WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

Vol. 10
No. 1
Spring
2015

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

Contents

EVALUATING SOCIAL TRACKING IN THE PRIMARY SCHOOL: EVIDENCE FROM THE LOMBARDY REGION (ITALY)[1]

Emanuela RAFFINETTI[2]
PhD, Post-Doc Research Fellow, Department of Economics,
Management and Quantitative Methods
Università degli Studi di Milano, Italy

E-mail: emanuela.raffinetti@unimi.it

Isabella ROMEO[3]
PhD, Post-Doc Research Fellow,
Department of Statistics and Quantitative Methods
University of Milano-Bicocca, Italy

E-mail: isabella.romeo@unimib.it

Abstract

Recently, the Italian schools were deeply affected by the "social tracking" phenomenon, intended as the process of segregating students into socio-economic classes. Typically, this phenomenon occurs within the lower secondary school. In such a perspective, the study reported in the paper is innovative, since addressed to investigate the actual presence of the social tracking phenomenon as an event starting from the primary school. For this purpose, we considered data provided by Invalsi (Istituto Nazionale per la Valutazione del Sistema di Istruzione e Formazione) with regard to students of the fifth grade of primary schools in the Lombardy region (Italy). The study was carried out following two different approaches. First, a preliminary descriptive analysis of the segregation phenomenon was carried out by computing the Gini coefficient of the the socio-economic status average at class level. Second, due to the usual hierarchical structure of educational data, multilevel models were considered with the aim of partitioning the pupils' socio-economic status variability within the student, class and school level. In this way, school and class social segregation indicators were obtained. Subsequently, a conditional multilevel model including school and class social segregation indicators as explanatory variables was built. Results underline that even though in general social tracking is not an actual threat for the Lombardy primary schools, a remarkable socio-economic heterogeneity among classes appears especially in some provinces of the Lombardy region.

Key words: social tracking phenomenon, class heterogeneity, Gini coefficient, segregation indices, multi-level modeling, Invalsi data

1. Introduction

Interest in evaluating the Italian education systems is manifest in a large number of recent publications and in the diffusion of standardized tests (e.g., Haladyna, 1991; Ballard and Bates, 2008). Typically, the content of these contributions focuses on the main pupils and schools' determinants affecting the learning levels of students. If on one hand the educational research field stresses the impact of such factors on the students' attainments, on the other hand only a few works addressed the issue of equal opportunity in education (i.e. each state must provide the same opportunities for everyone who attends school regardless of gender, race or nationality). Even though the Italian law imposes the "equity principle" which should be preserved by composing the most possible heterogeneous classes, recent studies highlight that the practice of segregating students with similar features is particularly widespread, especially in the lower secondary schools (e.g. Ferrer-Esteban, 2011). Such a sociological issue falls under the name of "formal tracking" phenomenon. In some cases, school staff may generate a great deal of selection by assigning children with similar achievement to the same classroom, in order to minimize teaching difficulty, or by placing all of the "problematic" students in a certain teacher's class because he is good at dealing with them. However, the segregation phenomenon can be generated in several ways and at different levels. Specifically, the increasing participation of the pupils' parents to the dynamics of the school is leading to a kind of "informal tracking" phenomenon, allowing families to influence the classroom composition in order to better respond to their social features, such as for instance their socio-economic status (e.g. Dupriez et al., 2008).

Social tracking gives rise to homogeneity within classes (social segregation) that in turn may come out in inequality of education opportunities (e.g., Checchi and Flabbi, 2007; Hindriks et al., 2010). Children with different family background, race and ability will have different access to knowledge. It was proven (for example, Loveless, 1999) that whether the curriculum is adjusted to better match ability level of students, while high ability students may receive a boosted achievement, low ability students may suffer from assignment to lower tracks. Thus, homogeneity within classes negatively affects disadvantages students. Classroom environment is then really important for student achievement, as stated by Hill and Rowe (1996): *"How much a student learns depends on the identity of the classroom on which the student is assigned"*. Indeed, a student's innate ability can affect his peers, not only through knowledge spillovers but also through his behavior. On the contrary, a student who has not learned self-discipline at home may bother the classroom.

The study presented in this paper is innovative since it attempts to explore the actual presence of the social segregation phenomenon in Italy as an event starting from the primary schools. Indeed, to the best of our knowledge, no research contributions illustrating the existence of an informal tracking phenomenon in the Italian primary schools are currently provided in literature. More precisely, our research question is the following. Since primary schools represent the first education compulsory stage after the kindergarten, the segregation process of kids can probably be encouraged by parents on the basis of their socio-economic features. Kindergarten has a relevant role in the process of contact among the families of kids. Thus, the pupils' families may wish that their children were kept together with their kindergarten friends, when accessing to the primary school.

The analysis was carried out on data provided by the National Evaluation Committee (Istituto Nazionale per la Valutazione del Sistema di Istruzione e Formazione, henceforth Invalsi). In Italy the National Evaluation Committee has been established with the specific aim of evaluating the Italian schools through the analysis of the students' achievement at different levels of education; second and fifth year of the primary school (age 7 and 10, respectively), first and third of the lower-secondary (age 11 and 13), second and fifth of the upper-secondary (age 15 and 18). The collection of such data started from the school year 2008-2009 and represents the first time that a law imposes a national evaluation by using standardized tests in all students population. Here, we considered a unique dataset that tracks the performance in Reading of students of the fifth grade of primary schools in the Lombardy region for the school year 2009-2010.

On the statistical point of view, our proposal was pursued through two different approaches. First, a preliminary investigation of the social tracking phenomenon was provided by resorting to a descriptive inequality index, the Gini coefficient, which is widely used for studying inequality in education attainments (e.g., Leckie et al., 2012). The Gini coefficient was computed by taking into account the class average value of the variable representing the socio-economic status (henceforth denoted by SES) over all the classes in every province of the Lombardy region. Second, to shed light on how the heterogeneity of the students' performance and SES are portioned out between school and class level, different multilevel models were considered both to properly take into account the hierarchical structure of data with pupils nested in classes and schools (e.g., Snijders and Bosker, 1999) and to define social segregation indices at school and class level. Finally, a conditional multilevel model with even the social segregation indices is performed.

The remainder of the paper is organized as follows. In Section 2 the examined Invalsi dataset is illustrated and some descriptive statistics provided. In Section 3 a preliminary analysis of the social tracking phenomenon is introduced by resorting to a descriptive approach based on the Gini coefficient. In Section 4 an overview of the proposed multilevel methodology is presented. In Section 5 school and class level social segregation indices are computed and commented. Section 6 is devoted to the discussion of the obtained results. Finally, Section 7 concludes.

2. Data

Our proposal is based on data coming from the survey led by Invalsi at the end of the school year 2009-2010 and referring to students of the fifth grade (students of about 11 years old). Coherently with our research scope, the variable under study is here detected by the pupils school achievement in Reading, expressed as the proportion of correct answers provided in the administered test by each student. Such data cover the whole population (it is not a sample) made up of 77.200 students belonging to 4.488 classes that in turn belong to 1.050 primary schools located in different provinces of the Lombardy region. The administered test is built on 41 multiple-choice items and is composed by two parts: the former is related to the comprehension of two texts and the latter is related to the grammar issues. The testing time is of one hour. The test reserves even a set of questions concerning the students' personal information (e.g. gender, ethnicity, grade retention and so on). Further information about the social, economic and cultural conditions of students are collected through additional questionnaires filled by the School Principals and students' parents. Variables considered for the analysis are enlisted below and include:

- demographic variables: i.e. gender, ethnicity, year of birth;

- sociocultural variables: in this case, a synthetic index, named SES is made directly available by Invalsi. It is computed analogously to the OECD's procedure, that is by considering the parents' occupation and education, possession of some kinds of goods such as, for instance, the availability of an encyclopedia or an Internet connection, the number of books at home and so on (Campodifiori et al., 2010);
- school variables: school size (number of students), type of school administration (private or public), number of female students, number of students repeating one or more grades and number of students belonging to ethnic minorities;
- geographical area of the school, specified in the provinces of the Lombardy region: Bergamo (BG), Como (CO), Lecco (LC), Lodi (LO), Milano (MI), Pavia (PV), Varese (VA), Brescia (BS), Cremona (CR), Mantova (MN) and Sondrio (SO).

A note about the type of school administration (private or public) is needed. For private school we mean schools with private involvement in managing and funding. Here, we only focused on private schools following the ministerial program and thus considered equivalent to the public ones.

Before proceeding to the construction of the statistical model, an analysis of missing data was done for all the variables that potentially may be included in it. The reference dataset is characterized by variables which present missing values at random. However, the main trouble appears with the pre-school (i.e. kindergarten attendance) variable whose lack of information is consistent, since missing values amount to the 10.4%. In such a context, the problem of missing data was easily solved by directly deleting the pre-school variable from the model. This is because, the ejection of the pre-school variable from the model found reason in its low contribution in explaining the Reading scores variability.

In order to provide more interpretable parameters, all the variables were standardized and a reference level was defined (e.g., Snijder and Bosker, 1999). Furthermore, to better clarify the role of the categorical variables included into the model and concerning the demographic characteristics of pupils (i.e. gender, ethnicity, and grade retention) and the school features (public or private status), a related description is presented in Table 1, where the corresponding reference categories are reported.

Table 1. Description of the pupil and school categorical variables

Variables	Description
Demographic	
Gender	Male (reference category); Female
Ethnicity	Italian (reference category); Ethnic minorities of first or second generation
Grade Repetition	Student that has not repeated a year (reference category, pupils born in 1998); Student that has reapeated at least a year (grade repetition)
Educational	
School Administration	Public (reference category); Private

With regard to class and school level, we considered variables representing the proportion of students being female, repeating one or more grades and belonging to ethnic minorities. These variables were already available in the dataset at school level and relate to students belonging to all grades in the school. On the contrary, variables at class level were derived as aggregation of individual covariates at class level. Thus, the latter are related only to students participating to the survey of the fifth grade. Moreover, the school and class average of the students' SES index were computed as aggregation of individual SES index.

The main key statistics about variables at class and school level are displayed in Table 2. It is worth noting that variables at school level were centered on the grand mean and variables at student level were centered on the school average. As shown in Table 2, the average score[4] amounts to 73.20 with a standard deviation equal to 16.63, the average percentage of female is the 49% at class level and the student SES average is 0.03 at class level and 0.04 at school level. In addition, almost the 9% of schools are private, the average percentage of ethnic minority students amounts to the 13% both at class and school level, while the average percentage of students repeating the year is the 3% at class level and smaller than the 1% at school level.

Table 2. Descriptive Statistics

		Number of units	Mean	St Dev	Min	Max
Class	Score in Reading	77,200	73.20	16.63	0.00	100.00
	% Ethnic Minorities	4,466	0.13	0.55	0.00	1.00
	Class mean SES	4,487	0.03	2.17	-2.05	2.16
	% Females	4,485	0.49	0.51	0.00	1.00
	% Student Repeating the year	4,487	0.03	0.23	0.00	1.00
	Class size	4,488	20.00	10.00	6.00	28.00
School	% Ethnic Minorities	1,050	0.13	0.18	0.00	0.83
	School mean SES	1,050	0.04	0.86	-1.28	2.04
	% Females	1,050	0.48	0.07	0.00	1.00
	% Student Repeating the year	1,050	0.003	0.01	0.00	0.03
	School size	1,050	532.00	428.00	28.00	1,338
	School Administration: Public	74,265	91.17			
	School Administration: Private	7,191	8.83			
	Province: BG	10,020	12.30			
	Province: BS	11,401	14.00			
	Province: CO	4,869	5.98			
	Province: CR	2,833	3.48			
	Province: LC	2,867	3.52			
	Province: LO	1,923	2.36			
	Province: MN	3,477	4.27			
	Province: PV	3,991	4.90			
	Province: SO	1,558	1.91			
	Province: VA	7,412	9.10			
	Province: MI	31,105	38.19			

3. Preliminary Analysis: the Gini coefficient

In the literature, a wide range of indices are proposed for assessing the actual presence of the social tracking phenomenon. As deeply discussed by Leckie et al. (2012), Hutchens (2004) and Reardon and Firebaugh (2002), researchers typically resort to descriptive indices such as, for instance dissimilarity and square root indices (e.g., Duncan and Duncan, 1955; Jenkins et al., 2008), in order to detect possible scenarios of inequality in education opportunity. Since our aim is not limited to detect the presence of inequality in opportunity but to measure its extent, within the large set of available descriptive indices, the Gini coefficient was considered (e.g., Gini, 1921). More in detail, the idea here is to provide a measure of the heterogeneity between classes in term of the socio economic status of students. For this purpose, we propose to consider as variable of interest the average SES at class level. For all the classes within each school and each province of the Lombardy region,

we computed the average value of the students' SES index. We remark that for every single student, the SES index ranges between -3 and +3. Thus, it is reasonable to believe
that the average SES at class level may take even negative values. In such a context, the reliability of the classical Gini coefficient may come less since requiring the considered variable to be characterized by non-negative values. Indeed, in case of negative values, the Gini coefficient may violate the normalization principle and thus take values greater than one. A solution to this problem was recently provided by Raffinetti et al. (2014), who introduced a new Gini coefficient adjusted for the presence of negative values. The new Gini coefficient, expressed as the ratio between the absolute mean difference $(1/N^2)\sum_{i=1}^{N}\sum_{j=1}^{N}|Y_i - Y_j|$ and $(2/N)\sum_{i=1}^{N}|Y_i|$, fulfills the normalization principle. This allows us to provide a measure of inequality in opportunity which can occur as a consequence of the class composition process conditioned to the pupils' socio-economic status. Indeed, if the Italian schools actually respected the legislative principle of "equal-heterogeneity" in the composition of classes, the Gini coefficient should be close to zero. This does not happen, as shown by results in Table 3, where the Gini coefficient of the average SES at class is reported for every province.

Table 3. Gini coefficient of the average SES at class level per province

Province	BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA
Gini coefficient	0.65	0.67	0.71	0.69	0.72	0.71	0.72	0.63	0.73	0.69	0.70

The Gini coefficient is greater than 0.60 in all the provinces. More precisely, over the 50% of the provinces presents a Gini coefficient greater than 0.70. The province of PV has the higher heterogeneity between classes with a Gini coefficient equal to 0.73. Such findings are made more evident by the boxplots in Figure 1 which show a remarkable variability of the average SES at class level in every province of the Lombardy region.

Even though these descriptive statistics seem to confirm the presence of the social tracking phenomenon, they are obtained without taking into account the hierarchical structure of classes nested in schools. Thus, the high heterogeneity between classes at province level may reflect the high heterogeneity between schools within the province. For this reason, one may assume this variability to be explained by the gaps across the territorial areas where schools are located. Indeed, schools located in more disadvantaged areas catch more disadvantaged students. Further investigations were carried out by distinctly computing the Gini coefficient of the SES average at class level within each school across all the provinces. Also in this case, the Gini coefficient reaches very high values, leading us to believe that heterogeneity between classes is a real threat for the equality in opportunity in the Italian primary schools. In order to validate such a conclusion, the multilevel modeling approach (e.g., Goldstein, 2011) was considered to take into account the complexity of the educational systems organized in school and class level. First, we assessed how the variability of SES portions out among the different considered levels in order to define segregation status indicators at class and school level, as suggested by Ferrer-Esteban (2011). Subsequently, we analyzed the partition of the variability of the scores in the Reading test among the different levels. Finally, a conditional multilevel model was built in order to evaluate the effects of both the SES index and segregation status indicators, after controlling for the aforementioned variables, with the purpose of detecting the actual presence of the social tracking phenomenon across the Lombardy primary schools.

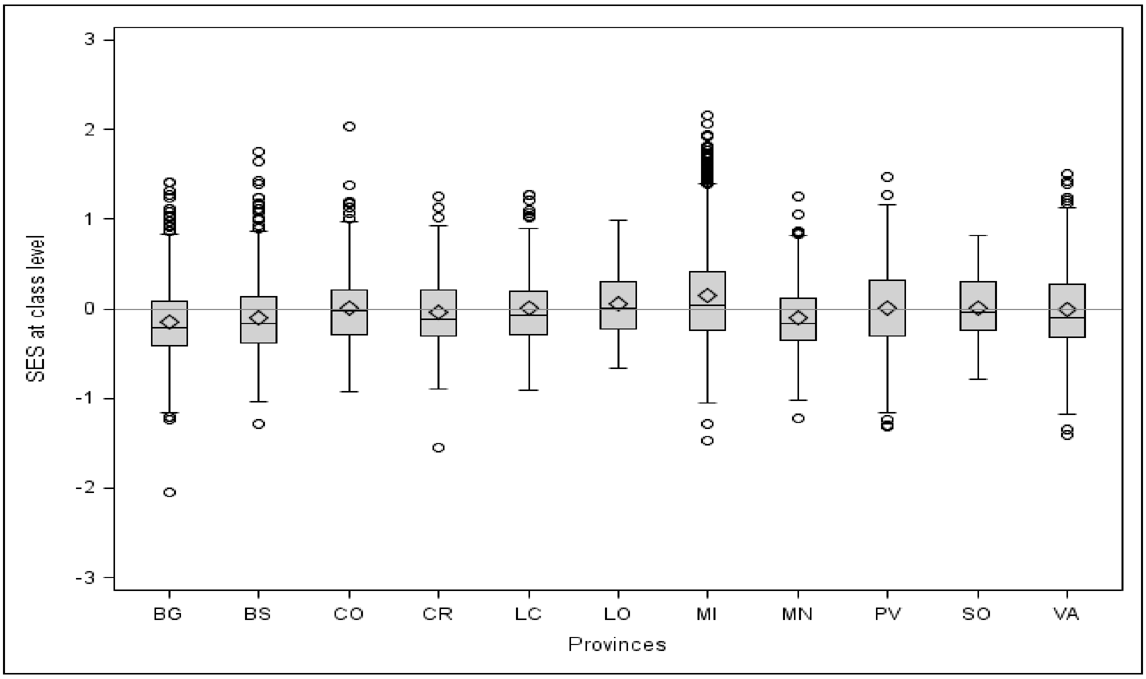


Figure 1. Boxplots of the average SES at class in every province of the Lombardy region

4. An overview about three-level models

As mentioned above, models suitable in treating hierarchical data are the multilevel models since they allow relationships to be simultaneously assessed at several levels (e.g., Snijders and Bosker, 1999), represented by pupils, classes and schools. Some details about the multilevel algebraic specification are briefly provided below.

Let us consider a three-level multilevel model in educational context, where level 1 is represented by students, level 2 by classes and level 3 by schools. The relationship between the i-th student's achievement, belonging to the j-th class, which in turn belongs to the k-th school, is expressed by:

$$y_{ijk} = \beta_{0ijk} + \beta_1 x_{ijk}\ , \qquad (1)$$

$$\beta_{0ijk} = \beta_0 + v_{0k} + u_{0jk} + e_{0ijk}\ , \qquad (2)$$

where: v_{0k} is the random effect at school level, an allowed-to-vary departure from the grand mean, u_{0jk} is the random effect at student level, a departure from the school effect and e_{0ijk} is the random effect at student level, a departure from the class effect within a school. The variance components at each level are defined as follows: variance between schools, $Var(v_{0k}) = \sigma_{v0}^2$; variance between students within classes within schools $Var(u_{0jk}) = \sigma_{u0}^2$; variance between students within classes within schools $Var(e_{ijk}) = \sigma_e^2$; and variance between classes $\sigma_{v0}^2 + \sigma_{u0}^2$. Different forms of variance shares are derived: the share of variance due to gaps between schools, corresponding to the intra-school correlation (level 3)

$ISC = \frac{\sigma_{v0}^2}{\sigma_{v0}^2+\sigma_{u0}^2+\sigma_e^2}$, and the share of variance due to differences between classes, corresponding to the intra-class correlation (level 2) $ICC = \frac{\sigma_{v0}^2+\sigma_{u0}^2}{\sigma_{v0}^2+\sigma_{u0}^2+\sigma_e^2}$.

Since multilevel models allow to decompose the variability of a specific phenomenon among the different involved levels, they provide information about the heterogeneity associated to each considered level. To have an idea of such heterogeneity, first the intra-class correlation coefficient (ICC) was computed. Indeed, through such a coefficient, the extent of the outcome variation related to gaps between units of each considered level was obtained. Secondly, a model with variables described in Section 2 was built to show both covariates really affecting the students' achievement and their impact. Furthermore, also segregation status indicators at class and school level were considered. The latter is identified as the between classes and schools variance, when an unconditional multilevel model for the SES variable is fitted. Finally, a comparison between the unconditional model (empty) and the conditional (full) model was introduced to show the contribution of the same model in explaining the performance variability at each level of the analysis. The residual variance located at different levels was interpreted as the result of unobserved factors, as discussed in more detail in Section 6.

5. Segregation indices

As suggested by Ferrer-Esteban (2011), social segregation at class and school level are typically measured trough the between-class variance and the the between-school variance, respectively.

A fully unconditional three-level model for the SES index allows to portion the SES variability among the considered level: within classes, between classes within schools and between schools. A high variability of SES between classes underlines more heterogeneity among different classes within the same school, meaning that classes are more homogeneous in respect of their social background. Conversely, a high SES variability between schools underlines more heterogeneity among schools, implying aggregation of students with similar social background within the school. These indicators give an idea of the extent both schools and classes within schools are socially dissimilar. Ferrer-Esteban (2011) analyzed the Italian secondary schools and found out that the SES variability at school level reaches a value of 32% in some Italian provinces, while the SES variability at class level reaches a value of 12%. Furthermore, they stressed that while the SES variability at school level is connected with the presence of metropolitan areas, the SES variability at class level has a clear pattern of territorial distribution that responds to a north-south gradient, with higher values of class segregation in the South of Italy.

For what concerns the primary schools we expected a remarkable SES variability between schools, given that this kind of school is particularly widespread across the Italian territory. For this reason, primary schools usually catch students of the area in which they are located. So, schools located in areas with more disadvantaged families will catch disadvantaged students. Furthermore, the high diffusion of the primary schools involves schools to be composed by one or few classes for each grade, leading to expect a low SES variability between classes. In Table 4 the segregation indicators at province and regional level for the Lombardy primary schools are reported.

Table 4. Social segregation indicators

Province	Between class variance (in %)	Between school variance (in %)
BG	2.24%	15.28%
BS	2.71%	14.85%
CO	5.33%	11.03%
CR	2.88%	11.76%
LC	2.20%	17.53%
LO	3.80%	12.36%
MI	3.46%	25.20%
MN	3.12%	7.06%
PV	4.47%	17.49%
SO	2.49%	8.53%
VA	3.67%	16.53%
Lombardy	3.23%	19.58%

In the Lombardy region, the variability of SES between schools is equal to 19.6%, while between classes it is equal to 3.2%. In particular, the Lombardy provinces highlight a SES variability at school level ranging between the 7.1% of Mantova and the 25.2% of Milano, and a SES variability at class level ranging between the 2.2% of Bergamo and Lecco and the 5.3% of Como and Pavia. These values, compared to the findings illustrated by Ferrer-Esteban (2011) across the whole Italy and for the lower secondary schools, are to be considered non-low. Indeed, it is well-known that the social segregation is a phenomenon appearing more marked in the lower secondary school and in the South of Italy. As expected, the metropolitan area of Milan presents a high variability of SES between schools. The SES variability between classes is low on average, but with non-low values for some provinces. To evaluate if such heterogeneity between schools and classes provides an actual impact on the students' achievement, a multilevel model built on the Reading score was considered. The related results are discussed in the following section.

6. Multilevel model results

The content of this section is focused on both the analysis of the partition of the test performance variability at individual and group level and identifying the presence of social segregation. Typically, in the education literature the study of variability in achievements is based on two-level models characterized by student and school level. Our aim is a little bit wider since addressed to identify the share of variability attached both to school and class level. The need of including the class level is supported by our research question, that is investigating the actual presence of social tracking within the Italian primary school. For this reason, a three-level model was applied to account for the class level.

In order to define the performance variability partition among the three involved levels, an empty model without explanatory variables was applied. The related results in terms of variance decomposition and intra-class correlation coefficient (ICC) are reported in Table 5.

Table 5. Variance decomposition of the implemented multilevel model – fifth grade of primary school

	Three level - School-Class-Student	
	Empty model	***ICC (%)***
Var. Between Schools	13.0	4.7
Var. Between Classes	21.3	7.7
Var. Within Classes	243.0	87.6
Total Var.	277.3	100.0

According to such findings, in primary school the well known prevailing variability in performance depends on the students' characteristics ($\sigma_e^2 = 243.0$). In addition, variability between classes is greater than variability between schools, in line with the evidence gathered from several studies (e.g., Hill and Rowe, 1996). Indeed, the former ($\sigma_{v0}^2 = 21.3$) is almost double with respect to the latter ($\sigma_{u0}^2 = 13.0$). Despite we expected a low variability between classes when only a grade is considered, we found a high percentage of this variability equal to more than the 7% of the total one. Variability between classes appears as the consequence of several factors which may be strictly related to grade, teacher effect and unobservable variables as well as peer effect.

To shed light on individual, class and school variables really impacting on the students' achievement, a conditional multilevel model was considered. Results about variables significance are displayed in Table 6. Variables marked by three asterisks in Table 6 are significant at a confidence level $\alpha = 0.01$, variables marked by two asterisks at a confidence level $\alpha = 0.05$ and finally, variables marked by only an asterisk at a confidence level $\alpha = 0.1$. Variables representing the percentage of female students, the percentage of students repeating a year, the size of class and school are non-significant in the analysis at both school and class level. The intercept value represents the Reading score for an "average" student who is defined as an Italian male student, with no grade retention, a SES index equal to the average value of all the students, and whose class presents a percentage of students belonging to ethnic minorities and an average SES index corresponding to the mean value of all the classes. Furthermore, his attended school is public, located in Milan and characterized by a percentage of students belonging to ethnic minorities and an average SES index equal to the mean value of all the schools. Such a student achieves a performance in Reading equivalent to 77.21 (i.e. an "average" student correctly answers to the 77% of the test). To be a male implies a reduction in Reading score equal to 1.64. As it is trivial to believe, when focusing on individual student variables, the consistent decrease of the Reading score is related to students belonging to ethnic minorities of first generation (-11 points). Student belonging to ethnic minorities of second generation provides a smaller decrease in Reading achievements, corresponding to 7.80 points. Definition of ethnic minorities of first and second generation is needed. Ethnic minorities of first generation are students born in their origin country from parents belonging to ethnic minorities, while ethnic minorities of second generation are students born in Italy from parents belonging to ethnic minorities. The same negative results on the Reading performance are associated to class and school variables concerning the percentage of students belonging to ethnic minorities. Obviously, also the grade retention involves a worsening of -7.43 points. Conversely, a positive trend in Reading performance is associated to the SES index. Indeed, an unitary increment in the SES index value provides an increase of 4.38 points in the Reading score. This happens also for the school and class SES index variable. Students attending a school in the province of Lodi, Mantova and Brescia reach worse results with respect to those attending a school in the province of Milan.

Table 6. Three-level Multilevel model Effects

Levels	Variables	Estimate
Student	Intercept	77.21***
	Gender (Female)	-1.64***
	Ethnic minority - First Generation (Ref. Italian)	-10.99***
	Ethnic minority - Second Generation (Ref. Italian)	-7.80***
	Grade Repetition	-7.43***
	Student SES	4.38***
Class	% Ethnic minority in class	-0.05**
	Class mean SES	0.98***
	% Female students in class	0.00
	% Repeating the year in class	0.03
	Class size	-0.04
School	% Ethnic minority at school	-0.08***
	School mean SES	3.18***
	% Female students in school	-0.01
	% Repeating the year in school	0.16
	School size	0.05
	Private school (Ref. Public)	-0.98
	School Segregation	-8.83*
	Class Segregation	-53.03
	Province: BG (Ref. MI)	0.26
	Province: BS (Ref. MI)	-0.93*
	Province: CO (Ref. MI)	1.37
	Province: CR (Ref. MI)	-1.31
	Province: LC (Ref. MI)	0.09
	Province: LO (Ref. MI)	-2.44**
	Province: MN (Ref. MI)	-2.75***
	Province: PV (Ref. MI)	0.24
	Province: SO (Ref. MI)	0
	Province: VA (Ref. MI)	0

Both school and class segregation indicators, computed in Section 5, were considered for every province and included as explanatory variables into the model. While the school segregation indicator is significant in the model, the class segregation indicator does not. This finding involves the presence of school segregation and the absence of class segregation in the Lombardy primary schools. Such a conclusion arises as the consequence of the widespread of the primary schools across the Italian territory. Indeed, since the primary schools usually receive students living in the area where the schools are located, the socio-economic status of the area strongly affects the socio-economic status of the school. The consistent SES variability between schools is in general an expected result which further stands out for the metropolitan area of Milan. In particular, it is worth noting that the effect of segregation on achievement is negative causing a reduction of over 8 points on the Reading performance.

To clarify the contribution of such variables in explaining the students' achievement variability in the Reading test, an analysis on the variance reduction at school and class level was carried out. As shown in Table 7, the full multilevel model provides a contribution of about only the 33.7% of the students' performance variability at school level, about the 17.7% at class level and about the 15% within class level. These outcomes have not to be considered as a failure of our proposed approach, since typically in the primary school the main share of variability in achievement is the consequence of non-observable students' characteristics such as, for instance, hours working on homework and/or students' interest in school matters.

Table 7. Decomposition of Variance

Variance	Empty Model	Conditional Model		Factors
	ICC	Reduction	ICC	
Between Schools	4.7%	33.7%	1.3%	Observed school factors
			2.5%	Other unobserved factors
Between Classes	7.7%	17.7%	1.3%	Observed class factors (class composition)
			6.2%	Teacher effect and other unobserved factors
Within Classes	87.6%	15.0%	13.3%	Observed individual factors
			75.5%	Other unobserved individual factors
Total	100.0%	16.1%	100.0%	

The explained variance has to be ascribable to the presence of observed factors at different levels, while the residual variance can be ascribable to the presence of unobserved factors. For instance, at class level the residual variance may include the impact of teacher and/or other unobserved factor. Definitely, observed school factors explain only the 1.3% of the performance variability. Unobserved school level factors account for the largest differences in variability in school performances, but in this case they capture just the 2.5% of the overall variability in achievements. Compositional factors at class level account for the 1.3% of the overall performance variability. Thus, it is reasonable to believe that, the impact of unobservable variables between classes (within schools) on gaps in achievement, is much more marked amounting to the 6.2%. Finally, the unobserved individual factors account for the 75.5% of the overall variability highlighting that the unobservable student's characteristics represent the largest differences in the non-explained variability.

7. Conclusions

In this paper we investigated the presence of the social segregation phenomenon by analyzing education data provided by Invalsi and concerning the achievement in the Reading test, obtained in the school year 2009-2010, by students attending the fifth grade of the primary schools in the Lombardy region (Italy). From this point of view, our study is innovative since it attempted to detect the social segregation phenomenon as an event starting from the primary schools. For this purpose, two different approaches were considered. First, a preliminary investigation of the social tracking phenomenon was provided by resorting to the Gini coefficient computed by taking into account the class average value of the SES variable over all the classes in every province of the Lombardy region. Results show a high heterogeneity between classes and would seem to validate the hypothesis of social tracking inside the primary schools. However, to account for the hierarchical data structure, a multilevel model was carried out. First of all, segregation indices at class and school level through a fully unconditional three-level model for the SES index were found out. Such indices are defined in terms of the SES index (representing the pupils' socio-economic background) variability among the considered levels (i.e. within classes, between classes within schools and between schools). Findings highlight that even though the SES index presents a low variability on average, such variability is consistent in value across some provinces. Then, a conditional multilevel model including both indicators of between class and school segregation as explanatory variables for every province was built. While the school segregation index is significant in the model, the class segregation index does not. These results suggest that the segregation phenomenon mainly occurs at school level, neglecting the actual threat of the social tracking phenomenon in the primary schools of the Lombardy region. However, from a descriptive point of view the presence of a consistent class heterogeneity is an evidence especially in some provinces of the Lombardy region. This issue encourages us to believe that such a phenomenon may represent an actual event in early education in the areas of Italy (South and Islands), where inequality in households' socio-economic status is known from the literature to be more marked.

Bibliography

1. Ballard, K. and Bates, A. **Making a connection between student achievement, teacher accountability and quality classroom instruction**, The Qualitative Report, Vol. 13, no. 4, 2008, pp. 560-580
2. Campodifiori, E., Figura, E., Papini, M. and Ricci, R. Campodifiori, E. **Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia** (in Italian), Invalsi, 2010
3. Checchi, D. and Flabbi, L. **Intergenerational mobility and schooling decisions in Germany and Italy: the impact of secondary school tracks,** IZA Discussion Paper No. 20(2876), 2007, pp. 210-217
4. Duncan, O. and Duncan, B. **A methodological analysis of segregation indexes**, American Sociological Review, Vol. 20, no. 2, 1955, pp. 210-217
5. Dupriez, V., Dumay, X. and Vause, A. **How do school systems manage pupils' heterogeneity?**, Comparative Education Review, Vol. 52, no. 2, 2008, pp. 245-273
6. Ferrer-Esteban, G. **Rapporto sulla scuola in Italia** (in Italian), Laterza, Bari, 2011
7. Gini, C. **Measurement of inequality and income**, The Economic Journal, Vol. 31, 1921, pp. 124-126
8. Goldstein, H. **Multilevel Statistical Models**, Wiley, 2011
9. Haladyna, T. **Raising standardized achievement test scores and the origins of test score pollution**, Educational Researcher, Vol. 20, no. 5, 1991, pp. 2-7
10. Hill, P. and Rowe, K. **Multilevel modelling in school effectiveness research**, School Effectiveness and School Improvement, Vol. 7, 1996, pp. 1-34
11. Hindriks, J., Verschelde, M., Rayp, G., and Schoors, K. **Ability tracking, social segregation and educational opportunity: evidence from Belgium**, Discussion Paper, Center for Operations Research and Econometrics, Bélgica, 2010
12. Hutchens, R. **One measure of segregation**, International Economic Review, Vol. 45, no. 2, 2004, pp. 555-578

13. Jenkins, S., Micklewright, J. and Schnepf, S. **Social segregation in secondary school: how does England compare with other countries?**, Oxford Review of Education, Vol. 34, no. 1, 2008, pp. 21-37
14. Leckie, G., Pillinger, R., Jones, K. and Goldstein, H. **Multilevel modeling of social segregation**, Journal of Educational and Behavioral Statistics, Vol. 37, no. 1, 2012, pp. 3-30
15. Loveless, T. **Will tracking reform promote social equity?**, Understanding Race, Class and Culture, Vol. 56, no. 7, 1999, pp. 28-32
16. Raffinetti, E., Siletti, E. and Vernizzi, A. **On the Gini coefficient normalization when attributes with negative values are considered**, Statistical Methods & Applications, DOI 10.1007/s10260-014-0293-4, 2014
17. Reardon, S. and Firebaugh, G. **Measures of multigroup segregation**, Sociological Methodology, Vol. 32, no. 1, 2002, pp. 33-67
18. Snijders, T. and Bosker, R. **Multilevel Analysis. An introduction to basic and advanced multilevel modeling**, London: Sage Publications, 1999

---

[1] **Acknowledgements**
Thanks are due to the Italian National Evaluation Committee (Invalsi - Istituto Nazionale per la Valutazione del Sistema di Istruzione e Formazione) for providing us with the data for this study.

[2] **Emanuela Raffinetti** has a Master Degree in Economics (Finance) at University of Pavia (Italy); International Postgraduate Master in Complex Systems at IUSS (Istituto Universitario Studi Superiori) of Pavia (Italy) and Ph.D. in Statistics at Bocconi University of Milan (Italy). Currently, she is Post-Doc Research Fellow at Università degli Studi di Milano (Italy). Her research activity concerns: dependence analysis, concordance and discordance measures, categorical variables treatment, models for ordinal variables, inequality measures in income distribution, statistical techniques for high-dimensional dataset, customer satisfaction assessment and evaluation of educational systems.

[3] **Isabella Romeo** has Bachelor Degree in Statistics at University of Milano-Bicocca (Italy); a Master Degree in Biostatistic and Applied Statistics at University of Milano-Bicocca (Italy), Ph.D. in Statistics at University of Milano-Bicocca (Italy). Currently, she is Post-Doc Research Fellow at University of Milano-Bicocca (Italy). Her research activity concerns: models for ordinal and categorical variables, causal inference in a counterfactual framework, policy evaluations, longitudinal analysis, education and labour market applications, statistical techniques to manage administrative data and high-dimension dataset.

[4] The score on Reading test corresponds to the total percentage of correct answers provided by the students. Thus, it lies between 0 and 100.

A LOGISTIC MODEL ON PANEL DATA FOR SYSTEMIC RISK ASSESSMENT – EVIDENCE FROM ADVANCED AND DEVELOPING ECONOMIES[1]

Smaranda CIMPOERU
PhD, University Assistant,
Department of Statistics and Econometrics,
Bucharest University of Econmic Studies, Romania

E-mail: smarandacimpoeru@yahoo.com; smaranda.cimpoeru@csie.ase.ro

Abstract
The present paper proposes a framework for developing a new early warning system (EWS) for identifying systemic banking risk and finding the macroeconomic indicators which turn to be the best indicators in predicting stressful situation in the economic environment. The research problem is very much debated in the specialty literature, as the exposure of the financial system is generally derived from deteriorating systemic conditions. We propose a logistic model applied on two panel data sets – advanced and emerging economies. Results are satisfactory, as apart from the GDP Growth or Debt level, as main triggers for financial stress situation, we also find the Output Gap as a significant early warning signal for predicting financial and economic crisis.

Keywords: *systemic risk, early warning systems, financial crisis, binary variables panel data*

1. Introduction

A well-functioning financial system is mandatory for an efficient economy. However, the fragility of financial systems can cause financial crisis and have significant impact in the real economy. The topic of financial crisis is highly relevant in terms of policy, as outlined by Kauko (2014). Crises trigger output losses and social costs, with an average production loss of 20% of annual Gross Domestic Product (GDP). It is very important to have a good understanding of the past crisis events, of the mistakes made and to learn the lessons from the crisis that happened over time because, as time showed us, history could repeat itself.

In the last twenty years, the world economy has been faced with a significant number of financial crises, from Latin America, to Asia, from Nordic Countries to East and Central European countries, it all culminated with the financial tsunami which burst in 2007. A new and critical need for the Early Warning Systems has appeared since 2008: an updated EWS that would correctly include in the model the way financial markets are affected by changes in risk factors and risk transmission. Since the Great Financial Crisis, it has come as an evidence the exposure to systemic risk is affected by propagation effects and links among financial institutions which are strongly determined by the structure of the financial system.

Considering the increased complexity of the financial systems and risk associated within, attention is drawn by the specialists that a new EWS tool should be used an orientation rather than a signaling technique. The main role and value of the EWS is providing a systemic overview and functioning as a monitor for the systemic risk. As mentioned in Gramlich et al (2010), the results of an EWS should not be overestimated. However, once critical signals are emitted, the supervisory authorities would need support "on the basis of an expected, but not yet realized, deterioration".

