New Methods for Exploring the Implications in the Evolution and Patterns of Romanian GDP

Vasile Paul Bresfelean and Mihaela Göndör

***Abstract*—** In the present article we extend our latest studies where we to explored implications in the progression and patterns of Romanian GDP. We employ data mining techniques, such as model trees and linear regression, over data extracted from statistics, socio-economic indicators tables and reports from the Eurostat, Romanian National Bank (NBR) and Romanian National Institute of Statistics, over the 2001-2010 period. We continue to present our investigation of GDP patterns spaced out from classic ways, through taxes on production, imports, on income and on wealth, employers' social contributions from various fields of economy (agriculture, commerce, constructions, industry, services, financial, banking, etc.) as well as salaries level, based on the data mining tasks.

***Keywords*—** GDP, Tax changes, Data mining, Model trees, Linear regression, M5.

I. Introduction

A. Objectives of our study

TAX changes used for influencing the Gross Domestic Product (GDP) have been a major public policy issue in recent years, moreover in crisis time. The major tax changes in Romania between 2009 and 2010 were passed amid firestorms of debate about their likely effects. Several policymakers [24] asserted that these changes would both reduce government deficit stimulating the economy in the short run and increase GDP in the long run. Others [24] argued that it would reduce the consumption and lower investor confidence, and thereby reduce GDP in both the short run and the long run.

As is presented in the literature review in this field, not only the policymaker's opinions differ so wide, but also the researchers'. Such a varying view of the effects of tax changes on the GDP reflects the fact that measuring fiscal policy's effects is very difficult [24], even if the tax changes main reason is to reduce budget deficit.

Rooted in these facts from the economic and research literature, we extend in this article our recent research [30] on the GDP patterns and progress in our country and its influence via salaries level, employers' social contributions from different fields of economy (agriculture, industry, commerce, construction, services, financial, banking, etc.) as well as taxes on production and imports.

Manuscript received September 25, 2011. A part of the research from this paper was supported by TE_316 CNCS research grant..

V.P. Bresfelean is with the Faculty of Economics and Business Administration, Babes-Bolyai University, Cluj-Napoca, Romania (e-mail: paul.bresfelean@econ.ubbcluj.ro).

M. Göndör is with the Faculty of Economics, Law and Administrative Sciences, at "Petru Maior" University Tîrgu Mures, Romania (email: mihaelagondor@yahoo.com).

B. Methodology of the study

We make use of data mining prediction methods, such as model trees algorithms and linear regression. The data employed consisted in 2001-2010 shareware statistics, socio-economic indicators tables and reports from the Romanian and European institutes.

C. Paper structure

The paper is divided in six sections. The first section is the introductive one, embodying the study objectives, methodology, the paper structure and a brief review of the literature in the research field. In the second section we present some theoretical grounds a review of existing arguments pro tax harmonization. The third section is dedicated to data mining studies in the field and its broad categories which are classification, association, estimation, clustering, description and prediction, highlighting the categories used in our present study. The practical part of this study begins with the fourth section which contains data used in the research being made of statistics, socio-economic indicators, tables and reports from the Romanian National Institute of Statistics, Eurostat and Romanian National Bank over the 2001-2010 periods. The fifth section embody the main part of our study, the GDP evolution based on model trees and linear regression and in the final section we present some conclusions and the directions of our future research in the field.

D. State of the art in the field

In the review of the macroeconomic literature we could find a several studies on the topic of GDP and its influence factors.

Regarding the effects of the tax structure on the economy's overall growth rate there has been much less work, most of researchers arguing low levels of influence. Moreover, the researchers who have argued significant levels of influence contradictorily concluded regarding the response of GDP to tax policy. In accordance with Buti, M. and P. van den Noord [22] the effects of changes in the level of taxes on real GDP are reasonable, the response being linear. At the same time, several studies suggested that the response of GDP to fiscal policy may be non-linear. [23]

According to some researchers the estimated elasticity of unemployment with respect to labor taxes ranges from zero. [28], [29]. On the contrary, some researchers argue that taxes can have a significant effect on the labor intensity and thus on its efficiency and productivity [26], [27]. Anyway, there exists too little empirical evidence or studies to support any of these view.

Some papers argue that the impact of labor taxes on unemployment is symmetric [26], [27]. On the contrary, some papers argue that it is asymmetric [25].

From the literature review we can conclude that many factors contribute to improve the labor productivity and based on this to influence the macroeconomic output, tax changes can play a role, but isolating the impact of tax changes on productivity and moreover on GDP is extremely difficult.

II. Theoretical Grounds

GDP represents the market value of all final goods and services generated inside a geographical entity in a certain period. Why it is called like this? Because [9]:

• Gross - the depreciation of the production and services capital is not deducted from the total GDP value;

• Domestic - it relates only to activities within a domestic economy regardless of nationality;

• Product – it refers to the economy's output, to the economic activities, meaning the goods and services that are being produced.

