leq 0 \quad \text{and from (Pₚ) results} \\
l_G \geq \theta \cdot l_G , \text{ie } l_G \cdot \theta > l_M \text{ or } l_G \cdot \theta_M \geq l_M \text{ ie } l_G \theta_M \geq l_M \theta_M \\
\text{Analogously, summing (2°) and (4°) we obtain:} \\
-\psi(l_M) - \psi(\theta l_M) & \geq - \psi(l_M) - \psi(l_M) \\
or by regrouping the terms we obtain:
\end{align*}
\]

\[
\begin{align*}
-\psi(l_M) - \psi(\theta l_M) + \psi(\theta l_M) & \leq 0 \quad \text{ie } f(l_M) - f(\theta l_M) \leq 0
\end{align*}
\]

Again, taking into account sentence 1, we obtain: \(l_M \geq \frac{l_M}{\theta} \text{ or } l_M \theta \geq l_L\), where from \(l_M \theta_M \geq l_L \theta_L\) and the sentence is demonstrated.
Proposition 4. If the local upstream restrictions are checked, then the global ascending restriction is checked.

Demonstration. We add the restrictions (1º) and (2º) and we obtain:

\[ \theta_G l_G - t_G - \psi(l_G) - \psi(l_M) \geq \theta_l l_L - t_L - \psi \left( \frac{\theta_M}{\theta_G} l_M \right) \]

\[ \theta_G l_G - t_G - \psi(l_G) \geq \theta_l l_L - t_L - \psi \left( \frac{\theta_M}{\theta_G} l_M \right) + \psi \left( \frac{\theta_G}{\theta_M} l_M \right) - \psi \left( \frac{\theta_G}{\theta_M} l_M \right) \]

which is checked if:

\[ \theta_G l_G - t_G - \psi(l_G) \geq \theta_l l_L - t_L - \psi \left( \frac{\theta_G}{\theta_M} l_M \right) \]

Proposition 5. If the downstream local restrictions are checked, then the downstream global restriction is checked.

Demonstration. We collect the relations (4º) and (5º) and we obtain:

\[ \theta_G l_G - t_G - \psi(l_G) \geq \theta l l_L - t_L - \psi \left( \frac{\theta_G}{\theta_M} l_M \right) \]

\[ \theta_G l_G - t_G - \psi(l_G) \geq \theta l l_L - t_L - \psi \left( \frac{\theta_G}{\theta_M} l_M \right) + \psi \left( \frac{\theta_G}{\theta_M} l_M \right) - \psi \left( \frac{\theta_G}{\theta_M} l_M \right) \]

which is checked if:

\[ \theta_G l_G - t_G \geq \theta l l_L - t_L - \psi \left( \frac{\theta_G}{\theta_M} l_M \right) \]

Since \( \theta_G l_G = \frac{\theta_G}{\theta_M} l_M \geq l_M \) according to the properties of the function \( f \), ie

\[ -f(\theta_G) + f(l_M) \geq 0 \]

By a convenient notation, namely:

\[ U_G \equiv \theta_G l_G - \psi(l_G) \]
\[ U_M \equiv \theta_M l_M - \psi(l_M) \]
\[ U_L \equiv \theta l l_L - t_L - \psi(l_L) \]

the incentive restrictions become:

\[ U_G \geq U_M - f \left( \frac{l_M}{\theta_G} \right) \]
\[ U_M \geq U_L - f \left( \frac{l_L}{\theta_M} \right) \]
\[ U_L \geq U_M + f \left( \frac{l_M}{\theta_M} \right) \]
\[ U_M \geq U_G + f \left( \frac{l_G}{\theta_G} \right) \]

We remove the variables \( t_L, l_M, t_G \) from the above transformations and the problem becomes successively:

First, the objective function will take the following form:

\[ \text{Max } U_G, U_M, U_L, l_M, l_G, l_L \left[ \Pi_L U(L) + \Pi_M U(M) + \Pi_G U(G) \right] \]

(maximizing expected utility)

Government Restriction (Budget Restriction) takes the following form:

\[ \Pi_G \left[ \theta_G l_G - \psi(l_G) + \Pi_M \left[ \theta_M l_M - \psi(l_M) \right] + \Pi_G \left[ \theta_G l_G - \psi(l_G) \right] - \left[ \Pi_G U(G) \right] \right] \]

\[ \Pi_M U_M + \Pi_G U_G \geq \underline{u}, \] where \( \underline{u} \) has the meaning of an expected minimum level of government revenue.
We will ignore the last two incitement restrictions (local descending restrictions) and finally we will show that the solution thus obtained also checks these restrictions.

If worse placed agents accept the contract, the better placed the contract accepts.