In the present paper we propose a logistic macroeconomic model for panel data with the aim of finding the macroeconomic leading indicators of distress. We carry out two models – for advanced and emerging economies and find which are the macroeconomic variables having the highest weight in the probability of a crisis. The rest of the paper is organized as follows. In section 2 we review literature in what concerns the construction of early warning systems for banking crisis – the role of the EWS, the main concepts and techniques used to model the systems, with reference to the latest findings in the literature; section 3 – gives an overview on the particularities of modeling binary outcomes for panel data – which is the methodology employed in the case study. In section four, we propose two models – one for advanced, one for emerging economies and find which are the macroeconomic indicators of systemic risk. The study is innovative as it includes data for both types of economies and also the whole period of the last 5 years since the burst of the Global Financial Crisis. Including the output gap variable in the list of signals is a new concept in the literature and proves to be a significant early trigger for systemic risk. Section five presents the conclusions.

2. Literature review – early warning systems for systemic banking crisis

As the cost of the most recent financial crisis was estimated at app. USD 12 trillion (reaching 20% of the GDP in most affected countries), the forward-looking instruments of supervisory banks gain more and more importance as the amplitude of financial crisis increases. With the crisis becoming more prominent, the literature on EWS models has grown significantly. However, the existing EWS models failed to predict the recent global crisis and this is mainly due to the fact that they do not fully reflect the way that financial markets are affected by changes in risk factors and risk transmission.

Basically, an early warning system (EWS) has the role of anticipating whether an economy will be affected by a financial crisis by developing a framework which would allow for predicting financial stressful situations. In the literature there are three approaches for constructing an Early Warning System for predicting banking crisis: the bottom-up approach, the aggregate approach and the macroeconomic approach. In the first approach mentioned, the probability of insolvency is estimated for each bank and the signal for systemic instability is triggered when the probability of insolvency becomes significant for a high proportion of the banking assets in the respective economy. For the second approach, the same model is applied to aggregate bank data instead of individual bank data. In what concerns the third approach, the attention is focused on establishing a relationship between economy wide variables, based on the fact that a number of macroeconomic variables are expected to affect the financial system and reflect its condition. The third approach will also be used in the case study that we are proposing in the paper.

Gramlich et al. (2010) make a critical review of earlier EWS literature and highlight the main components of a EWS risk model:

- Risk measures – stress assessment; in the literature this can take the form of a binary index (Kaminsky, Reinhart – 1999; Edison – 2003), three-state index (Bussiere; Fratzscher – 2002) or continous index (Illing Liu – 2003, Hanschel, Monnin – 2005);
- Risk factors – risk indicators – usually chosen between micro risks, macro risks (most cited being the work of Reinhart, Rogoff – 2009) and structural risks;
- The risk model – a theory on how to combine the risk measures and the risk factors. Basically there are two approaches for this: the leading indicator (or signal theory) and data-focused regression models.

The approaches of the EWS models are mainly statistical driven. First models are proposed by Diebold and Rudebusch (1989) for constructing economic indexes. The technique was adapted by Kaminsky and Reinhart (1999) who propose the signal approach: a potential crisis is signaled when a risk factor exceeds a predefined threshold. The threshold is adjusted to balance type I errors (model failed to predict crises when they actually take place) and type II errors (models wrongly predicts crises that do not occur). This technique has also been approached by Borio (2002, 2009). Demirguc-Kunt and Detragiache (1998) are the first to use regression analysis for evaluating the predictive power of risk factors. In their later study (2005), the compare both techniques and conclude that the logit model is the most suitable in assessing financial risk. We also note, the neuro-fuzzy approach of Lin et al (2006) for identifying the drivers of currency crisis and find that this artificial intelligence tool improves the prediction of crisis. Still, the black-box pattern of these methods remains a disadvantage for understanding the big picture of the crisis mechanisms.

Other approaches from the specialty literature include : a non-parametric method based upon K-means clustering to predicting crisis events (Fuertes, Kalotychou, 2004) - in their study they find that the optimal model can be constructed based on the decision-makers preferences regarding the desired trade-off between missed defaults and false alarms; Kalman filter estimation of state space models (Mody, Taylor, 2003) – with the aim of extracting a measure of regional vulnerability for emerging economies; factor model with Markov regime switching dynamics (Chauvet, Dong, 2004) for the prediction of nominal exchange rates in the East Asian countries.

3. Binary outcome models – particularities for panel data

Considering that in our case, the dependent variable takes the form of a binary variable (presence or absence of the crisis event), we will turn our attention to the binary choice models. In this case, the model will have the following form :

$$Prob\ (Y = 1|x) = F\ (x, \beta)$$
$$Prob\ (Y = 0|x) = 1 - F\ (x, \beta)$$

where x is the vector of explanatory factors and β is the vector of parameters that reflect the changes in x on the probability. The problem that arises is to find a suitable model for the function F. If we would use the familiar liner regression model, we would encounter a series of problems. First of all, the disturbances in the model would be heteroscedastic due to the restriction imposed to have the dependent variable 0 or 1. Assuming that this problem can be solved by a GLS estimation, a more serious problem is that we cannot be assured that the predictions in the model will look like probabilities. That is the main reason for which we have to use another type of function, that would have the following properties:

$$\lim_{x'\beta \to +\infty} Prob\ (Y = 1|x) = 1$$

$$\lim_{x'\beta \to -\infty} Prob\ (Y = 1|x) = 0$$

As stated in Greene, in principle, any "proper, continuous probability distribution defined over the real line will suffice". If the normal distribution is being used, the probit model is obtained:

$$Prob\ (Y = 1|x) = \int_{-\infty}^{x'\beta} \Phi(t)\, dt = \Phi\ (x'\beta)$$

Due to its mathematical advantages, the logistic distribution is also often used, determining the logit model:

$$Prob\ (Y = 1|x) = \frac{e^{x'\beta}}{1 + e^{x'\beta}} = \Lambda\ (x'\beta)$$

where $\Lambda\ (\cdot)$ indicates the logistic cumulative distribution function. The question arises on which one of the two models to use. The two distributions have similar bell shaped distributions, with the difference that the tails are heavier in the logistic one. The logistic distribution tends to give larger probabilities to $Y = 1$ for extremely small values of $x'\beta$ than the normal distribution would. Or otherwise said, the conditional probability approaches 0 or 1 at a slower rate in logit than in probit. One would expect to obtain different predictions from the two models if the sample contains very few favorable cases (Y's equal to 1) or very few un-favorable cases (Y's equal to 0). "There are practical reasons for favoring one the other in some cases for mathematical convenience, but it is difficult to justify the choice of one distribution or another on theoretical grounds" (Greene). Most applications would state the models generally give similar results, with the limitations expressed before.

An important thing to note for logit and probit models is that the parameters in the model are not necessarily the marginal effects like in the classical regression models. This happens because the marginal effect of a regressor in the logit model depends not only on the coefficient of that regressor, but also on the value of all regressors in the model. For computing marginal effects, we can evaluate the expression for the samples means of the data or evaluate the marginal effects at every observation and use the sample average of the individual marginal effects.

The literature dedicated to the binary choice models for panel data is rapidly growing. An overview is given in Greene (2011). We distinguish between random and fixed effects models by the relationship existing between the unobserved, individual specific hetereogeneity and the vector of regressors. The effect model has the following form:

$$y_{it}^{*} = x_{it}'\beta + v_{it} + u_i, i = 1, \dots, n; t = 1, \dots, T_i$$

$$y_{it} = 1\ if\ y_{it}^{*} > 0, and\ 0\ otherwise.$$

As per Greene (2011), the assumption that u_i is unrelated to x_{it} produces the random effects model. However, this places a restriction on the distribution of the heterogeneity. If the model permits correlation between u_i and x_{it} , then we have a fixed effects model. The disadvantage of the fixed effect model is that the maximum likelihood estimator becomes inconsistent, while in the random effects model strong assumptions regarding heterogeneity should be made.

4. Case study

In this part of the paper, we propose a framework that could be used a starting point for developing an early warning signals system comprising macroeconomic indicators for monitoring and maintaining financial stability in an economy.

In the first part we describe the data used. As data sources we relied on macroeconomic data publicly available at World Bank and International Monetary Fund. Due to significant discrepancy regarding data availability across countries, but also based on particularities of emergent versus advanced economies, we decided to split the initial sample into two data sets. That is, one data set contains the information for the advanced economies: Austria, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Sweden and Belgium. The variables included for this sample are: Cash deficit, GDP growth, Exports, Stocks, Inflation, Output Gap and Debt. The observation period is 1990 – 2012, that is the entire panel for advanced economies contains 315 observations in 15 groups. The second sample will include data on emerging economies: Bulgaria, Czech Republic, Croatia, Hungary, Iceland, Israel, Lithuania, Poland, Romania, Slovak Republic, Slovenia, Latvia. The variables included for the emergent countries sample are: M2 growth, GDP growth, Exports, Stocks, Inflation. Fewer variables are included due to issues regarding data availability. That is also one of the reasons the observation period is reduced to 1995 – 2012. Another reason for reducing the observation period is the particularities of the emergent economies included in the sample, economies which are mainly from the ex-communist bloc and in the first years of the 1990s developed abnormal values of the macroeconomic indicators. Total panel for the emergent economies contains 216 observations in 12 groups. In the next table we present a detailed description of the indicators included, as they are given on the official sites cited.

The dependent variable used in the model is a binary variable and takes the value 1 if the country has been reported as experiencing a banking crisis in the respective year. Data for the banking crises has been taken from official sources in IMF (Leaven, 2008 and further extended).

Table 1. Indicators description

Indicator	Indicator Description	Observations
GDP growth (annual %)	Annual percentage growth rate of GDP at market prices based on constant local currency. Aggregates are based on constant 2005 U.S. dollars. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources.	Observation period for advanced economies 1990 – 2012; for emergent economies 1995 – 2012.
Cash surplus/deficit (% of GDP)	Cash surplus or deficit is revenue (including grants) minus expense, minus net acquisition of nonfinancial assets. This cash surplus or deficit is closest to the earlier overall budget balance (still missing is lending minus repayments, which are now a financing item under net acquisition of financial assets).	Observation period for advanced economies 1990 – 2012.
Inflation, consumer prices (annual %)	Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at	Observation period for advanced economies 1990 – 2012; for emergent economies

Indicator	Indicator Description	Observations
	specified intervals, such as yearly. The Laspeyres formula is generally used.	1995 – 2012.
Money and quasi money growth (annual %)	Average annual growth rate in money and quasi money. Money and quasi money comprise the sum of currency outside banks, demand deposits other than those of the central government, and the time, savings, and foreign currency deposits of resident sectors other than the central government. This definition is frequently called M2. The change in the money supply is measured as the difference in end-of-year totals relative to the level of M2 in the preceding year.	Observation period for emergent economies 1995 – 2012.
Exports of goods and services (% of GDP)	Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments.	Observation period for advanced economies 1990 – 2012; for emergent economies 1995 – 2012.
Stocks traded, total value (% of GDP)	Stocks traded refers to the total value of shares traded during the period. This indicator complements the market capitalization ratio by showing whether market size is matched by trading.	Observation period for advanced economies 1990 – 2012; for emergent economies 1995 – 2012.
Output gap (% of potential GDP)	Output gaps for advanced economies are calculated as actual GDP less potential GDP as a percent of potential GDP.	Observation period for advanced economies 1990 – 2012.
General government net debt (% of GDP)	Net debt is calculated as gross debt minus financial assets corresponding to debt instruments. These financial assets are: monetary gold and SDRs, currency and deposits, debt securities, loans, insurance, pension, and standardized guarantee schemes, and other accounts receivable.	Observation period for advanced economies 1990 – 2012.

Source: World Bank Data, International Monetary Fund

4.1. Estimation results for the advanced economies

Before estimating the model, we analyze a graphic representation of the variables included. Although all variables experienced a drop in the 2007 – 2008 period, the most representative evolution is the one of the GDP growth. The graphs for the first panels are reproduced in Figure 1. We notice the evolution of the GDP growth for Greece which remains on a descendent path, although the rest of the economies experience a drop in the GDP growth in 2008 followed by a modest recovery in the next years.

Next step is to test the stationarity of the time series included. For this, we apply specific unit – root tests for panel data. For consistency of results we use four tests: Levin – Lin – Chen , Breitung, Im – Pesaran – Shin, Hedri LM Test. In the first two tests the null hypothesis is that the panels contain unit roots with the alternative hypothesis that panels are stationary, while in the last two tests the null hypothesis is that all panels contain unit roots with the alternative hypothesis that some panels are stationary.

Results are presented in Figure 2 (example for a unit root estimation output – results for the test Levin – Lin – Chen applied to GDP growth) and in tables 2, 3 and 4 which summarize the statistics and p-values for the four tests, for all variables included in the analyze.

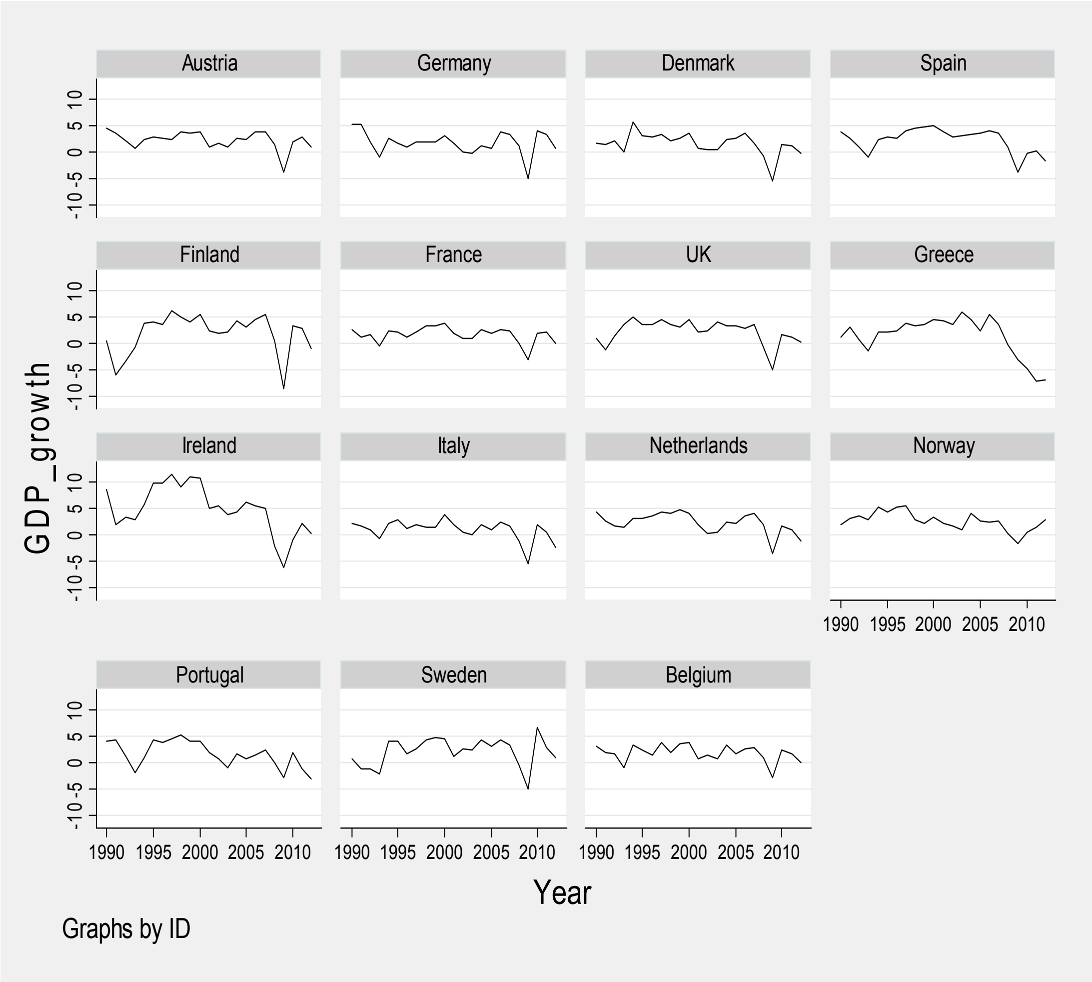


Figure 1. GDP growth evolution in the period 1990 – 2012 for advanced economies

```
Levin-Lin-Chu unit-root test for gdp_growth
-------------------------------------------
Ho: Panels contain unit roots                Number of panels  =     15
Ha: Panels are stationary                    Number of periods =     23

AR parameter: Common                         Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:      Bartlett kernel, 9.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -9.5966
 Adjusted t*         -4.6545        0.0000
------------------------------------------------------------------------------
```

Figure 2. Results of test Levin – Lin – Chu for the GDP growth

Table 2. Results of Unit Root Tests for Cash – Deficit and GDP Growth

Unit Root Test	Statistic	P-Value	Unit Root Test	Statistic	P-Value
Levin-Lin-Chu*	-4,6678	0,0000	Levin-Lin-Chu*	-4,6545	0,0000
Breitung*	-7,5674	0,0000	Breitung*	-5,8305	0,0000
Im-Pesaran-Shin **	-3,2867	0,0005	Im-Pesaran-Shin **	-4,6086	0,0000
Hadri LM test **	7,8664	0,0000	Hadri LM test **	9,6757	0,0000

*null hypothesis panels contain unit roots / alternative hypothesis that panels are stationary

**null hypothesis that all panels contain unit roots / alternative hypothesis that some panels are stationary

Table 3. Results of Unit Root Tests for Exports and Stocks

Unit Root Test	Statistic	P-Value	Unit Root Test	Statistic	P-Value
Levin-Lin-Chu*	-1,6862	0,0459	Levin-Lin-Chu*	-3,9922	0,0000
Breitung*	-0,4516***	**0,3258*****	Breitung*	-3,2177	0,0006
Im-Pesaran-Shin **	-2,8915***	0,0019***	Im-Pesaran-Shin **	-1,6853	0,0460
Hadri LM test **	16,5181 ***	0,0000 ***	Hadri LM test **	23,9276	0,0000

*null hypothesis panels contain unit roots / alternative hypothesis that panels are stationary

**null hypothesis that all panels contain unit roots / alternative hypothesis that some panels are stationary

*** including time trend; if the time trend component were not included, the series would contain unit roots

Table 4. Results of Unit Root Tests for Inflation and Output Gap

Unit Root Test	Statistic	P-Value	Unit Root Test	Statistic	P-Value
Levin-Lin-Chu*	-8,7241	0,0000	Levin-Lin-Chu*	-4,3547	0,0000
Breitung*	-1,3878***	**0,0826*****	Breitung*	-2,9368	0,0017
Im-Pesaran-Shin **	-6,8070	0,0000	Im-Pesaran-Shin **	-2,4560	0,0070
Hadri LM test **	24,7587	0,0000	Hadri LM test **	6,1074	0,0000

*null hypothesis panels contain unit roots / alternative hypothesis that panels are stationary

**null hypothesis that all panels contain unit roots / alternative hypothesis that some panels are stationary

*** including time trend; if the time trend component were not included, the series would contain unit roots

Considering the results above, we can conclude, based on all four tests applied that the variables: Cash Deficit, GDP growth, Stocks, Inflation and Output Gap are stationary (for a significance level of maximum 5%). However, for variable exports, the results of the test show the presence of the unit root if trend component is not included (null hypothesis cannot be rejected – Table 4) and for variable Debt all tests have associated p-values larger than 0.1 concluding that the series is not stationary. For these two variables we take the first difference of the variables and obtain that the resulted series are stationary. Results are summarized in table 5.

Table 5. Results of Unit Root Tests for D(Exports) and D(Debt)

Unit Root Test	Statistic	P-Value	Unit Root Test	Statistic	P-Value
Levin-Lin-Chu*	-8,5419	0,0000	Levin-Lin-Chu*	-2,1655	0,0152
Breitung*	-9,3639	0,0000	Breitung*	-5,3703	0,0000
Im-Pesaran-Shin **	-8,5182	0,0000	Im-Pesaran-Shin **	-4,8709	0,0000
Hadri LM test **	-1,2660***	0,1027***	Hadri LM test **	6,3528	0,0000

*null hypothesis panels contain unit roots / alternative hypothesis that panels are stationary

**null hypothesis that all panels contain unit roots / alternative hypothesis that some panels are stationary

*** including time trend

Next, we begin estimating the models. As stated before, we have the option of estimating logit or probit models with the random or fixed effects (random effects possible only for logistic models). Considering the nature of the "early warning signals" model we are proposing, we include in our list of variables all the variables with lagged for two periods.

However, we obtain that only three of them are significant, that is : the GDP growth with lag one, the Output gap with lag two and the first difference of variable Debt with lag one.

Table 6. Results for the estimation of the logistic model for advanced economies (random effects)

```
Random-effects logistic regression             Number of obs      =       315
Group variable: id                             Number of groups   =        15

Random effects u_i ~ Gaussian                  Obs per group: min =        21
                                                              avg =      21.0
                                                              max =        21

Integration method: mvaghermite                Integration points =        12

                                               Wald chi2(10)      =     49.11
Log likelihood  = -82.941931                   Prob > chi2        =    0.0000
```

bc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
cash_deficit	-.0684804	.0846879	-0.81	0.419	-.2344657	.0975049
gdp_growth	-1.630855	.2672036	-6.10	0.000	-2.154564	-1.107146
stocks	.0132952	.0060939	2.18	0.029	.0013514	.025239
inflation	.0941213	.1648881	0.57	0.568	-.2290534	.417296
outputgap	1.320462	.3052972	4.33	0.000	.7220906	1.918834
d_exports	.0791391	.114939	0.69	0.491	-.1461373	.3044155
d_debt	.1446802	.0426859	3.39	0.001	.0610174	.228343
gdp_growth						
L1.	-1.321017	.2696935	-4.90	0.000	-1.849607	-.7924275
outputgap						
L2.	-1.227266	.2600205	-4.72	0.000	-1.736897	-.7176349
d_debt						
L1.	.0907671	.0441878	2.05	0.040	.0041606	.1773736
_cons	1.843755	.9490806	1.94	0.052	-.016409	3.703918
/lnsig2u	1.019219	.6333285			-.2220818	2.26052
sigma_u	1.664641	.5271323			.8949022	3.096462
rho	.457198	.1571719			.1957724	.7445347

Likelihood-ratio test of rho=0: chibar2(01) =   18.28 Prob >= chibar2 = 0.000

Results for the logit model with random effects are presented in table 6. The model is valid, considering the likelihood-ratio test for rho (p-value = 0.0000). Considering the p-values of the variables included in the sample, at a 0.05 significance level, the following variables are significant: GDP growth and GDP growth lagged one period (both coefficients with negative signs, as expected): Stocks (positive sign), Output Gap and Output Gap lagged two periods (first with a positive sign and second with a negative sign); first difference of the governmental debt and first difference of the governmental debt lagged with one period (both coefficients positive).

The results show that cash deficit, inflation level and variation in exports are not significant early warning signs for predicting crisis. The signs of the significant variables are related to economic theory. A decrease in the GDP growth and the increase in the output gap are the most significant early warning signs for the advanced economies. Also, an increase in the variation of governmental debt (one year prior to crisis) and the increase in volumes of stocks traded can be viewed as early warning indicators, but with smaller contributions to the probability of a crisis appearance.

Apart for this model, we also estimate (using same variables) a logit model with fixed effects. The results of the estimation are presented in a comparative manner in Table 7 below.

Table 7. Comparative results of Logit models (fixed / random effects)

Variable	Coefficient	Std. Error	Model
GDP Growth	-1.6308	0.2672	Logit Random Effects
	-1.6779	0.2690	Logit Fixed Effects
Stocks	0.0132	0.0060	Logit Random Effects
	0.0078	0.0071	Logit Fixed Effects
Output Gap	1.3204	0.3052	Logit Random Effects
	1.6748	0.3572	Logit Fixed Effects
D(Debt)	0.1446	0.0426	Logit Random Effects
	0.1459	0.0396	Logit Fixed Effects
GDP Growth (L1)	-1.3210	0.2696	Logit Random Effects
	-1.4626	0.2862	Logit Fixed Effects
Output Gap (L2)	-1.2272	0.2600	Logit Random Effects
	-1.3160	0.2732	Logit Fixed Effects
D (Debt) (L1)	0.0907	0.0441	Logit Random Effects
	0.1310	0.0463	Logit Fixed Effects

The results are similar for the two types of models. However, considering the estimated probabilities of the model (probability that the outcome is positive), we conclude that the random effects model is much more suitable for the underlying data. The post estimation results are in Table 8. As per IMF statistics used, the only countries that did not experience crisis in 2009 from the advanced economies selected are Finland and Norway. That is, the probability estimated for Norway is very good, but the one estimated for Finland is associated to a crisis situation, although the country has not been reported as so. We also note, the low probability reported for Sweden, although the country has been reported as affected by the crisis. Greece, Italy and Ireland, as well as Portugal have estimated probabilities very close to one – these being the countries the most affected by the crisis, thus with the level of the macroeconomic variables most eroded.

Table 8. Post estimation results for the logistic model – advanced economies (random effects)

Country	**Year**	**Exp Prob**
Austria	2009	0.5582764
Germany	2009	0.8579196
Denmark	2009	0.9491678
Spain	2009	0.7292508
Finland	2009	0.8606029
France	2009	0.9507059
UK	2009	0.9999521
Greece	2009	0.9997895

Country	Year	Exp Prob
Ireland	2009	0.9878655
Italy	2009	0.9950404
Netherlands	2009	0.6415532
Norway	2009	0.0431651
Portugal	2009	0.9893235
Sweden	2009	0.2477488
Belgium	2009	0.5460162

4.2. Estimation results for the emergent economies

The graphic representation of the GDP growth's evolution for the emergent economies in the panel is found in Figure 3 below. The graph analyze is similar with the one that we had for the advanced economies. However, we note some particularities – the countries from the former communist block experienced a drop GDP also in the period 1995 – 1996, due to transition period. Also, Baltic Countries (Latvia and Lithuania) experienced the most severe drops in GDP in the crisis years, as can be easily observed from Figure 3.

In Table 9 and 10 we have the results for the unit root tests applied to the variables M2 growth, GDP growth, Exports, Stocks and Inflation. We find that M2 growth, GDP growth and Stocks are all stationary. For Inflation, all tests (except the Breitung test) confirm the that the variable is stationary. However, considering that for Exports, the null hypothesis that panels contain unit roots cannot be rejected for three of the four tests, we decide to use the first difference of exports in the model – where we accept the stationarity of the variable in three out of four tests (results of unit root tests before and after differentiation are presented in Table 11).

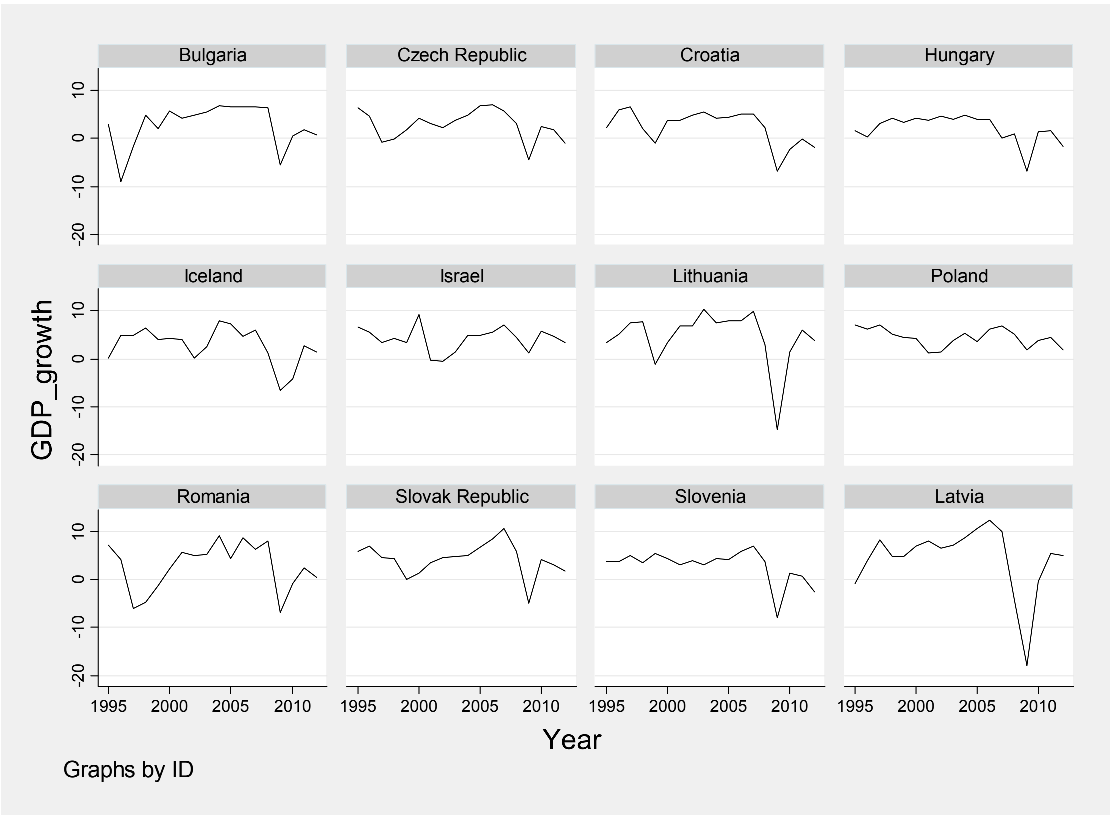


Figure 3. GDP growth evolution in the period 1995 – 2012 for advanced economies

Table 9. Results of Unit Root Tests for M2 Growth and GDP Growth

Unit Root Test	Statistic	P-Value	Unit Root Test	Statistic	P-Value
Levin-Lin-Chu	-3.4923	0.0002	Levin-Lin-Chu	-4.8273	0.0000
Breitung	-2.3557	0.0092	Breitung	-5.5205	0.0000
Im-Pesaran-Shin	-4.1504	0.0000	Im-Pesaran-Shin	-3.7274	0.0001
Hadri LM test	7.7165	0.0000	Hadri LM test	2.0937	0.0181

Table 10. Results of Unit Root Tests for Stocks and Inflation

Unit Root Test	Statistic	P-Value	Unit Root Test	Statistic	P-Value
Levin-Lin-Chu	-4.9641	0.0000	Levin-Lin-Chu	-5.5655	0.0000
Breitung	-3.4122	0.0003	Breitung	0.1903	0.5754
Im-Pesaran-Shin	-1.7552	0.0396	Im-Pesaran-Shin	-4.6811	0.0000
Hadri LM test	6.3827	0.0000	Hadri LM test	3.3155	0.0005

Table 11. Results of Unit Root Tests for Exports and Variation in exports (first difference)

Unit Root Test	Statistic	P-Value	Unit Root Test	Statistic	P-Value
Levin-Lin-Chu	-1.6202	0.0526	Levin-Lin-Chu	-5.0471	0.0000
Breitung	0.5703	0.7158	Breitung	-6.9226	0.0000
Im-Pesaran-Shin	0.3539	0.6383	Im-Pesaran-Shin	-5.5524	0.0000
Hadri LM test	20.7622	0.0000	Hadri LM test	-0.7522	0.7740

In what follows, we proceed to the same steps as for the sample of advanced economies. We estimate the logistic model – with random and fixed effects. As we did previously, we include in the list of variables all the variables lagged for two periods. This time, we obtain that only the GDP growth with lag one, the variation of exports with lag two are significant in the model. The results of the estimation with random effects are presented in Table 12.

The model is valid, considering the likelihood-ratio test for rho (p-value = 0.001). Considering the p-values of the variables included in the sample, at a 0.05 significance level, the following variables remain significant: M2, GDP growth, Variation in Exports and GDP growth lagged one period (both coefficients with negative signs, as expected). Inflation level and stocks, as well as the variation in exports lagged with two periods are not significant early warning signs for predicting crisis in the case of emerging economies. The signs of the significant variables are related to economic theory. A decrease in the GDP growth or a decrease in the money supply can be considered the most significant early warning signals for the emergent economies. In Table 13 we present the post-estimation results for the random effects logistic model. We notice that the model give weaker results than the one for the advanced economies. This could be mainly due to the lower number of variables included in the model. The expected probabilities for the Baltic Countries (Lithuania, Latvia) are, as expected, the most close to one, as these are countries which experienced the most dramatic fall in the economy (as also shown from the graph).

Table 12. Results for the estimation of the logistic model for emerging economies (random effects)

```
Random-effects logistic regression              Number of obs      =       180
Group variable: id                              Number of groups   =        12

Random effects u_i ~ Gaussian                   Obs per group: min =        15
                                                               avg =      15.0
                                                               max =        15

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(7)       =     27.20
Log likelihood  = -56.892479                    Prob > chi2        =    0.0003
```

bc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
m2_growth	-.1219335	.0436757	-2.79	0.005	-.2075362	-.0363308
gdp_growth	-.3857805	.104889	-3.68	0.000	-.5913591	-.1802019
d_exports	.1089936	.0505533	2.16	0.031	.0099109	.2080763
stocks	-.0069754	.0180328	-0.39	0.699	-.0423191	.0283683
inflation	-.0020855	.0561766	-0.04	0.970	-.1121897	.1080186
gdp_growth						
L1.	-.1313661	.0790617	-1.66	0.097	-.2863241	.0235919
d_exports						
L2.	.0880516	.0593502	1.48	0.138	-.0282727	.2043759
_cons	.4819724	.7813385	0.62	0.537	-1.049423	2.013368
/lnsig2u	1.478609	.832565			-.1531886	3.110406
sigma_u	2.094478	.8718945			.9262655	4.736048
rho	.571448	.2038912			.2068471	.8720892

```
Likelihood-ratio test of rho=0: chibar2(01) =    10.26 Prob >= chibar2 = 0.001
```

Table 13. Post estimation results for the logistic model – emerging economies (random effects)

Country	Year	Exp Prob
Bulgaria	2009	0.5983654
Czech	2009	0.7480773
Croatia	2009	0.9052944
Hungary	2009	0.8786198
Iceland	2009	0.9756301
Israel	2009	0.1613406
Lithuania	2009	0.9918979
Poland	2009	0.1706047
Romania	2009	0.7642968
Slovak Republic	2009	0.3201566
Slovenia	2009	0.5925208
Latvia	2009	0.9994201

5. Conclusions

In the present paper, we propose a framework to be used for developing an Early Warning System for assessing systemic risk. We find important insight regarding the macroeconomic variables that could be considered early triggers of banking distress. On one hand, for advanced economies, the cash deficit, the variation in exports and inflation are not significant signals for situation of crisis, while for emerging economies, inflation and value of stocks traded turn out to have no prediction power for predicting crisis (a note should be made here that the indicator value of stocks traded is significant for the advanced economies – this could be explained by the still immature stock market in emerging economies). On the other hand, the evolution on GDP growth is the most important signal for a crisis situation, that is one year prior to crisis eruption. Moreover, the paper ads important contribution to the specialty literature by considering the Output Gap in the model – which is find to be a significant trigger for the inefficiency of the economy and a good predictor of crises. The model has very good estimates of the probability of default, confirming the set of most affected economies by the Financial Crisis (Greece, Italy, Ireland, Portugal, Baltic Countries) and stable economies – the Nordic Countries.

Paper is subject to further development – quarterly data could be used instead on annually for a more dynamic picture of the crisis development; also, instead of the binary variable, a continuous index for banking or financial stability would offer much more information for the economy's evolution.

References

1. Borio, C. and Lowe, P. **Assessing the Risk of Banking Crises,** Bank for International Settlements Quarterly Review, 2002, pp. 43-54
2. Bussiere, M. and Fratzscher, M. **Towards a new early warning system of financial crises,** European Central Bank Working Paper, No. 145, Frankfurt, 2002
3. Chauvet, M. and Dong, F. **Leading indicators of country risk and currency crises: the Asian experience,** Economic Review, Federal Reserve Bank of Atlanta, Issue Q1, 2004, pp. 25-37
4. Davidescu, A. A. M. **Evaluating the relationship between official economy and shadow economy in Romania. A Structural Vector Autoregressive approach,** Journal of Social and Economic Statistics, vol.3, no.2, 2014, pp. 57-65
5. Demirgüç-Kunt, A. and Detragiache, E. **The Determinants of Banking Crises in Developing and Developed Countries,** IMF Staff Papers, Vol. 45, No. 1, 1998, pp. 81-109
6. Diebold, F. X. and Rudebusch, G. D. **Scoring the Leading Indicators,** Journal of Business, Vol. 62, No. 3, 1989, pp. 369-391
7. Edison, H. **Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System,** International Journal of Finance and Economics, Vol. 8, Iss. 1, 2003, pp. 11-53
8. Fuertes A. and Kalotychou, E. **Elements in the Design of an Early Warning System for Sovereign Default,** Computing in Economics and Finance, Society for Computational Economics, Vol. 231, 2004
9. Gramlich, D., Miller, G.L., Oet, M.V. and Ong, S.J. **Early warning systems for systemic banking risk: critical review and modelling implications,** Banks and Bank Systems , Volume 5, Issue 2, 2010
10. Greene, W.H. **Econometric Analysis – 7th Edition,** Prentice Hall, 2011

11. Hanschel, E. and Monnin, P. **Measuring and Forecasting Stress in the Banking Sector: Evidence from Switzerland,** Bank for International Settlements Working Paper, Basel, No. 22, 2005
12. Illing, M. and Liu, Y. **An Index of Financial Stress for Canada,** Bank of Canada Working Paper, Ottawa, No. 14, June 2003
13. Kaminsky, G. and Reinhart, C. **The Twin Crises: The Causes of Banking and Balance-of-Payments Problems,** American Economic Review, Vol. 89, No. 3, 1999, pp. 473-500
14. Kauko, K. **How to forsee banking crises? A survey of the empirical literature,** Economic Systems, 2014 in press
15. Laeven, L. and Valencia, F. **Systemic Banking Crises: a New Database,** IMF Working Paper No. 08, 2008, pp. 1-78
16. Miricescu, E. C. **Investigating the determinants of long-run sovereign rating,** Financial Studies, Volume 18, issue 3, 2014, pp. 25-32
17. Mody A. and Taylor M.P. **Common Vulnerabilities,** Centre for Economic Policy Research (CEPR) Discussion Papers 3759, 2003
18. Moscalu, M. **The impact of interest rate spreads for Euro denominated loans on the leverage ratio of Romanian listed companies,** Proceedings of the 16th International Scientific Conference Finance and Risk, Vol. I, Bratislava, Publishing House Ekonom, 2014, pp. 138–145
19. Zaman, G., Goschin, Z., Partachi, I. and Herteliu, C. **The contribution of labour and capital to Romania's and Moldova's economic growth,** Journal of applied quantitative methods, 2(1), 2007, pp. 179-185
20. Zaman, G., Goschin, Z., and Herteliu, C. **Analysis Of The Correlation Between The Gdp Evolutions And The Capital And Labor Factors In Romaniam,** Romanian Journal for Economic Forecasting, 2(3), 2005, pp. 5-21

[1] **Acknowledgement**
This work was cofinanced from the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/134197 „Performance and excellence in doctoral and postdoctoral research in Romanian economics science domain”.