Although the GDP is at the intersection of demand, production and income, it also has some deficiency points: [17]

 ignores the role of non-market activities without a well established market price;

 consequently underestimates the actual output;

 does not capture income disparities, which are an important indicator of the citizens' well-being (owing to the fact that GDP is an aggregate measure).

 fails to recognize harmful externalities caused by ecological degradation and resource exhaustion.

Another argument is that GDP can be enhanced merely by increasing the price that will be paid (e.g. by raising quality) for its goods and services, or by increasing the quantity of goods or services that the economy generates [9].

The basic idea behind the relationship between tax changes and tax macroeconomic output is that changes in tax rates have two effects on revenues: the arithmetic effect and the economic effect. The arithmetic effect is simply that if tax rates are lowered, tax revenues will be lowered by the amount of the decrease in the rate. The reverse is true for an increase in tax rates. The economic effect, however, recognizes the positive impact that lower tax rates have on labor and thus on macroeconomic output, by providing incentives for work and business, generating the increase of the tax base. Raising tax rates has the opposite economic effect by discouraging the participation in the taxed activities. The arithmetic effect always works in the opposite direction from the economic effect.

Because the effects of tax changes on the output (GDP) are often correlated with other factors, its disentangling is inherently difficult.

III. Data mining studies in the field

While data mining emerged during the late 1980s and developed during 1990s [6], in the last years it grew to an extensive importance and became a multidisciplinary field processing vast amount of operational data stored by many organizations. Data mining can be defined [21] as the extractions of implied, formerly indefinite, and potentially valuable information from data, such as strong patterns that will possible generalize to make precise predictions on future data.

In one of several classifications, data mining tasks were divided into broad categories [6]: classification, association, estimation, clustering, description, prediction. In the following subsections we present each one of them, highlighting the categories used in our studies: [30]

A. Classification

The classification methods are projected [14] for learning dissimilar functions which map each item of the chosen data into a predefined set of classes. They can routinely forecast the class of other unclassified data of the training set, if given the set of predefined classes, attributes, and the training set. The representation of the uncertain value by its expectation value and treated as a certain data, is a perceptive way of managing uncertainty in classification [7]; then the classification algorithms can be unswervingly utilized.

B. Association

Association rules are in a way comparable to the classification rules, with some exceptions: [21]

• have the means for attribute prediction, not just the class,

• are able to predict combinations of attributes.

• are not intended to be used together as a set, as classification rules are.

Different association rules normally forecast different things [21], and also convey different regularities that bring about the dataset.

C. Estimation

Estimation methods [8] originated from statistics, namely estimators, of the unknown model quantities that can produce reliable estimates when utilized over sample data. They can be divided into: [8]

• Point estimate methods – the quantity is expected with a precise value,

• Confidence interval methods – the quantity is probable to have a high frequency of lying within a region, typically an interval of the real line.

D. Clustering

Clustering is regularly seen as part of unsupervised learning, and can often be applied when there is no reachable information regarding the bond of data items with predefined classes [10]. It was described [18] as the process of organizing objects in a database into groups or clusters, that objects within the same cluster have a high degree of similarity, while objects belonging to different clusters have a high degree of dissimilarity.

E. Description

Description can be considered another type of data mining task. Occasionally, the data mining issue is to basically describe what is happening in a complex database [2]. It can include knowing the people the products or applicable processes and which constitute the database [3].

F. Prediction

The prediction [5] is intended to be the best possible estimate of the actual value given the data available. It was generally synonymous with regression of some form in statistics, the essential idea [4] being that a model is created that maps values from predictors implying that the lowest error takes place in making a prediction.

a) Linear regression

Linear Regression is one of the simplest form of regression which includes one predictor and a prediction, and their relationship between that can be mapped on a two dimensional space [4]:

- Y axis - the records plotted for the prediction values,
- X axis - the records for the predictor values.

Linear regression should be taken into consideration [21] when both the outcome, or class, and the attributes are numeric.

In effect, regression models all fit the same general pattern [1]:

- There are a number of independent variables, which, when taken together, produce a result - a dependent variable.

- The regression model is then used to predict the result of an unknown dependent variable, given the values of the independent variables.

The principal idea [21] behind the linear regression is to outline the class as a linear combination of attributes (a_i), with prearranged weights (w_i):

$$x = w_0 + w_1 a_1 + w_2 a_2 + \ldots + w_k a_k \qquad (1)$$

In most recent studies, researchers [20] proposed new algorithms for online linear regression whose efficiency guarantee to satisfy the requirements of the KWIK (Knows What It Knows) framework. They improved on the complexity bounds of the current state-of-the-art procedure in this setting and also explored several applications of the algorithm for learning compact reinforcement-learning representations.

b) Model trees

Trees used for numeric prediction resemble to decision trees [21], with some exceptions, and can be classified into:

- regression trees – which store at each leaf a class value representing the average value of instances that reach the leaf;
- model trees – which store at each leaf a linear regression model that predicts the class value of instances that reach the leaf. They have the ability to add background knowledge to the model [11], an important evolution from the time series and other classic statistical methods.