The initial optimization problem (P1) simplifies (fewer restrictions) and becomes:

\[
\text{Max } U_G, U_M, U_L, l_G, l_M, l_L \quad \text{s.t.} \quad (P2) \quad \Pi_L U(u_L) + \Pi_M U(u_M) + \Pi_G U(u_G) \\
U_G \geq U_M - f\left(\frac{L}{\theta_G}\right) \\
U_M \geq U_L - f\left(\frac{L}{\theta_M}\right) \\
U_L \geq 0, U_M \geq 0, U_G \geq 0, \\
l_L \geq 0, l_M \geq 0, l_G \geq 0
\]

We associate the KUHN-TUCKER multipliers \(\lambda, \mu, \rho\) of the three constraints and construct Lagrange’s function as follows:

\[
L(u, u_M, u_L, l_G, l_M, l_L; \lambda, \mu, \delta) = \Pi_L U(u_L) + \Pi_M U(u_M) + \Pi_G U(u_G) + \lambda [U_G - \Psi(l_G) - U_L] + \\
\mu [U_M - \Psi(l_M) - U_L] + \rho [U_L - \Psi(l_L) - U_G] + \\
\delta \left[\lambda U_G - \mu U_M + f\left(\frac{L}{\theta_G}\right) + \delta \left[U_M - U_L + f\left(\frac{l_M}{\theta_M}\right)\right]\right]
\]

The KUHN-TUCKER conditions of first order (necessary and sufficient according to the properties of utility functions) and \(f\) are written:

\[
\frac{\partial L}{\partial u_L} = \Pi_L u'(u_L) + \lambda \Pi_L - \delta = 0 \quad (11^a) \\
\frac{\partial L}{\partial u_M} = \Pi_M u'(u_M) - \lambda \Pi_M - \mu + \delta = 0 \quad (12^a) \\
\frac{\partial L}{\partial u_G} = \Pi_G u'(u_G) - \lambda \Pi_G + \mu - \delta = 0 \quad (13^a) \\
\frac{\partial L}{\partial l_L} = \lambda \Pi_L[l_G - \Psi'(l_G)] + \delta \frac{1}{\theta_L} f'(\frac{l_L}{\theta_L}) = 0 \quad (14^a) \\
\frac{\partial L}{\partial l_M} = \lambda \Pi_M[l_M - \Psi'(l_M)] + \mu \frac{1}{\theta_M} f'(\frac{l_M}{\theta_M}) = 0 \quad (15^a) \\
\frac{\partial L}{\partial l_G} = \lambda \Pi_G[l_G - \Psi'(l_G)] = 0 \quad (16^a)
\]

The final result is included in the following theorem:

**Theorem 1.** All KUHN- multipliers are strictly positive and the level of optimal effort is given by the equality:

\[
\theta_G = \Psi'(l_G) \\
u_G = u_M - f\left(\frac{l_M}{\theta_M}\right) \\
u_M = u_L - f\left(\frac{l_L}{\theta_L}\right)
\]
Conclusion

Determining the optimal levels $\tilde{e}_G, \tilde{e}_M, \tilde{e}_L$ and the variables $\tilde{u}_G, \tilde{u}_M, \tilde{u}_L$ allows to write optimal taxes according to the formula:

$\tilde{t}_G = \theta_G \tilde{e}_G - \tilde{u}_G - \Psi(\tilde{e}_G)$

$\tilde{t}_M = \theta_M \tilde{e}_M - \tilde{u}_M - \Psi(\tilde{e}_M)$

$\tilde{t}_L = \theta_L \tilde{e}_L - \tilde{u}_L - \Psi(\tilde{e}_L)$

In this paper (study), the authors have departed from the practical necessity of optimal taxation of effort. In current activity, optimal effort taxation involves several states. From the study, analysis and demonstration, it is clear that there is a well-articulated mathematical model presented in the article’s content which best describes deterministically the way in which a balance between allocative efficiency and distribution is ensured. The article did not intend to do a practical study as, first of all, the theoretical relationship shown in the article must first be demonstrated and then used, applied in concrete cases. For a country’s economy, such a mathematical model is useful because it refers to a number of issues that are important in terms of how best-effort taxation can be applied. It is easy to apply the expected mathematical model to a concerted situation based on the data that is encountered in an economy.

References

4. Anghelache Constantin and Madalina Gabriela Anghel, 2016, Bazele statisticii economice. Concepte teoretice și studii de caz, Editura Economică, București
5. Anghelache Constantin and Madalina Gabriela Anghel, 2016, Econometrie generală. Concepte, teorie și studii de caz, Editura Artifex, București
6. Anghelache Constantin and Madalina Gabriela Anghel, 2014, Modelare economică. Concepte, teorie și studii de caz, Editura Economică, București
7. Anghelache Constantin, 2008, Tratat de statistică teoretică și economică, Editura Economică, București
8. Arsene Catâlin and Marin Dumitru, 2007, Modelarea deciziei pe piețele financiare, Editura ASE, București
37. Marston, R.C., 1984, "Real Wages and the Term of Trade: alternative Indexation Rules for an Open Economy", Journal of Money, Credit, and Banking, 16, 285-301
39. Muthuraman Kumar and Sunil Kumar, 2006, "Multidimensional portfolio optimization with proportional transaction costs", Mathematical Finance, 16(2), 301–335
46. Wilmott Paul, 2001, Paul Wilmott Introduces Quantitative Finance, John Wiley & Sons, Chichester