STOCHASTIC OPTIMIZATION USING INTERVAL ANALYSIS, WITH APPLICATIONS TO PORTFOLIO SELECTION[1]

Silvia DEDU
PhD, Bucharest University of Economic Studies, Department of Applied Mathematics;
Romanian Academy, Institute of National Economy

E-mail: silvia.dedu@csie.ase.ro

Florentin SERBAN
PhD candidate, University of Bucharest, Faculty of Mathematics and Computer Science
PhD, Bucharest University of Economic Studies, Department of Applied Mathematics

Abstract

In this paper we study a class of optimization problems under uncertainty, with parameters modeled by stochastic random variables. Interval analysis and multiobjective stochastic programming concepts are introduced. Then these two concepts are combined to build a stochastic programming model, with the coefficients of the constraints and the coefficients of the objective function modeled by interval numbers and discrete interval random variables. This model can be used to solve a portfolio optimization problem.

Keywords: *interval analysis, multiobjective stochastic programming, uncertainty, optimization*

1. Introduction

The input parameters of the mathematical programming model are not exactly known because relevant data are inexistent or scarce, difficult to obtain or estimate, the system is subject to changes, and so forth, that is, input parameters are uncertain in nature. This type of situations are mainly occurs in real-life decision-making problems. These uncertainties in the input parameters of the model can characterize by interval numbers or random variables with known probability distribution.

The occurrence of randomness in the model parameters can be formulated as stochastic programming (SP) model. SP is widely used in many real-world decision-making problems of management science, engineering, and technology. Also, it has been applied to a wide variety of areas such as, manufacturing product and capacity planning, electrical generation capacity planning, financial planning and control, supply chain management, dairy farm expansion planning, macroeconomic modeling and planning, portfolio selection, traffic management, transportation, telecommunications, and banking

An efficient method known as two-stage stochastic programming (TSP) in which policy scenarios are desired for studying problems with uncertainty. In TSP paradigm, the decision variables are partitioned into two sets. The decision variables which are decided before the actual realization of the uncertain parameters are known as first stage variables. Afterward, once the random events have exhibited themselves, further decision can be made by selecting the values of the second-stage. The formulation of two-stage stochastic programming problems was first introduced by Dantzig [5]. Further it was developed by Barik [2] and Wallcup [15].This article proposes such of stochastic programming

2. Interval Analysis

Interval analysis was introduced by Moore [10] . The growing efficiency of interval analysis for solving various real life problems determined the extension of its concepts to the probabilistic case. Thus, the classical concept of random variable was extended to interval random variables, which has the ability to represent not only the randomness character, using the concepts of probability theory, but also imprecision and non-specifficity, using the concepts of interval analysis. The interval analysis based approach provides mathematical models and computational tools for modeling data and for solving optimization problems under uncertainty.

The results presented in this chapter are discussed in more detail in [1, 11],

Let x^L, x^U be real numbers, $x^L \le x^U$.

Definition 2.1. An *interval number* is a set defined by:

$$X = [x] = \left\{ x \in \mathbf{R} \,|\, x^L \le x \le x^U;\ x^L, x^U \in \mathbf{R} \right\}.$$

Remark 2.1. We will denote by $[x]$ the interval number $\left[x^L, x^U\right]$, with $x^L, x^U \in \mathbf{R}$.

We will denote by $\mathbf{IR}$ the set of all interval numbers.

Definition 2.2. Let (Ω,K,P) be a probability space and **IR** be the set of the real intervals. An interval random variable [X] is an application [X]: Ω→**R**, defined by:
$[X](\omega) = [X^L(\omega), X^U(\omega)]$, where X^L, X^U: Ω→**R** are random variables, such that $X^L \le X^U$ almost surely. We say that the interval random variable [X] is a discrete interval random variable if it takes values in a finite subset of the set of the real numbers Otherwise we say that [X] is a continuous interval random variable.

Definition 2.3. The *product between the real number a and the interval number* $[x]$ is defined by:

$$a \cdot [x] = \{a \cdot x | x \in [x]\} = \begin{cases} \left[a \cdot x^L, a \cdot x^U\right], & \text{if } a > 0 \\ \left[a \cdot x^U, a \cdot x^L\right], & \text{if } a < 0 \\ [0], & \text{if } a = 0 \end{cases}.$$

Let $[x] = \left[x^L, x^U\right]$ and $[y] = \left[y^L, y^U\right]$ be interval numbers, with $x^L, x^U, y^L, y^U \in \mathbf{R}$.

Definition 2.4. The *equality* between interval numbers is defined by:

$[x] = [y]$ if and only if $x^L = y^L$ and $x^U = y^U$.

Definition 2.5.

The *summation* of two interval numbers is defined by: $[x] + [y] = \left[x^L + y^L, x^U + y^U\right]$.

The *subtraction* of two interval numbers is defined by: $[x] - [y] = \left[x^L - y^U, x^U - y^L\right]$.

The *product between* two interval numbers is defined by:

$$[x] \cdot [y] = \left[\min\left\{x^L y^L, x^L y^U, x^U y^L, x^L y^U\right\}, \max\left\{x^L y^L, x^L y^U, x^U y^L, x^L y^U\right\}\right]$$

If $0 \notin [y]$, then $\frac{1}{[y]}$ is defined by: $\frac{1}{[y]} = \left[\frac{1}{y^U}, \frac{1}{y^L},\right]$

If $0 \notin [y]$, then the division between two interval numbers is defined by:

$$\frac{[x]}{[y]} = \left[\min\left\{\frac{x^L}{y^U}, \frac{x^U}{y^L}\right\}, \max\left\{\frac{x^L}{y^U}, \frac{x^U}{y^L}\right\}\right].$$

Definition 2.6. $[x] \le [y]$ if f $x^L \le y^L$ and $x^U \le y^U$;

3. Stochastic Programming

3.1. Multiobjective Stochastic Programming

In stochastic or probabilistic programming some or all of the parameters of the optimization problem are described by stochastic or random variables rather than by deterministic quantities. In recent years, multiobjective stochastic programming problems have become increasingly important in scientifically based decision making involved in practical problem arising in economic, industry, healthcare, transportation, agriculture, military purposes, and technology. Mathematically, a multiobjective stochastic programming problem can be stated as follows:

$$\begin{aligned} &\max \quad z^t = \sum_{j=1}^{n} c_j^t]x_j,\ t = 1,2,\dots,\ T \\ &\text{subject to} \quad \sum_{j=1}^{n} a_{ij} x_j \le b_i,\ i = 1,2,\dots,\ m_1 \\ &\qquad \sum_{j=1}^{n} d_{ij} x_j \le b_{m_1+i},\ i = 1,2,\dots,\ m_2 \\ &\qquad x_j \ge 0,\ j = 1,2,\dots,\ n \end{aligned} \tag{3.1}$$

where the parameters $a_{ij},\ i = 1,2,\dots,m_1,\ j = 1,2,\dots,n$ and $b_i,\ i = 1,2,\dots,m_1$ are discrete random variables with known probability distributions. The rest of the parameters $c_j^t,\ j = 1,2,\dots,n,\ t = 1,2,\dots,T$, $d_{ij},\ i = 1,2,\dots,m_2,\ j = 1,2,\dots,n$ and $b_{m_1+i},\ i = 1,2,\dots,m_2$ are considered as known intervals.

3.2. Multiobjective Two-Stage Stochastic Programming

In two-stage stochastic programming (TSP), decision variables are divided into two subsets:

(1) a group of variables determined before the realizations of random events are known as first stage decision variables, and

(2) another group of variables known as recourse variables which are determined after knowing the realized values of the random events.

A general model of TSP with simple recourse can be formulated as follows [3,9,15]:

$$\max \ \bar{z} = \sum_{j=1}^{n} c_j x_j - E\left(\sum_{i=1}^{m_1} q_i |y_i| \right)$$

$$\text{subject to } y_i = b_i - \sum_{j=1}^{n} a_{ij} x_j, \ i = 1,2,..., m_1$$

$$\sum_{j=1}^{n} d_{ij} x_j \le b_{m_1+i}, \ i = 1,2,..., m_2 \qquad (3.2)$$

$$x_j \ge 0, \ j = 1,2,..., n$$

$$y_i \ge 0, \ i = 1,2,..., m_1$$

where $x_j, j=1,2,...,n$ and $y_i, i=1,2,...,m_1$ are the first stage decision variables and second stage decision variables respectively.

Further, $q_i, i=1,2,...,m_1$ are defined as the penalty costs associated with the discrepancies between $\sum_{j=1}^{n} a_{ij} x_j$ and b_i and E is used to represent the expected value of a random variable.

Multiobjective optimization problems appear in most of the real life decision making problems. Thus, a general model of multiobjective stochastic programming model (3.1) can be stated as follows:

$$\max \ z^t = \sum_{j=1}^{n} c_j^t x_j - E\left(\sum_{i=1}^{m_1} q_i^t |y_i| \right), \ t = 1,2,...,T$$

$$\text{subject to } y_i = b_i - \sum_{j=1}^{n} a_{ij} x_j, \ i = 1,2,..., m_1$$

$$\sum_{j=1}^{n} d_{ij} x_j \le b_{m_1+i}, \ i = 1,2,..., m_2 \qquad (3.3)$$

$$x_j \ge 0, \ j = 1,2,..., n$$

$$y_i \ge 0, \ i = 1,2,..., m_1$$

4. Random Interval Multiobjective Two-Stage Stochastic Programming

Optimization model incorporating some of the input parameters as interval random variables is modeled as random interval multiobjective two-stage stochastic programming (RIMTSP) to handle the uncertainties within TSP optimization platform with simple recourse. Mathematically, it can be presented as follows:

$$\max \ \hat{z}^t = \sum_{j=1}^{n} \left[c_j^t\right] x_j - E\left(\sum_{i=1}^{m_1} q_i^t \left|y_i\right|\right), \ t = 1,2,...,T$$

$$\text{subject to } y_i = [B_i] - \sum_{j=1}^{n} [A_{ij}] x_j, \ i = 1,2,..., \ m_1$$

$$\sum_{j=1}^{n} \left[d_{ij}\right] x_j \le \left[b_{m_1+i}\right], \ i = 1,2,..., \ m_2 \qquad (4.1)$$

$$x_j \ge 0, \ j = 1,2,..., \ n$$

$$y_i \ge 0, \ i = 1,2,..., \ m_1$$

where $x_j, j = 1,2,...,n$ and $y_i, i = 1,2,...,m_1$ are the first stage decision variables and second stage decision variables respectively. Further, $\left[c_j^t\right], \ j = 1,2,...,n, \ t = 1,2,...,T$ are the costs associated with the first stage decision variables and $q_i^t, i = 1,2,...,m_1, t = 1,2,...,T$ are the penalty costs associated with the discrepancie between $\sum_{j=1}^{n}[A_{ij}]x_j$ and $[B_i]$ of the *k*th objective function. The left hand side parameter $[A_{ij}]$ and the right hand side parameter $[B_i]$ are discrete interval random variables, with known probability distributions and between *E* is used to represent the expected value associated with interval random variables.

5. Conclusions

The new approach based on interval analysis provides mathematical models and computational tools for modeling the imprecision of financial data and for solving decision making problems under uncertainty. This article have proposed a random interval multiobjective two-stage stochastic programming

References

1. Alefeld, G., Herzberger, J., **Introduction to Interval Computation**, Academic Press, 1983
2. Barik, S.K, Biswal, M.P. and Chakravartay, D. **Multiobjective Two-Stage Stochastic Programming Problems with Interval Discrete Random Variables**, Advances in Operations Research, Article ID 279181, 2012
3. Birge J.R. and Louveaux F. **Introduction to Stochastic Programming,** Springer, New York, 1997
4. Chinneck J.W. and Ramadan K. **Linear programming with interval coefficients,** Journal of the Operational Research Society, Vol. 51, No. 2, 2000, pp. 209-220
5. Dantzig, G.B **Linear programming under uncertainty**, Management Science 1, pp. 3-4,1955
6. Hansen, E. and Walster, G.W. **Global optimization using interval analysis**, Marcel Decker Inc., 2004
7. Ileanu, B. V., Isaic-Maniu, A. and Herteliu, C. **Intellectual capital components as causes of regional disparities. A case study in Romania,** Romanian Journal of Regional Science, Vol. 3, No. 2, 2009, pp. 39-53

8. Isaic-Maniu, A. an Herteliu, C. **Ethnic and religious groups in Romania-Educational (co) incidences,** Journal for the Study of Religions and Ideologies, No. 12, 2005, pp. 68-75
9. Kambo, N.S. **Mathematical Programming Techniques, Affiliated East-West Press,** New York, 1997
10. Moore, R.E. **Methods and Application of Interval Analysis**, SIAM Philadelphia, 1979
11. Nobibon, F.T. and Guo, R. **Foundation and formulation of stochastic interval programming,** PhD Thesis, African Institute for Mathematical Sciences, Cape Town, South Africa, 2006
12. Raducanu, A. M., Feraru, V., Herteliu, C. and Anghelescu, R. **Assessment of The Prevalence of Dental Fear and its Causes Among Children and Adolescents Attending a Department of Paediatric Dentistry in Bucharest,** OHDMBSC, Vol. 8, No. 1, 2009, pp. 42-49
13. Shapiro, A., Dentcheva D. and Ruszczynski A. **Lectures on Stochastic Programming: Modeling and Theory**, Book Series: MOS-SIAM Series on Optimization 9, 2009, pp. 1-436
14. Toma, A. and Dedu, S. **Quantitative techniques for financial risk assesment: a comparative approach using different risk measures and estimation methods**, Procedia Economics and Finance, Vol. 8, 2014, pp. 712-719
15. Walkup, D.W. and Wets R.J.B. **Stochastic programs with recourse,** SIAM Journal on Applied Mathematics 15, 1967

[1] **Acknowledgement**
This paper has been financially supported within the project entitled "Horizon 2020 - Doctoral and Postdoctoral Studies: Promoting the National Interest through Excellence, Competitiveness and Responsibility in the Field of Romanian Fundamental and Applied Scientific Research", contract number POSDRU/159/1.5/S/140106. This project is co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013. Investing in people!

BOUNDS TEST APPROACH FOR THE LONG RUN RELATIONSHIP BETWEEN SHADOW ECONOMY AND OFFICIAL ECONOMY. AN EMPIRICAL ANALYSIS FOR ROMANIA[1]

Adriana AnaMaria DAVIDESCU[2]
PhD, Lecturer, Department of Statistics and Econometrics,
Bucharest University of Economic Studies,
Researcher, National Scientific Research Institute
for Labour and Social Protection, Bucharest, Romania

E-mail: adrianaalexandru@yahoo.com

Abstract

The paper aims to investigate the nature of the relationship between the shadow economy (SE) and recorded GDP for the case of Romania using Pesaran et al.(2001) bounds tests approach for cointegration for the period 2000-2010. The size of Romanian shadow economy is estimated using a revised version of the currency demand approach based on autoregressive distributed lag (ARDL) approach to cointegration analysis. To investigate the long-run causal linkages and short-run dynamics between shadow economy and recorded GDP, ARDL cointegration approach is applied.

The ARDL causality results revealed only the existence of a long-run unidirectional causality that runs from shadow economy official economy, revealing a negative relationship betwwen them on long-run. In addition, the CUSUM and CUSUMSQ tests confirm the stability of causal relationships.

Keywords: *shadow economy, currency demand approach, economic development, ARDL cointegration approach, CUSUM, CUSUMQ tests*

1. Introduction

The impact of the shadow economy on overall economic performance was investigated in various studies (Dell'Anno, 2003; Schneider and Klinglmair, 2004; Schneider, 2005; Dell'Anno, 2008; Halicioglu and Dell'Anno, 2009).

Klinglmair and Schneider (2006),Giles (1997a, b), and Giles et al. (2002) pointed out that an increase in the size of shadow economy will affect the tax base, leading to lower official growth. Also, an increased size of SE will be very attractive for workers from official sector, creating unfair competition between unofficial and official firms (Enste, 2003). Hidden activities favor corruption and link with criminal activities.

At opposite side, SE can creates positive effects to official economy, creating an extra-added value that Schneider and Enste(2000) consider that can be spent in the official economy, estimating that at least two-third of the income earned in unofficial market is spent in the official economy. Smith (2002) argues that shadow economy have a positive effect on employment, helping some individuals that otherwise will be unemployed and so the unofficial sector may represent a social buffer in the countries with high unemployment rate. Giles (1997a, 1997b, 1999a) and Giles and Tedds (2000) carried out one of the most relevant technique-Granger causality approach in New Zealand and Canada, revealing a significant Granger causality that runs from official economy to unofficial one.

Schneider (2005) quantifies the relationship between SE and official economy, pointing out that the degree of economic development has relevant implications on both sectors. The empirical results have pointed out the existence of a negative between SE and the official economy for developing countries and a positive relationship for industrialized and transition countries, revealing that SE is pro-cyclical for developing economies and countercyclical for developed and transition countries.

In this study, I adopted the definition of Schneider (2006) and Schneider et al. (2010) regarding the shadow economy[3] and the subject of the paper do not deal with typical underground, economic (classical crime) activities, which are all illegal actions that fit the characteristics of classical crimes like burglary, robbery, drug dealing and also exclude the informal household economy which consists of all household services and production.

The main empirical results regarding the Romanian shadow economy are obtained by both national and international studies using different estimation methods and are presented in table 1.

Tabel 1. The size of Romanian shadow economy (% of official GDP)

Authors	Approach	Period	Size of SE (min-max)
Albu(2003, 2008, 2010, 2011)	Discrepancy between actual and desired income	1995-2007	14.6%-22.3%
Institutul Național de Statistică	Labour input method	1998-2009	14.5%-23.5%
Johnson(1997, 1998)	Physical input method	1990-1995	18.0%-28.3%
Lacko(1999)	Physical input method	1990-1995	20.9%-31.3%
Schneider et al.(1998, 2000, 2002, 2004, 2005, 2006, 2007, 2009)	DYMIMIC model and currency demand approach	1990-2005	26.2%-37.4%
Schneider et al.(2010)	MIMIC model	1999-2006	34.4%-36.7%

As Schneider and Enste (2000) stated, no approach is exempt from criticism, the empirical results being different. So, if according to National Institute of Statistics, the informal activity represents between 14.5% and 23.5% of official GDP, Schneider et al.(2010) estimates the size of shadow economy in Romania to overcome the threshold of 35% of official GDP.

The paper aims to investigate the relationship between the size of the shadow economy (SE) and official economy for the case of Romanian data using bounds test approach and ARDL causality analysis for quarterly data covering the period 2000-2010. The size of Romanian shadow economy was estimated using using a revised version of the currency demand approach based on bounds testing approach to cointegration and error correction models, developed within an autoregressive distributed lag (ARDL) framework. A detailed description of the shadow economy estimation is presented in (Davidescu and Dobre 2013).

The empirical results of currency demand approach based on ARDL models emphasizes that there is a general downward trend in the size of the shadow economy as % of official GDP for the period 2000-2010 with an highlight on two low periods, 2003Q1 and 2008Q4.Thus, the size of the shadow economy as % of official GDP measures approximately 45% at the end of 2000 and achieving the value of 37.4% in the last quarters of the period. The estimates are in line with the last empirical studies[4].

It is important to note that because of its undetectable nature and character, it is nearly impossible to measure precisely the size of economic activities taking place in the informal economy of any country in the world, whether developed or less developed. Given this, any theoretical or empirical inference derived from these results should always be regarded as an approximation. In the face of these difficulties, the results drawn from these estimates should be interpreted with due reserve, given the limitations of the methods.

The paper is divided three sections presenting the data, the methodology and the main econometrical results.

2. The relationship between shadow economy and official economy for the case of Romania

Official economic situation plays a crucial role in people's decision to work or not in the informal sector (Bajada and Schneider, 2005; Schneider et al., 2010). In a booming official economy, people have a lot of opportunities to earn a good salary and even extra money. This is noi the case of a economy in recession, when people try to compensate the loss of income from formal economy through involvement in the informal economy.

The shadow economy manifests both positive and negative effects on official economy. The studies of Frey and Weck-Hannemann (1984), Loayza (1996), Kaufmann si Kaliberda (1996), Eilat and Zinnes (2000), Schneider and Enste (2000), Ott (2002), Dell'Anno (2003), Dell'Anno, Gomez and Alañón (2007) and Dell'Anno (2007) argue the existence of a negative effect of shadow economy on GDP growth, based upon the idea that unofficial activities, by creating unfair competition, interfere negatively with the market allocation.

A negative correlation between the size of informal sector and the growth rate of official real GDP per capita for 14 Latin American countries is also found by Loayza(1996), while the same conclusion has been drawed be Eilat and Zinnes (2000) in 24 transition countries, revealing that a one-dollar fall in official GDP was associated with a 31-percent increase in the size of the SE.

Kaufmann and Kaliberda (1996) estimated that for "every 10 percent cumulative decline in official GDP, the share of the irregular economy in the overall increases by almost 4 percent" (ibidem, p. 46).The 76 countries survey conducted by Schneider and Enste's (2000) pointed out that a growing SE has a negative impact on official GDP growth.

At other side, the shadow economy may manifest an positive impact on GDP growth, creating markets, increasing financial resources, enhancing entrepreneurship, and transforming social institutions, economic, legal and necessary capital accumulation (Asea, 1996). The positive realtionship between shadow economy and official one was revealed in studies such as: Adam and Ginsburgh (1985), Giles (1999), Giles and Tedds (2002), Tedds (2005), Schneider and Hametner (2007), Chatterjee, Chaudhuri and Schneider (2003), Dell'Anno (2008), Bovi and Dell'Anno (2007), Dell'Anno and Halicioglu (2010), Schneider and Klinglmair (2004) and Brambilla (2008).

Schneider and Enste (2000) considers the informal economy creates additional value that can be spent in the economy. The informal economy provides employment opportunities to certain individuals who would otherwise be unemployed and provide services to low income people who are involved in informal production activities. Thus, it represents a "social buffer" for countries with high unemployment. Adam and Ginsburg (1985) found a positive relationship between the growth of the SE and the official economy under the assumption of low probability of enforcement.

Enste (2003) argues that SE stimulates economic development in transition countries. He considers the shadow economy as an incentive to develop both the entrepreneurial spirit and a constraint to limit an excessive growth of the government activities. Schneider (2003) emphasizes that UE, stimulating higher competition, leads to more efficient resource allocation on both sides of economy.

Also Dell'anno(2008) has analysed the relationship between unofficial economy (UE) and official GDP, reaveling a positive correlation is found between unofficial and official GDP, SE being considered as beneficial to sustain economic growth.

Halicioglu and Dell'Anno(2009) estimated the size of unrecorded economy (SE) of Turkey over the period 1987-2007 using a revised version of the currency demand approach and analyzed the relationship between UE and recorded GDP (gross domestic product) revealing that causality runs from the recorded GDP to the SE.

In Latin American countries, the study of Maloney(1999) revealed empirical evidence on substantial flow of workers back and forth between formal and informal employment. Galli and Kucera (2003) assess that "informal employment serves as a macroeconomic buffer for formal sector employment over the course of business cycles, with informal employment expanding during downturns and contracting during upturns (ibidem, p. 17)".

In 2005, Schneider considers that the effects of SE on the official economic growth are conditioned to the degree of economic development, revealing a negative relationship for low-income countries and a positive one in industrialized and transition countries. The explanation was that in high-income countries citizens are overburdened by taxes and regulation so that an increasing SE stimulated the official economy as the additional income earned in the SE was spent in the official sector. On the contrary, for low-income countries, an increasing SE "erodes the tax base, with the consequence of a lower provision of public infrastructure and basic public services with the final consequence of lower official economy" (Schneider, 2005, p. 613).

A valuable paper that traits the relationship between official and unofficial economy for the ASEAN from 1996 to 2013 is written by Vo and Pham (2014) who finds that when the official economy is proxied by the GDP growth or the GDP per capita growth, the unofficial economy negatively contributes to the official economy.

2.1. Data

In the econometrical demarche of the investigation of the relationship between shadow and official economies, it has been used quarterly data covering the period 2000:Q1 to 2010:Q2.

The size of Romanian shadow economy as % of official GDP has been obtained using a revised version of the currency demand approach based on bounds testing approach to cointegration and error correction models, developed within an autoregressive distributed lag (ARDL) framework.A detailed description of the shadow economy estimation is presented in (Davidescu & Dobre, 2013).

The empirical results of currency demand approach based on ARDL models emphasizes that there is a general downward trend in the size of the shadow economy as % of official GDP for the period 2000-2010 with an highlight on two low periods, 2003Q1 and 2008Q4.Thus, the size of the shadow economy as % of official GDP measures approximately 45% at the end of 2000 and achieving the value of 37.4% in the last quarters of the period. The estimates are in line with the last empirical studies[5].

The official economy was quantified using real official gross domestic product(2000=100) expressed in millions RON taken from Tempo database of National Institute of Statistics.

The graphical evolution of the shadow economy versus official economy reveals the existence of a negative relationship between variables, intermediate as intensity quantified by a value of -0.65 of correlation coefficient.

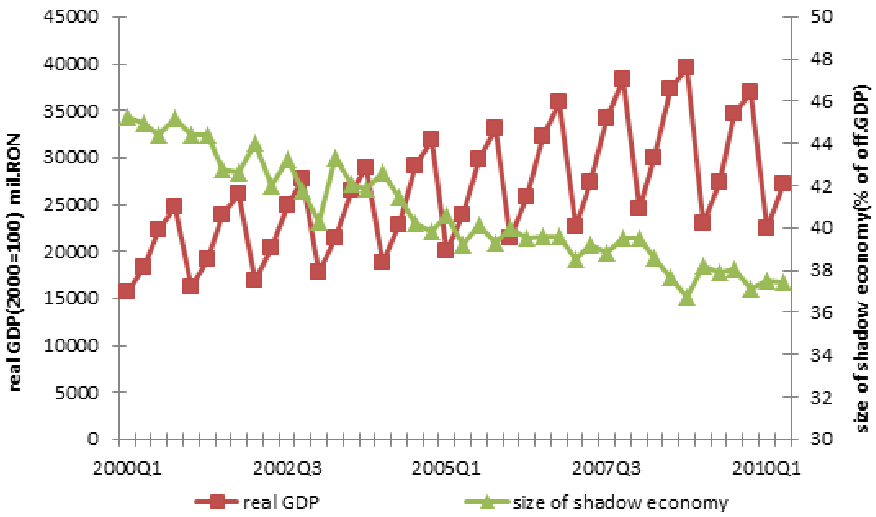


Figure 1. The size of the shadow economy vs. official economy in Romania
Source: Tempo database of National Institute of Statistics

The aim of the paper is to investigate the nature of the relationship between official economy and the size of the Romanian shadow economy and to identify the direction of causality between them using ARDL cointegration and causality approach.

2.2. Methodology

The non-stationary analysis is realised using the the unit root tests (The Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP)). The bounds test approach were applied in prder to verify the possible relationship between these two variables, having the advantage that the regressors can have different order of integration.

The models that describe the relationship between these two variales are:

$$official\_economy_t = \alpha_1 + \beta_1 \cdot SE_t + \varepsilon_{1t} \quad (1)$$

$$SE_t = \alpha_2 + \beta_2 \cdot official\_economy_t + \varepsilon_{1t} \quad (2)$$

where: SE_t is the size of Romanian shadow economy as % of official GDP obtained through ARDL models; the official economy is cuantified using is real GDP expressed in prices of 2000; α_1, α_2 are constants; $\varepsilon_{1t}, \varepsilon_{2t}$ are the disturbance terms.

The first step in the ARDL approach to cointegration is to estimate the following relationship using the OLS estimation technique:

$$\Delta official\_economy_t = a_0 + \sum_{i=1}^{m} a_{1i} \Delta official\_economy_{t-i} + \sum_{i=0}^{m} a_{2i} \Delta SE_{t-i} + \\ + a_3 \cdot official\_economy_{t-1} + a_4 \cdot SE_{t-1} + \varepsilon_{1t} \quad (3)$$

$$\Delta SE_t = b_0 + \sum_{i=1}^{m} b_{1i} \Delta SE_{t-i} + \sum_{i=0}^{m} b_{2i} \Delta official\_economy_{t-i} + \\ + b_3 \cdot SE_{t-1} + b_4 \cdot official\_economy_{t-1} + \varepsilon_{2t} \quad (4)$$

where: Δ is the difference operator; SE_t is the size of Romanian shadow economy as % of official GDP;official economy is expressed using real GDP(2000=1000; ε_{1t} and ε_{2t} are are the disturbance terms ; "m" lags.

The first part of equations (3)-(4) with a_{1i}, a_{2i} and b_{1i}, b_{2i} represents the short-run dynamics of the models and the second part with a_3, a_4 and b_3, b_4 represent the long-run phenomenon.

The null hypothesis in the first equation (3) is $H_0 : a_3 = a_4 = 0$, which means the non-existence of a long-run relationship against the alternative $H_1 : a_3 \neq a_4 \neq 0$ meaning that there is a long-run relationship. In the second equation (4), the null is $H_0 : b_3 = b_4 = 0$ against the alternative $H_1 : b_3 \neq b_4 \neq 0$ which states that we have cointegration. The F tests for the joint significance of the coefficients on the one period lagged levels of the variables is compared with the F critical taken from Pesaran[6] (2001) or Narayan[7] (2005).

Once cointegration is confirmed, we move to *the second stage* and estimate the long-run coefficients of the level equations (1)-(2) and the short-run dynamic coefficients via the following ARDL error correction models[8]:

$$\Delta official\_economy_t = \gamma_0 + \sum_{i=1}^{m} \gamma_{1i} \Delta official\_economy_{t-i} + \\ + \sum_{i=0}^{n} \gamma_{2i} \Delta SE_{t-i} + \gamma_3 ECT_{t-1} + \varepsilon_{1t} \quad (5)$$

$$\Delta SE_t = \lambda_0 + \sum_{i=1}^{m} \lambda_{1i} \Delta SE_{t-i} + \sum_{i=0}^{n} \lambda_{2i} \Delta official\_economy_{t-i} + \\ + \lambda_3 ECT_{t-1} + \varepsilon_{2t} \quad (6)$$

where: SE_t, $official\_economy_t$ are the variables analysed; Δ is the difference operator and ECT_{t-1} is one lag error correction term that maust be negative, γ_3, λ_3 are the adjustement speed to the equilibrium after a shock. The coefficients $\gamma_{1i}, \gamma_{2i}, \lambda_{1i}, \lambda_{2i}$ are the coefficients for the short-run dynamics of the model's convergence to equilibrium, and $\varepsilon_{1t}, \varepsilon_{2t}$ are the error terms. To ascertain the goodness of fit of the ARDL models, diagnostic and stability tests are conducted. The diagnostic test examines the serial correlation, functional form, normality, and heteroscedasticity associated with the model. Parameter stability is important since unstable parameters can result in model misspecification (Narayan and Smith, 2004). The stability of parameters is tested using the Cusum and CusumQ tests.

The third stage includes conducting standard Granger causality tests augmented with a lagged error-correction term. A statistically significant ECT term implies long-run causality running from all the explanatory variables towards the dependent variable.

An augmented form of Granger causality test is involved to the error-correction term and it is formulated in a bi-variate p-th order vector error-correction model (VECM) which is as follows:

$$\Delta SE_t = c_1 + \varphi_{11}^{p}(L)\Delta SE_t + \varphi_{12}^{q}(L)\Delta official\_economy_t + \delta_1 \cdot ECT_{t-1} + \varepsilon_{1t} \quad (7)$$

$$\Delta official\_economy_t = c_2 + \varphi_{21}^{p}(L)\Delta official\_economy_t + \varphi_{22}^{q}(L)\Delta SE_t + \delta_2 \cdot ECT_{t-1} + \varepsilon_{2t} \quad (8)$$

where:

$$\varphi_{11}^{p}(L) = \sum_{i=1}^{P_{11}} \varphi_{11,i}^{p} L^i,\ \varphi_{12}^{p}(L) = \sum_{i=0}^{P_{12}} \varphi_{12,i}^{p} L^i,\ \varphi_{21}^{p}(L) = \sum_{i=1}^{P_{21}} \varphi_{21,i}^{p} L^i\ \ \varphi_{22}^{p}(L) = \sum_{i=0}^{P_{22}} \varphi_{22,i}^{p} L^i$$

SE_t, $official\_economy_t$ are the analysed variables; Δ denotes the difference operator. L denotes the lag operator, where (L)$\Delta Y_t = \Delta Y_{t-1}$, ε_{1t} and ε_{2t} are the disturbance terms.
Iinto a matrix form, the Granger causality looks as follows:

$$\begin{bmatrix} \Delta SE_t \\ \Delta official\_economy_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} \varphi_{11}^{p} & \varphi_{12}^{p} \\ \varphi_{21}^{p} & \varphi_{22}^{p} \end{bmatrix} \begin{bmatrix} \Delta SE_{t-i} \\ \Delta official\_economy_{t-i} \end{bmatrix} + \\ + \begin{bmatrix} \delta_1 ECT_{t-1} \\ \delta_2 ECT_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (9)$$

where: Δ is a difference operator, ECT is the error-correction term from ARDL model, c_i (i = 1, 2) is constant and ε_i (i = 1, 2) are the disturbance terms. The optimal lag length p is based on the Akaike Information Criterion. Long-run causality can be revealed through the significance of the lagged *ECTs* by t test, while short-run causality is validated by using F-statistic or Wald test.

2.2. Empirical results

The main goal of the study is to investigate the nature of the relationship between the shadow economy and the official economy and to identify any possible direction of causality between them. The analysis of stationarity using Dicley-fuller tst revealed that all series are integrated on the same order, I(1).

Forthmore, we investigated the possibility of cointegration between the shadow economy and official one using the bounds tests within the ARDL modeling approach. The optimal lag length[9] required in the bounds test cointegration test has been selected on the both SBC and AIC Information Criteria.

The lag order selected by AIC in the model in which official economy is the dependent variable is $p = 2$ if a trend is included and $p = 4$ if not and those selected by SBC is $p = 2$ irrespective of whether a deterministic trend term is included or not. In view of the importance of the assumption of serially uncorrelated errors for the validity of the bounds tests, the lag $p = 2$ has been selected.

In the model in which shadow economy is the dependent variable, the lag order selected by AIC and SBC is 1, irrespective of whether a deterministic trend term is included or not.

A bounds F-test was applied to equation (4) for shadow economy and official economy to establish a long-run relationship between the variables under the three scenarios: with restricted deterministic trends (F_{IV}), with unrestricted deterministic trends (F_V) and without deterministic trends (F_{III}) and with all intercepts unrestricted. The results are presented in Table 2.

Table 2. The Bounds Test for Co-integration

	With Deterministic Trends			Without Deterministic Trend		
Variables	F_{IV}	F_V	t_V	F_{III}	t_{III}	Conclusion
Off. economy and SE						
$F_{off\ ec}$ (off_ec / SE)						Ho Accepted
p = 2*			-1.65a		-	
3			-1.30a		-	
4			-1.57a		-1.82a	
5			-0.81a		-1.44a	
6					-1.41a	
7					-1.65a	
F_{SE} (SE / off_ec)						Ho Rejected
p = 1*			-5.47c		-4.40c	
2			-3.57b		-2.71b	
3			-2.53a		-1.73a	
4			-2.93a		-1.67a	

Note: Akaike Information Criterion (AIC) and Schwartz Criteria (SC) were used to select the number of lags required in the co-integration test. p shows lag levels and * denotes optimum lag selection in each model as suggested by SBC. F_{IV} represents the F statistic of the model with unrestricted intercept and restricted trend, F_V represents the F statistic of the model with unrestricted intercept and trend, and F_{III} represents the F statistic of the model with unrestricted intercept and no trend. t_V and t_{III} are the t ratios for testing $\sigma_{1Y} = 0$ in equation (4) and $\varpi_{1X} = 0$ in Equation (5) respectively with and without deterministic linear trend. ^{a} indicates that the statistic lies below the lower bound, ^{b} that it falls within the lower and upper bounds, and ^{c} that it lies above the upper bound(Katircioglu, 2009).

The cointegration test under the bounds framework involves the comparison of the F and t statistics against the critical values of F and t for ARDL approach presented in table 3 for the three different scenarios.

Table 3. Critical Values for ARDL Modeling Approach

	90% level		95% level		99% level	
k = 1	I (0)	I (1)	I (0)	I (1)	I (0)	I (1)
F_{IV}	4.05	4.49	4.68	5.15	6.10	6.73
F_V	5.59	6.26	6.5	7.30	8.74	9.63
F_{III}	4.04	4.78	4.94	5.73	6.84	7.84
t_V	-3.13	-3.63	-3.41	-3.95	-3.96	-4.53
t_{III}	-2.57	-2.91	-2.86	-3.22	-3.43	-3.82

Source: Pesaran(2001) for F-statistics pg.300-301 and for t-ratios pg.303-304.

Note: (1) k[10] is the number of independent variables in ARDL models (Erbaykal, 2008), F_{IV} represents the F statistic of the model with unrestricted intercept and restricted trend, F_V represents the F statistic of the model with unrestricted intercept and trend, and F_{III} represents the F statistic of the model with unrestricted intercept and no trend. (2) t_V and t_{III} are the t ratios for testing a3 = 0 in Equation (4) and b3= 0 in Equation (5) respectively with and without deterministic linear trend (Katircioglu, 2009).

Using equations (4)-(5)-each variable is considered as dependent variable in the calculation of the and t-ratios.

When official economy is the dependent variable, the values of t-ratios for each lag lies below the lower bound for all lags, revealing that there is not a level official economy equation, irrespective of trend restrictions. When shadow economy is the dependent variable, for lag 1 irrespective of trend impositions, the values of t-ratios lies outside the 0.01 critical value bounds, and reject the null hypothesis that there is no level shadow economy equation.

Overall, the bounds test results support the existence of a mutual long-run relationship between SE and official economy.

Having cointegrated relationships in bounds tests, the ARDL approach can be now adopted to estimate the level relationship. On the Akaike Selection Criterion, the selected ARDL order is 6 for the official economy and 0 for SE without deterministic trend.

The empirical estimates of level relationship for the ARDL error corection model(lags: 5, 0) revealed that the estimated parameters are statistically significant and the model shows that official economy have inelastic but negative coefficients. In the long run period, the long run elasticity (coefficient of offical economy) is statistically significant. (Prob. =0.00). All five lagged changes in shadow economy are statistically significant, further justifying the choice of p=5.

The equilibrium correction coefficient is estimated as -0.90 (0.173) which is reasonably large and highly significant at 1% level. This shows that Romanian shadow economy coverge to its long run level by 90% by the contribution of official economy. The intercept is not statistically significant and the lagged coefficients in the short term are inelastic, but not totally statistically significant.

The adjusted R^2 is 0.60 suggesting that such error correction model fit the data reasonably well. In addition, the computed F-statistics clearly reject the null hypothesis that all regressors have zero coefficients for all cases. Importantly, the error correction coefficient carries the expected negative sign and are highly significant in both cases. This helps reinforce the finding of cointegration.