There were numerous applications of the model trees in practice found in the research literature [30]. For example [11], the continuous numeric prediction techniques which build model trees and apply linear regression at the terminal nodes, were used to characterize resource consumption in a computers system. Models were built using production data from financial institutions in collaboration with domain experts.

Potts [15] developed incremental methods for growing linear model trees: a node splitting rule, a stopping rule and a method for pruning. His algorithm was empirically tested in three domains and compared with existing incremental, instance-based and batch methods.

Loh et al. [13] considered the extrapolation and interpolation errors of linear model tree when used for prediction. They proposed several ways to restrict the size of the errors and demonstrated that the solutions were effective in reducing the average mean squared prediction error.

Other researchers chose the domain of microphone forensics [12] and illustrated that media forensics could profit from the arrangement between statistical pattern recognition (using supervised classification) and unweighted information fusion. Their practical results showed an accuracy increased to 100%, by using a carefully selected fusion strategy and multiclass classifiers: model trees and linear logistic regression models.

Further studies [19] presented techniques to efficiently mine common embedded and induced subtrees from a database of XML documents. They proposed an efficient approach to tackle the complexity of mining embedded subtrees by utilizing a novel embedding list representation, and introducing the maximum level of embedding constraints [19].

IV. EMPIRICAL EVIDENCE

Table 1. Real GDP in Romania during 2001-2010 (billion Euros).

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
GDP	40.2	44.8	48.4	50.3	58.9	78.3	121.2	136.8	115.9	122.0

Source: NBR Annual Report [32]

Table 2. Total receipts from taxes and social contributions in Romania during 2001-2010 (percentage of GDP)

2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
28.9	28.5	28.1	27.7	28.5	29.2	29.8	28.8	27.7	28.1

Source: Eurostat 2011 [36]

Table 3. Taxes on production and imports in Romania during 2001-2010 (percentage of GDP)

2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
11.3	11.6	12.3	11.7	12.9	12.8	12.3	11.7	10.7	12.1

Source: Eurostat 2011 [36]

According Eurostat Definition [36] “Taxes on production and imports (ESA95 code D.2) consist of compulsory, unrequited payments, in cash or in kind which are levied by general government, or by EU institutions, in respect of the production and importation of goods and services, the employment of labor, the ownership or use of land, buildings or other assets used in production. In ESA95, taxes on production and imports comprise taxes on products and other taxes on production”.

Table 4. Taxes on income, wealth, etc. in Romania during 2001-2010 (percentage of GDP).

2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
6.4	5.8	6.0	6.4	5.3	6.0	6.7	6.7	6.5	6.2

Source: Eurostat 2011

According Eurostat Definition [36] “current taxes on income, wealth, etc. (ESA95 code D.5) cover all compulsory, unrequited payments, in cash or in kind, levied periodically by general government and by the rest of the world on the income and wealth of institutional units, and some periodic taxes which are assessed neither on the income nor the wealth. In ESA95, current taxes on income, wealth, etc. are divided into taxes on income and other current taxes.”

Based on above presented empirical evidence, we intend to examine the effects of taxes on production, imports, on income and on wealth, employers’ social contributions from diverse fields of economy (agriculture, industry, commerce, construction, services, financial, banking, etc.) as well as salaries level on GDP during the considered time.

V. GDP Evolution Based on Model Trees and Linear Regression

In our continuing research we sought GDP patterns and future evolution in our country, and its influence via salaries level, employers’ social contributions from different fields of economy: agriculture, industry, commerce, construction, services, financial, banking etc., as well as taxes on production and imports. The data employed consisted in shareware statistics, socio-economic indicators tables and reports from the Romanian and European institutes over the 2001-2010 period.

Let’s consider the minimal function of the real GDP growth:

$$\Delta Y_t = \alpha + \beta \Delta T_t + \varepsilon_t \qquad (2)$$

$$\varepsilon_t = \sum_{i=1}^{j} \varepsilon_t^i \qquad (3)$$

Where:

Y_t - the logarithm of real GDP,

ΔT - tax changes (discrete events)

ε_t - disparate factors which affect real GDP besides tax changes (public expenditures, monetary policy, crisis, disasters, etc.)

The (x) minimal function ignores the dynamics (although tax changes do not affect GDP only in the current quarter).

Let’s consider the determinants of tax changes:

$$\Delta T_t = \sum_{i=1}^{j} k_t^i \varepsilon_t^i + \sum_{m=1}^{n} \delta_t^m \qquad (4)$$

Where:

δ - Additional influences on tax policy.

Additional influences on tax policy means tax changes generated by political reasons or policymakers subjective perceptions.