Finally, we tested the direction of causality within the conditional Granger causality tests using the ARDL mechanism as a long-run context. The F-statistics for the short-run causations and the t statistics of ECTs for the long-run causations must be statistically significant to achieve Granger causality between the shadow economy and official economy.

Table 4. Results of Granger Causality

Dependent Variable	F-statistics [probability values] Official economy$_t$	SE_t	t-stat (prob) for ECT_{t-1}
Official economy$_t$	-	()	-1.40 [0.18]
SE_t	()	-	-1.80* [0.09]

* denote the rejection of null hypothesis respectively at 0.10 levels.

The empirical results reveal the existence of a long-run unidirectional causality that runs from official economy to shadow economy but in the short run, the lack of F-statistics results does not support short-run causations. We have a Granger causality for long-run period, because the t-statistics for ECT(error correction term) is statistically significant at 10% levels.

Next, we examine the stability of short-run and long-run coefficients, performing the CUSUM and CUSUMQ stability tests for the AIC-based error correction models. The tests applied to the residuals indicate the absence of any instability of the coefficients because the plots of the CUSUMQ and CUSUM statistic are confirmed within the 5% critical bounds of parameter stability.

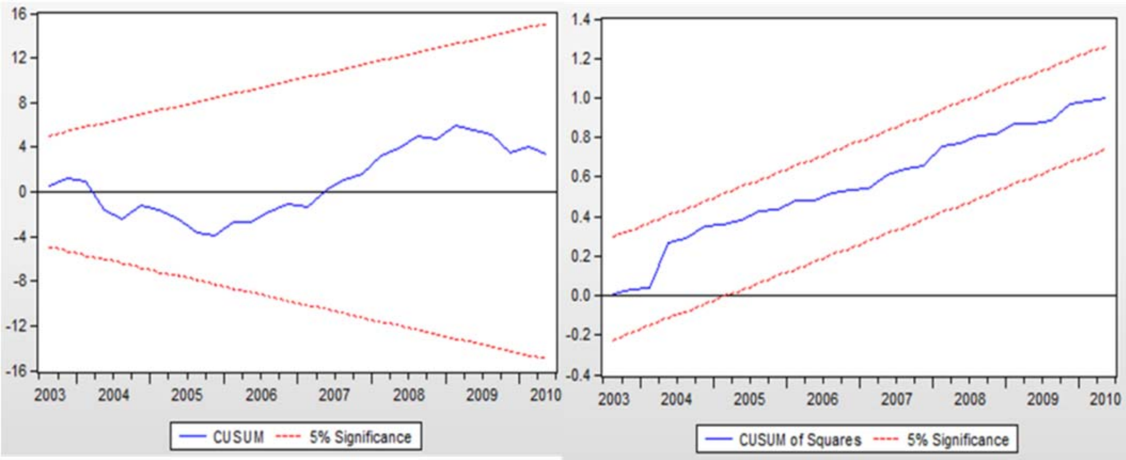


Figure 2. Plots of CUSUM and CUSUMSQ Statistics for Coefficient Stability for the relationship between shadow economy and official economy

Conclusions

In this paper, we investigated the relationship between official economy and the size of the Romanian shadow economy using bounds test approach and ARDL causality analysis for quarterly time series data from 2000-2010. The size of Romanian shadow economy is estimated using a revised version of the currency demand approach based on autoregressive distributed lag (ARDL) approach to cointegration analysis. A detailed description of the estimation process is described in Davidescu and Dobre (2013). The size of the shadow economy as % of official GDP measures approximately 45% at the end of 2000 and achieving the value of 37.4% in the last quarters of the period.

Cointegration test results does not support any short-run relationship between official economy and shadow economy but in the long-run official economy have a negative effect on shadow economy, when it is taken into account a significance level of 10%.

The ARDL causality results revealed the existence of a uni-directional causality that runs official economy to the shadow economy, but only on long-run. The empirical results are in line with the studies of Eilat and Zinnes (2000) for 24 transition countries and Kaufmann and Kaliberda (1996) who estimate a negative impact of official GDP on the size of the shadow economy, mentioning that a decline in official GDP, will lead to an increase in the size of the shadow economy.

References

1. Albu, L.L., Ghizdeanu, I. and Stanica, C. **Spatial Distribution of the Informal Economy. A Theoretical and Empirical Investigation,** SCIENZE REGIONALI, FrancoAngeli Editore, vol. 0(1), 2011, pp. 63-80
2. Davidescu A.A. and Dobre, I. **Revisiting the Relationship between U.S. Shadow Economy and the Level of Unemployment Rate using Bounds Test Approach for Cointegration and Causality**; Economic Computation and Economic Cybernetics Studies and Research, Vol. 46, No. 2, 2012, pp.91-104
3. Dobre, I. and Davidescu, A. **Long-run demand for money and the size of shadow economy in Romania:An application of ARDL model**, Economic Computation and Economic Cybernetics Studies and Research, Vol. 47, No. 3, 2013, pp. 91-110
4. Eilat, Y. and Zinnes, C. **The Evolution of the Shadow Economy in Transition Countries: Consequences for Economic Growth and Donor Assistance**. CAER II Discussion Paper No. 83, Harvard Institute for International Development. Cambridge, MA, 2000
5. Giles, D.E.A. and Tedds, L.M. **Taxes and the Canadian Underground Economy**, Canadian Tax paper 106, Toronto, Canadian Tax Foundation, 2000
6. Katircioglu, S.T. **Tourism, Trade and Growth: The Case of Cyprus**; Applied Economics, Vol. 41, No. 21, 2009, pp. 2741-2750
7. Maer-Matei, M.M. **Measures Of Occupational Mismatch**. SEA-Practical Application of Science, Vol.5, 2014, pp. 425-430
8. Narayan, P. K. and Smyth, R. **The Relationship between the Real Exchange Rate and Balance of Payments: Empirical Evidence for China from Co-integration and Causality Testing**. Applied Economic Letters, Vol. 11, 2004, pp. 287–291
9. Pesaran, M. H., Shin, Y. and Smith, R. J. **Bounds Testing Approaches to the Analysis of Level Relationships**; Journal of Applied Econometrics, Vol. 16, 2001, pp. 289–326
10. Schneider, F. and Buehn, A. **Shadow economies and corruption all over the world: revised estimates for 120 countries,** Economics - The Open-Access, Open-Assessment E-Journal, Kiel Institute for the World Economy, vol. 1, No. 9, 2007, pp. 1-53
11. Schneider, F., Buehn, A. and Montenegro, C. **Shadow Economies All over the World: New Estimates For 162 Countries From 1999 To 2007**, Working Papers wp322, University of Chile, Department of Economics; 2010
12. Strat, V.A. **What Happened with the Attractiveness of the Romanian Counties for FDI during the Period 2001 – 2012?,** Journal of Applied Quantitative Methods, Vol. 9, No. 4, 2014, pp.22-39
13. Vo, D. and Ly, T. **Measuring the Shadow Economy in the ASEAN Nations: The MIMIC Approach**, International Journal of Economics and Finance, Vol. 6, No. 10, 2014, pp.139 – 149
14. Zaman, G., Goschin, Z., Partachi, I. and Herteliu, C. **The contribution of labour and capital to Romania's and Moldova's economic growth,** Journal of applied quantitative methods, 2(1), 2007, pp. 179-185
15. Zaman, G., Goschin, Z., and Herteliu, C. **Analysis Of The Correlation Between The Gdp Evolutions And The Capital And Labor Factors In Romaniam,** Romanian Journal for Economic Forecasting, 2(3), 2005, pp. 5-21
16. Zamfir, A.M., Mocanu, C., Maer-Matei, M.M. and Lungu, E.O. **Immigration and Integration Regimes in EU Countries,** Journal of Community Positive Practices, Vol. 14, No. 1, 2014, pp. 104-115
17. *** Quarterly National Accounts database, Eurostat.
18. *** Quarterly Government Finance Statistics database, Eurostat.
19. *** Quarterly Interest Rates database, Eurostat.
20. *** Quarterly Monetary and Financial Statistics database, Eurostat.

21. *** Tempo database, National Institute of Statistics, www.insse.ro
22. *** Monthly Bulletins of National Bank of Romania, 2000-2010, www.bnr.ro

---

[1] **Acknowledgements**
This work was supported from the European Social Fund through Sectorial Operational Programme Human Resources Development 2007–2013, project number POSDRU/ 159/1.5/S/142115, project title "Performance and Excellence in Postdoctoral Research in Romanian Economics Science Domain".

[2] **Adriana AnaMaria DAVIDESCU (ALEXANDRU)** has graduated the Faculty of Cybernetics, Statistics and Economic Informatics in 2006. She holds a PhD diploma in Economics from 2011 and currently she is lecturer within the Department of Statistics and Econometrics from the Faculty of Cybernetics, Statistics and Economic Informatics and scientific researcher III within National Scientific Research Institute for Labour and Social Protection. Her main topics are analysis of informal economy, economic growth, and unemployment and also labour market studies. She is the author of more than 24 articles in international journals of which 6 in ISI journals, 15 articles published in volumes of international scientific conferences recognized in the country and abroad. She has also participated to more than 26 scientific national and international conferences and 4 summer schools. She has a reach experience in the field of applied statistics and econometrics working in different research projects of which 2 projects in quality of project manager and various as member of research team.

[3] this paper addresses the concept of shadow economy as defined by Schneider (2006) and Schneider et al. (2010) and does not trait the informal sector.

[4] For Schneider et al. (2010), the size of shadow economy in % of official GDP, using the DYMIMIC model is 34.4% in 2000, 35.4% in 2002, 35.9% in 2004, 36.2% in 2005 and 36.7% in 2006.

[5] For Schneider et al. (2010), the size of shadow economy in % of official GDP, using the DYMIMIC model is 34.4% in 2000, 35.4% in 2002, 35.9% in 2004, 36.2% in 2005 and 36.7% in 2006.

[6] Pesaran et al. (2001) have generated critical values using samples of 500 and 1000 observations.

[7] Narayan (2005) argued that these critical values are inappropriate in small samples which are the usual case with annual macroeconomic variables. For this reason, Narayan (2005) provides a set of critical values for samples ranging from 30 to 80 observations for the usual levels of significance.

[8] The Optimal ARDL models are specified on a basis of a set of criteria (Schwarz, Akaike).

[9] The maximum duration of lags for both models has been taken as 7.

[10] k is the number of repressors for the dependent variable in the ARDL models.

THE APPLICATION OF GREY SYSTEM THEORY IN PREDICTING THE NUMBER OF DEATHS OF WOMEN BY COMMITTING SUICIDE- A CASE STUDY

Kalyan MONDAL[1]
MSc, Birnagar High School (HS), Birnagar, Ranaghat, , West Bengal, India

E-mail: kalyanmathematic@gmail.com

Surapati PRAMANIK[2]
PhD, Department Of Mathematics,
Nandalal Ghosh B.T. College, Panpur, West Bengal, India
Corresponding author

E-mail: sura_pati@yahoo.co.in

Abstract:

Sexual harassment, dowry problem, torture, importation of girls, kidnapping, rape and other social problems are forced a woman to commit suicide. These risk factors include man dominated social structure, insecurity of woman, unequal priority level of man and woman, family problems, un-employment etc. Indian constitution offers equal rights for male and female. So the problem of woman suicide becomes a complicated one that restricts the development of country and threatens for the parallelism of male-female ratio. Considering the complexity and uncertainty of the influencing factors on woman suicides, suicide forecasting can be regarded as a grey system with unknown and known information, so it can be analyzed by grey system theory. Grey models require only a limited amount of data to estimate the behavior of unknown systems. In this paper, the original predicted values of woman suicides are separately obtained by the GM (1, 1) model, the Verhulst model and the GM (2, 1) model. The results obtained from these models on predicting woman suicide show that the forecasting accuracy of the GM (1, 1) is better than the Verhulst model and the GM (2, 1) model. Then, the GM (1, 1) model is proposed to predict woman suicide in Indian context.

Key words: Woman suicide, Grey system theory, GM (1, 1) model, Verhulst model, GM (2, 1) model, Forecasting

1. Introduction

Indian civilization is one of the greatest civilizations in the world history. Women suicide is found in the great epic like the Mahabharata and the Ramayana. Committing suicide is a multidimensional, multifaceted malaise. At present India is a developing nation. Indian constitution offers the same rights of man and woman. With the development of economy overall demands regarding all sphere of life of a woman are increasing day by day. Now, in urban life, women have to lead first life in order to meet the demand of her family and other reasons. As a result, they are affected both mentally and physically such as high blood pressure, high blood sugar, stress, hypertension, mental depression etc. In rural and urban sections, there have been an increasing number of cases such as sexual harassment, dowry problem, mental and physical torture, importation of girls, kidnapping, rape, divorce, love affairs, cancellation or the inability to get married (in accordance with the system of arranged marriages in India), illegitimate pregnancy, extra-marital affairs, family conflicts, family problems, illness high expectation, and other unknown problems. These factors are the main causes behind committing suicide. Many young girls lose their deep love affairs and take maximum decision of committing suicide [1][3]. Eight suicides per day are occurred due to poverty and dowry dispute [1]. One suicide out of every five suicides was committed by a housewife [1]. However, the occurrence of woman suicidal cases [1, 2] reflects a rising tendency as a result of the quick growing of alertness. Though the occurrence of woman suicidal case is occasional, it can be predicted scientifically based on the related statistical indexes. Accurate prediction of the woman suicide is important not only for government's policy, but also for social organizations that are devoted to deal with woman's problems.

Grey system theory proposed by Deng [3] in 1982 is a powerful theory for dealing with partially known and partially unknown information. The concept of the grey system theory is used in several fields such as rainfall prediction [4], industry [5], business [6] and geological systems studies [7], environmental studies [8], decision making [9], etc. As an essential part of grey system theory, grey forecasting models [10] are popularly used in time -series forecasting because of its simplicity and ability and high precision to characterize an unknown system by using a few data points [11, 12].

In recent years, the grey system theory has been widely used to forecast in various fields such as grey prediction model for traffic demand [13], electricity demand [14], and internet access population [15].

In review of literature, no prediction model for women suicide is still found. In this paper, the original predicted values of woman suicides are separately obtained by using the GM (1, 1) model [16], the Verhulst model [17] and the GM (2, 1) model [18]. The results of these models on predicting woman suicide are compared. Then, the GM (1, 1) model is proposed to predict woman suicide accidents in Indian context.

Rest of the paper is organized as follows: Section 2 presents mathematical presentation of three grey prediction models. Section 3 is devoted to present case study in Indian context. Section 4 presents concluding remarks.

2. Mathematical Presentation Of Prediction Models

2.1. The GM (1, 1) Model [16]

The most commonly used grey forecasting model is GM (1, 1), which indicates that one variable is employed in the model. The first order differential equation is adopted to match the data generated by the accumulation generating operation (AGO).

For the algorithm of GM (1, 1), the raw data sequences is presented as follows:

$$X^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\right\} \quad (1)$$

Here n is the total number of modeling data. The AGO formation of $X^0(1)$ is defined as follows:

$$X^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\right\} \quad (2)$$

Here,

$$x^{(1)}(k) = \sum_{j=1}^{k} x^{(0)}(j),\ k = 1,2,\ldots, n \quad (3)$$

The GM (1, 1) model can be formed by establishing a first order differential equation for $X^{(1)}(k)$ as follows:

$$\frac{dX^{(1)}}{dt} + a X^{(1)} = u \quad (4)$$

Here, the parameters a, u are called the developing coefficient and grey input respectively.

In practical, the parameters a, u are not calculated directly from the equation (4). Therefore, the solution of the equation (4) can be obtained by using the least square method as follows:

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{u}{\hat{a}}\right]e^{-ak} + \frac{u}{\hat{a}} \quad (5)$$

Here $\hat{a} = [a,u]^T = \left(B^T B\right)^{-1} B^T Y$ and

$$B = \begin{bmatrix} -\frac{1}{2}\left(x^{(1)}(1) + x^{(1)}(2)\right) & 1 \\ -\frac{1}{2}\left(x^{(1)}(2) + x^{(1)}(3)\right) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left(x^{(1)}(3) + x^{(1)}(4)\right) & 1 \end{bmatrix} \quad (6)$$

$$Y_N = \left(X^{(0)}(2), X^{(0)}(3), \cdots, X^{(0)}(n)\right)^T \quad (7)$$

Applying the inverse accumulated generation operation (IAGO), the obtained solution is presented by:

$$\hat{x}^{(0)}(k) = \left[x^{(0)}(1) - \frac{u}{a}\right]\left(1 - e^{a}\right)e^{-a(k-1)} \quad (8)$$

Here $\hat{x}^{(1)}(1) = x^{(0)}(1)$ and k = 2, 3,..., n.

2.2 The Grey Verhulst Model [17]

The Verhulst model [17] was first introduced by a German biologist Pierre Franois Verhulst. The main purpose of Velhulst model is to restrict the whole development for a real system. For an initial time sequence,

$X^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\right\}$, the initial sequence $X^{(0)}$ is used to construct the Verhulst model directly as follows:

$$\frac{dX^{(0)}}{dt} + a X^{(0)} = u\left(X^{(0)}\right) \quad (9)$$

Here a presents the development coefficient and u denotes the grey action quantity. The solution of the parameter vector $\hat{a} = [a,u]^T$ can be obtained by using the least square method.

Here $\hat{a}=\left[(A\vdots B)^{T}(A\vdots B)\right]^{-1}(A\vdots B)^{T}Y$

and,

$$A=\begin{bmatrix} -\frac{1}{2}(x^{(0)}(1)+x^{(0)}(2)) \\ -\frac{1}{2}(x^{(0)}(2)+x^{(0)}(3)) \\ \vdots \\ -\frac{1}{2}(x^{(0)}(n-1)+x^{(0)}(n)) \end{bmatrix};\ B=\begin{bmatrix} \left[\frac{1}{2}(x^{(0)}(1)+x^{(0)}(2))\right]^2 \\ \left[\frac{1}{2}(x^{(0)}(2)+x^{(0)}(3))\right]^2 \\ \vdots \\ \left[\frac{1}{2}(x^{(0)}(n-1)+x^{(0)}(n))\right]^2 \end{bmatrix} \quad (10)$$

$$Y=\left[x^{(0)}(2)-x^{(0)}(1),\ x^{(0)}(3)-x^{(0)}(2),\ \cdots,\ x^{(0)}(n)-x^{(0)}(n-1)\right]^{T} \quad (11)$$

The re-solution of (9) can be presented as follows:

$$\hat{x}^{(0)}(k+1)=\frac{ax^{(0)}(1)}{ux^{(0)}(1)+(a-ux^{(0)}(1))e^{ak}} \quad k = 0, 1, 2,\dots, n \quad (12)$$

2.3 The GM (2, 1) Model [18]

The GM (2, 1) model is a single sequence second-order linear dynamic model and is fitted by differential equations.

Let us assume that an original sequence $X^{(0)}$ be

$X^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\right\}$.

A new sequence $X^{(1)}$ is generated by the AGO as follows:

$X^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\right\}$, here

$$X^{(1)}(k) = \sum_{j=1}^{k} x^{(0)}(j),\ k = 1, 2,\dots, n \quad (13)$$

Now the differential equation of GM (2, 1) model can be presented as follows:

$$\frac{d^2X^{(1)}}{dt^2}+a\frac{dX^{(1)}}{dt}=u \quad (14)$$

$$\hat{a}=[a,u]^T=\left(B^T B\right)^{-1}B^T Y$$

$$B=\begin{bmatrix} -x^{(0)}(2) & 1 \\ -x^{(0)}(3) & 1 \\ \vdots & \vdots \\ -x^{(0)}(n) & 1 \end{bmatrix};\ Y=\begin{bmatrix} (x^{(0)}(2)-x^{(0)}(1)) \\ (x^{(0)}(3)-x^{(0)}(2)) \\ \vdots \\ (x^{(0)}(n)-x^{(0)}(n-1)) \end{bmatrix} \quad (15)$$

From the equation (14), we have

$$\hat{x}^{(1)}(k+1)=\left(\frac{u}{a^2}-\frac{x^0(1)}{a}\right)e^{-ak}+\frac{u}{a}(k+1)+\left(x^0(1)-\frac{u}{a}\right)\left(\frac{1+a}{a}\right) \quad (16)$$

The prediction values of original sequence can be obtained by applying inverse AGO to $\hat{x}^{(1)}$ as follows:

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k),\ \text{here } k=1, 2,\dots, n \quad (17)$$

3. Case Study

In this section, the GM (1, 1) [16], the Verhulst model [17] and the GM (2, 1) [18] are used for comparison. The woman suicide data [1, 2] in India from 2008 to 2013 is used to demonstrate the effectiveness and practicability of the models. The data of women suicide in 2006-2010 is presented to form the three grey prediction models and the data of women suicide from 2011 to 2013 is used as data set to compare the accuracy of the three prediction models.

The evaluation criterion is the mean relative percentage error (MRPE), which measures the percentage of prediction errors. MRPE can be presented as follows:

$$\mathrm{MRPE}=\frac{1}{n}\sum_{k=1}^{n}\left[\left|x^{(0)}(k)-\hat{x}^{(0)}(k)\right|/x^{(0)}(k)\right] \tag{18}$$

In the GM (1, 1) model values of the essential terms are presented as follows:

$$B=\begin{bmatrix}-44488 & 1\\ -44788 & 1\\ -45193 & 1\\ -45780 & 1\\ -47601 & 1\\ -49801 & 1\\ -51048 & 1\end{bmatrix};\ Y=\begin{bmatrix}44750\\ 44825\\ 45560\\ 46000\\ 49201\\ 50400\\ 51695\end{bmatrix};\ \hat{a}=[a,u]=[-0.02,\ -20]^{T} \tag{19}$$

In the Verhulst model values of the essential terms are presented as follows:

$$(A\vdots B)=\begin{bmatrix}-44488 & 1979182144\\ -44788 & 2005964944\\ -45193 & 2042407249\\ -45780 & 209580840\\ -47601 & 2265855201\\ -49801 & 2480139601\\ -51048 & 2605898304\end{bmatrix}\quad Y=\begin{bmatrix}525\\ 75\\ 735\\ 440\\ 3201\\ 1199\\ 1295\end{bmatrix}\quad \hat{a}=[a,u]=[-0.01,\ 1/4428000]^{T} \tag{20}$$

In the GM (2, 1) model values of the essential terms are presented as follows:

$$B=\begin{bmatrix}-44750 & 1\\ -44825 & 1\\ -45560 & 1\\ -46000 & 1\\ -49201 & 1\\ -50400 & 1\\ -51695 & 1\end{bmatrix},\ Y=\begin{bmatrix}525\\ 75\\ 735\\ 440\\ 3201\\ 1199\\ 1295\end{bmatrix}\quad \hat{a}=[a,u]=[-0.21,\ -9000]^{T} \tag{21}$$

The real and forecasted values are shown in 'Table1' to compare the three model accuracy and relative error. The corresponding calculated results (the mean error in the different stage) are shown in Table2.

Table1 indicates that the GM (1, 1) prediction model is smaller than the others by comparing the relative error. From Table2, it is seen that the MRPE of the GM (1, 1) model, the Verhulst model and the GM (2, 1) from 2011 to 2013 are 1.360%, 2.007% and 1.503%, respectively. The effectiveness and accuracy of GM (1, 1) model is higher than the Verhulst model and the GM (2, 1) model.

Table1. Model values and prediction error of the woman suicidal case in India

	Year	Real Value	GM(1,1)		Verhulst		GM(2,1)	
			Model value	Error R (%)	Model value	Error R (%)	Model value	Error R (%)
Model set up stage	2006	44225	44225	0	45128	-2.04	44225	0
	2007	44750	45098	-0.78	46067	-2.94	46103	-3.02
	2008	44825	45989	-2.6	47048	-4.96	46542	-3.83
	2009	45560	46898	-2.94	48071	-5.51	47084	-3.35
	2010	46000	47825	-3.97	49139	-6.82	47751	-3.81
Post set up stage	2011	49201	48771	0.87	50256	-2.14	48576	1.27
	2012	50400	49736	1.32	51424	-2.03	49592	1.60
	2013	51695	50721	1.88	52649	-1.85	50846	1.64

Table2. Error results for the different prediction models

Stage	GM(1,1)	Verhulst	GM(2,1)
	MRPE (%)	**MRPE (%)**	**MRPE (%)**
2008-2010	2.058	4.454	2.802
2011-2013	1.36	2.007	1.503

The comparison of Table1 and Table2 show that the GM (1, 1) model and the GM (2, 1) model have the better forecasting precision in 2006-2010, but the GM (1, 1) prediction model offers the lowest post-forecasting errors and it is more suitable to make a short-term prediction, so the GM (1, 1) model is used to predict women suicide for 2014 and 2015 in India. In Table 3, the comparison between the real values and predicted values obtained from GM (1, 1) model (see Figure1) for women suicides in India.

Table3. The result of forecasting

Year	2010	2011	2012	2013	2014	2015
Real values	46000	49201	50400	51695	---	---
GM (1, 1) Model values	47825	48771	49736	50721	51552	52253

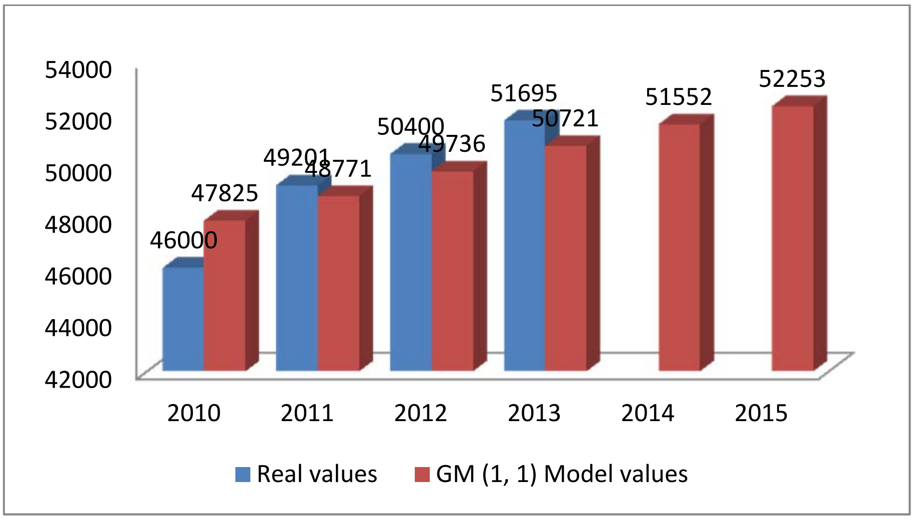


Figure 1. Comparison of real values and predicted values obtained from GM (1, 1) model

4. Conclusion

Committing suicide is an important issue in social context. This paper demonstrates how the grey system theory deals with prediction problem with incomplete or unknown information with large sample. In this paper, we compare the performance of the accuracy of the three grey forecasting models to predict women suicide in India. This paper demonstrates that performance of the GM (1, 1) model in prediction is better than the other two prediction models because it has the merits of both simplicity of application and high forecasting precision. Therefore, we suggest to using the GM (1, 1) model to predict the number of suicides in India and other countries for planning and other issues.

References

1. Guoa, Z., Song, X. and Ye, J. **Verhulst model on time series error corrected for Port through put forecasting,** Journal of East Asia Society of Transport Study, Vol. 6, 2005, pp. 881-891
2. Julong, D. **Control problems of grey system,** System Control Letters, Vol. 1, No. 5, 1982, pp. 288-294
3. Julong, D., **Introduction to grey system theory,** Journal of Grey System, Vol. 1, 1989, pp. 1-24
4. Kayacan, E., Kaynak, O. and Ulutas, B. **Grey system theory- based models in time series prediction,** Expert Systems with Application, Vol. 37, 2010, pp. 1784-1789
5. Lan, J. and Cheng, H. **The grey system and prediction of geological and mineral resources,** Mathematical Geology, Vol. 24, No. 6, 1999, pp. 653-662
6. Lin, Y.H., Chiu, C.C., Lin, Y.J. and Lee, P.C., **Rainfall prediction using innovative grey model with the dynamic index,** Journal of Marine Science Technology, Vol. 21, No. 1, 2013, pp. 63-75
7. Luo, Y. and Che, X., **Improvement and application of initial value of non-equidistant GM (1,1) Model,** International Journal of Computer Science Issues, Vol. 10, No. 2, 2013, pp. 113-118
8. Mohammadi, A., Moradi, L., Talebnejad, A. and Nadaf, A. **The use of grey system theory in predicting the road traffic accident in Fars province in Iran,** Australian Journal of Business Management Research, Vol. 9, 2011, pp. 18-23
9. Niu, W., Zhai, Z., Wang, G., Cheng, J. and Guo, Y., **Adaptive multivariable grey prediction model,** Journal of Information Computer Science, Vol. 8, No. 10, 2011, pp. 1801-1808
10. Pramanik, S. and Mukhopadhyaya, D., **Grey relational analysis based intuitionistic fuzzy multi-criteria group decision-making approach for teacher selection in higher education,** International Jourrnal of Computer Application, Vol. 34, No. 10, 2011, pp. 21-29
11. Quanping, H. and Xiaoyi, Y. **Base a EMD-grey model for textile export time series prediction,** International Journal of Data Theory Application, Vol. 6, No. 6, 2013, pp. 29-38
12. Tseng, F.M. and Tzeng, G.H. **The Comparison of four kinds of prediction methods: ARIMA, fuzzy time series, fuzzy regression time series and grey forecasting: an example of production value forecasting of the mechanical industry in Taiwan,** Journal of Chinese Grey System Association, Vol. 2, No. 2, 1999, pp. 83-98
13. Wu, W.Y. and Chen, S.P., **A prediction method using the grey model GM (1, n) combined with the grey relational analysis: a case study on internet access population forecast,** Applied Mathematics and Computation, Vol. 169, No. 1, 2005, pp. 198-217
14. Yang, C.M., Chen, J.C., Peng, L.P., Yang, J.S. and Chou, C.H., **Earthquake-caused landslide and grey prediction for vegetation recovery,** Botanical Bulletin of Academia Sinica, Vol. 43, 2002, pp. 69-75
15. Yang, Z. and Zhang, Y. **Traffic demand forecasting for EGCS with grey theory based multi-model method,** International Journal of Computer Science Issues, Vol. 1, No. 1, 2013, pp. 61-67

16. Zhou, P., Ang, B.W. and Poh, K.L. **A trigonometric grey prediction approach to forecasting electricity demand,** Energy Conversion Management, Vol. 31, 2006, pp. 2839-2847
17. * * * Accidental deaths and suicides in India, **National crime records bureau ministry of home affairs government of India.** R. K. Puram, New Delhi, 2010, http://ncrb.gov.in , accessed 11 July 2014
18. * * * **Suicides in India,** 2012, http:// *ncrb.nic.in/CD-ADSI-2012/suicides-11.pdf* , accessed 12 July 2014

---

[1] **Kalyan Mondal** (M.Sc., B. Ed.) passed B. Sc. Honors and M. Sc in Mathematics in 2003 and 2005 respectively from the University of Calcutta and University of Kalyani. Currently, he is an assistant Teacher of mathematics at Birnagar High School (HS), Birnagar, Ranaghat, Nadia, Pin Code: 741127, West Bengal, India. He has coauthored more than six research papers. His field of research interests includes fuzzy goal programming, grey system theory, intuitionistic Fuzzy decision making and neutrosophic decision making.

[2] **Dr. Surapati Pramanik** (Ph. D., M.SC., M. Ed.) did his B. Sc. and M. Sc. in Mathematics from University of Kalyani. He received Ph. D. in Mathematics in 2010 from Bengal Engineering and Science University (BESU) Shibpur, India. He is currently an Assistant Professor of Mathematics at the Nandalal Ghosh B. T. College, Panpur, P.O.-Narayanpur, West Bengal, India. He has authored/co-authored more than 50 research papers in international journals; He has published one mathematics method book from Aheli publishers. He coauthored five books for B. Ed. Courses from Aheli Publisher, Kolkata, India. His research interests include operations research and optimization, soft computing, grey system theory, neutrosophic decision making, rough sets, mathematics education, comparative education, international relation.

[3] Codification of references within text:

[1]	* * * Accidental deaths and suicides in India, **National crime records bureau ministry of home affairs government of India.** R. K. Puram, New Delhi, 2010, http://ncrb.gov.in , accessed 11 July 2014
[2]	* * * **Suicides in India,** 2012, http:// *ncrb.nic.in/CD-ADSI-2012/suicides-11.pdf* , accessed 12 July 2014
[3]	Julong, D. **Control problems of grey system,** System Control Letters, Vol. 1, No. 5, 1982, pp. 288-294
[4]	Lin, Y.H., Chiu, C.C., Lin, Y.J. and Lee, P.C., **Rainfall prediction using innovative grey model with the dynamic index,** Journal of Marine Science Technology, Vol. 21, No. 1, 2013, pp. 63-75
[5]	Tseng, F.M. and Tzeng, G.H. **The Comparison of four kinds of prediction methods: ARIMA, fuzzy time series, fuzzy regression time series and grey forecasting: an example of production value forecasting of the mechanical industry in Taiwan,** Journal of Chinese Grey System Association, Vol. 2, No. 2, 1999, pp. 83-98
[6]	Quanping, H. and Xiaoyi, Y. **Base a EMD-grey model for textile export time series prediction,** International Journal of Data Theory Application, Vol. 6, No. 6, 2013, pp. 29-38
[7]	Lan, J. and Cheng, H. **The grey system and prediction of geological and mineral resources,** Mathematical Geology, Vol. 24, No. 6, 1999, pp. 653-662
[8]	Yang, C.M., Chen, J.C., Peng, L.P., Yang, J.S. and Chou, C.H., **Earthquake-caused landslide and grey prediction for vegetation recovery,** Botanical Bulletin of Academia Sinica, Vol. 43, 2002, pp. 69-75
[9]	Pramanik, S. and Mukhopadhyaya, D., **Grey relational analysis based intuitionistic fuzzy multi-criteria group decision-making approach for teacher selection in higher education,** International Jourrnal of Computer Application, Vol. 34, No. 10, 2011, pp. 21-29
[10]	Niu, W., Zhai, Z., Wang, G., Cheng, J. and Guo, Y., **Adaptive multivariable grey prediction model,** Journal of Information Computer Science, Vol. 8, No. 10, 2011, pp. 1801-1808
[11]	Julong, D., **Introduction to grey system theory,** Journal of Grey System, Vol. 1, 1989, pp. 1-24
[12]	Kayacan, E., Kaynak, O. and Ulutas, B. **Grey system theory- based models in time series prediction,** Expert Systems with Application, Vol. 37, 2010, pp. 1784-1789
[13]	Yang, Z. and Zhang, Y. **Traffic demand forecasting for EGCS with grey theory based multi-model method,** International Journal of Computer Science Issues, Vol. 1, No. 1, 2013, pp. 61-67
[14]	Zhou, P., Ang, B.W. and Poh, K.L. **A trigonometric grey prediction approach to forecasting electricity demand,** Energy Conversion Management, Vol. 31, 2006, pp. 2839-2847
[15]	Wu, W.Y. and Chen, S.P., **A prediction method using the grey model GM (1, n) combined with the grey relational analysis: a case study on internet access population forecast,** Applied Mathematics and Computation, Vol. 169, No. 1, 2005, pp. 198-217
[16]	Luo, Y. and Che, X., **Improvement and application of initial value of non-equidistant GM (1,1) Model,** International Journal of Computer Science Issues, Vol. 10, No. 2, 2013, pp. 113-118
[17]	Guoa, Z., Song, X. and Ye, J. **Verhulst model on time series error corrected for Port through put forecasting,** Journal of East Asia Society of Transport Study, Vol. 6, 2005, pp. 881-891
[18]	Mohammadi, A., Moradi, L., Talebnejad, A. and Nadaf, A. **The use of grey system theory in predicting the road traffic accident in Fars province in Iran,** Australian Journal of Business Management Research, Vol. 9, 2011, pp. 18-23

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

METHODS OF MEASURING CORE INFLATION IN INFLATION TARGETING COUNTRIES

Ion PARTACHI
PhD, University Professor,
Academy of Economic Studies, Kishinev, Moldova

E-mail: ipartachi@ase.md

Vitalie MOTELICA
Academy of Economic Studies, Kishinev, Moldova

Abstract:

This article tackles the issue of the transitory effects on growth of prices and the main methods for the calculation of monetary policy relevant inflation in inflation targeting countries. In order to have an indicator which would capture the medium term inflation pressures, several measures of core inflation are considered. Besides that this article mentions the main advantages and disadvantages of the different measures of medium term inflation pressures. Most of the central banks which are implementing the inflation targeting regime are using core inflation indicators based on the exclusion of certain components, however there are several central banks which are using statistical measures of core inflation such as the trimmed mean and weighted median. This article also describes the existing core inflation indicators as well as the main features of the trimmed mean and weighted median for the Republic of Moldova. Even though the obtained indicators of core inflation have different values, they have similar trajectories.

Key words: core inflation, statistical measures, trimmed mean, weighted median, monetary factors

1. Introduction

Within the process of achieving their main objective of price stability, central banks are usually monitoring, analyzing and forecasting the dynamics of the Consumer Price Index, even though there are many other indicators which can provide information about the change of prices such as Producer Price Indices, GDP and consumption deflators etc. The reason behind this is the fact that the CPI has several advantages such as the fact that it is known to the public, it is based on the expenditures made by the households and it is disseminated on a monthly basis just several days after the end of the reference period. Besides its advantages, CPI has several shortcomings. It doesn't reveal the growth of prices which is caused by monetary factors. In other words, some price changes might be determined by some sectorial shocks. These have transitory effects on the general price level and are not considered a part of the inflationary process. In this way, food prices might grow due to some bad weather conditions which determined a bad harvest or the fuel prices might grow once the fiscal authority increases the excise taxes. The direct effect of these changes is temporary and is not a part of the general inflation trend.

A solution for detecting the trend of the inflation in the economy which would reveal the monetary inflation and which would provide value added to the decision makers would be the elimination of the high frequency data and keeping just the low frequency data, in other words calculation of trends. However this procedure would decrease the timing and the relevancy of the recent information and it wouldn't be of much use for the decision makers[2].

Another way to tackle this information is by excluding some volatile components from the overall index. For example, the food and energy prices are determined to a great extent by supply shocks caused by weather conditions and the decrease of supply of oil when some conflicts in Middle East arise. So, by excluding these components from the overall index, it can be obtained an index which would reflect better the inflation trends in the economy. However, this method involves some shortcomings as well. There is no certainty that changes in food prices do not contain useful information on inflation trends in the economy. The removal could result in the removal of valuable information for decision makers. Furthermore, there are other factors other than food prices and fuel prices that could compromise attempts to measure the increase in prices due to monetary factors. The resulting index does not necessarily have a clear picture of inflationary trends.

In order to overcome these issues, there have been developed new approaches that are based more on statistical procedures rather than on the characteristics of certain components, such as the trimmed mean and weighted median. These methods have the advantage of eliminating of different irregularities in data while still keeping the important information which was eliminated by traditional approaches.