From the literature in the research field we have systemized four different types of tax changes motivation: a countercyclical change, a tax change relative to a spending change, a change for balancing the budget (for a deficit reduction), and a change to encourage long-run growth. [24]

In order to use the minimal function of the real GDP growth we have to discuss some of the national tax system properties. Have there been changes over 2001-2010 periods in the number, size, and specific motivation of tax changes? We have to examine the trends in the number, size, and motivation for tax changes for 2001-2010 and finally to examine how our series on tax changes compares with the change in real GDP over the same periods.

Once established the conceptual framework, we need to identify legislated tax changes, and, if possible, to identify the motivation for each change. Finally, based on these identifications, we need to determine the impact of the tax changes on GDP.

A. M5 generated tree

In the first stages, we utilized the M5 algorithm [30], due to its functionality: [16]

• it generated trees whose leaves were associated to multivariate linear models;

• the nodes of the tree were selected over the attribute that maximized the anticipated error reduction as a function of the standard deviation of output parameter.

The chosen data mining software was Weka, who’s M5 algorithm called M5P built a decision tree dividing the attribute space in an orthohedric clusters, with the border parallel to the axis [16]. The model trees were simply transformed into rules, every branch having a certain condition [16]:

`attribute ≤ value or attribute > value` (5)

e.g. wage_costs_industry<=14039.9
or wage_costs_industry >14039.9

The building of the tree that was used to predict possible GDP future (Appendix II) can be described as follows [21]:

- tree was followed down to a leaf in the regular way, using the instance's attribute values to make course-plotting choices at all nodes;
- the leaf enclosed a linear model founded on some of the attribute values;
- model evaluation for the test instance to yield a raw predicted value.

During the mining process was also applied an appropriate smoothing calculation, with the intend of reducing the discontinuities that took place between adjacent linear models [21]:

$$p' = \frac{np + kq}{n + k} \qquad (6)$$

Where:
p′ - prediction passed up to the next higher node,
p - prediction passed to this node from below,
q - value predicted by the model at this node,
n - number of training instances reaching the node below,
k - smoothing constant.

The model M5P resulted tree [30] (Fig. 1 and Appendix II), visibly reveals that for the value of the GDP the wage_costs_agriculture factor (total wage costs in agriculture per quarter in Romania, in millions RON, between 2000-2011) is considered an important indicator that represents a significant pattern of the national activities. Subsequent to the root, a larger value of the wage_costs_industry (total wage costs in industry) is used to develop the Linear Models 2, and a lower value for the Linear Model 1. In the same manner, the variable wage_costs_constructions (total wage costs in constructions) is used to develop the Linear Models 3 and 4. Based on the level of each model's indicator, there can be observed that models 2 and 3 are more representatives for the value of the GDP.

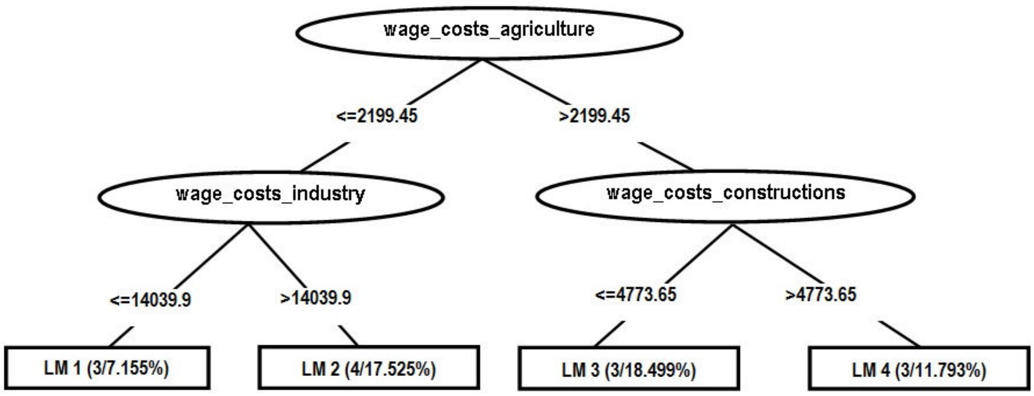

Fig.1 M5P generated 1st model tree [30]

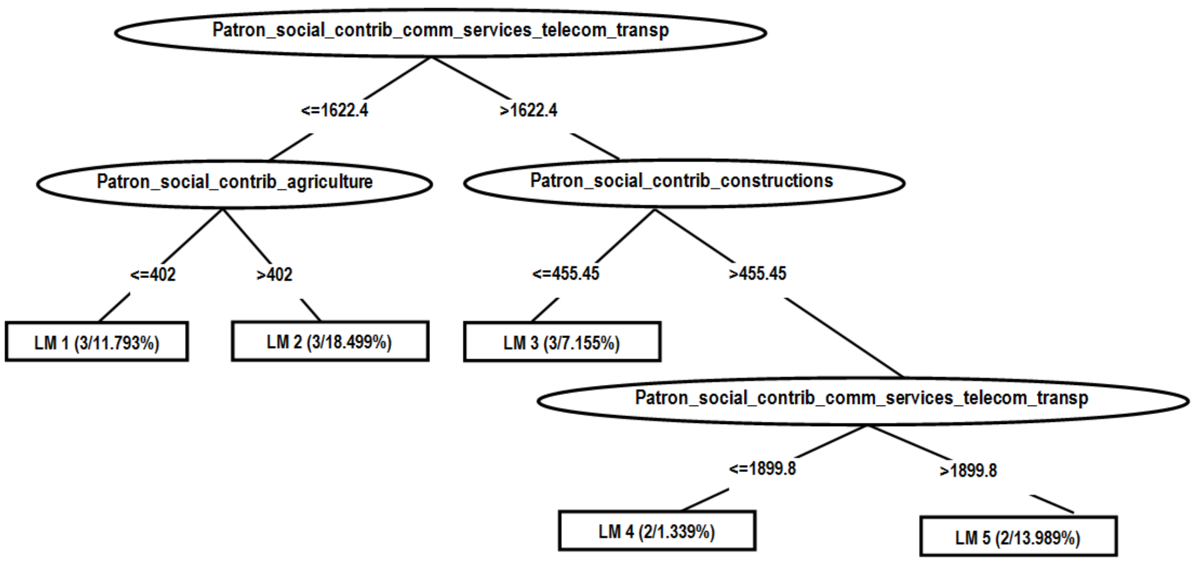

Fig.2 M5P generated 2nd model tree

The second model (Fig.2 and Appendix IV), reveals that for the value of the GDP the patron_social_contrib_comm_services_telecom_transp factor (employers' social contributions per quarter in Romania, millions RON, between 2000-2011 in commerce, services, telecommunications and transport) may be the root indicator of the national activities. Subsequent to the root, a larger value of the patron_social_contrib_agriculture (employers' social contributions per quarter in Romanian agriculture) is used to develop the Linear Models 2, and a lower value for the Linear Model 1. In the same manner, a lower value of patron_social_contrib_constructions (employers' social contributions per quarter in Romanian constructions) is used to develop the Linear Models 3, and again, patron_social_contrib_comm_services_telecom_transp factor appears as an ending node, determining Lineal Models 4 and 5. Based on the level of each model's indicator, there can be observed that models 2 and 5 are more representatives for the value of the GDP, followed by LM1.

B. Linear regression model

Data mining isn't just about outputting a single number [1], but about identifying patterns and rules and rather to create a model that allow to detect patterns, predict output, and come up with conclusions backed by the data. In applying the Linear regression in Weka we utilized the M5's method which [21]:

- stepped throughout the attributes,
- removed the attribute with the minimum standardized coefficient until no improvement was observed in the estimate of the error, given by the Akaike information criterion.

The linear regression equation results by Weka, after running linear regression over the 1st dataset based mainly on employers' social contributions (Appendix I), were as follows [30]:

```
GDP =
    -17.7142 * patron_social_contrib_agriculture +
     6.8501 * patron_social_contrib_industry +
     8.8856 * patron_social_contrib_other_service +
     1.6562 * Gross_operating_surplus +
     1.524  * production&import_taxes +
  -34801.6475
```

The interpretation of the patterns and conclusions that the model tells us: [30]

• patron_social_contrib_other_service (services employers' social contributions per quarter in Romania, millions RON, between 2001-2010) greatly influences the GDP, followed in a great measure by the patron_social_contrib_industry (industry employers' social contributions).

• gross_operating_surplus and also the production&import_taxes influence, but in a smaller manner the GDP evolution.

• In contrast, the patron_social_contrib_ agriculture (agriculture employers' social contributions) pessimistically influences the GDP value, due to the negative coefficient in front of this variable. This is a major problem in the Romania's real market economy for the reason that a large portion of the agricultural work force is employed in the black market, thus avoiding a correct evidence, and also a major part of social contributions and taxes.

The linear regression equation results by Weka, after running linear regression over the 2nd dataset based mainly on total wage costs on each field (Appendix III), were as follows:

```
GDP =
     6.3325 * wage_costs_agriculture +
     4.4261 * wage_costs_constructions +
     3.0635 * wage_costs_comm_services_telecom_trsp +
     1.3504 * wage_costs_other_service +
     1.7974 * production&import_taxes +
   13626.095
```

The interpretation of the patterns and conclusions that this model tells us:

• wage_costs_agriculture (total costs with salaries in Romanian agriculture per quarter, millions RON, between 2001-2010) greatly influences the GDP, followed closely by the wage_costs_constructions (total costs with salaries in Romanian constructions) and wage_costs_comm_services_ telecom_trsp (total costs with salaries in Romanian commerce, services, telecommunication and transport)

• wage_costs_other_service and also the production& import_taxes influence, but in a smaller manner, the GDP evolution.

VI. CONCLUSIONS

In this article we presented a part of the study on the Romanian economy, namely examining the patterns and evolution of the Gross Domestic Product. For this we turned to a number of national statistical data and reports collected during 2001-2010 by the National Institute of Statistics, Eurostat and Romanian National Bank. The data used for the research knew a prior processing, so as to suit the best method of analysis. With the intention of finding new patters in the Romanian GDP, we used some representative processes for data mining prediction methods, such as model trees and linear regression.