In this article we will present the main ways to handle the transitory effects on inflation by investigating what types of core inflation are used by some other central banks in the region as well as by calculating and analyzing the dynamics of the core inflation excluding the food, fuel and regulated prices as well as the core inflation determined using statistical methods such as the trimmed mean and weighted median on CPI data for Republic of Moldova.

The paper is organized as follows. Section 2 provides a brief literature review highlighting some important findings of other authors regarding core inflation measures. In section 3 we present some insight on the data and the main methodology in addressing core inflation measures in our study. Section 4 presents the main results of our research. The paper ends with section 5 where conclusions are provided.

2. Literature review

The main shortcomings of the CPI in addressing the monetary policy relevant inflation was tackled in M.F. Bryan and S. G. Cecchetti paper "Measuring core inflation" (1993) [2]. They state that it is difficult to measure the monetary inflation as a monetary phenomenon because of the non-monetary events that can temporarily produce noise in the data. They list some alternative solutions to this issue such as low-frequency trends or excluding certain components from the overall CPI index based on the assumptions that they are most affected by the noise. They also come with some statistical measures of core inflation such as median and the trimmed mean. They state that these measures are robust to the presence of many types of noise. Besides that they tried to evaluate the usefulness of their proposed measures of core inflation for monetary policy by assessing which of them is mostly correlated with the money growth. They find that the statistical methods are superior to CPI in several respects such as the fact that they have higher correlations with past money growth and provide improved forecasts for future inflation.

M. Silver [9] outlined the many approaches and methods to the measurement of core inflation and the many approaches to judging the preferred measure. His research shows that different measures of core inflation yield different results, that is, that choice of measure matters. Furthermore, he states that different approaches to the choice of measure yield different results and, even for the same approach to choice, the preferred measure may differ across countries, and even within a county for different time periods. Choice of measure should, in principle, be data-driven for each country based on appropriate criteria. According to his paper, exclusion-based methods are found to be not optimal according to the criteria selected by the monetary authorities. The choice of the method should be data driven, so that the methods adopted are tailored to the features of the evolution of that country's economy and so that the choice of measures can be justified on an objective, transparent basis.

After evaluating several candidate series that have been proposed as core measures of consumer price index (CPI) inflation and personal consumption expenditure (PCE) inflation for the United States, R. Rich and C. Steindel [8] concluded that policy would be best served by recognizing that core measures differ in the quality and nature of the insights they can provide about the dynamics of inflation and to draw from this varied information for guidance.

M. A. Wynne also reviews various approaches to the measurement of core inflation that have been pro-posed over the years using the stochastic approach to index numbers as a unifying framework [12]. According to his paper, there is no theoretical ideal for a monetary measure of core inflation. He concludes that before choosing a measure of core inflation one needs to specify what it is one wants the measure for. Depending on the reason behind it, different methods would be appropriate.

According to B. Meyer and G. Venkatu [6] trimmed-mean inflation statistics diagnose the most volatile monthly price changes as noise and "trim" them from the price-change distribution, leaving a clearer inflation signal behind. These measures systematically remove sources of noise on a monthly basis, rather than ad hoc exclusionary measures such as the ex food and energy ("core") CPI. They tried to find whether median CPI was the appropriate measure of trimmed - mean inflation statistic to use as a measure of underlying inflation. Besides the symmetric trims, they also tried to use the asymmetric ones. They conclude that median CPI is generally a better forecaster of future inflation over policy-relevant time horizons than the headline and core CPI.

3. Data and Methodology

3.1 Data

For calculating different types of core inflation we use the monthly data concerning change of prices and CPI component weights starting from 2009 until September 2014 available from the statistical authority. The weights of the CPI components are determined by the statistical authority on a yearly basis based on the Survey concerning income and expenditures of households in the previous year.

3.2 Core inflation by exclusion of certain components

Core inflation can be calculated by excluding certain components of the CPI which are considered to have volatile behavior, are determined by central or local governments or are subject to frequent supply shocks. In other words, the excluded items are believed to be beyond the control of the monetary policy. The resulting index can be obtained using the formulae below:

$$CII = \frac{\sum_{i=1}^{n} w_i * iw_i - \sum_{j=1}^{m} w_j^{ex} * iw_j^{ex}}{\sum_{i=1}^{n} w_i - \sum_{j=1}^{m} w_j^{ex}} \qquad (1)$$

where:

CII – core inflation index;

w_i – the weight of the item in the CPI basket;

iw_i – price index of an item in the CPI basket;

w_j^{ex} – the weight of the item excluded form the CPI basket;

iw_j^{ex} – price index of an item excluded from the CPI basket;

i – goods and services included in the CPI index;

j – goods and services that are excluded from the CPI index for calculation the CII;

n – number of goods and services that are part of CPI basket;

m – number of goods and services that are excluded from the CPI during calculation of the CII.

3.3. Statistical measures of core inflation –trimmed mean and weighted median

The trimmed mean of α - percent is calculated by the formula nr. 2:

$$\bar{x}_\alpha = \frac{1}{1-2\frac{\alpha}{100}} \sum_{i \in I_\alpha} w_i x_i \ , \qquad (2)$$

where

I_α is $\frac{\alpha}{100} \lhd W_i \lhd (1-\frac{\alpha}{100})$

$W_i \equiv \sum_{j=1}^{i} w_j$

w is the weight of the component, x is the monthly price change of the component

In general, the method involves:

- ranking ascending price increases for each period with corresponding weights.
- then, we add the previous weights for each price increase previously ordered.
- next, the monthly increases for which the sum of their weights is less than α will be excluded,
- the same thing will happen with monthly increases for which the sum of the corresponding weights is greater than 100-α.

- the trimmed mean will be calculated then as a weighted average of the remaining components.

The weighted median is an extreme case of the trimmed mean, so it represents the growth of the component which is situated in the middle of the increasingly ordered distribution. Thus, half of the weighted monthly increases are above the weighted median and half are below. Therefore, the median is calculated according to the previous procedure, except that it is the first price change whose cumulative weight is greater or equal to 50 %.

3.4. Determining the optimal level of exclusion of information for the trimmed mean approach

In order to determine the optimal level of exclusion of information from both ends of the distribution for a given time, several indicators according to the formulae nr. 2 will be calculated using various exclusion rates (0 percent which is actually CPI until the median - which excludes all items except the observation in middle of the distribution).For each of these indicators we will determine the root mean squared error against the trend of inflation (RMSE , formulae nr. 3). The optimal index is considered the one for which the error is the smallest. To determine the trend of CPI inflation we will apply the Hondrick-Prescot Filter on CPI data (lambda =16600) (figure nr. 4).

$$RMSE(\alpha) = \sqrt{\frac{\sum_{t=1}^{k} (T_t(\alpha) - \overline{\pi})^2{}_t}{k}}, \qquad (3)$$

where

$T_t(\alpha)$- the index of the trimmed mean α % at time t

$\overline{\pi}$ - trend of CPI inflation determined by using the Hondrick Preskot filter on CPI

k- number of observations

4. Results

4.1. Measures of core inflation in other central banks that target inflation

In international practice, usually the objective of monetary policy is price stability which implies a moderate amount of inflation as measured by CPI. Thus, central banks assess and communicate the effectiveness of their actions depending on whether the overall inflation is close to the proposed level in the medium term. However, given the aforementioned problems on this index, the monetary policy authorities monitor various measures of core inflation when taking decisions in a timely way and with the desired impact in order to contain inflation. These are supposed to allow exclusion of transitory effects and to identify the trends of inflation due to monetary factors. The diversity of the measures of core inflation can be observed by studying inflation reports published by central banks in other countries in the region.

Czech Republic switched to inflation targeting regime in December 1997. From January 2010 the target is 2% ±1 percentage point [5]. The Inflation Report published by the Czech National Bank denotes that CPI includes Net inflation and regulated prices. The net inflation is decomposed in food prices, fuel prices and Adjusted net inflation [15].

Central Bank of England adopted the IT strategy in October 1992. The current target is point target of 2% annual rate of inflation. Central Bank of England, similarly, practice exclusion method to identify the transitory effects on inflation. Thus, in the inflation reports published by the bank, CPI inflation is decomposed in food prices, fuel prices, the prices of services, education, energy and gas prices and the component other. Previously, in some older editions of the Bank of England inflation reports one can find the dynamics of other measures of core inflation such as median and trimmed mean of 15% [13].

The National Bank of Poland adopted the IT regime in 1998. Since 2004 it has a target of 2.5%±1 percentage point. NBP considers several measures of core inflation in its reports. In addition to the traditional method of exclusion of some default volatile components (inflation without the volatile prices, without food and energy prices, without regulated prices) volatile, it also uses the trimmed mean of 15% [16].

The National Bank of Romania switched to inflation targeting in 2005. Starting from 2013 it has a 2.5% ±1 percentage point. In the Inflation Reports published by the National Bank of Romania, there are shown three measures of core inflation. The component CORE 1 is the difference between total inflation and administered prices. Component CORE 2 also excludes volatile prices and the component adjusted CORE2 results from the exclusion of volatile prices, the regulated and the tobacco and alcoholic beverages from total inflation. The volatile prices components include vegetables, fruits and eggs [14].

4.2. Core inflation measures in Republic of Moldova

In late 2009 the National Bureau of Statistics of the Republic of Moldova adopted the methodology of the calculation of core inflation index [7]. According to it, core inflation index is calculated using the method of exclusion. The NBS calculates four measures of core inflation:

1. Total CPI excluding food and beverages
2. Total CPI excluding products and services with regulated prices
3. Total CPI excluding fuel prices
4. Total CPI excluding food and beverage products and services with regulated prices, fuel prices.

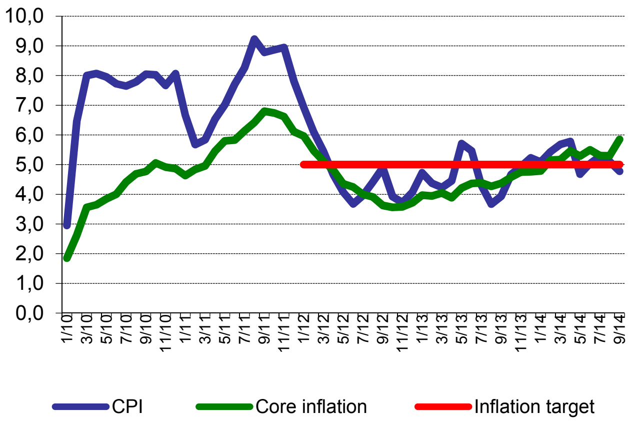


Figure 1. CPI and core inflation, yoy, %

Although the National Bureau of Statistics publishes several measures of core inflation, the Inflation Reports published by NBM reveal the dynamics of the core inflation index that is excluding regulated prices, prices of food and fuel prices.

In 2010 the National Bank started to create the necessary pre-conditions for implementing the inflation targeting regime. The inflation reports published by the bank reveal the dynamics of core inflation index which is calculated by excluding from total inflation prices of food and drinks, regulated prices and fuel prices.

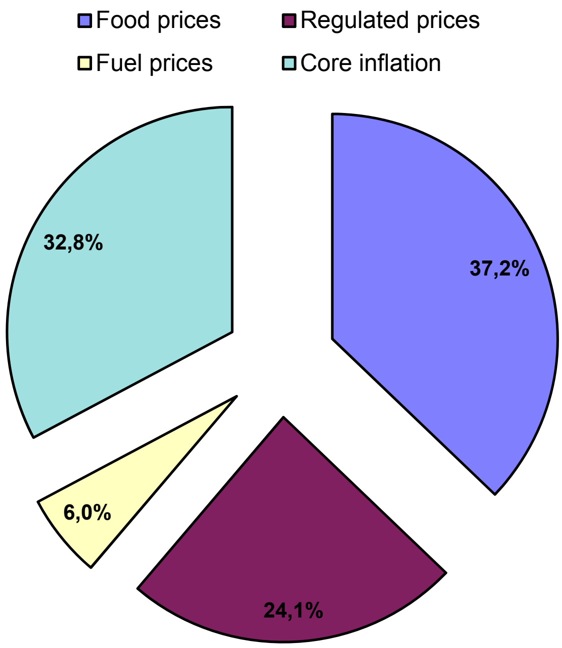


Figure 2. CPI structure in Moldova (2014)

Since 2010, the annual evolution of the above-mentioned core inflation indicator was characterized by a significantly lower volatility than that of the overall inflation (figure 1). At the same time it oscillated closer to the medium-term inflation target of 5.0 percent ± 1.5 percentage points, its average being 4.7 percent in the period. Core inflation has seen an upward trend with business recovery after the crisis of 2009. In the end of 2011, along with the slowdown in economic activity, it reversed its previous trend decreasing from 6.8 percent in September to 3.6 percent in October 2012. Starting from the beginning of 2013, core inflation recorded a slightly upward trend increasing up to a value of 5.8 percent in September 2014, driven mostly by monetary policy measures undertaken to prevent annual inflation from leaving the target range.

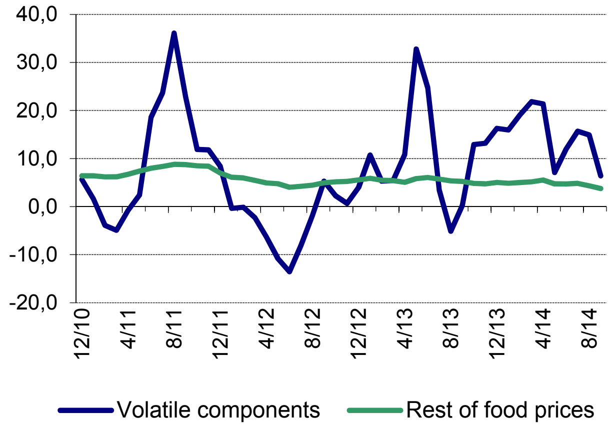


Figure 3. Food prices, yoy, %

However, the core inflation measure presented above has several drawbacks. It excludes, in addition to regulated prices and fuel prices, completely the component food and drinks. As a result, the core inflation component has a weight of about 32.8 percent of the total CPI (in 2014) (figure 2) which is quite low compared with core inflation indexes monitored in other countries with similar regimes. So, according to the above mentioned procedure more than two thirds of the information is excluded from core inflation measure analyzed in the Inflation Report

The exclusion of the food prices from the core inflation is usually justified by the fact that they have a high volatility driven largely by supply-side factors and not the medium-term inflation trend. However, in the structure of the food price component can be identified very volatile subcomponents such as prices of vegetables, fruits, eggs which dynamics is indeed mostly caused by transitory effects. The structure of the food prices also contains some less volatile components which includes most of the processed foods and are not as sensitive to agro-meteorological conditions. These might be largely influenced by aggregate demand (see figure 3). Thus, they might present useful information about medium-term inflation trends in the economy and might not be excluded from the measure of core inflation. Also, in this way the share core inflation in CPI structure would significantly increase.

4.3. The trimmed mean and weighted median measure for CPI data from Moldova

Given the fact that most core inflation indicators published by the statistical authority in Moldova are calculated by the method of exclusion of pre-determined components and at the moment there is not an alternative core inflation index, next we will provide the trimmed mean and weighted median for the consumer prices in Moldova based on the formulae nr. 2.

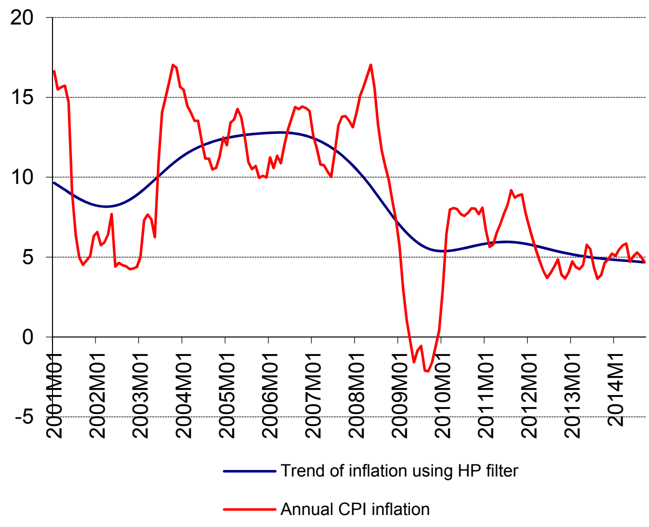


Figure 4. CPI inflation and the trend of inflation

The information on the RMSE suggests that the optimal measure of the trimmed mean, i.e. the closest to the trend of inflation is the one for which 10 % from each end is truncated. This means that at the upper and at the lower end of the distribution we will exclude 10 percent of observations on price changes (figure no. 5).

The annual dynamics of the trimmed mean (figure 6) is slightly different from the annual growth rate of core inflation calculated by the method of exclusion. Thus, although in early 2010 they had similar values, the trimmed mean recorded a faster increase in the first quarter of 2010 which determined a higher trajectory than that of the traditional core inflation index. suggesting higher inflationary pressures form the aggregate demand compare to the second indicator.

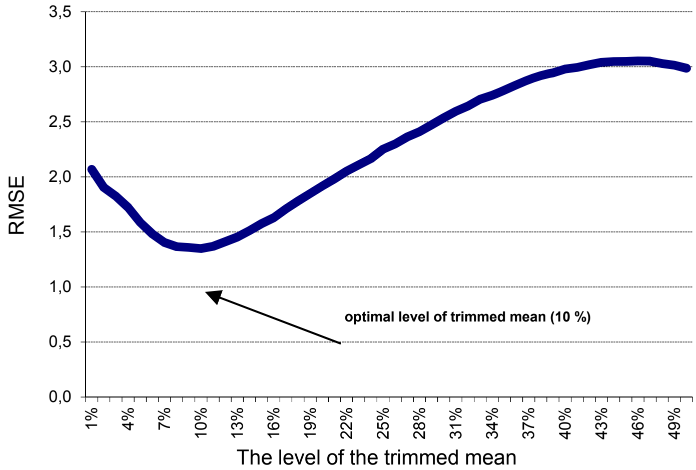


Figure 5. RMSE

In the first quarter of 2011 the trimmed mean experienced a pronounced downward evolution, while the traditional core inflation had a stable dynamics. After this episode, by the end of 2011 both indicators had similar increasing dynamics signaling pressures from increasing demand on prices. However, the overall path of the trimmed mean was lower than that of the core inflation calculated by the exclusion method(aprox. 1 penrcentage point). In 2012, both indicators had a downward trajectory due to decrease in the economic activity and the difference between core inflation calculated by the truncated mean method and calculated by the method of exclusion had been maintained. In late 2012, both above mentioned indicators started a moderate increasing path which lasted till the end of the sample (3rd quarter 2014). However the difference between the two recorded a slight increase.

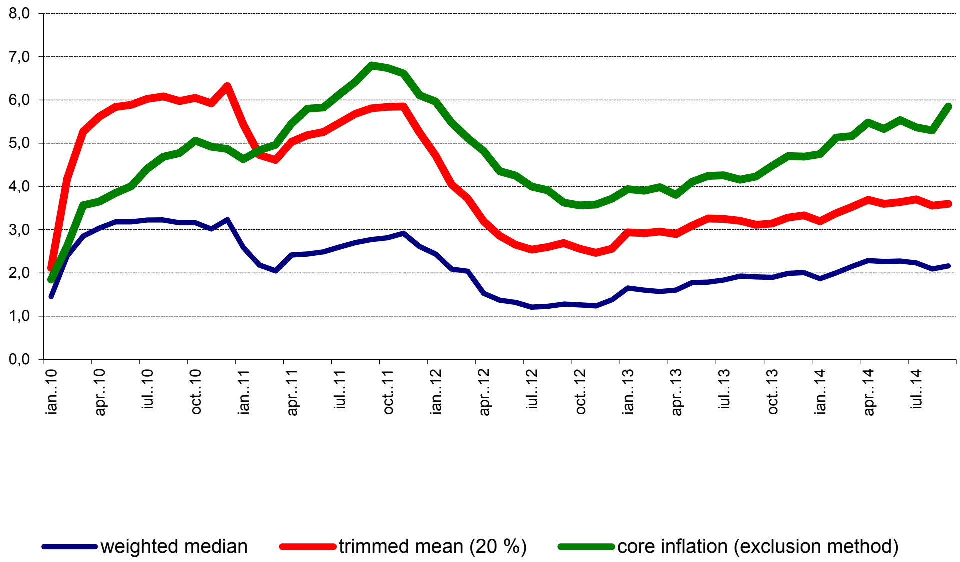


Figure 6. Core inflation measures using exclusion method and statistical methods

In case of the weighted median (figure 6) in 2010, it had a similar pattern to that of core inflation calculated by the method of exclusion. After significant reduction in first quarter of 2011, and a more modest increase to the end of the year, the weighted median trajectory was significantly lower than that of the other indicator. Towards the end of 2011, the difference between the two measures of core inflation was about. 4.0 percentage points. This difference, however, decrease in 2012 and early 2013 to approx. 2.5 percentage points. The weigted median started a moderate upward trend in early 2013 similar to the core inflation calculated by the exclusion method and similar to the trimmed and by the end of the 3rd quarter 2014 it reached 2 %. The basic idea of the 2 alternative core inflation indicators is that they suggest lower pressures on inflation coming from the aggregate demand compare to the traditional core inflation measure.

5. Conclusion

This article tackles some of the main issues the policymakers face when monitoring the price dynamics within the inflation targeting regime. The Consumer Price index, besides important information on medium term inflation trends, might still contain information determined by transitory effects or some measurement errors. Therefore, it is of high interest to have a so called core inflation index that would be useful for taking the right decisions to contain inflation within the medium term target.

The measures of core inflation calculated by the method of exclusion of certain pre-determined components whose dynamics is mostly driven by external factors, by the decisions of authorities or which exhibited a very volatile behavior over the history are more commonly used by central banks implementing inflation targeting regime. However, these indicators have several drawbacks and the most important of them were mentioned within the article such as the fact that these methods might exclude important information from the CPI data. The index which is left after the exclusion can be in the end a small part of the initial CPI index. Furthermore, this index can also still include some components which are not a part of the medium term inflation process and are not relevant for policy makers. In this way it can sometimes provide an inaccurate view on inflationary pressures caused by monetary factors.

This article suggests that there are alternative methods to the traditional exclusion procedures, such as statistical methods for determining core inflation, the trimmed mean and weighted median. According to these measures, the excluded components differ in each period, and their exclusion criterion is determined by certain statistical properties, in this case how far the respective component is from the central tendency in a certain period, and does not contain any economic reasoning.

Given the fact that the inflation reports published by the National Bank of Moldova denote the dynamics of a core inflation index which is determined by the exclusion of the food prices, fuel prices and regulated prices form the CPI index, the trimmed mean and the weighted median of inflation might be an important additional source of information for policymakers in Republic of Moldova. Even though these indicators of core inflation have different values, they have similar trajectories over the sample analyzed in the article compare to the traditional core inflation indexes calculated for Moldova .

As a conclusion, the trimmed mean and the weighted median for Moldova could present useful information about inflationary trends that might be missed by traditional core inflation measures and should be considered as an additional source of information to guide decision making in the process of keeping the overall inflation in the inflation target band.

Bibliography

1. Bike, L. and Stracca, L. **A persistence weighted measure of core inflation in the euro-area,** European Central Bank, 2007
2. Bryan, F. Stephen, G. and Cecchetti, M. **Measuring Core Inflation,** 1993
3. Cutler, J. **Core inflation in UK,** Bank of England, 2001
4. Griffiths, D. **Core inflation measures produced in New Zealand,** 2009
5. Hammond, G. **State of the art of inflation targeting,** Handbook no. 29, Centre for Central Banking Studies, Bank of England, 2012
6. http://en.wikipedia.org/wiki/Truncated_mean
7. http://www.statistica.md/public/files/Metadate/alte/Metodologia_Inflatia_de_baza.pdf
8. Rich, R. and Steindel, C. **A review of Core inflation and an evaluation of its measures,** Federal Reserve Bank of New York, 2005
9. Silver, M. **Core inflation measures and statistical issues in choosing among them,** 2006
10. Zaman, G., Goschin, Z., and Herteliu, C. **Analysis Of The Correlation Between The Gdp Evolutions And The Capital And Labor Factors In Romaniam,** Romanian Journal for Economic Forecasting, 2(3), 2005, pp. 5-21
11. Zaman, G., Goschin, Z., Partachi, I. and Herteliu, C. **The contribution of labour and capital to Romania's and Moldova's economic growth,** Journal of applied quantitative methods, 2(1), 2007, pp. 179-185
12. Wynne, M.A. **Core Inflation: A review of Some Concepptual Issues,** 1999
13. www.bankofengland.co.uk
14. www.bnr.ro
15. www.cnb.cz
16. www.nbp.pl

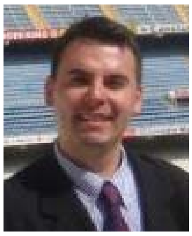
CONNECTIONS BETWEEN WILL TO EMIGRATE AND ATTACHMENT THEORY – A DATA MINING APPROACH[1]

Angel-Alex HAISAN[2]
PhD, Faculty of Economics and Business Administration,
Babes-Bolyai University, Cluj-Napoca, Romania

E-mail: haisan-angel@hotmail.com

Vasile Paul BRESFELEAN[3]
PhD, Lecturer, Faculty of Economics and Business Administration,
Babes-Bolyai University, Cluj-Napoca, Romania

E-mail: paul.bresfelean@econ.ubbcluj.ro



Abstract:
Many studies have been carried out in the last years in Romania, due to the scale of the phenomenon, on the factors that are responsible for the migration decision of people and the consequences of this act.
The family is being generally accepted as the nucleus of a society and that's why is very important to create a medium where it can develop and grow in a healthy way. Family is one of the 12 domains of our main study direction, quality of life and it still is the first options when it comes to individual's support.
A new aspect that should be brought into attention should be represented by the recent findings in the emigration research domain, which show that there is a significant relationship between the will to emigrate and an unresolved attachment status.
So, we would like to see if indeed the emigration decision is based on financial motivations, as we've discovered so far, or it also has ties with other indicators.

Key words: family, attachment theory, emigration, data mining, teachers

1. Introduction

Many studies have been carried out in the last years in Romania, due to the scale of the phenomenon, on the factors that are responsible for the migration decision of people and the consequences of this act. We can encounter in the research literature, as major domains of research, different approaches, such as economical [1,18], sociological [28, 40], educational [41, 22] or psychological [46, 37] directions, each of them having their own sub directions that most of the times intersect each other, thus needing an interdisciplinary approach.

Although much has been written regarding the migration process of Romanians in general, very few studies approach the problem on a smaller scale, such as socio-professionals categories. We can talk about the brain-drain problem here, which like all major concepts, has been discussed in the beginning without analyzing a certain category and only in the last couple of years it began to know a more focused approach, the attention of the research community being mainly drawn by the migration of specialists from health care and IT. As a consequence to that, in one of our past studies we have tried to approach the migration subject from a different category's point of view, namely teachers [21]. We've tried to discover, by employing data mining techniques, the motivations that reside behind the emigration decision of this category based on their marital status and found out that the economical factor is of great importance for the married without children and unmarried ones while for the ones that are married with children the family comes first.

The family is being generally accepted as the nucleus of a society and that's why is very important to create a medium where it can develop and grow in a healthy way. Also, family is one of the 12 domains of our main study direction, quality of life and it still is the first options when it comes to individual's support [15]. Regardless if we are talking about close or enlarged family, its members offer comfort to the individual whether we are considering its moral, financial or other needs. Another aspect that should be brought into the equation should be that recent findings in the emigration research domain show that there is a significant relationship between the will to emigrate and an unresolved attachment status [47]. Attachment theory refers to the way a human responds within relationships after he has been heart, separated from its loved ones or threatened, this way being formed in early life, before we can talk, based on how reliable, responsive and understanding our caregiver is [10]. Having these in mind, we are set to analyze the will to emigrate of our subjects by dividing them into groups depending on the vital status of their parents.

2. Problem Formulation

The necessary data for conducting the following study has been extracted from the general data base generated by our main research direction, quality of life of pre-university professors from Cluj-Napoca. This direction, being one that covers all domains of one's life, generates a very large quantity of data and includes aspects about close and enlarged family. The questionnaire, especially developed for this research, using EQLS's approach with 12 domains: health, job, income, education, family, social involvement, housing, environment, infrastructure, personal safety, leisure and life satisfaction, has been distributed to all physical education professors from pre-university schools from Cluj-Napoca, the most important city in the Nord-West development region, second largest in Romania and at the same time one of the largest university centers in the country. After centralizing the results, we've counted 105 valid answers from a total of 149 potential respondents, thus having a response rate of 70,46%.

After a first analysis we've identified several indicators as potential material for future studies, because of their unusual high values. Among these indicators the will to emigrate was by far the one that registered the highest values, 38% of the respondents that offered valid answers saying they would like to emigrate and if this wasn't concerning enough, 22% of these stated that they would emigrate anywhere. As a consequence to this we've started to analyze this particular indicator in relation with others that we've considered they could influence it. First on our list was to see how this indicator varies depending on the marital status of our respondents, so in relation with their close family, because a lot of issues could come from this decision, including relational problems with their life partner, or even worse psychological traumatizing experiences or sentimental deprivation for children [21]. We then continued with its analysis in relation with the financial status of our respondents, starting from the premises that the decision to emigrate has strong ties with the way they cope with daily financial needs [24]. In the present paper, after revising the research literature, we would like to return to the family indicator but from another perspective, enlarged family, more precisely how the emigration indicator varies among our subjects in relation with the vital status of their parents, mostly because all of our subjects declared that have a good and very good relation with their families, family support being an important part of the quality of life assessment and also because it seems that there are strong ties between the will to emigrate and attachment problems, that occur in early life mainly because the caregiver, usually the parents, don't respond well to the needs of the child. So we would like to see if indeed the emigration decision is based on financial motivations, as we've discovered so far, or it also has ties with other indicators.

As a result, we were set to analyze which could be the indicators that had the most influence in the decision to emigrate, how these decisions differentiate based on the vital status of our respondent's parents and if there's a connection with the attachment theory. In order to achieve this, we've split the study group into four categories by taking into consideration their answers to their parent's vital status:

1. Their mother and father are alive – coded as "mom_yes_dad_yes" for compatibility with the utilized software;
2. Their mother is alive and their father is dead – coded as "mom_yes_dad_no" for compatibility with the utilized software;
3. Their mother is dead and their father is alive – coded as "mom_no_dad_yes" for compatibility with the utilized software;
4. Their mother and father are dead – coded as "mom_no_dad_no" for compatibility with the utilized software.

The fifth category, "NA", comprised of the ones that haven't answered to the income question, was excluded because it didn't bear any relevance.

3. State of the Art Research

3.1. Emigration and family

Confucius said "The strength of a nation derives from the integrity of the family". So, in accordance with this statement if we were to change something in our country we should begin by promoting and supporting the concept of a united family.

Kent Hoffman, which is one of the founders of a parental training program based on the theory of attachment, stated at one of his seminars held in Romania that the dynamic of family from ex communist countries was very affected and distorted by this doctrine, but at the same time those hard times managed to unite it and strengthen the connections between its members.

If communism is a thing of the past, at least on paper, our society is confronted nowadays with new challenges. One of the toughest it seems to be the impossibility to offer young capable people a stable and healthy environment where they can develop, start a family and live a decent life. Recent studies have surfaced worrying results such as the fact that the majority of the ones that chose to emigrate belong to the age interval 20-35, which represents basically the ideal fertile period and more so, over 60% of these are females [42].

These findings can explain the low fertility rates that Romania faces in present. Also, another problem could be represented by the fact that a large number of highly qualified workers, doctors, nurses, engineers, informaticians, choose to emigrate [35, 39].

Having in mind these drawing attention findings, some researchers tried to identify the reasons behind peoples will to emigrate and found that most of them are seeking a better salary and a higher standard of living [4]. Although several studies have shown that there are some benefits, mostly economical, for the families that have one or both adults members implicated in an emigration process [3], the vast majority of the research literature states the same conclusion: long term effects on family are devastating and it usually begins with the alienation of its members and ends with scission, the most affected being the children which can develop unhealthy psychological behaviors [13, 38, 6]. If adults can, at some point, get over the pain caused by a separation, children on the other hand will remain with scars that will be transmitted to further generations, this being the reason why it is so important to take care of our children within the family unit.

The ideal situation, towards which we should aspire, would be to raise our children in a happy stable family, because it has been scientifically proven that happy raised children manage to do better in life on all areas when they benefit of the care of both parents [27], without thinking about the money. Happiness is free and paradoxically at the same time expensive. It can be achieved for free by doing small things that cost you noting, like a kind word, offering a helping hand, moral support or by simply being there, or you can work hard to earn it, although with a different outcome, as researchers from the Warwick University have discovered [50].

3.2. Data mining

Data mining is defined as the "nontrivial process of extracting valid, previously unknown, comprehensible and useful information from large databases" [36]. Fayyad et al., [16] identified that for the knowledge to be discovered through this technique some preparatory steps need to be fulfilled: data cleaning, data reduction, data transformation, data mining and pattern evaluation.

Classical data mining techniques were mostly used to collect data and later became a tool for analyzing large quantity of data [2]. In the last years, its role has extended beyond initial borders, today being encountered in almost all fields, such as marketing, banking, medicine, astronomy, education, sociology, etc., thus becoming an important tool in decision making processes. Its utilization basically flourished due to the fact that almost every life domain became data-intensive [48].

In the research literature we have found many studies that use data mining techniques covering various fields: security technology [33], manufacturing [26], banking [14], management [45], sport [34], medicine [29], transport [49], etc., but very few that used data mining techniques upon traditional education. Instead, we could observe that this method is nowadays extensively used in e-learning [25, 43, 11, 52, 31, 32]. We have managed in some of our past studies to associate it with traditional education by successfully establishing raw connections between various indicators concerning life of pre-university schools teachers. This offered us unique inside perspectives with a direct impact on the national educational system [21, 23, 24].

The data mining method most used in our research was classification learning and allowed us to automatically learn models [51]. We have based our approach towards classification on decision trees, mainly due to the fact that they can operate under supervision by being provided with the actual outcome for each of the training examples. The models have been used for scanning the data in order to generate trees and make predictions.

The instances are classified by decision trees based on their feature values, each node being a feature in an instance to be classified and each branch a value the node can obtain [30]. The main advantages of this method are that it creates models that are easy to understand and missing values within the data don't affect them [5], although due to the fact that it only permits single dependent variable it can create certain restrictions [44].

For the present classification learning experiment, J48 and J48graft methods (developed from C4.5 algorithm) have been employed, with the help of the very popular Weka 3 open source GNU software for machine learning [51].

3.3. Attachment theory

John Bowlby, who is considered to be the father of attachment theory, defined it as "one specific and circumscribed aspect of the relationship between a child and caregiver that is involved with making the child safe, secure and protected" [7].

The emergence of this theory dates back to the late '40's, when Bowlby began with the help of James Robertson to observe hospitalized and institutionalized children that have been separated from their parents[12]. After some studies he concluded, based on empirical evidence, that a small child in order to grow up mentally healthy "should experience a warm, intimate, and continuous relationship with his mother, or permanent mother substitute, in which both find satisfaction and enjoyment" [9]. Also, he emphasized on the role of social networks and economy as factors which influence well developed functioning relationships between mother and child stating in one of his books that "children are absolutely dependent on their parents for sustenance, so in all hut the most primitive communities, are parents, especially their mothers, dependent on a greater society for economic provision. If a community values its children it must cherish their parents" [17].

Unfortunately the negative outcome of this behavior has long term repercussions on the individual because "the initial relationship between self and others serves as blueprints for all future relationships" [8].

So, returning to the context of our study, the vast majority of our respondents lived most of their life or have been raised for the first years of their life in a communist society. Family in the communism period although was proclaimed in the official ideology as "basic cell of the society", wasn't just a simple propaganda but a justification of the intervention of the state in private space, in order to gain control by destroying its traditional values [19]. The extensive character of the communist economy, in the context of a forced industrialization, required growth of human workforce. So, along with the exploitation of the rural workforce, came the concept of women emancipation, which in order to function required for women to be relieved of family duties [19].

If we analyze these statements with the ones in the previous paragraph we start to see that some connections begin to form. A lot of our subjects want to emigrate and not all of them have financial reasons behind, as we've discovered in our previous studies. Children with attachment problems have a tendency to run away from home [20] and as long as this wish of alienation persists even after reaching adult age and no other reasons such as financial exist, we can consider emigration a form of "run away". In this regard, a recent study has analyzed from the attachment's theory perspective the Dutch and Belgian immigrants from California and found a significant relationship between unresolved attachment status and being an immigrant [47].

4. Results

4.1. Mother alive, father alive

We proceed by analyzing the first category, where both parents of our respondents are alive. This is the most numerous group with a number of 51 subjects, from which 29 wish to emigrate, the average age being 33.29. After employing data mining techniques we've obtain the following decision tree, which is graphical represented in figure number 1.