In the first part of the article we reviewed some of the issues and debates on the calculation of GDP and its prospects, then, in the second section we highlighted a classification of data mining methods, focusing on those used in research. In the last part we described a part of our practical contribution to the conducted study, offering two models that were generated and also their interpretation. We found that during the investigated period, services employers' social contributions per quarter in Romania greatly influenced the GDP, followed in a great measure by the industry employers' social contributions; Gross_operating_surplus and also the production&import_taxes influence, but in a smaller manner the GDP evolution. In contrast, the agriculture employers' social contributions pessimistically influences the GDP value, due to the negative coefficient in front of this variable. During the examined period Romanian policymakers often mention that the motivation for tax changes is to reduce the budget deficit. As a result we can conclude that in Romania, the tax and social contribution changes are not focused for stimulate economic growth.

Moreover, we can conclude that the increases of the taxes on production and on income during the year 2009, determined the reduction of GDP, even if it had led to a reduction of the budget deficit.

In the continuation of our studies we desire to expand the research in banking and capital markets through other statistical and data mining methods.

APPENDIX

Appendix I .

===== Run information =====
Scheme:weka.classifiers.functions.LinearRegression-S0-R 1.0E
Relation: tabel GDP3_MIC2
=====Classifier model (full training set) =====
Linear Regression Model

GDP =
-17.7142 * patron_social_contrib_agriculture +
6.8501 * patron_social_contrib_industry +
8.8856 * patron_social_contrib_other_service +
1.6562 * Gross_operating_surplus +
1.524 * production&import_taxes +
-34801.6475

Time taken to build model: 0.01seconds

===== Evaluation on training set =====
=== Summary ===

Correlation coefficient	0.9972
Mean absolute error	1547.1756
Root mean squared error	1773.3887
Relative absolute error	7.3114 %
Root relative squared error	7.5345 %
Total Number of Instances	13

Appendix II

===== Run information =====
Scheme:weka.classifiers.trees.M5P -U -R -M 4.0
Relation: tabel GDP3_MIC
Test mode:evaluate on training data

===== Classifier model (full training set) =====
M5 pruned regression tree:

wage_costs_agriculture <= 2199.45 :
| wage_costs_industry <= 14039.9 : LM1 (3/7.155%)
| wage_costs_industry > 14039.9 : LM2 (4/17.525%)
wage_costs_agriculture > 2199.45 :
| wage_costs_constructions <= 4773.65 : LM3 (3/18.499%)
| wage_costs_constructions > 4773.65 : LM4 (3/11.793%)

LM num: 1
GDP =
+ 96048.0667
LM num: 2
GDP =
+ 113413.675
LM num: 3
GDP =
+ 137959.6333
LM num: 4
GDP =
+ 159131.3333
Number of Rules : 4
Time taken to build model: 0.02seconds

=== Evaluation on training set ===
=== Summary ===

Correlation coefficient	0.9891
Mean absolute error	2889.6744
Root mean squared error	3470.2179
Relative absolute error	13.6556 %
Root relative squared error	14.7437 %
Total Number of Instances	13

Appendix III

=== Run information ===

Scheme:weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8
Relation: tabel GDP3_MIC-weka.filters.unsupervised.attribute.Remove-R9-weka.filters.unsupervised.attribute.Remove-R7
Test mode:evaluate on training data

=== Classifier model (full training set) ===

Linear Regression Model

GDP =
6.3325 * wage_costs_agriculture +
4.4261 * wage_costs_constructions +
3.0635 * wage_costs_comm_services_telecom_transp +
1.3504 * wage_costs_other_service +
1.7974 * production&import_taxes +

13626.095

Time taken to build model: 0seconds
=== Evaluation on training set ===
=== Summary ===

Correlation coefficient	0.994
Mean absolute error	2008.1799
Root mean squared error	2566.8599
Relative absolute error	9.4899 %
Root relative squared error	10.9057 %
Total Number of Instances	13

Appendix IV
=== Run information ===

Scheme:weka.classifiers.trees.M5P -N -R -M 4.0
Relation: tabel GDP3_MIC2
Test mode:evaluate on training data

=== Classifier model (full training set) ===

M5 unpruned regression tree:
(using smoothed linear models)

```
patron_social_contrib_comm_services_telecom_transp <= 1622.4 :
|   patron_social_contrib_agriculture <= 402 : LM1 (3/11.793%)
|   patron_social_contrib_agriculture >  402 : LM2 (3/18.499%)
patron_social_contrib_comm_services_telecom_transp >  1622.4 :
|   patron_social_contrib_constructions <= 455.45 : LM3 (3/7.155%)
|   patron_social_contrib_constructions >  455.45 :
|   |   patron_social_contrib_comm_services_telecom_transp <= 1899.8 : LM4 (2/1.339%)
|   |   patron_social_contrib_comm_services_telecom_transp >  1899.8 : LM5 (2/13.989%)
```