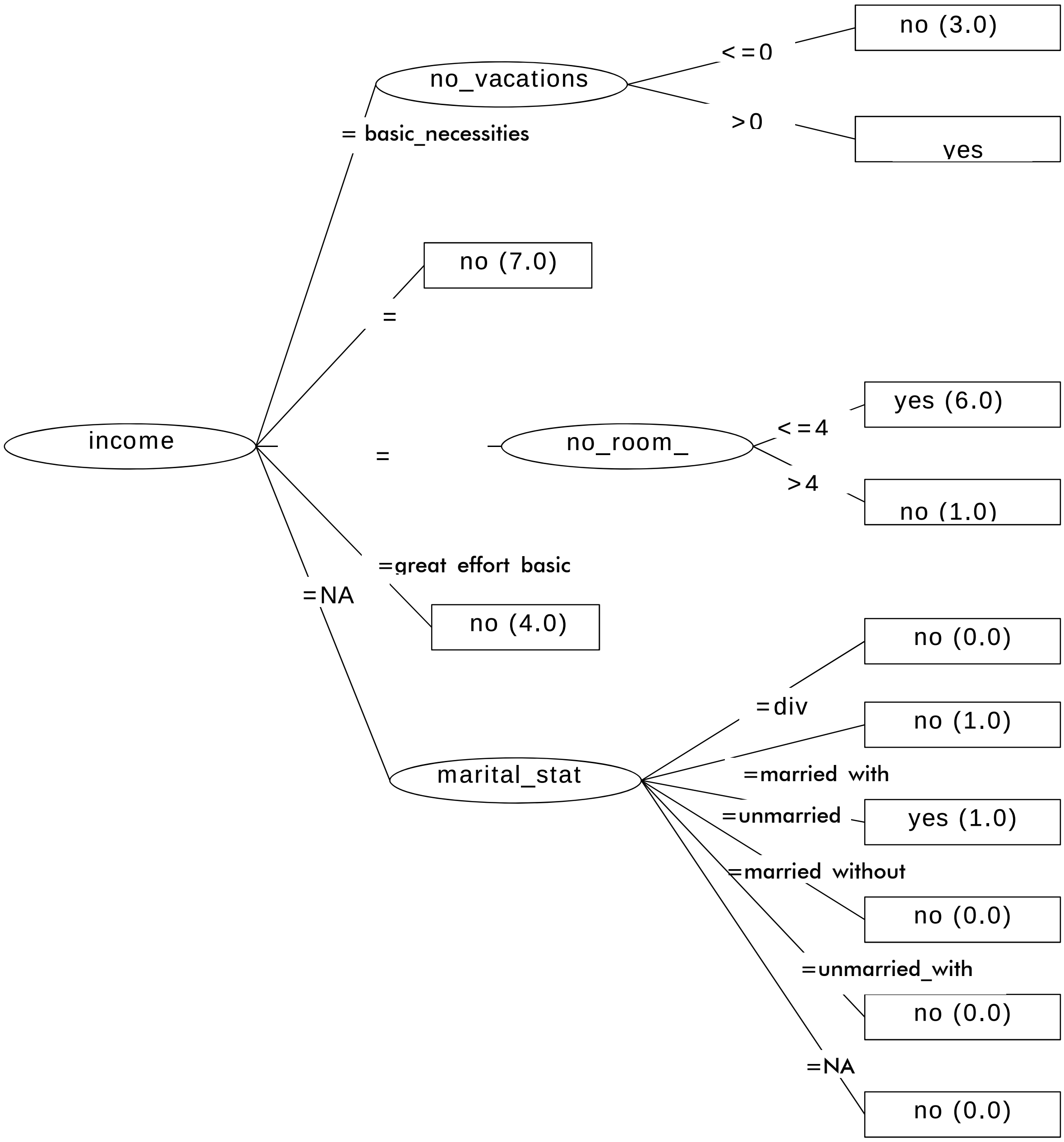


Figure 1. J48 decision tree based on "mom_yes_dad_yes" group

As we can see, the main indicators that influence the will to emigrate for this group are of financial type. So, the respondents that are most likely to emigrate are of 3 types: a) the ones that have an income that only covers their basic necessities and managed to spend more than 0 vacations in the last five years in a resort from Romania or abroad with 19 persons. These were initially divided, by their answers to the will to emigrate, in 7 that didn't want and 12 that did. So, our program, as we can see from the bellow graph, identified 4 persons from the ones that initially stated they don't wish to emigrate as in fact being potential candidates for emigration; b) the ones that consider that their income doesn't cover even their basic necessities and have an apartment with less than 4 rooms with 7 persons. In this group's case our program kept the initial distribution of the respondents based on their questionnaire answers with 6 that would wish to emigrate and 1 not; c) the ones that didn't respond to the income indicator and are unmarried with 1 person for which our program hasn't found any other connections.

Concerning direct connections with attachment issues, our program didn't identified any, because indicators that could suggest such a thing like "help from parents" or "members of living unit", weren't taken into consideration. So, in order to find some connections we've dug deeper and isolated for each of the 3 groups the ones that have been identified by the program as potential candidates for emigration and found out that: for the a) group 4 persons from total 16 ones have a special cohabitation situation, in the sense that they are still living with their parents and more than that, all of them are unmarried and have ages very close and over 30, so it's possible that these particular subjects could experiment attachment problems. The help that they are receiving from parents consists only in food; for the b) and c) groups we could not find any connections.

4.2. Mother alive, father dead

We've continued by analyzing the second category, in which case only the mother of our respondents is alive. This group has a number of 25 subjects with an average age of 45,96. As we've discovered in our previous studies, once the age increases the will to emigrate decreases and this is also true for this group, because only 5 persons want to emigrate. Another possible explanation for the low number of persons that wish to emigrate would be that only their mother is alive and they choose to stay around and help. After running the program on the data, we've identified that for this group the indicators that count the most, when talking about emigration, are the ones related to work and family. So, the most likely to emigrate belong to 4 categories: a) the ones that have a second job and evaluate the educational system as being one of a poor quality with 2 respondents; b) the ones that have a second job and didn't respond to the question regarding the quality of the educational system with one person; c) the ones that don't have a second job and live with their life partner and parents with one person; d) the ones that don't have a second job and live with their child with one person.

As we can see from the graphic bellow, our program discovered some direct connections between one of the indicators that could signalize attachment problems "members of living unit" and will to emigrate. After the isolation of these two persons, we've observed something very interesting: the one that doesn't have a second job and lives with his life partner and parents answered negative to the will to emigrate question, but the program, as we can see, indentified him as a potential emigrant; and the one that declared he lives only with his children, his marital status is married, although he had the divorced or widow options to choose from, so we would be inclined to think that these two persons have some problems of attachment.

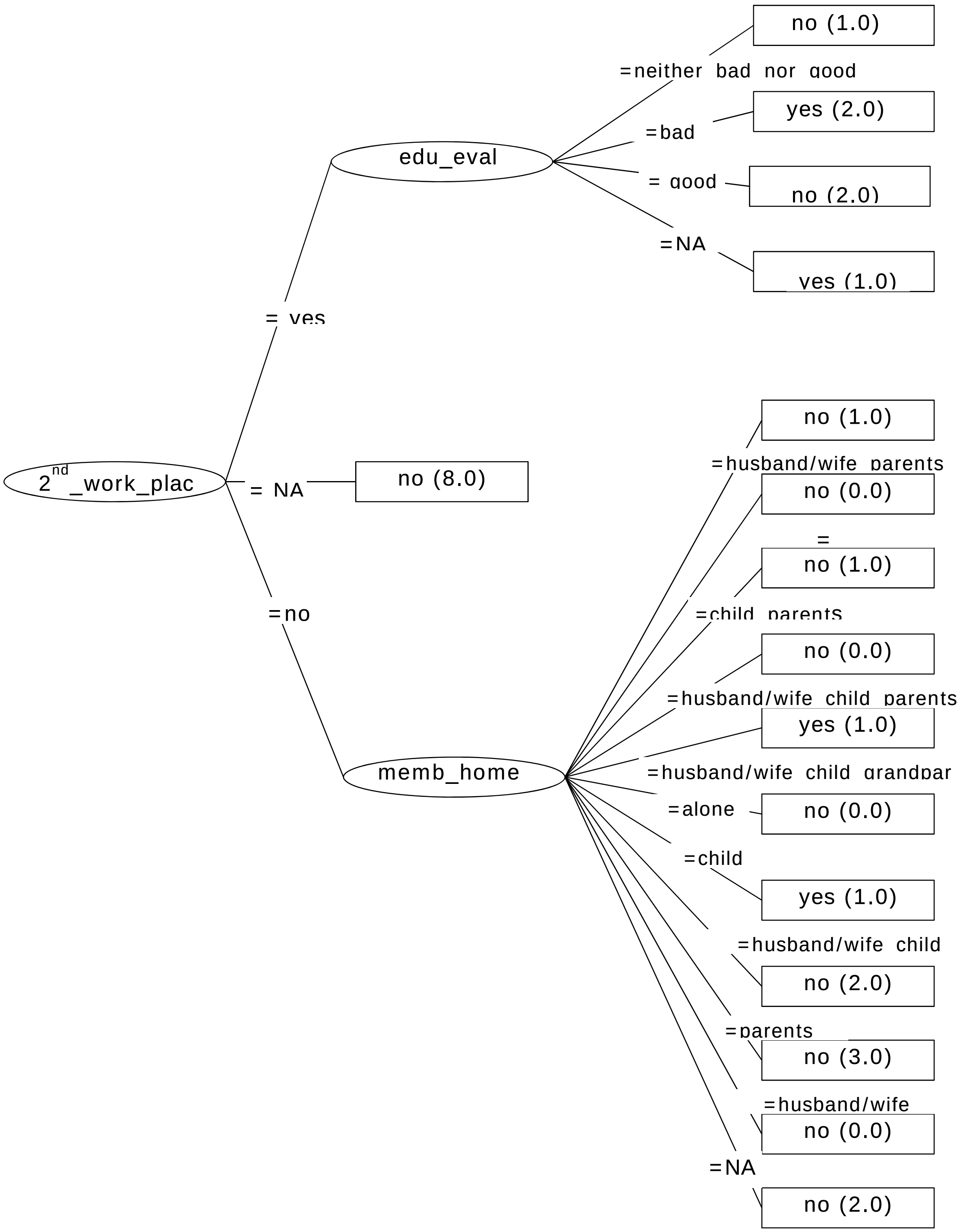


Figure 2. J48 decision tree based on "mom_yes_dad_no" group

In order to find connections between attachment theory and will to emigrate for the other 3 persons belonging to the a) and b) groups, we've proceeded as in the case of the precedent category and isolated them. The findings were interesting, because one of two persons that were identified by the program to belong to the a) group presents similar characteristics as the ones from the a) group in "mother alive, father alive" category. He's unmarried with the age of 31 and lives with his parents, in this case only with his mother. The help received from parent consists in durable goods.

4.3. Mother dead, father alive

The third category is represented by the ones that declared their mother is dead and their father is alive and it is the smallest of our study with only 3 respondents. Average age continues to increase reaching 50,3. Our program has identified for this category as principal indicator in the decision to emigrate gender, factors that could be related to attachment problems not being directly taken into consideration. Although it doesn't bare much relevance, because of the small number of respondents, we would like to proceed with a small analysis. So, as we can see from the bellow graph, if they are females they do not want to emigrate and if they are males they do. After isolating the one that wants to emigrate we also couldn't find any relations to the indicators that could signal an attachment problem.

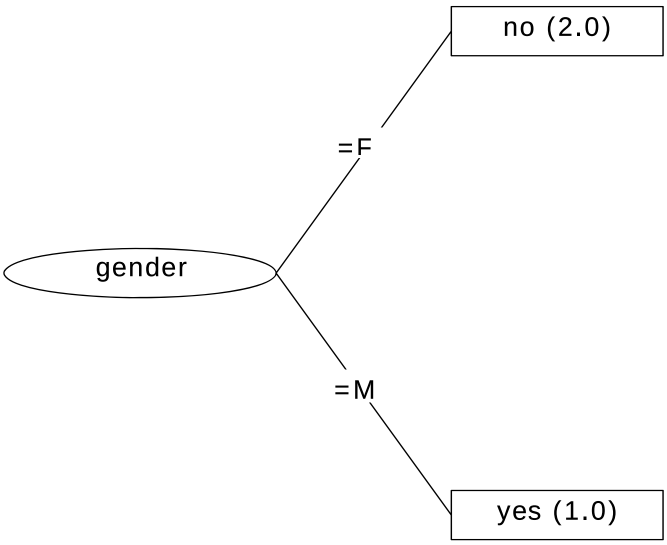


Figure 3. J48 decision tree based on "mom_no_dad_yes" group

4.4. Mother dead, father dead

Last group of our study is represented by the ones that declared that none of their parents are alive. It has 22 respondents and an average age of 57,72. For this group the indicators that count the most in the emigration decision, as we can see from the bellow graph, are ones of financial type. Initially 4 respondents from this group stated they would like to emigrate, but our program has identified, based on their answers to the others indicator, that one of them actually wouldn't. The most likely to emigrate belong to 2 categories: a) the ones that own an apartment and land with 2 persons; and b) the ones that own an apartment and a car and their income doesn't even cover their basic necessities with one person. Due to the fact that both parents of the subjects included in this category are dead it will be irrelevant to pursue a connection between will to emigrate and attachment theory.

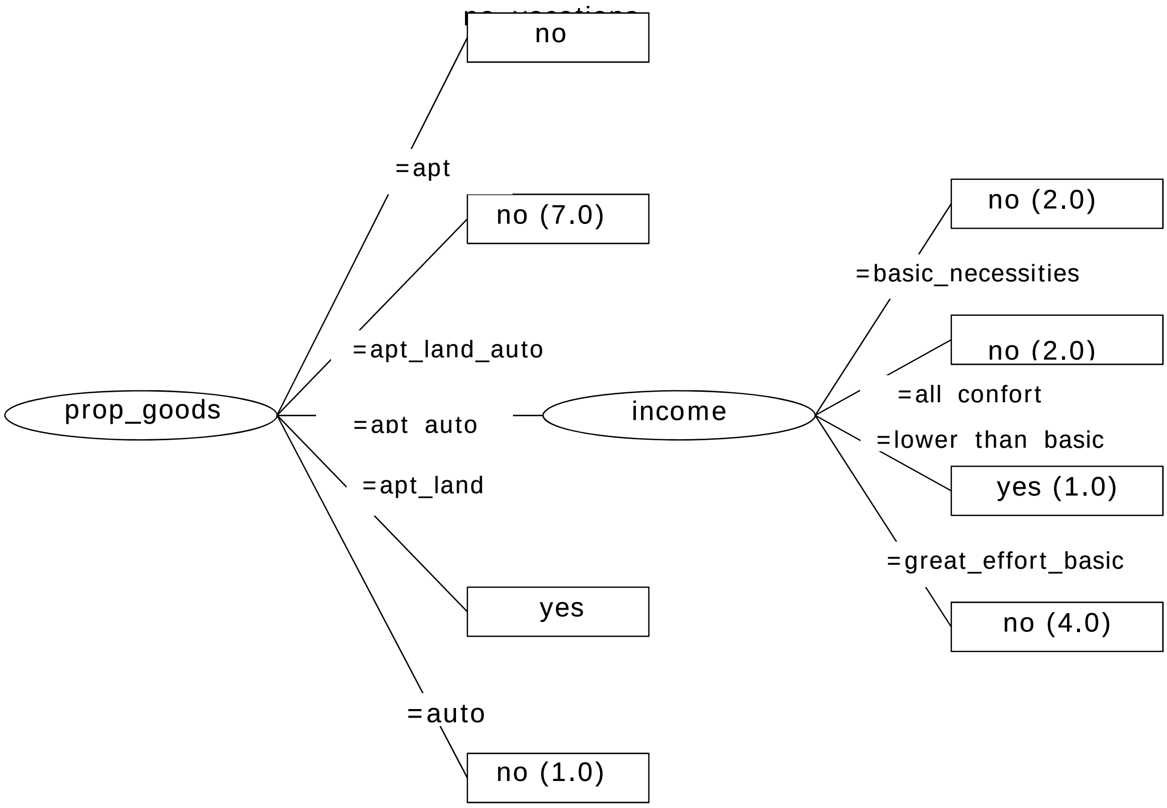


Figure 4. J48 decision tree based on "mom_no_dad_no" group

5. Conclusions

Although the relevance of this study can be easily contested, mainly due to the low number of respondents and the presence of not so many indicators related to the attachment theory, our goal wasn't necessarily to obtain hard evidence but to establish a conceptual framework and method for future studies.

Having this in mind we declare ourselves satisfied with the results and can affirm that data mining methods helped us on one hand to directly identify indicators that count for our respondents in the decision to emigrate based on the vital status of their parents and on the other hand to narrow down possible special cases for further analysis of the indicators that could denote problems in relation with attachment theory. In fact, we consider this method very promising because it managed to find in such a small number of respondents some cases that can be correlated to our main hypothesis.

The four categories chosen are in accordance with the findings of the studies conducted by the National Institute of Statistics which state that females benefit of a higher life expectancy then males and it is demonstrated, in our study, by the number of respondents for the first three categories. So, the very low number of respondents for the "mother alive, father dead" category would be somehow justified.

We continue by synthesizing the results given by the use of the data mining method. For the first category "mother alive, father alive" the main reasons for emigrating would be of financial type. For the second category "mother alive, father dead", the reasons change and refer to work and family. For the third "mother alive, father dead" category although the number of respondents was very low the indicator that counted in the decision to emigrate or not was gender. For the last category "mother dead, father dead", as if it was a circle of life, the reasons return to financial ones.

We conclude with the findings related to the attachment theory. As a general rule, where our program didn't find any direct connections, we've proceeded with an individual analysis of the ones that were identified by the program with a wish to emigrate. For the first category "mother alive, father alive" no direct connections were found, so we've isolated the respondents and identified four possible drawing attention cases, mainly because of their cohabitation status, age and marital status. For the second category "mother alive, father dead" our program found direct connections through the indicator "members of living unit". Interestingly enough is the fact that this particular respondent, found by the program, has a very similar situation with the ones identified by us in our analysis of the first group, he's 31 years old, lives with his mother and is unmarried. For the third category "mother alive, father dead" neither the program nor us could find any connections, we think mainly due to the low number of respondents. The last category "mother dead, father dead" didn't bear any relevance in this case.

References

1. Ailenei, D. and Bunea, D. **Labour market flexibility in terms of internal migration,** Annals of University of Oradea, 2010, pp. 153-158
2. Andronie, M. and Crisan, D. **Commercially Available Data Mining Tools used in the Economic Environment,** Database Systems Journal, 2010, pp. 45-54
3. Antman, F. **The Impact of Migration on Family Left Behind,** Bonn: IZA, 2012
4. Asociatia Nationala a Birourilor de Consiliere pentru Cetateni. **Romanii si migratia fortei de munca in Uniunea Europeana,** Bucharest, ANBCC, 2005
5. Berson, A, Smith, S. and Thearling, K. **Building data mining applications for CRM,** USA: McGraw Hill, 2000
6. Botezat, A. and Pfeiffer, F. **The Impact of Parents Migration on the Well-Being of Children Left Behind – Initial Evidence from Romania,** ZEW - Centre for European Economic Research Discussion Paper, 2014
7. Bowlby, J. **Attachment and loss: retrospect and prospect,** American Journal of Orthopsychiatry, 1982
8. Bowlby, J. **Attachment theory, separation anxiety and mourning,** American Handbook of Psychiatry, 1975, pp. 292-309
9. Bowlby, J. **Maternal care and mental health,** World Helath Organization, 1951
10. Bowlby, J. **O baza de siguranta. Aplicatii clinice ale teoriei atasamentului,** Bucharest, Ed. Trei, 2011
11. Bresfelean, V.P. **Implicatii ale tehnologiilor informatice asupra managementului institutiilor universitare,** Cluj-Napoca, Risoprint, 2008
12. Bretherton, I. **The Origins of Attachment theory: John Bowlby and Mary Ainsworth,** Developmental Psychology, 1992, pp. 757-775
13. Ciuperca, N. **http://ciupercaniculina.blogspot.ro/2009/05/efectele-emigrarii-asupra-familiei.html**
14. Divsalar, M., Roodsaz, H., Vahdatinia, F., Norouzzadeh, G. and Behrooz, A. **A Robust Data-Mining Approach to Bankruptcy Prediction,** Journal of Forecasting, 2012, pp. 504–523
15. European Foundation for the Improvement of Living and Working Conditions. **EURLife,** http://www.eurofound.europa.eu/areas/qualityoflife/eurlife/index.php.

16. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. **From Data Mining to Knowledge Discovery in Databases,** American Association for Artificial Intelligence, 1996, pp. 37-54
17. Feig, C. **The Confluence of Attachment Style, Perceived Social Support, and Role Attainment in Women Experiencing Postpartum Mood Disorders,** Dissertations, 2012
18. Frunza, R., Maha, L.G. and Mursa, C.G. **Reasons and Effects of the Romanian Labour Force Migration in European Union Countries,** CES Working Papers, 2009, pp. 37-62
19. Ghebrea, G. **Regim social-politic si viata privata. Familia si politica familiala in Romania,** Bucharest, Ed. Universitatii, 2000
20. Golder, S., Gillmore, M., Spieker, S. and Morrison, D. **Substance Use, Related Problem Behaviors and Adult Attachment in a Sample of High Risk Older Adolescent Women,** Journal of Child and Family Studies , 2005, pp. 181-193
21. Haisan, A.-A. **Disfunctionalitati in sistemul educational national,** Cluj-Napoca, Presa Universitara Clujeana, 2013
22. Haisan, A.-A. **Life quality of physical education and sport teachers,** The 6TH International Conference "Perspectives in the Science of Human Movement", Cluj-Napoca, Faculty of Physical Education and Sport - UBB, 2012, pp. 143-148
23. Haisan, A.-A., and Bresfelean, V.P. **A Data Mining Examination on the Romanian Educational System - teachers Viewpoint,** International Journal of Mathematical Models and Methods in Applied Sciences, Vol. 7, no. 3, 2013, pp. 277-285
24. Haisan, A.-A., and Bresfelean, V.P. **A Data Mining Survey on the Factors that Influence Emigration Decisions among Romanian Teachers based on their Incomes,** Journal of Social and Economic Statistics, 2014, pp. 38-52
25. Hanna, M. **Data mining in the e-learning,** Campus-Wide Information Systems, 2004
26. Harding, J.A., Shahbaz, M. and Kusiak, A. **Data mining in manufacturing: A review,** Journal of Manufacturing Science and Engineering - ASME, 2006, pp. 969-976
27. Institute for Family Studies. **Strong Families, Sustainable Societies,** http://ifstudies.org/strong-families-sustainable-societies/
28. Ionescu, I. **The migration of Romanians in European Community,** Review of Research and Social Intervention, 2008, pp. 23-35
29. Kharat, A. **Data mining in radiology,** The Indian journal of radiology & imaging, 2014
30. Kotsiantis, S.B., Zaharakis, I.D. and Pintelas, P.E. **Machine learning: a review of classification and combining techniques,** Artificial Intelligence Review, 2006, pp. 159-190
31. Lile, A. **Analyzing E-learning systems using educational data mining techniques,** Mediterranean Journal of Social Sciences, 2011, pp. 403-419
32. Maiorana, F, Mongioj, A. and Vacc, M. **A Data Mining E-learning Tool: Description and Case Study,** Proceedings of the World Congress on Engineering . London: WCE, 2012
33. McCarthy, J. **Security and data mining,** InfoWorld, 2003, pp. 48-51
34. McKeever, S. **Sports data mining,** Hanover, Informs, 2012
35. Moldoveanu, R, and Chiricescu, A. **http://www.evz.ro/detalii** http://www.evz.ro/detalii/stiri/cei-10000-de-medici-emigrati-va-saluta-de-peste-mari-si-tari-954499.html.
36. Olaru, C., and Wehenkel, L. **Data mining,** Computer Applications in Power IEEE, 1999, pp. 19-25
37. Paduraru, M. E. **Romania - Emigration's Impact on Families and Children,** Journal of Community Positive Practices, 2014, pp. 27-36
38. Pescaru, M. **Consecințele migrației familiei contemporane asupra creșterii și educării copiilor,** Reconstruind Socialul. Riscuri si solidaritati noi - Prima Conferinta Internaţionala a Societatii Sociologilor din Romania, Cluj-Napoca, Facultatea de Sociologie si Asistenta Sociala, 2010, pp. 1-12
39. Petcana, A. M. **Financiar - O treime dintre romani vor sa plece din Romania,** http://www.gandul.info/financiar/aproape-doua-treimi-dintre-romani-cumpara-doar-strictul-necesar-de-alimente-10754023

40. Petrescu, R. M., Zgura, I.D. and Bac, D.P. **Descriptive Analysis Of The International Migration Phenomenon In Romania Between 1991 And 2008,** Annals of Faculty of Economics, 2011, pp. 288-294
41. Pociovalisteanu, D.-M. **Migration for Education Nowadays,** Annals of the „Constantin Brancusi" University of Targu Jiu, 2012, pp. 82-87
42. Pritulescu, R. **http://www.adevarul.es/stiri/social,** http://www.adevarul.es/stiri/social/cum-dispare-populatia-romaniei
43. Romero, C. and Ventura, S. **Educational Data Mining: A Review of the State of the Art,** IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 40, no. 6, 2010, pp. 601-618
44. Shah, S, Roy, R. and Tiwari, A. **Technology Selection for Human Behaviour Modelling in Contact Centres,** Decision Engineering Report Series, Cranfield University, 2006
45. Shen, C. **Data Mining the Data Processing Technologies for Inventory Management,** Journal of computers, 2011
46. Timofti, I. C. **Romanians' attitudes towards emigration,** Journal of Psychological and Educational Research, 2011, pp. 117-123
47. van Ecke, Y., Chope, R. and Emmelkamp, P. **Immigrants and attachment status: Research findings with Dutch and Belgian immigrants in California,** Social Behavior and Personality, 2005, pp. 657-674
48. Venkatadri, M. and Reddy, C. **A Review on Data mining from Past to the Future,** International Journal of Computer Applications, 2011, pp. 19-22
49. Wallander, J, and Makitalo, M. **Data mining in rail transport delay chain analysis,** International Journal of Shipping and Transport Logistics, 2012, pp. 269-285
50. Wilson, C. and Oswald, A. **How Does Marriage Affect Physical and Psychological Health? A Survey of the Longitudinal Evidence,** Bonn: IZA, 2005
51. Witten, I., Eibe, F. and Hall, M. **Data Mining: Practical Machine Learning Tools and Techniques,** Burlington, Morgan Kaufmann, 2011
52. Zhou, M. **Data Mining and Student e-Learning Profiles,** E-Business and E-Government (ICEE). Guangzhou: IEEE, 2010, pp. 5405-5408

Appendices

Appendix 1 - Mother alive, father alive

```
=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 1 -A
Relation:     mother alive, father alive
Instances:    51
Attributes:   18
        marital_stat
        age
        sex
        no_child
        support_parent
        no_room_apt
        prop_goods
        memb_home
        marriage
        achivment
                income
        will_emigr
        no_vacations
        2nd_work_place
        soc_traject
        edu_eval
        fin_retrib
```

```
        prof_eval
Test mode: evaluate on training data

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

J48 pruned tree
------------------
income = basic_necessities
|   no_vacations <= 0: no (3.0)
|   no_vacations > 0: yes (16.0/4.0)
income = great_effort_basic: yes (16.0/6.0)
income = all_confort: no (7.0)
income = lower_than_basic
|   no_room_apt <= 4: yes (6.0)
|   no_room_apt > 4: no (1.0)
income = NA
|   marital_stat = div: no (0.0)
|   marital_stat = married_with: no (1.0)
|   marital_stat = unmarried: yes (1.0)
|   marital_stat = married_without: no (0.0)
|   marital_stat = NA: no (0.0)
|   marital_stat = unmarried_with: no (0.0)

Number of Leaves  :     12
Size of the tree :      16
Time taken to build model: 0.01 seconds

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

Correctly Classified Instances          41               80.3922 %
Incorrectly Classified Instances        10               19.6078 %
Kappa statistic                          0.5771
Mean absolute error                      0.3315
Root mean squared error                  0.3809
Relative absolute error                 67.5203 %
Root relative squared error             76.9048 %
Total Number of Instances               51

=== Detailed Accuracy By Class ===

                         TP Rate   FP Rate   Precision   Recall  F-Measure   ROC   Area  Class
                         0.545     0         1           0.545   0.706       0.846       no
                         1         0.455     0.744       1       0.853       0.846       yes
Weighted Avg.    0.804   0.258     0.854     0.804       0.79    0.846

=== Confusion Matrix ===

  a  b   <-- classified as
 12 10 |  a = no
  0 29 |  b = yes
```

Appendix 2 - Mother alive, father dead

```
=== Run information ===

Scheme: weka.classifiers.trees.J48 -U -M 2
Relation:     mother alive father dead
Instances:    25
Attributes:   18
                   marital_stat
        age
        sex
        no_child
        support_parent
        no_room_apt
        prop_goods
        memb_home
        marriage
        achivment
```

```
             income
        will_emigr
        no_vacations
        2nd_work_place
        soc_traject
        edu_eval
        fin_retrib
        prof_eval
Test mode: evaluate on training data

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

J48 unpruned tree
------------------
2nd_work_place = yes
|   edu_eval = good: no (2.0)
|   edu_eval = neither_bad_nor_good: no (1.0)
|   edu_eval = bad: yes (2.0)
|   edu_eval = NA: yes (1.0)
2nd_work_place = no
|   memb_home = husband/wife: no (0.0)
|   memb_home = NA: no (2.0)
|   memb_home = parents: no (3.0)
|   memb_home = husband/wife_child: no (2.0)
|   memb_home = husband/wife_parents: no (1.0)
|   memb_home = child: yes (1.0)
|   memb_home = alone: no (0.0)
|   memb_home = husband/wife_child_grandparents: yes (1.0)
|   memb_home = husband/wife_child_parents: no (0.0)
|   memb_home = child_parents: no (1.0)
memb_home = husband/wife_nephew: no (0.0)
2nd_work_place = NA: no (8.0)

Number of Leaves: 16
Size of the tree:       19
Time taken to build model: 0 seconds

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

Correctly Classified Instances          25              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances               25

=== Detailed Accuracy By Class ===

                  TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                    1         0          1          1        1                   1        no
                    1         0          1          1        1                   1        yes
Weighted Avg.       1         0          1          1        1                   1

=== Confusion Matrix ===

  a  b   <-- classified as
 20  0 |  a = no
  0  5 |  b = yes
```

Appendix 3 - Mother dead, father alive

```
=== Run information ===

Scheme: weka.classifiers.trees.J48 -U -M 1
Relation:     mother dead father alive
Instances:    3
Attributes:   18
```

```
                    marital_stat
        age
        sex
        no_child
        support_parent
        no_room_apt
        prop_goods
        memb_home
        marriage
        achivment
            income
        will_emigr
        no_vacations
        2nd_work_place
        soc_traject
        edu_eval
        fin_retrib
        prof_eval
Test mode:evaluate on training data

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

J48 unpruned tree
------------------
sex = F: no (2.0)
sex = M: yes (1.0)

Number of Leaves  :     2
Size of the tree :   3
Time taken to build model: 0 seconds

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

Correctly Classified Instances         3          100    %
Incorrectly Classified Instances       0            0    %
Kappa statistic                        1
Mean absolute error                    0
Root mean squared error                0
Relative absolute error                0    %
Root relative squared error            0    %
Total Number of Instances              3

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall    F-Measure   ROC Area  Class
               1         0         1           1         1           1         no
               1         0         1           1         1           1         yes
Weighted Avg.  1         0         1           1         1           1

=== Confusion Matrix ===

 a b   <-- classified as
 2 0 | a = no
 0 1 | b = yes
```

Appendix 4 - Mother dead, father dead

```
=== Run information ===

Scheme: weka.classifiers.trees.J48 -U -M 2
Relation:     mother dead father dead
Instances:    22
Attributes:   18
                    marital_stat
        age
        sex
        no_child
        support_parent
        no_room_apt
```

```
        prop_goods
        memb_home
        marriage
        achivment
                income
        will_emigr
        no_vacations
        2nd_work_place
        soc_traject
        edu_eval
        fin_retrib
        prof_eval
Test mode: evaluate on training data

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

J48 unpruned tree
------------------
prop_goods = apt: no (3.0/1.0)
prop_goods = apt_auto
|   income = basic_necessities: no (2.0)
|   income = great_effort_basic: no (4.0)
|   income = all_confort: no (2.0)
|   income = lower_than_basic: yes (1.0)
prop_goods = apt_land_auto: no (7.0)
prop_goods = apt_land: yes (2.0)
prop_goods = auto: no (1.0)

Number of Leaves  :     8
Size of the tree :      10
Time taken to build model: 0 seconds

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

Correctly Classified Instances          21               95.4545 %
Incorrectly Classified Instances         1                4.5455 %
Kappa statistic                          0.8308
Mean absolute error                      0.0606
Root mean squared error                  0.1741
Relative absolute error                 19.2771 %
Root relative squared error             45.0273 %
Total Number of Instances               22

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall    F-Measure   ROC Area  Class
                 0.75      0          1          0.75       0.857      0.986    yes
                 1         0.25       0.947      1          0.973      0.986    no
Weighted Avg.    0.955     0.205      0.957      0.955      0.952      0.986

=== Confusion Matrix ===

  a  b   <-- classified as
  3  1 |  a = yes
  0 18 |  b = no
```

[1] **Acknowledgements**
This paper has been financially supported within the project entitled "Horizon 2020 - Doctoral and Postdoctoral Studies: Promoting the National Interest through Excellence, Competitiveness and Responsibility in the Field of Romanian Fundamental and Applied Scientific Research", contract number POSDRU/159/1.5/S/140106. This project is co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013. Investing in people!

[2] Mr. Haisan Angel-Alex (born in Piatra Neamt, 18.04.1983) has a PhD in sociology and currently is a postdoctoral researcher at the Institute of National Economy, where he is conducting an interdisciplinary research on the quality of life of pre-universitary teachers from Romania, under the title "Reality or myth? Geographical positioning and its effects on socio-professional categories".

[3] Mr. Bresfelean Vasile Paul (born in Bistrita, 3.12.1979) is a Lecturer 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).

A STUDY OF SURVIVAL MODELLING IN DIALYSIS PATIENTS APPLYING DIFFERENT STATISTICAL TOOLS[1]

Diana-Silvia ZILISTEANU
MD, PhD, Assistant Professor,
Carol Davila University of Medicine and Pharmacy, Bucharest, Romania

E-mail: diana@nefrolog.ro

Ion Radu ZILISTEANU
PhD, Bucharest University of Economic Studies, Romania

Mihai VOICULESCU
MD, PhD, Professor,
Carol Davila University of Medicine and Pharmacy, Bucharest, Romania

Abstract:
The aim of this study is to develop a model for survival probabilities of incident dialysis patients based on demographic, clinical and biological characteristics. We used statistical methods, mainly Cox regression, and 2 statistical tools: SPSS version 21 and Excel. During the first stage, data were analysed using SPSS software, edition 21. We performed survival analysis using Cox proportional hazard regression test, to assess the relationship of explanatory variables with survival time. Second stage of analysis was performed using Excel computations based on the results provided by Cox analysis. Starting from the basal risk curve, and applying the coefficients derived from the Cox regression analysis, the hazard curve was calculated for any combination of values for the variables included in the equation. Based on these elements, we constructed an Excel model for survival simulation.

Key words: statistics, SPSS, survival analysis, Cox proportional hazard regression, mathematical model

1. Introduction

Utility of mathematical algorithms applied for assessing the risk of future negative health outcomes emerged from the Framingham Risk Score, which was first developed based on data obtained from the Framingham Heart Study in order to estimate the 10-year risk of developing coronary heart disease. Because risk scores give an indication of the likely benefits of prevention, they are useful for both the individual patient and for the clinician in helping decide whether lifestyle modification and preventive medical treatment, and for patient education.

Morbidity and mortality in patients with chronic kidney disease included in the replacement of renal function program is influenced by a number of factors related both to the patient (age, sex, renal disease and comorbidities) and the "quantity and quality" of nephrological care in the predialysis period. O'Hare et all (2005) describe the impact of age on the mortality risk in chronic renal failure.

1.1. Working hypothesis

The aim of this study is to develop a model that can estimate the probability of survival of incident dialysis patients based on demographic, clinical and biological characteristics recorded at the time of dialysis initiation. Upon this model, we can make recommendations for optimization and planning care for a patient with chronic renal failure before and after inclusion in chronic dialysis.

The working hypothesis of the study is that late referral to the nephrologist adversely affect the survival of patients with chronic kidney disease included in the dialysis program. In this paper, we propose to assess the impact that can be attributed to nephrology referral on mortality in incident dialysis patients, analyzing the effect of this factor, controlling for other factors that may influence survival of these patients.

1.2. Material and method

Patients with chronic kidney disease, hospitalized in Nephrology Department, Fundeni Clinical Institute, aged over 18 years, incident in the dialysis program between January 1st, 2007 and July 1st, 2012, were included in this analysis. Follow-up period ended on September 1st, 2014. The main indicator of the evolution of patients was survival from the time of inclusion in the dialysis program.

For the study patients, we recorded the following data:

- Demographical data: date of birth, gender, age at initiation of dialysis;
- Etiology of renal disease: hypertensive nephropathy, diabetic nephropathy, tubulo-interstitial nephropathy, primitive or secondary glomerular disease, genetic diseases (including autosomal dominant polycystic disease, Alport syndrome), systemic vasculitis (systemic lupus erythematosus, ANCA vasculitis, Henoch Schonlein etc), multiple myeloma or amyloidosis; cases where etiology was unknown were also recorded;
- Type of dialysis (hemodialysis or peritoneal dialysis) and access method used for renal replacement (central venous or arteriovenous fistula for hemodialysis or peritoneal dialysis catheter);
- Nephrology monitoring interval in months from the time the patient was evaluated for the first time in a nephrology service until entry into dialysis;
- Clinical manifestations (hyperhydration status, presence of pericarditis, heart failure, arrhythmias, pleural effusion, pulmonary infections, neurological, digestive manifestations, bleeding syndrome);
- Biological parameters recorded at the time of dialysis initiation: glomerular filtration rate estimated using the CKD-EPI formula (ml/min/1.73m2), hemoglobin (g/dL), leucocytes (count/mmc), platelets (count/mmc), sideremia (mcg/dl), serum ferritin (ng/mL), serum sodium (mmol/ L), potassium levels (mmol/L) total serum calcium (mg/dL) serum phosphate

(mg/dL), an intact parathyroid hormone PTH (pg/mL), serum albumin (g/dL), blood pH and serum bicarbonate concentrations (mmol/L).

2. Statistical analysis

The study was performed on a total of 430 patients included in dialysis, hospitalized between January 2007 and July 2012. Survival data were collected until September 2014.

2.1. Identification of survival indicators

During the **first stage**, data were analyzed using SPSS software, edition 21. As shown by Kleinbaum and Klein (2005), we performed survival analysis using Cox proportional hazard regression test, to assess the relationship of explanatory variables to survival time. Cases were considered censored (value status = 0) if the patient was alive or lost to follow-up during the study period, while deceased patients were considered cases that met the study goal (value status = 1), as explained by Sedgwick (2011). We applied a Cox regression sequence using different control variables to identify indicators that significantly affect survival, gradually eliminating the variables for which we did not obtained significant values.

The final model for survival (Table 1) included the following variables that significantly influence survival: age ($p < 0.0001$), heart failure ($p = 0.001$), bleeding syndrome ($p = 0.003$), diagnosis of multiple myeloma / amyloidosis ($p < 0.0001$) serum albumin ($p < 0.0001$). Although we did not obtain statistically significant values for the variable coefficient logarithm-referral logR ($p = 0.252$), we included this variable in the final model in order to shape the effect of the nephrological monitoring on survival of incident dialysis patients.

Table 1. The final Cox regression analysis including the identified variables influencing significantly the survival: age, heart failure, hemorrhagic syndrome, serum albumin, etiology of multiple myeloma/amyloidosis, and length of referral period

Survival indicators	B	SE	Wald	df	Sig.	Exp(B)
Age	0,040	0,007	34,541	1	0,000	1,041
Heart failure	0,647	0,196	10,855	1	0,001	1,910
Hemorrhagic syndrome	0,694	0,230	9,089	1	0,003	2,001
Serum albumin	-0,676	0,158	18,267	1	0,000	0,508
Multiple myeloma/amyloidosis	1,845	0,290	40,569	1	0,000	6,328
logR	-0,061	0,053	1,314	1	0,252	0,941

By applying Cox regression analysis, we developed the basal risk curve, which expresses the probability of death at a certain time for patients who survived until that time. This probability is not constant over time, permanently changing depending on the time the analysis is done. Based on the risk curve, survival curve is calculated directly by the arithmetic operation, which does not require other parameters.

To model mathematically the chances of survival at a certain time, we used the above results obtained by determining statistically significant variables Cox analysis, to create a survival function.

The relationship between the survival curve S(t) and the cumulative hazard curve CumH is exponential, and is given by the ecuation:

$$S(t) = \exp(-CumH)$$

The hazard function h(t) for a given combination of characteristics (values of explanatory variables) is the product of:

- The basic hazard function (the baseline hazard), h0 (t),
- Exponential of linear sum of the products of the values of explanatory variables (x1, x2, ..., xn) and the corresponding coefficients (β_1, β_2, ..., β_n).

Thus, the hazard function becomes:

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)$$

Using SPSS analysis, we determined:

- The basal cumulative hazard curve;
- Coefficients for each explanatory variable, β_1, β_2, ..., β_n.

Based on that information, one can calculate the cumulative curve hazard function and the survival curve for every combination of values of the explanatory variables.

Hazard or survival curves are well estimated by the following type of equations:

$$\hat{h}(t) = a \times t^b \Leftrightarrow \ln(\hat{h}(t)) = \ln(a) + b\ln(t)$$

This is visible when we exemplify the hazard or survival curve through a logarithmic scale chart. In such a graph, a linear relationship between the natural logarithm of the hazard curve, ln(h(t)), natural logarithm of the scale of the time, ln(t) shows the relationship expressed in the equation above.

The use of equations for computing the values on survival curve corresponding to a certain value on time axis and to a certain combination of explicative values, even if introduces some errors towards using the values obtained from data, has the advantage to highlight the main characteristics of survival curve and allows the focus on them, and the visual comparison between curves corresponding to different cathegories of patients are eased by eliminating non-essential variations.

2.2. Development of survival model

Second stage of analysis was performed using Excel computations based on the results provided by Cox analysis.

Using Cox analysis, we identified the following explanatory variables, which are statistically significant:

- X1 = patient age (in years);
- X2 = age of referral to the nephrologist (months x 10);
- X3 = presence of heart failure (categorical variable, which can take values of 0 or 1, signifying the absence or presence of disease);
- X4 = presence of bleeding syndrome (categorical variable, which can take values of 0 or 1, signifying the absence or presence of disease);
- X5 = serum albumin (10 x g / dl);
- X6 = multiple myeloma / amyloidosis (categorical variable, which can take values of 0 or 1, signifying the absence or presence of disease).

Also, the βi coefficients obtaind from Cox regression are:

β1 = 0.040;
β2 = -0.061;
β3 = 0.647;
β4 = 0.694;

β5 = -0.068;
β6 = 1.845.

Empirical survival curves can be described by mathematical functions that are more flexible, because they require knowledge of a small number of parameters; derived mathematical function well approximated survival curve which was empirically obtained (r^2 = 98%) (Figure 1). For this reason, we further applied the mathematical function for building survival curves.

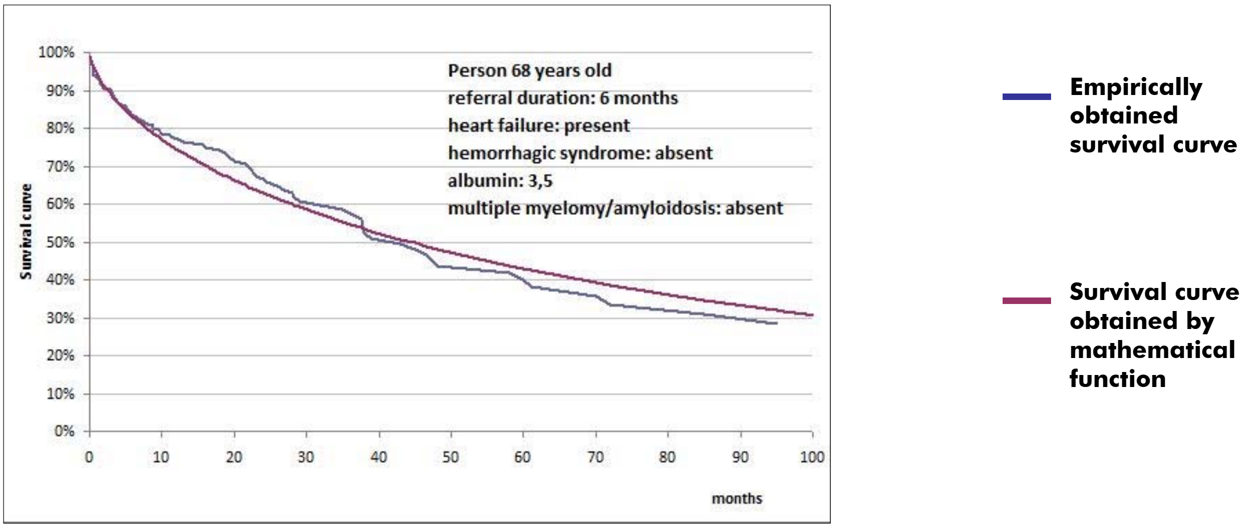


Figure 1. Correspondence between empirically obtained survival curve and curve obtained by the mathematical function

Starting from the basal risk curve, and applying the coefficients derived from the Cox regression analysis, the hazard curve can be calculated for any combination of values for the variables included in the equation.

Based on these elements, we constructed an Excel model for survival calculation (Figure 2), with the following advantages over SPSS output:

- Is more flexible than SPSS output, which can only express the curves for categorical variables, but not modeling for continuous variables;
- Allows a higher resolution analysis of the relationship between survival and variables of interest; thus, can analyze the impact of small changes in clinical and biological indicators on survival.

		coef
age	68	0,040
ln_referral	1,79	-0,061
referral duration (months x 10)	60	
heart failure	1	0,647
hemorrhagic syndrome	0	0,694
serum albumin (x 10)	35	-0,068
multiple myelomy/amyloidosis	0	1,845
survival treshold (months)	12	
survival odds for 12 months	74%	

Figure 2. Excel model for survival estimation

The Excel model for survival estimation allows different simulations based on combinations of values of included variables, even for hypothetical cases that have no counterpart in the database of patients in the study group. Thus, one can choose different values of continuous variables included (age, length of reference, serum albumin level), while selecting various combinations of categorical variables indicating the presence/absence of a diagnosis of multiple myeloma/amyloidosis and of the uremic complications (heart failure, bleeding syndrome).

Based on the selected combinations of values, Excel model estimates the survival chance calculated as a percentage, for a certain survival threshold.

We assessed whether increasing the length of nephrological care before dialysis, by earlier referral, can improve the survival handicap given by the presence of heart failure or bleeding syndrome after initiation of dialysis (Figure 3).

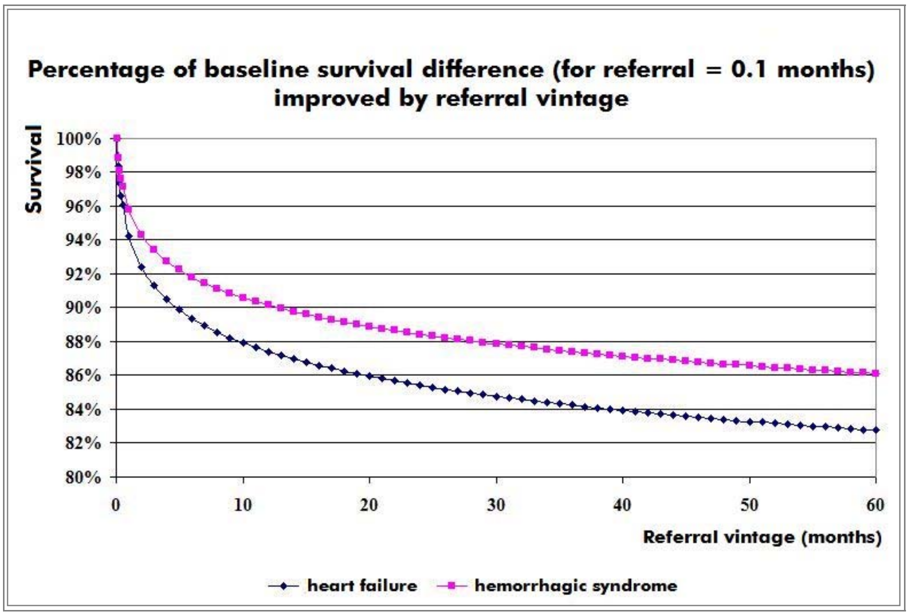


Figure 3. Percentage of initial survival difference between patients who have a certain status (heart failure or hemorrhagic syndrome) versus those without that clinical condition, which can be recovered by earlier referral to nephrologist. It is considered that at baseline (reference length = 0.1 months) there is a survival difference of 100% between the two categories

Considering that, for the referral length of 0.1 months, the difference in survival is 100%, we found the following:

- For patients with heart failure, the difference in survival drops to 94% if was referred at 1 month before dialysis, to 89% for the referral vintage of 6 months, at to 87% for referral of 12 months;
- For patients with bleeding syndrome, the difference in survival drops to 96% for 1 month referral, to 92% for the 6 months referral, and becomes 90% for 12 months referral vintage.

Therefore, we can say that for a patient with heart failure syndrome or bleeding syndrome at the time of initiation of dialysis, his chances of survival improve more so as he was referred earlier to the nephrologist.

3. Discussions and conclusions

The problem of late referral to nephrologist and initiation of renal replacement therapy in the emergency situation is extremely serious, considering that in Romania the number of dialysis patients incident has gradually increased in recent years, exceeding the number of 3000 in 2011 (from 1933 in 2007 to 3161 in 2011), as reported by Romanian Renal Registry. Medical care of these patients requires significant human and material costs, while being associated with a high mortality rate in short, medium and long term, as shown by Van Biesen (1999), Obialo (2005) and Black (2010). This justifies the need for coherent health policies related to chronic kidney disease, as shown by the report published by Levey and colleagues (2009). According to Vassalotti (2010) and McCullough (2011), a successful program has been promoted during the last years in United Stated, proving that a community-based screening approach can address disparities in chronic kidney disease.

This study emerged from the necessity of estimating the risk of future negative health outcomes for patients with chronic kidney disease included in the replacement of renal function program, based on influences by a number of factors related both to the patient (age, sex, renal disease and comorbidities) and the length of nephrological care in the predialysis period. Our results are similar with other of studies in the literature. Thus, Khan et all (2005) showed that consistent nephrology care may be more important than previously thought, especially because the frequency and severity of uremic complications increase as patients approach dialysis. This was supported also by Jones et al (2006), who showed the different decline in kidney function before and after nephrology referral and the effect on survival in moderate to advanced chronic kidney disease.

The mathematical model we developed is based on survival data in our group. Based on this model, we demonstrated that early referral can contribute to the partial recovery of handicap given by the unfavorable profile of a patient. This model we have developed, by estimating the chance of survival in patients enrolled in chronic renal dialysis program, could become a useful tool for scoring the severity of clinical and biological status in chronic renal patients. Future research will focus on expanding the patients' database in order to create a better approximation of survival chances based on cited parameters.

However, the utility of such mathematical model can be extended beyond the study in which was originally designed. This model can be considered a template for further survival analysis in different patients' categories, using diverse indicators and variables. Of great interest to the medical field would be the creation of modular software that can be used independently by each physician as a tool for tailored estimation of the risk score for an individual patient, by applying specific characteristics of each subject.

References

1. **Annual Report 2011 of Romanian Renal Registry**. Ministry of Health - Hospital Nephrology "Dr Carol Davila", Bucharest, Romania, 2012, available at http://www.srnefro.ro/media/RRR.Raport.2011.pdf
2. Black, C., Sharma, P., Scotland, G., McCullough, K., McGurn, D., Robertson, L., and Smith, C. **Early referral strategies for management of people with markers of renal disease: a systematic review of the evidence of clinical effectiveness, cost-effectiveness and economic analysis,** NIHR Health Technology Assessment programme: Executive Summaries, Vol. 14, No. 21, 2010

3. Jones, C., Roderick, P., Harris, S. and Rogerson, M. **Decline in kidney function before and after nephrology referral and the effect on survival in moderate to advanced chronic kidney disease,** Nephrology Dialysis Transplantation, Vol. 21, 2006, pp. 2133-2143
4. Khan, S., Xue, J., Kazmi, H., Gilbertson, D., Obrador, G., Pereira, B. and Collins, A. **Does predialysis nephrology care influence patient survival after initiation of dialysis?** Kidney International, Vol. 67, 2005, pp.1038–1046
5. Kleinbaum, D.G and Klein, M. **Survival Analysis. A Self-Learning Text**. 2nd ed. New York, Springer, 2005
6. Levey, A. S., Schoolwerth, A. C., Burrows, N. R., Williams, D. E., Stith, K. R. and McClellan, W. **Comprehensive public health strategies for preventing the development, progression, and complications of CKD: report of an expert panel convened by the Centers for Disease Control and Prevention,** American Journal of Kidney Diseases, Vol. 53, No. 3, 2009, pp. 522-535
7. McCullough, P. A., Vassalotti, J. A., Collins, A. J., Chen, S. C. and Bakris, G. L. **National Kidney Foundation's Kidney Early Evaluation Program (KEEP) annual data report 2009: executive summary,** American Journal of Kidney Diseases, Vol. 55, No. 3, 2010, S1-S3
8. Obialo, C. I., Ofili, E. O., Quarshie, A. and Martin, P. C. **Ultralate referral and presentation for renal replacement therapy: socioeconomic implications,** American journal of kidney diseases, Vol. 46, No. 5, 2005, pp. 881-886
9. O'Hare, A. M., Bertenthal, D., Covinsky, K. E., Landefeld, C. S., Sen, S., Mehta, K. and Walter, L. C. **Mortality risk stratification in chronic kidney disease: one size for all ages?,** Journal of the American Society of Nephrology, Vol. 17, No. 3, 2006, pp. 846-853
10. Sedgwick, P. **Survival (time to event) data: censored observations,** British Medical Journal; Vol. 343, 2011, d4816
11. Sedgwick, P. **Cox proportional hazards regression,** British Medical Journal, Vol. 347, 2013, f4919
12. Van Biesen, W., De Vecchi, A., Dombros, N., Dratwa, M., Gokal, R., LaGreca, G.,and Lameire, N. **The referral pattern of end-stage renal disease patients and the initiation of dialysis: a European perspective,** Peritoneal dialysis international, Vol. 19, Suppl. 2, 1999, pp. S273-S275
13. Vassalotti, J.A., Li, S., McCullough, P.A. and Bakris G.L. **Kidney early evaluation program: a community-based screening approach to address disparities in chronic kidney disease**. Seminars in Nephrology, Vol. 30, 2010, pp. 66–73
14. Zilisteanu, D.S. **Late nephrology referral and impact on morbidity and mortality of patients with chronic renal disease,** Ph.D. Thesis, Carol Davila University of Medicine and Pharmacy, Bucharest, 2013

[1] **Acknowledgment**
This paper was co-financed from the European Social Fund, through the Sectorial Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/138907 "Excellence in scientific interdisciplinary research, doctoral and postdoctoral, in the economic, social and medical fields - EXCELIS", coordinator The Bucharest University of Economic Studies.

THE REDISTRIBUTIVE EFFECT OF THE ROMANIAN TAX-BENEFIT SYSTEM: A MICROSIMULATION APPROACH[1]

Eva MILITARU
Postdoctoral fellow, Bucharest University of Economic Studies, Romania
Researcher, National Research Institute for Labour and Social Protection, Romania

E-mail: evamili@yahoo.com

Abstract:
This paper attempts to investigate the income distribution of Romanian households, focusing on the role of the tax-benefit system in income redistribution. We evaluate the redistributive effect by estimating income inequality changes due to tax-benefit components. We use EU-SILC microdata and the EUROMOD microsimulation model to simulate income components. The results point out that income inequality is considerably reduced through the tax-benefit system, as a great deal of income is redistributed among households. The analysis of the income components that contribute to inequality reduction emphasizes that pensions, personal income taxes and social benefits are in favour of inequality reduction, while social contributions act the opposite way. Our results are sensitive to social and fiscal policy changes.

Key words: Income distribution, Redistributive effect, Tax-Benefit System, Income inequality, Microsimulation

1. Introduction

The aim of this paper is the investigation of household income distribution in Romania, with focus on the role of the tax-benefit system in income redistribution. The evolution of income distribution in Romania has encountered many changes during recent years. We remark average household income growth before the economic crisis (up to 2008), income decline during crisis and slight recovery during the most recent years (2012-2013). It seems that the poorer households have benefited more from pre-crisis positive economic developments and have lost lower proportions of their incomes during the economic crisis, as compared to higher income households. The unequal changes along the income distribution have shaped a more equal income distribution in 2013 compared to 2007. Besides the economic developments which had a direct influence on household income levels, the changes that took place in the tax-benefit system (as social and fiscal policy response to crisis) have a serious impact on household income developments.

The paper attempts to evaluate the redistributive effect of the Romanian tax-benefit system, by estimating income inequality changes due to tax-benefit components. It covers the period between 2007 and 2013. The results indicate that half of the income inequality before taxes and transfers is reduced through the tax-benefit system. We find out that the economic crisis has led to the decline of income inequalities, as richer households have lost more of their market incomes; but income inequality has dropped also due to tax-benefit system changes that were adopted in order to cope with the emerging situation. We use the EUROMOD microsimulation model in order to split the household income into income components (i.e. social benefits: pensions, means-tested benefits, non means-tested benefits, etc.; taxes: personal income tax, social insurance contributions) and assess the role of each of these components in the redistribution of income. The redistribution of income through the tax-benefit system is evaluated by calculating an indicator derived from the Gini coefficient of pre and post social transfers and taxes. We concentrate on the income components which are responsible for the differences between the two measures, by decomposing the Gini coefficient by income source.

The rest of the paper is organized as follows. We continue with a brief overview of the general framework with respect to recent empirical findings concerning the evolution of income distribution in Romania. Then, we focus on the description of methodology, data and indicators used. The following section summarizes the most important findings concerning the estimation of the redistributive effect of the tax-benefit system in Romania. The paper ends with some concluding remarks.

2. General framework

During the last decades there has been a great interest in the measurement of the redistributive effect of social benefits and fiscal systems and of the contribution of each income component to redistribution. Starting with Kakwani (1977a, 1977b) who has laid the foundations for the measurement of income redistribution through the difference between the pre and post taxes and transfers Gini coefficients, a large strand of the literature in this field has focused on theoretical issues regarding measurement, but also on effective assessment of income redistribution through the tax-benefit system.

We mention as follows several recent relevant studies dealing with household income distribution in Romania. Most of these studies were focusing on the estimation and explanation of income inequalities, and very few are concerned with the effects of the tax-benefit system on income distribution.

One of the most relevant studies concerning the estimation and analysis of income inequalities can be attributed to Molnar (2010) who has decomposed income inequality by groups of main household characteristics. Her results show that the most important elements driving income inequalities between groups of households are education and labour market status. A decomposition exercise has been employed also by Zamfir et al. (2008) who have investigated the impact of remittances sent by Romanians working abroad on income inequalities between and within urban and rural areas. Their results show that remittances have driven the decline of income inequalities both between and within rural and urban areas. Dachin and Mosora (2012) have studied the inequalities driven by the regional distribution of household income and shown that the most relevant factors driving the unequal distribution of income by regions are the employment structure by economic activities and the prevalence of subsistence agriculture.

Concerning the effects of the country's economic development on income distribution, we mention Militaru and Stroe (2010) who have investigated the income dynamics in Romania between 2000 and 2007 using a growth incidence curve approach. Their findings clearly show that the economic growth has been pro-poor, meaning that the average income growth of poor households has been more substantial than the income growth of the rest of the households. Both households from rural and urban areas have been affected by crisis, but not equally. This is an issue addressed by Dachin and Sercin (2012) who concluded that household income in rural areas is less affected by crisis, compared to household income in the urban area. This can be explained by the different structure of household income by income sources between the two areas, the consumption from own-resources and the prevalence of informal income in the rural area as well.

The effectiveness of social policies in reducing income inequalities has been investigated by Precupetu (2013) who has focused on income inequalities in Romania after 1990. The concern on the tax-benefit system's effect on income distribution in Romania is very recent though. For example, Voinea and Mihaescu (2009) have measured the changes in the income distribution due to the income tax reform that took place in 2005, shifting from a progressive to a flat rate personal income tax and showed that only the richest 20% are clear winners of this reform. Avram et al. (2012) in their study on the distributional impact of fiscal consolidation measures taken in Romania (and other eight EU countries) during the recent economic crisis have shown that richer households have lost higher proportions of their incomes than poorer households, as a result of the above-mentioned measures. A similar analysis has been carried out by De Agostini et al. (2014), but they have measured the effects of all changes in the tax-benefit system (not limited to fiscal consolidation). They have concluded that in Romania the changes in the tax-benefit system were progressive, in the sense that their distributional effects were mainly beneficial for the bottom of the income distribution. Avram, Levy et al. (2014) have studied the redistributive effect of the tax-benefit system and found out that in Romania, unlike in most of the EU countries, social contributions increase income inequalities mostly due to higher limits set on contribution base.

3. Methodological issues

3.1. Methodology and data

We base our analysis on microdata from the European Union Survey on Income and Living Conditions (EU-SILC). The data is collected annually and it is nationally representative for the Romanian population. We use data collected during the 2008 and 2010 surveys, the income reference years being 2007, respectively 2009. Using updating factors by detailed income components (i.e. change in the average value of an income component between the year of the data and the current/ policy year), we adjust the value of the income variables from 2007 to 2008 and from 2009 to 2010-2013. Other variables (demographic, household size and composition, labour market variables) are kept constant to the survey years. We estimate the direct, static effect of the tax-benefit system on household income distribution.

We make use of a tax-benefit microsimulation instrument, namely the tax-benefit microsimulation model EUROMOD. The model comprises the Romanian tax-benefit policy rules for 2007-2013 and is built on EU-SILC data. The model can simulate the entitlement to cash social benefits (i.e. in-kind benefits are not taken into account) and tax and social contribution liabilities. The implemented tax-benefit policy rules are those in place at the middle of the year (i.e. the 30th of June), being assumed that no changes have occurred during the rest of the year. In the interpretation of results from microsimulations using

EUROMOD, one should bear in mind this very important issue. In our case, several relevant changes in the Romanian tax-benefit system took place during the second half of the year, they being effectively implemented in EUROMOD in the following year's rules. The model assumes 100% benefit take-up (exception in the case of the minimum guaranteed income) and no tax evasion. Asset tests that condition the entitlement for means-tested benefits are not simulated due to the lack of adequate information (EUROMOD Country Report: Romania, 2007-2009, 2009-2010, and 2009-2013).

The household disposable income is calculated as the sum between the original income (i.e. market or gross income) and the social transfers, minus direct taxes. The social transfers (benefits) are split into three categories: pensions, means-tested benefits (i.e. beneficiaries have to comply with some eligibility criteria regarding income levels below a threshold, often differentiated by household size, number of children, etc.; the beneficiaries may also be subject to asset tests for some of the benefits), non means-tested benefits (such as the state allowance for children, etc.). The category of direct taxes includes the flat-rate personal income tax (together with the tax allowance) and the social insurance contributions paid by the employees, self-employed (and pensioners) in order to cover the risks of retirement, sickness, unemployment, work-accidents, etc. The household size and age structure is taken into account by using the modified OECD equivalence scale. Thus, household income is adjusted and each household member is assigned the same amount of income.

3.2. Indicators

In order to measure the redistributive effect of the tax-benefit system in Romania, we use the common approach proposed by Kakwani (1977a, 1977b), who has suggested the assessment of the size of the income redistribution (RE) through the social benefit and tax system by the difference between the Gini coefficients of pre-fiscal income (no social benefits and taxes) (G_X) and post-fiscal income (G_N):

$$RE = G_X - G_N \tag{1}$$

The Gini coefficient measures the income inequality by the area between the Lorenz curve and the equality line. A progressive tax-benefit system moves the Lorenz curve towards the equality line; therefore the income inequality will be lower in this case. The redistributive effect is larger for greater average tax rates and greater progressivity. Atkinson (1980) and Plotnick (1981) pointed out that the tax-benefit system induces, besides the movement of the Lorenz curve, the re-ranking of individuals/ units, which can be measured by the difference between the Gini and the concentration coefficient of post taxes and transfers income. A few years later, Kakwani (1984) has decomposed the redistributive effect into vertical (progressivity) and re-ranking terms. In other words, the redistributive effect is reduced by the changes in the new ranking of individuals/ households which occurred in the post- tax and transfers system (see formula (2) below):

$$RE = V^K - R^{AP} \tag{2}$$

where, V^k is the Kakwani vertical effect and R^{AP} the Atkinson-Plotnick index of re-ranking.

The vertical effect can be computed as below:

$$V^K = G_x - D_N^x = \frac{t^x P_T^k}{1 - t^x} \tag{3}$$

where t_x is the average tax rate and P_T^k is the progressivity of the tax-benefit system (named the Kakwani index of progressivity).

The re-ranking effect is the difference between the Gini coefficient of post-taxes and transfers (G_N) and the concentration coefficient of post-tax and transfers income (D_N^x):

$$R^{AP} = G_N - D_N^x \tag{4}$$

We estimate the redistributive effect of the Romanian tax-benefit system between 2007 and 2013 and then, we decompose the effect into vertical and re-ranking effect. The Gini coefficient is decomposed by income source in order to estimate the contribution of each income component to income inequality, following the approach described in Lerman and Yitzhaki (1985) and in Stark, Taylor and Yitzhaki (1986), which allows the calculation of the impact that a marginal change in a particular income source will have on inequality. The influence of an income component on total income inequality depends on the importance of the income source with respect to the total income *(S_k)*, the extent of equality/ inequality in the distribution of that income source *(G_k)* and on the correlation of the income source with the total income distribution (R_k) (see formula (5)).

$$G = \sum_k S_k G_k R_k \tag{5}$$

Using the above decomposition we estimate the effect that 1% change in income from source *k* will have on total income inequality, as:

$$\frac{S_k G_k R_k}{G} - S_k \tag{6}$$

This approach concerning the measurement and decomposition of the redistributive effect of the tax-benefit system has been most recently used by Verbist and Figari (2014), and the decomposition of inequality by income source has been employed by Avram et al. (2014). These papers follow a comparative framework and analyse groups of EU countries. The contribution of our paper is that the methodology is applied on Romania, for the period between 2007 and 2013 and is focused on the dynamics of the redistributive effect, explaining the impact of certain changes that took place in the tax-benefit system on the size of the redistributive effect.

4. Main findings

4.1. Income distribution and the structure of the tax-benefit system, 2007-2013

Between 2007 and 2013, the household income dynamics has been strongly influenced by the economic downturn which became visible in Romania by the end of 2009. The average household income (real income, adjusted with the consumer prices index, reference 2007) has dropped in 2009 by approximately 5%. The negative developments of household incomes continued during the next two years, but the pace of decline was smoother than in 2009 (see Fig. 1).

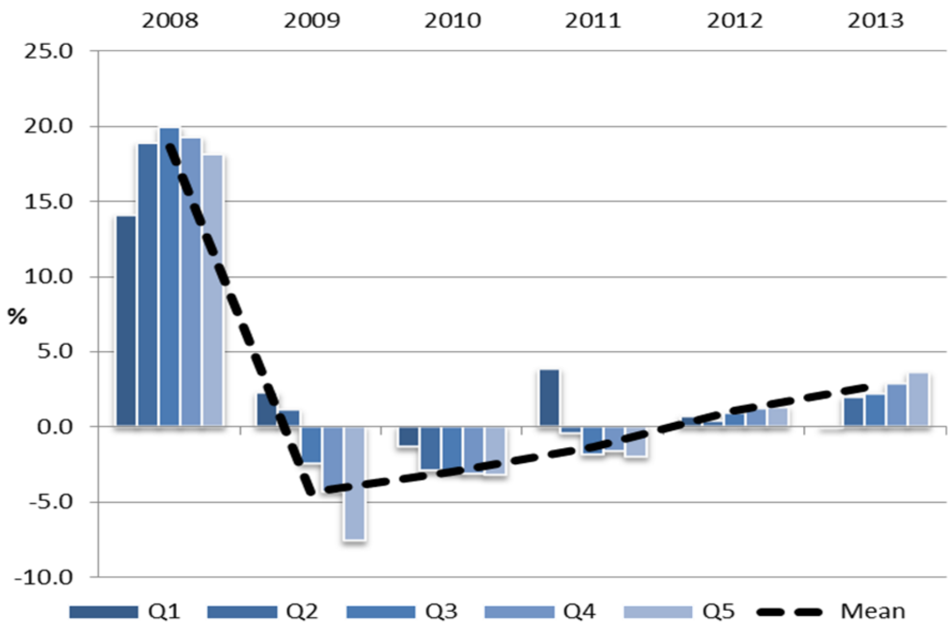


Figure 1. Annual percentage change in the average household disposable income, by quintile groups, %

Source: own calculations using EU-SILC, EUROMOD ver. G.1.0
Note: incomes are adjusted with the consumer prices index, reference year 2007; quintiles are constructed based on the equivalised household disposable income.

However, the developments were uneven along the income distribution (by quintile groups, each quintile comprises 20% of the population), the middle and the upper quintiles have benefited more from the economic growth in 2008, but also lost more during the crisis than the bottom quintiles, who have managed to preserve their levels of income from one year to another (except for the year 2010). This is mostly due to important changes in the tax-benefit system, the so-called "austerity measures" aiming fiscal consolidation, but also helping the worse off population. The fiscal policy changes that took place in 2010 and 2011 seem to have had a positive impact on household disposable incomes, while some of the changes in the social benefit system had a positive impact on household disposable income (i.e. changes in the means-tested benefits) and others a negative effect (i.e. the decrease of the unemployment benefit and the changes in the rules for the child raising allowance). Overall, the changes in the tax-benefit system seem to be progressive, as the bottom of the income distribution is advantaged in terms of income losses.

The Romanian tax-benefit system's largest component is public pensions. The pensions' share in the average household disposable income accounts for around 23-28%, slightly changing with the years. The other social benefits, either means-tested or not, do not exceed 8% of the household disposable income. The direct taxes, which consist of personal income tax and social insurance contributions account for almost 30% of the household disposable income. The social insurance contributions are designed to cover contingencies such as old-age, sickness, unemployment, work accidents, etc. and are paid by employees and self-employed. Additionally, pensioners with pension levels exceeding a statutory threshold pay the health insurance contribution.

As it can be seen in the figure bellow (Fig.2), the structure of the tax-benefit system has not changed considerably between 2007 and 2013. We notice though an increased share of social benefits in 2010. This can be explained by the increase of the income eligibility thresholds for some means-tested benefits (i.e. minimum social pension, social assistance benefit). There was also a decline in the share of social contributions after 2011, most likely as a result of the introduction of an upper ceiling to the social insurance contribution of employees and self-employed, and the introduction of lower limits to health insurance contribution for all population (active population and pensioners). In 2011, the share of social benefits has contracted, consequence of the following changes: decrease of the unemployment benefit, maximum threshold set for the child raising benefit and the policy rules were changed, increase of the child raising incentive, the allowance for the new-born children was abolished, the income thresholds and the amounts of the means-tested family benefit and the means-tested heating benefit have been changed. We should note that some of the above mentioned changes took place during the second half of the year 2010, being part of the austerity measures, but according to EUROMOD rules, they are implemented in 2011 (see the previous section on methodology of the paper).

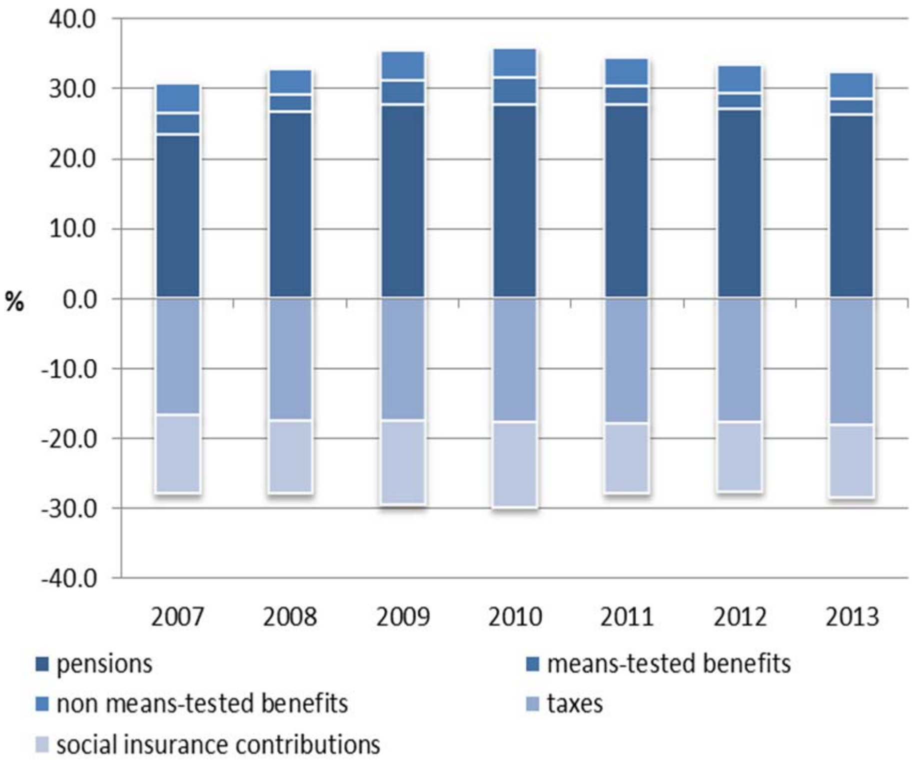


Figure 2. Structure of the tax-benefit system, % of household disposable income, 2007-2013
Source: own calculations using EU-SILC, EUROMOD ver. G.1.0

It is important to mention that the structure of the tax benefit-system varies a lot by quintile groups constructed based on the equivalised household disposable income. Naturally, the bottom quintile (1st quintile) relies on means-tested benefits to a much greater extent than the other parts of the income distribution. On the other hand, the upper quintiles (4th and 5th quintiles) are paying a higher proportion of their disposable income as personal income taxes and social insurance contributions (see Fig. 3). This picture points towards a progressive tax-benefit system, where poorer households benefit more from social transfers and the richer pay more taxes, but the size of the redistribution is to be treated in the next sub-section.

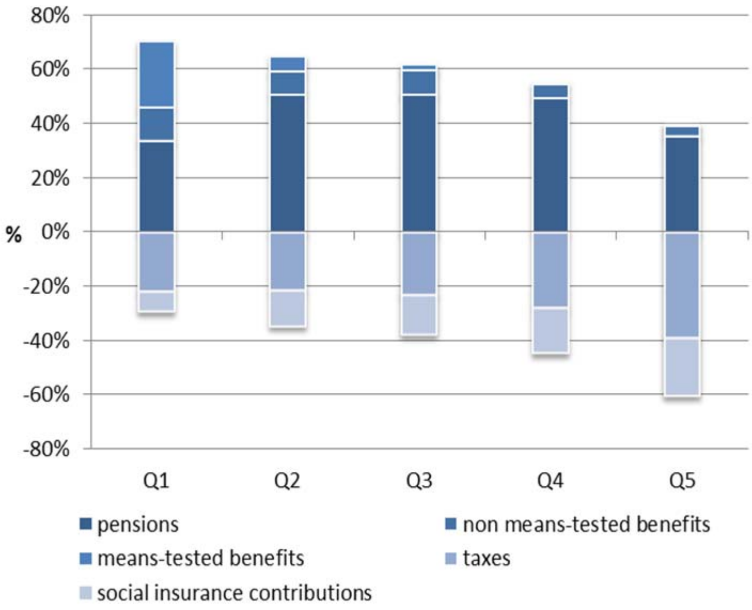


Figure 3. Structure of the tax-benefit system, by quintile groups, % of household disposable income, 2013
Source: own calculations using EU-SILC, EUROMOD ver. G.1.0

4.2. Redistribution of income through the tax-benefit system

We have measured the redistributive effect of the tax-benefit system as a whole, by the difference between the Gini coefficient of pre and post taxes and transfers, as described in detail in the section on methodological issues. The results are presented in the figure below (see Fig. 4). More than half of the income inequality before taxes and transfers (i.e. original or market income) is reduced through the tax-benefit system. It seems that the economic crisis has led to the decline of income inequalities, as richer households have lost more of their market income, but also due to the tax-benefit system changes that were adopted in order to cope with the crisis. Thus, the redistributive effect of the tax-benefit system was highest in 2011 and lowest in 2009.

We have decomposed the redistributive effect into vertical effect and re-ranking effect, the idea behind being that the vertical effect is actually reduced by the re-ranking of individuals that has occurred in the post-tax and transfers system. As it can be seen in the figure below (Fig. 4), the re-ranking effect resulted from the redistribution of income lowers the total redistributive effect by approximately 40%. Only in 2009 and 2010, the re-ranking effect has exceeded 50% of the total redistributive effect. Nevertheless, the dynamics of the redistributive effect is strongly driven by the vertical equity term.

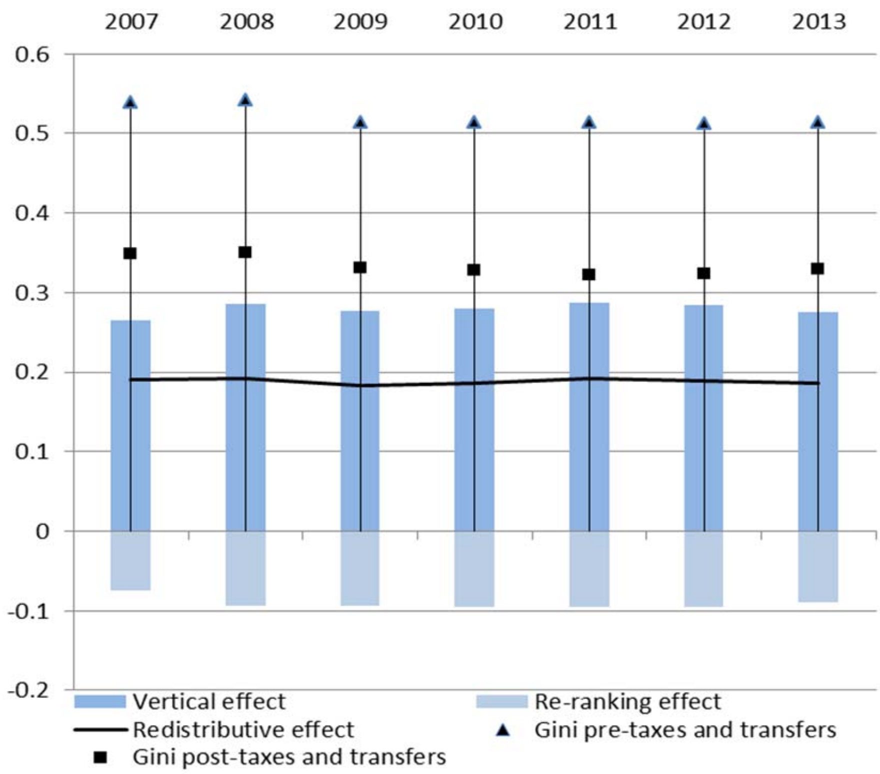


Figure 4. Redistributive effect of the tax-benefit system, 2007-2013
Source: own calculations using EU-SILC, EUROMOD ver. G.1.0

The decomposition of the Gini coefficient by income source shows that pension income is the most important driver for inequality reduction from all income components. This is because pensions are the largest component of the tax-benefit system and are more equally distributed among the whole population than other income components (except for the personal income tax and the social insurance contributions). The means-tested benefits contribute to the reduction of income inequality due to their negative correlation with the distribution of total income, as the lower part of the income distribution benefits more from these transfers.