LM num: 1
GDP =
+ 132674.8745

LM num: 2
GDP =
+ 131666.6984

LM num: 3
GDP =
+ 118842.5201

LM num: 4
GDP =
+ 119894.0605

LM num: 5
GDP =
+ 119840.5094

Number of Rules : 5

Time taken to build model: 0.01seconds
=== Evaluation on training set ===
=== Summary ===

Correlation coefficient	0.9311
Mean absolute error	14803.537
Root mean squared error	17758.6811
Relative absolute error	69.9562 %
Root relative squared error	75.4502 %
Total Number of Instances	13

REFERENCES

[1] Abernethy M., *Data mining with WEKA*, Part1: Introduction and regression, IBM Corporation, 27 Apr 2010 http://www.ibm.com/developerworks/opensource/library/os-weka1/index.html

[2] Asegehgn L.W., *The Application of Data Mining in Crime Prevention: the case of Oromia Police Commission*, Thesis Addis Ababa University, 2003

[3] Berry M. and Linoff G. (1997). *Data mining techniques: For marketing, sales and customer support.* New York. John Wiley and Sons, Inc

[4] Berson A., Smith S., Thearling K*., Building Data Mining Applications for CRM*, McGraw-Hill Companies, 1999

[5] Chakrabarti S. et al., *Data mining: know it all*, Morgan Kaufmann know it all series,Elsevier Inc., 2009

[6] Dwivedi R. K. and Bajpai R., "Data Mining Techniques for dynamically Classifying and Analyzing Library Database Convention" on *Automation of Libraries in Education and Research Institutions*, CALIBER, 2007, 477-485

[7] Ge J., Xia Y., Nadungodage C., "UNN: A Neural Network for Uncertain Data Classification", *Proceedings of PAKDD* (1) 2010, 449-460

[8] Giudici P., Figini S., Applied Data Mining for Business and Industry, Second Edition, John Wiley & Sons, Ltd., 2009

[9] Goossens Y. et al., "Alternative progress indicators to Gross Domestic Product (GDP) as a means towards sustainable development", European Parliament's Committee on the Environment, *Public Health and Food Safety*, October 2007

[10] Grira N, Crucianu M, Boujemaa N. "Unsupervised and Semi-supervised Clustering: a Brief Survey". *Report of the MUSCLE European Network of Excellence* (FP6); August 15, 2005.

[11] Heisig S., Moyle S., "Using model trees to characterize computer resource usage". In *Proceedings of WOSS '04*, 2004, 80-84

[12] Kraetzer C., Schott M., and Dittmann J., "Unweighted fusion in microphone forensics using a decision tree and linear logistic regression models". In *Proceedings of MM&Sec '09*, ACM, USA, 2009, 49-56

[13] Loh W.Y., Chen C.-W., and Zheng W., "Extrapolation errors in linear model trees". *ACM Trans. Knowl. Discov.* Data 1, 2, August 2007

[14] Olson D.L., Delen D., *Advanced Data Mining Techniques*, Springer-Verlag Berlin Heidelberg, 2008

[15] Potts D., "Incremental learning of linear model trees". In *Proceedings of ICML '04*. ACM, New York, NY, USA, 84

[16] Rodriguez D., Cuadrado J.J., Sicilia Urbán M.A., Ruiz R., "Segmentation of Software Engineering Datasets Using the M5 Algorithm". In *Proceedings of International Conference on Computational Science* (4)'2006, 789-796

[17] Raghavendran S., and Jain N., "GDP metric – Time for a makeover?" *Global Economic Outlook*, A Deloitte Research publication, 3rd Quarter, 2011

[18] San OM, Huynh VN, Nakamori Y. "An Alternative Extension of The K-Means Algorithm for Clustering Categorical Data". *Int. J. Appl. Math. Comput. Sci.*, Vol. 14, No. 2, 2004, 241–247

[19] Tan T., Hadzic F., Dillon T.S., Chang E., Feng L., 2008. "Tree model guided candidate generation for mining frequent subtrees from XML documents". *ACM Trans. Knowl. Discov.* Data 2, 2, Walsh T.J., Szita I., Diuk C., Littman M.L., "Exploring compact reinforcement-learning representations with linear regression". *Proceedings UAI '09*, 591-598

[20] Witten I.H., E. Frank, Hall M.A., *Data mining : practical machine learning tools and techniques.-3rd ed.*, Morgan Kaufmann, Elsevier, 2011

[21] Buti, M. and P. van den Noord (2004), "Fiscal Policy in EMU: Rules, discretion and political incentives", European Commission, *Economic Papers* no. 206.

[22] Francesco Giavazzi, Tullio Jappelli, Marco Pagano, "Searching for non-linear effects of fiscal policy: Evidence from industrial and developing Countries", *European Economic Review 44* (2000) 1259-1289, http://paper.blog.eonet.jp/Giavazzi_et_al_2000.pdf

[23] Christina D. Romer , David H. Romer, *The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks, Working paper*, University of California, Berkeley, 2008

[24] Tino Berger, Gerdie Everaert, *Is the Impact of Labour Taxes on Unemployment Asymmetric?*, Working paper, Institute for International Economics, University Munster & SHERPPA, Ghent University, 2010

[25] Pissarides, C. A., "*The impact of employment tax cuts on unemployment and wages: the role of unemployment benefits and tax structure*". European Economic Review 42 (1), 1998, 155-183.

[26] Daveri, F., Tabellini, G., 2000. "Unemployment, growth and taxation in industrial countries". *Economic Policy* 15 (30), 47{104.

[27] Blanchard, O., Wolfers, J., 2000. "The role of shocks and institutions in the rise of European unemployment: the aggregate evidence". *Economic Journal* 110, 1-33.

[28] Layard, R., Nickell, S., Jackmann, R., *Unemployment, Macroeconomic Performance and the Labour Market*. Oxford University Press.2005.

[29] Bresfelean V.P., "Data Mining and Model Trees Study on GDP and its Influence Factors", *Recent Advances in Applied&Biomedical Informatics and Computational Engineering in Systems Applications*, Proceedings of IAASAT 2011 Florence, Italy, 2011, 401-406

[30] Hsu C-H, Lee T-Y, and Kuo H-M. "Mining the body features to develop sizing systems to improve business logistics and marketing using fuzzy clustering data mining". *WSEAS Transactions on Computers*. 8, 7 (July 2009), 1215-1224

[31] Mamcenko J. and Beleviciute I., "Data mining for knowledge management in technology enhanced learning". *In Proceedings of the 6th conference on Applications of electrical engineering* (AEE'07), Metin Demiralp and Nikos E. Mastorakis (Eds.). WSEAS, 115-119

[32] Pinto F.M. and Santos M.F., "Considering application domain ontologies for data mining". *WSEAS Transactions on Information Science and Applications*. 6, 9 (September 2009), 1478-1492

[33] Wanqing L., Lihua M., and Dong W., "Data mining based on rough sets in risk decision-making: foundation and application". *WSEAS Transactions on Computers* 9, 2 (February 2010), 113-123

[34] Zubi Z.S., "Using text mining techniques in electronic data interchange environment". *WSEAS Transactions on Computers* 9, 8 2010, 832-846.

[35] ***Eurostat,2011 *http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/*

[36] *** National Bank of Romania, Statistics *http://www.bnro.ro/Statistica-87.*

Bresfelean Vasile Paul (Bistrita, 3.12.1979) is an Assistant Professor PhD at Babes-Bolyai University of Cluj-Napoca, Romania. His educational background includes: PhD in Cybernetics and Economic Statistics (Babes-Bolyai University of Cluj-Napoca, Romania, 2009), Master of Science in Business Information Systems (Babes-Bolyai University of Cluj-Napoca, Romania, 2004), Bachelor of Economics (Babes-Bolyai University of Cluj-Napoca, Romania, 2003).

His earlier work experience comprises: American Work Experience intercultural program – work and travel activities in the US (2001-2002), Computer Science Teacher at the "Iulian Pop" College in Cluj-Napoca (2003-2004), Teaching and Research Assistant at Babes-Bolyai University (2004-2008). Amongst his fields of interest and didactic activities are: Data Mining, ERP, Databases, Decision Support Systems. He is author and co-author of 10 articles indexed in ISI Thomson database, and several other international and national publications, and presently manager of a national PNII_TE research grant.

Assist. Prof. Bresfelean is an active member of the professional society: INFOREC - Association for Promoting Education for Economic Informatics.

Mihaela Göndör (Suceava, 24.04.1965) is an Associate Professor PhD at "Petru Maior" University Tîrgu Mures, Romania. Her educational background includes: PhD in Economics-Financial Management (Al.I.Cuza University, Iasi, Romania, 2003) Master Degree in Local Public Administration Management, (Academy of Economic Studies, Bucharest,Romania, 2001), Bachelor of Economics (Al.I.Cuza University, Iasi, Romania, 1987).

Her previous work experience comprises: Tax Consultant- liberal profession (2011), PhD Lecturer (1995-2005), Senior Adviser Corporate Taxes (1998-2003), Associate Manager (1991-10.1998), Economist (1987-1991). She is author and co-author of 6 books, 43 university courses, 13 professional research papers published in international journals, 44 publications in international conferences volumes and proceedings. She is in charge of the following academic courses: Public Finance, Comparative Fiscal Policies, Monetary and Financial International Relations, Banking Risk and Performance.

Prof. Göndör is an active member of the professional society: Chamber of Tax Consultants, Romania. She is awarded with a postdoctoral scholarship from the Romanian Academy with the theme "Discretionary and nondiscretionary in fiscal policy of European Union States"