The personal income tax is decreasing income inequality due to its distribution and strong negative correlation with total income distribution. However, the size of the effect is lowered by the nature of the tax rate, this being a flat-rate tax. The non means-tested benefits have lower impact on income inequality. As expected, their share in total household disposable income is the lowest. The social insurance contributions have acted in the sense of inequality reduction after 2011, as a result of several important changes that took place in 2011 in the social insurance system. On one hand, an upper ceiling was introduced for the social insurance contribution paid by employees and self-employed which could have increased income inequality, but this was counterbalanced by the introduction of lower limits to health insurance contribution in the case of pensioners, thus the overall effect being in favour of inequality reduction. The dynamics of the marginal effect of personal income tax shows a decline in the contribution of the income tax to income inequality reduction.

During the first years of economic crisis (2009-2010), the means-tested benefits have strongly acted as to decrease income inequalities.

5. Conclusions

The paper has attempted to study the income distribution of Romanian households, concentrating on the structure of the tax-benefit system and on the effects on the income redistribution of income components and of the system as a whole.

Our analysis covers the period between 2007 and 2013 and is based on annual nationally representative microdata from the European Union Survey on Income and Living Conditions (EU-SILC). We use the tax-benefit microsimulation model EUROMOD in order to simulate the components of the tax-benefit system. In order to measure the redistributive effect of the tax-benefit system in Romania, we use the approach proposed by Kakwani (1977a, 1977b) and we assess the size of the income redistribution through the social benefit and tax system by the difference between the Gini coefficients of pre-fiscal and post-fiscal income. We decompose the redistributive effect into vertical and re-ranking effect. In order to establish the contribution of each income component to income inequality, we decompose the Gini coefficient by income source, following the approach described in Lerman and Yitzhaki (1985) and in Stark, Taylor and Yitzhaki (1986), which allows the calculation of the impact on income inequality of a marginal change in a particular income source.

Our results show that between 2007 and 2013, the household income dynamics has been strongly influenced by the economic downturn which became visible in Romania by the end of 2009. The average household income has dropped in 2009 and the negative developments have continued for the next two years, but the pace of decline was smoother than in 2009. Starting from 2012, we notice a slight increase in the average level of household income. Though, the developments were unequal along the income distribution, the middle and the upper quintiles have benefited more from the economic growth in 2008, but also lost more during the crisis than the bottom quintiles, who generally have managed to preserve their levels of income. This latter result is mostly due to important changes that took place in the tax-benefit system.

With respect to income redistribution, the results indicate that income inequality before taxes and transfers is reduced to half through the tax-benefit system. During the economic crisis, richer households have lost more of their market income. This is reflected in the reduction of the original income inequality (before taxes and transfers). Additionally, the role of the tax-benefit system was considerable in income inequality reduction, due to changes that were adopted in order to cope with the economic crisis. The decomposition of the redistributive effect into vertical effect and re-ranking effect shows that the redistributive effect is reduced by the re-ranking of households that has occurred in the post-tax and transfers system. The re-ranking of households lowers the total redistributive effect by approximately 40%. In 2009 and 2010, the re-ranking effect has exceeded 50% of the total redistributive effect. However, the dynamics of the redistributive effect is mainly driven by the vertical equity term.

The decomposition of the Gini coefficient by income source has shown that pensions, which account for the larger part of the tax-benefit system, play the most important part in income redistribution, while social insurance contributions increase income inequalities (especially before 2011). The personal income tax is redistributive, though its effect is not substantial due to its flat-rate. Means-tested and non means-tested social benefits are conducive to income inequality reduction. During the economic crisis, the means-tested benefits have been the most influential on decreasing income inequality.

References

1. Atkinson, A. **Horizontal Inequity and the Distribution of the Tax Burden**. In Aaron, H. and Boskin, M. eds. "The Economics of Taxation", Washington D.C. The Brookings Institution, 1980
2. Avram, S., Figari, F., Leventi, C., Levy, H., Navicke, J., Matsaganis, M., Militaru, E., Paulus, A., Rastrigina, O. and Sutherland, H. **The distributional effects of fiscal consolidation in nine EU countries**. Research Note 1/2012 of the European Observatory on the Social Situation and Demography, European Commission, 2012
3. Callan, T., Leventi, C., Levy, H., Matsaganis, M., Paulus, A. and Sutherland, H. **The distributional effects of austerity measures: a comparison of six EU countries**. Research Note 2/2011 of the European Observatory on the Social Situation and Demography, European Commission, 2011
4. Dachin, A. and Mosora, L.C. **Influence factors of regional household income disparities in Romania**, in Journal of Social and Economic Statistics, no.1, vol. 1, 2012.
5. Dachin, A. and Sercin, A. **Effects of the economic crisis on rural household incomes in Romania**, Paper presented at the 3rd International Symposium "Agrarian Economy and Rural Development - realities and perspectives for Romania", Bucharest, Romania, 2012
6. Davidescu (Alexandru), A. A. **Estimating the size of romanian shadow economy a labour approach**, Journal of Social and Economic Statistics, vol. 3, no. 3, 2014, pp. 25-37
7. De Agostini, P., Paulus, A., Sutherland, H. and Tasseva, I. **The effect of tax-benefit changes on income distribution in EU countries since the beginning of the economic crisis**. Research note 03/2013 of the Social Situation Monitor, European Commission, 2013
8. Immervoll, H., Levy, H., Lietz, C., Mantovani, D., O'Donoghue, C., Sutherland, H. and Verbist, G. **Household incomes and redistribution in the European Union: quantifying the equalizing properties of taxes and benefits** in Papadimitriou, D. B. (Ed.) "The Distributional Effects of Government Spending and Taxation", London, Palgrave Macmillan, 2006
9. Kakwani, N. **Measurement of tax progressivity: an international comparison**, in Economic Journal, Vol. 87, 1977, pp. 71-80
10. Kakwani, N. **Application of Lorenz Curves in Economic Analysis**, in Econometrica, Vol. 45, No. 3, 1977, pp. 719-727
11. Molnar, M. **Income distribution in Romania**, MPRA Paper No. 30062, 2010
12. Militaru, E. and Stroe, C. **Poverty and Income Growth: Measuring Pro-Poor Growth in the Case of Romania**, Proceedings of the 11th WSEAS Mathematics and Computers in Science Engineering, WSEAS Press, 2010
13. Precupetu, I. and Precupetu, M**. Growing inequalities and their impacts in Romania. Country Report**, GINI Growing inequalities' impact, 2013

14. Plotnick, R. **A Measure of Horizontal Inequity, in Review of Economics and Statistics**, Vol. 63, 1981, pp. 283-288
15. Reynolds, M. and Smolensky, E. **Public Expenditures, Taxes, and the Distribution of Income: The United States**, New York, Academic Press, 1977
16. Sutherland, H. **EUROMOD: the tax-benefit microsimulation model for the European Union**. In Gupta, A. and Harding, A. (Eds.) "Modelling our future: population ageing, health and aged care", International Symposia in Economic Theory and Econometrics Vol. 16, Elsevier, Amsterdam, 2007, pp. 483-488
17. Stroe, C., Militaru, E., Avram, S. and Cojanu, S. **EUROMOD Country Report: Romania 2007-2010**, ISER, University of Essex, Colchester, 2012
18. Verbist, G. **Redistributive effect and progressivity of taxes: an international comparison across the EU using EUROMOD**, EUROMOD Working Paper Series: EM5/04, 2004
19. Voinea, L. and Mihaescu, F. **The impact of the flat tax reform on inequality – the case of Romania**, Romanian Journal of Economic Forecasting, Vol. 4, 2009
20. Zamfir, A.M., Mocanu, C., Militaru, E. and Pirciog, S. **Impact of Remittances on Income Inequalities in Romania,** In Schuerkens, U. (Eds.) "Globalization and Transformations of Social Inequality", Routledge Taylor& Francis Group, New York, London, 2010, pp. 58-75

[1] **Acknowledgment**
This paper was co-financed from the European Social Fund, through the Sectorial Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/138907 "Excellence in scientific interdisciplinary research, doctoral and postdoctoral, in the economic, social and medical fields -EXCELIS", coordinator The Bucharest University of Economic Studies.

SURVEY REGARDING RESISTANCE TO CHANGE IN ROMANIAN INNOVATIVE SMEs FROM IT SECTOR[1]

Eduard Gabriel CEPTUREANU
PhD, Assistant Professor,
Bucharest University of Economic Studies, Romania

E-mail: eduard_ceptureanu@yahoo.com

Abstract:
Unfortunately, few changes predominantly generate positive effects involving major effort and costs are often not far short of expectations. Why efforts to implement the changes result in failure or do not match the expected results? We will try to formulate a response based on achieving an investigation on a sample of 819 SMEs innovative IT Romanian order: (i) identify the types of resistance to change prevailing in the analyzed companies; (ii) identify change management tools used to reduce resistance to change; (iii) proliferation substantiate future directions of change management in Romanian.

Key words: change management, innovative SMEs, resistance to change

1. Introduction

Resistance to change issue is based on a set of logical reasons arising from the third law of Newton's dynamics that every movement always meets resistance forces. To overcome resistance to change we must answer at least two questions:

- what are the causes of resistance to change?

and

- how to work on these causes to eliminate or substantially reduce?

Before we attempt to answer these questions, we consider useful to present the opinion of Rick Maurer, author of "Beyound The Wall Of Resistance". According to it, the base resistance are two sets of elements that represent two distinct levels:

- Level 1- such information-logical, visible, relatively easier to see and countered;
- Level 2- personal and emotional, that often people do not flaunt it, to be discovered, evaluated and addressed specific means.

This postulate is reflected by P. Senge showing the life cycle of change through a curve that unrealized potential for growth and development due to resistance to change manifested in various forms:

American researcher I. Ansoff [1] notes that resistance to change is "a multiaspectual phenomenon generating unexpected obstacles in the process of organizational change and instability thus introducing unexpected efforts in the process. At the same time, is an expression of irrational behavior of organization members who refuse to recognize the new dimensions of reality and ignore the logical arguments. "

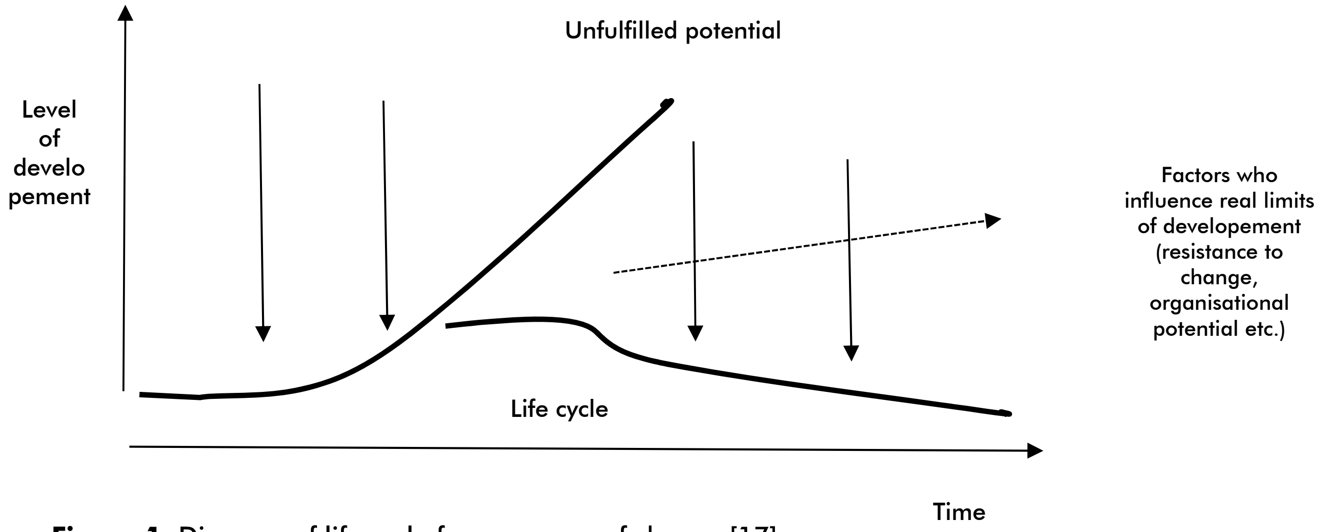


Figure 1. Diagram of life cycle for processes of change [17]

Based on the statements, we can consider that (Ceptureanu, 2012):

- resistance organization is a permanent phenomenon generated by the tendency of a system to maintain a relatively stable equilibrium inside and organizational change is perceived as a destabilizing phenomenon;
- resistance to change should not be seen solely as a negative reaction because, given the appearance of objectivity creates the prerequisites necessary to test the viability of new ideas;
- although resistance to change is objective and has a legitimate source is subjective element of the system (organization) - the man who has a major importance in the development activity, fulfilling both the role of "organizer" (by behavior, initiative, incentives) and the "destabilizing". given that the source of opposition to change a form subjective element of the system, as the source of this phenomenon objectively analyzing subjective reasons such as: fear of the new and inertia are presented in several forms. why people are hostile to something new are very different and are limited to the following associations: property damage, loss of current status, new responsibilities, limitation of rights, liquidation function, increase the volume and complexity of work, loss of moral advantages (status, authority, power), replacing old methods of work, formal and informal relationships, feelings of incompetence for new tasks, functions.

This particular form of organizational behavior - resistance can occur in two forms:

- Active: when the manager hears, sees, understands why a negative feedback and take steps to change it;
- Passive (hidden / masked) when nobody open disagreement, but no changes are not implemented (no resistance or "deformation").

Even transformations routine daily occurring in the coordination of any business, such as launching new products, forming interactions or new systems are often accompanied by tension, disagreement, stress - in other words by resistance . If this is the case of reduced scale transformations, we can imagine how hard it is to achieve major changes involving changes in formal and informal structures, such as: restructuring the organization, merger, managerial reengineering, culture change etc [16].

An analytical research conducteb by Ovidiu Nicolescu [13] identify the most frequent sources of resistance to change, which refers both to those directly involved in changing and changing context. In Figure 2 we present the main sources of resistance to change:

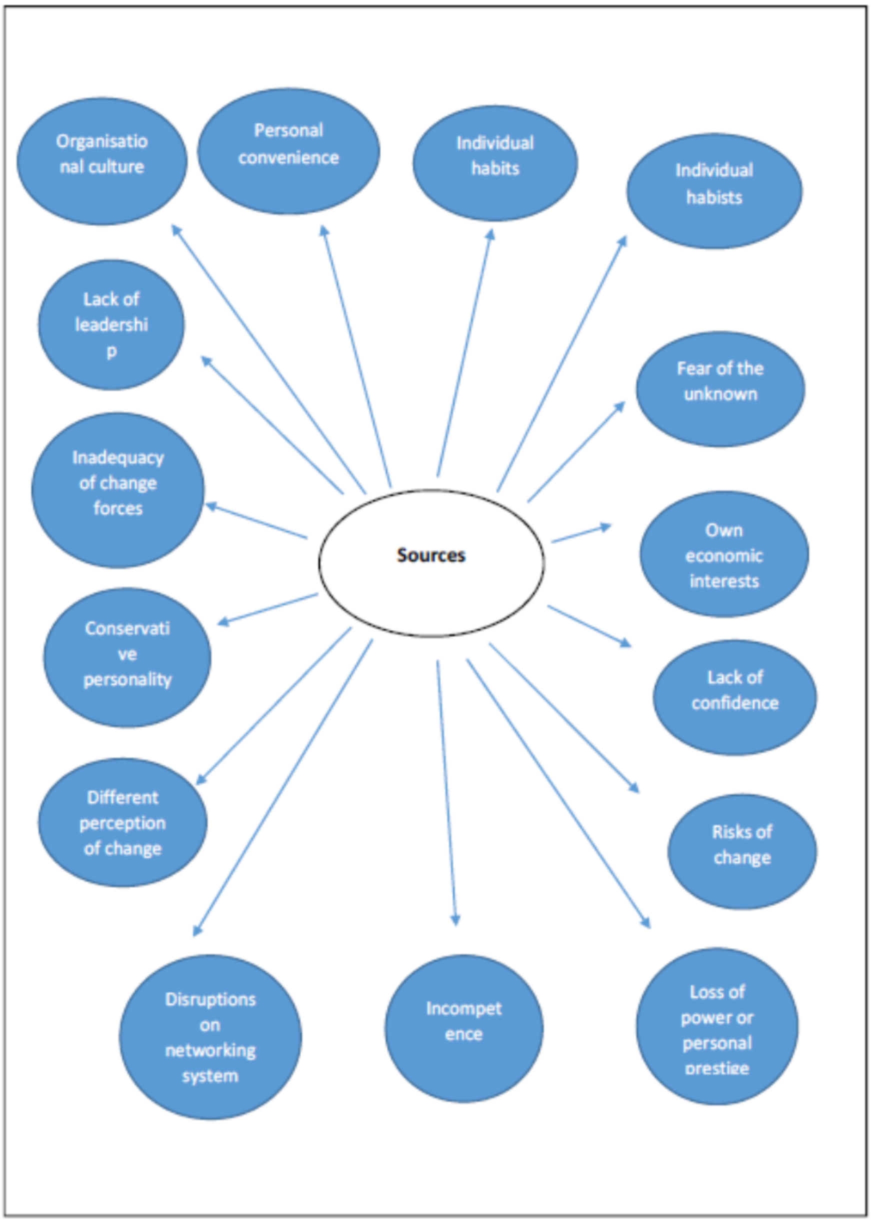


Figure 2. The main sources of resistance to change by employees

Without specifically insist further explained briefly what is each potential source of resistance to change:

- **personal convenience** is a factor that is found in a certain proportion to each person. At the level of each of us is manifested with a certain intensity tend to save available forces, not always use them to make something new, mulţumindune with what we have, with the current situation, even if not the best or favorable for us. The expression of this situation is devoted to "anything goes".

- **individual habits**. Over time, each person has formed certain habits, resulting from the specific personality and background conditions involved. There is a tendency not to give up our habits and organizational changes that are involved always affects some of our habits.
- **the fear of the unknown**. No matter how strong a person psychologically, how much confidence in itself and in those around him, changing and its promoters, always appears a sense of anxiety and fear. The stronger it is, the resistance to change is more intense.
- **own economic interests**. Sometimes expected changes may cause a decrease in meeting our economic interests in the organization - salary, bonus, bonuses, access to machinery spaces protocol etc. Such situations are strong motivations for the persons concerned to oppose, to "resist" change.
- **lack of confidence in change and / or those who promote it**. Whenever a person involved in the change process does not trust those who promote or does not believe in its success, it will manifest itself, consciously or unconsciously, a certain resistance. To prepare people for change and promoting its prestige and possessing the ability to exchange helps eliminate the inhibitor of change.
- **the risks involved in change**. When a person certain risks associated with the expected change in personal, group or organization, even if its promoters trusts and the end result, he will show some restraint or opposition to engage in change.
- **loss of power and / or reducing personal prestige**. Such motivation to resist the change applies particularly to managers and specialists, people in formal or informal power and prestige are intrinsic components of their work. Naturally, when I see that the change envisaged will diminish their power and prestige, they will be tempted to block this change.
- **incompetence**. Organizational change always causes changes in different proportions in employee tasks and how to do. In situations where employees do not have the knowledge to achieve them, it is likely that these changes seek to avoid or to reduce as much.
- **disrputions on networking system** Disruption of the person within the organization. Each employee is integrated in a micro office in the organization, being in some work and personal relationships with other people. When the employee is satisfied with it, and the change will affect the relational context and position within it, it will tend not to get involved and do not favor this change.
- **different perceptions of change**. Presentation by the managers of change that will achieve is not always perceived in the manner intended by them. The employees who develop different perceptions of the objectives, content, implications and effects of change, is likely not generate the same motivation for change sometimes occur even motivations antischimbare generating passivity or even resistance to their implementation.
- **conservative personality**. A proportion of the population in any country, is characterized by native tendencies to avoid new, the lock, excessive cantonându the past and present. The ability to take risks, tolerance for ambiguity inherent in innovation, resistance to stress are reduced. Employees who fall into this category -

and they are not few - will always tend to block change, or at least not to get involved in their operationalization. They must apply special treatment, especially for strategic change, large-scale.

- **inadeqacy of change forces**. As noted, in any organization there are forces that resist change generated by previous factors. Countering their organizational level is done by generating forces that promote and encourage change, higher premiums. If not done this superiority, perceived by employees and other stakeholders, their resistance to change will be more intense.
- **lack of leadership**. Multiple internal sources, intrinsic resistance to change, you have listed, can be removed and / or substantially diminished when those people show their impact on a strong leader influential promoter of innovations consistently. Whenever there is such a leader, employees will manifest insufficient responsiveness, passivity and even resistance to expected changes. The leader is a driving force for successful change.
- **organisational culture**. Although it is an external factor in relation to persons involved in changing organizational culture strongly influences their attitude towards change. Companies that possess organizational culture focused on innovation, effort, team spirit, obtaining performance from employees will induce a favorable attitude change, thus diminishing the explicit and implicit resistance to change.

Naturally, this factors are not exhaustive, but only a selection of the most intense and frequent, occurring in firms in general, including those in Romania (Ceptureanu, 2010).Resistance to change is a natural psychological reaction caused by the action of any of the factors listed above. People always need a certain level of stability and safety, and the change involves a new situation of uncertainty that causes a feeling of uncertainty and therefore it is likely that employees feel vulnerable in several respects (risk taking, committing mistakes, s. a.).

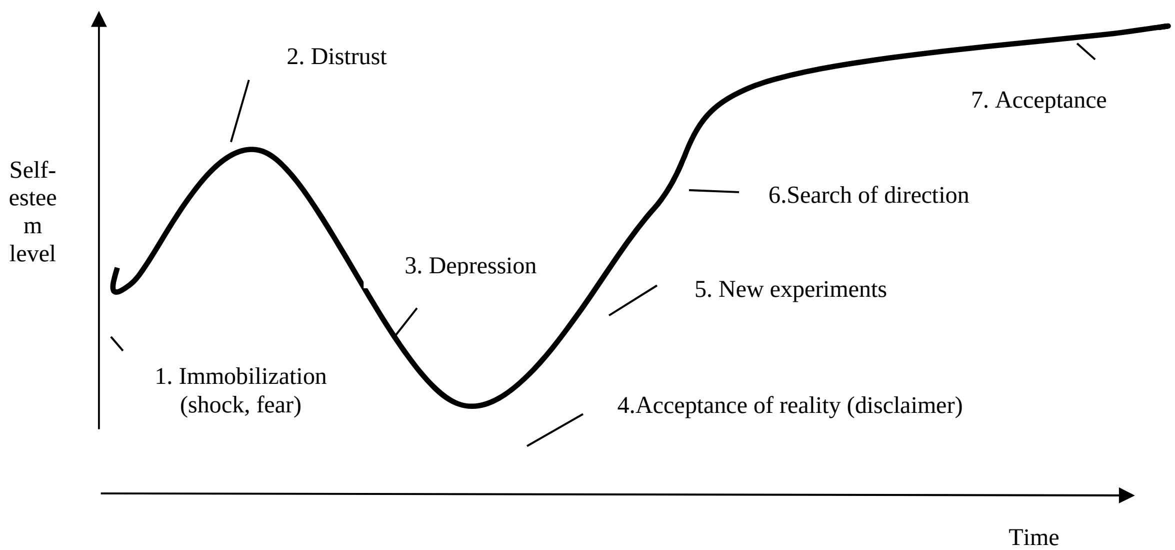


Figure 3. Changes suffered by self-esteem during transitions [5]
Note: 1,2,3, - negative reactions to change; 4- neutral reaction; 5,6,7 - positive reactions to change

Very few people are prepared to give up ideas for your loved obvious risks. Difficult to give up something very specific for the human being and it happens because it seems quite dangerous to give a firm foothold and you head into the unknown (Ceptureanu, 2012). Every instinct of human logic, emotion of self-preservation and oppose this action extremely risky. From the point of view of psychology whose criteria do not necessarily reflect those of logic, these events are easier to understand. The vast majority of people under risk losing the flexibility of thinking. Preventing and resolving resistance to change depends on the ability to understand the reactions of individuals in such situations vary depending on a variety of criteria: mentality, character and culture. Thus, some want new and are pleased transformation, while others feel fear and exhibit resistance to loss of the status quo. It is possible that ambivalence to get more complex aspects: people may welcome the change and at the same time, to show resistance to its implementation (Masssa, 2008).

When reacting to a significant change in people, according to L. Clarke follows a predictable pattern of response - was called "transition curve" (Figure 3.) Showing an individual's reaction to change in a period of time.

As we see, the beginning of the transition process that involves changes are negative aspects related to the perception of change, followed by adjustment period, which lasts differently to different people, depending on the individual flexibility.

According to the American consultant J. Kotter [9] differ tangled emotions that occur change as anger, pessimism, arrogance, pride, cynicism, panic, fatigue, distrust, anger and emotions that help to achieve that change: positive trust, optimism, results orientation, satisfaction the positive results achieved, incentives, concern, excitement, hope. Also the author emphasizes, in particular, the need to act on emotions cause people to change itself and later to change things can change. Emphasizing and arguing prevailing social aspects of change for a successful outcome researcher suggests the following method of working with people: **SEE -> FEEL -> CHANGE**, i.e. employees must be shown opportunities and threats in a convincing manner and particularly the EU would achieve it aware of the need for change and actually achieve it (Lester, 2001).

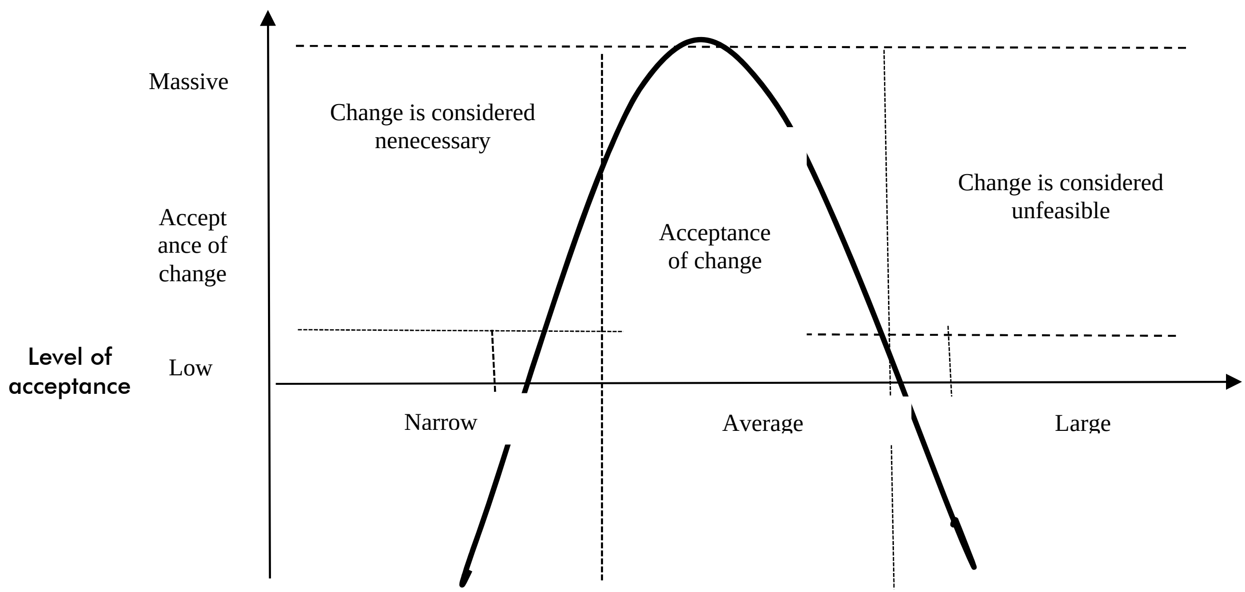


Figure 4. Dependence between dimensions of change and levels of acceptance [8]

According to G. Johns [8] in general, there are two reasons that "justify" change (Figure 4):

1. The change is not necessary because there is only a small discrepancy between the condition of the moment and the ideal state of the organization;
2. The change cannot be achieved because there is a discrepancy between the present state too large of things or requirements.

As we see in the figure as the size change (their size or depth) is higher, the change is more disagreeable, and the same reaction when forming its dimensions are small. Indeed, it is hard to convince and to convince others that "good is the enemy of good" and that perfection has no limits (Prusak, 2007). The middle is the only approved: the magnitude of change coincides with the needs, desires for change and the potential that we can use.

2. Change management research on innovative Romanian SMEs

2.1. Sample size and structure of SMEs

To analyse the trends, motives and peculiarities of change management in ITC innovative Romanian SMEs, we use a survey database that was collected by Romanian National Trade Registration Office- main legal entity with function of keeping the register of trade. The survey targeted SMEs, defined as enterprises with 1-249 employees, and also large companies and was implemented by means of computer-assisted telephone interviewing. Data collection was done over a 2 month period during September- October 2014. To reliably identify trends only respondents with long tenure and representing enterprises that systematically innovate and implement change, were selected. The survey therefore started with screening questions. Respondents first indicated if their company had developed implementation of change management processes and at least one innovation in the past year. This could either be a product, process-, organizational- or marketing-related innovation as defined by the Oslo manual (a set of integral guidelines for the collection of innovation data, see OECD, 2005). Secondly, respondents had to be involved at least in one implementation of change management process during the last 5 years. In this way, the screening ensured that respondents all represented SMEs with systematic efforts in change and they were in a position to adequately judge if and how change processes had developed over the past years. The sample was represented only by representatives of ITC domain (generate by difficulties to identify innovative SMEs on Romanian economy) and disproportionally stratified across four size classes (0–9, 10-49, 50-249 employees) (official EU classification of SMEs) and > 250 employees. Enterprises with less than 10 employees (micro-enterprises) were not excluded since they generally have limited identifiable innovation activities and this population usually contains many start-ups who are very innovative in order to survive on the market. Interviewers explicitly asked for those who were responsible for implementation of change, i.e. small business owners, general managers or staff managing new business development activities.

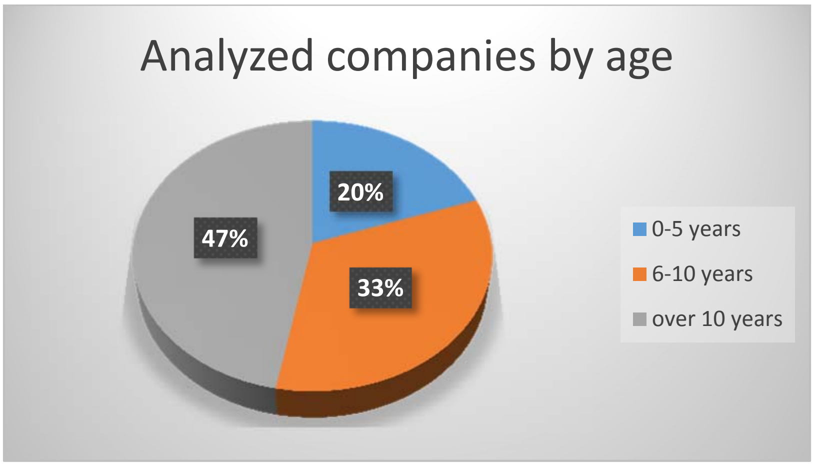
Distribution by Romanian counties

No.	Counties	Number of companies
1	Alba	1
2	Arad	3
3	Argeș	4
4	Bacău	1
5	Bihor	16

6	Bistrița-Năsăud	2
7	Botoșani	2
8	Brașov	32
9	București	435
10	Buzău	2
11	Călărași	1
12	Cluj	83
13	Constanța	12
14	Covasna	2
15	Dolj	10
16	Galați	5
17	Giurgiu	2
18	Gorj	2
19	Harghita	2
20	Hunedoara	3
21	Iași	39
22	Ilfov	20
23	Maramureș	7
24	Mureș	13
25	Neamț	4
26	Olt	1
27	Prahova	19
28	Satu Mare	7
29	Sibiu	25
30	Suceava	3
31	Timiș	54
32	Tulcea	1
33	Vaslui	2
34	Vâlcea	4
	Total	**819**

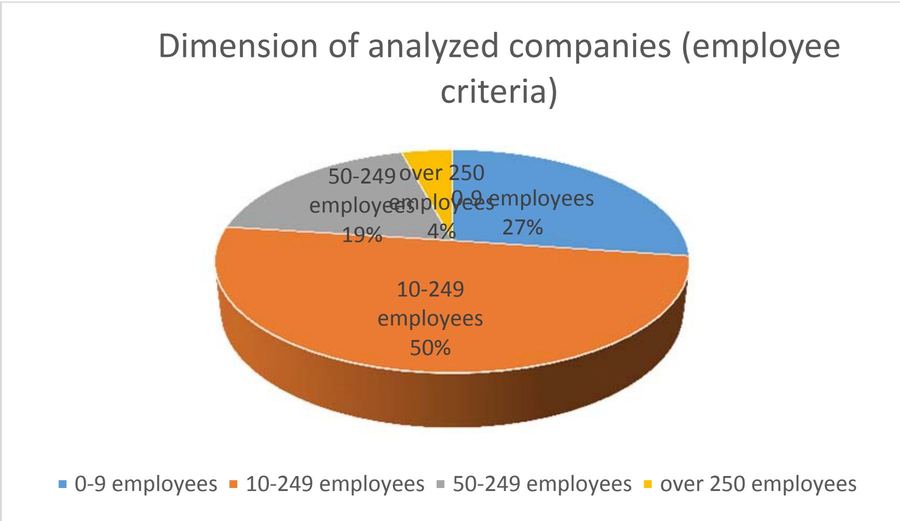
Given the age of SMEs (Figure 1), most of the companies that were the subject of research were older the 10 years (47%), followed by enterprises between 6-10 years (33%) and those established in the last 5 years (20%).

0-5 years	6-10 years	over 10 years
161	273	385
20%	33%	47%



Source: own research

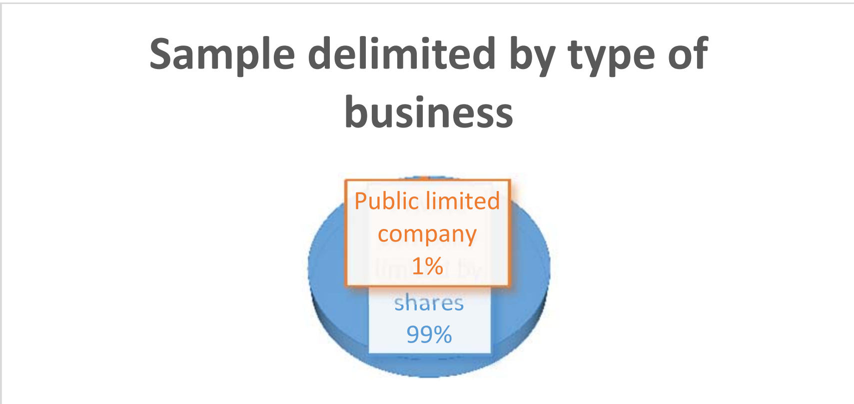
Considering the size of the organizations, as shown in Figure 2, small enterprises represents 50% of the SMEs surveyed, microenterprises account for 27% and midsize companies have a rate of 19% .We also consider a sample of 4% of large companies in order to simulate accurately the conditions of Romanian economy.



0-9 employees	10-249 employees	50-249 employees	over 250 employees
223	408	154	34
27%	50%	19%	4%

Source: own research

As regards the legal form of SMEs, 99% of companies are private companies limited by shares and 1% public limited companies. See Figure 3.



Source: own research

Type of business

Private company limited by shares	Public limited company
808	11
99%	1%

Given the NACE codes, the structure of companies is as follows: 54.9% of companies NACE code principal- 6201 (Activities to develop custom software (software-oriented client), 20.9% CAEN 6202 (consultancy activities information technology), 1.2% -

NACE 6203 (management activities (administration and operation) of calculation), 9.9% - NACE code 6209 (Other information technology service activities), 10,9% - NACE 6311 (data processing, hosting and related activities), 1,2% - NACE 6312 (activities of web portals) and 6391 and 1% mainly operate on CAEN code 6399 and 6391 (Other information service activities).

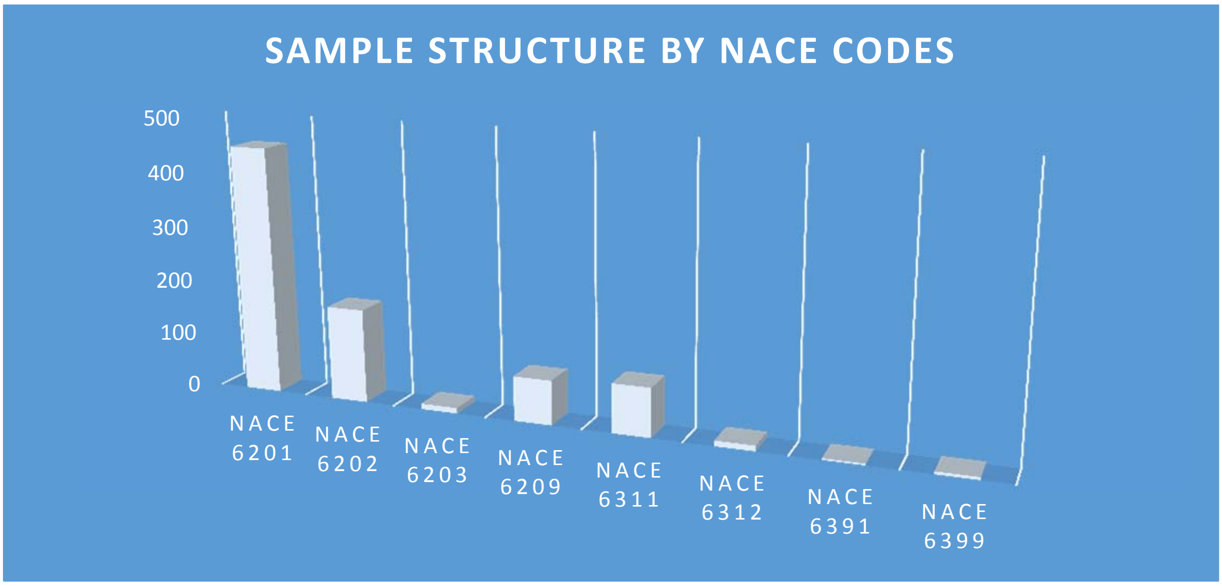


Figure 5. Sample structure by NACE code
Source: own research

2.2. Information about the change processes in investigated companies

Table 1. Survey variables

Variables (partial approach)	**Findings**
Correlation between change and survival of the organization	74,5% of respondents agree with the statement that the change provides better conditions for survival of the company in the medium and long term
The level of involvement of organizational subdivisions in the process of change	Regarding organizational structures involved in the change process we emphasis the role of Sales Department (33,3%), followed by R&D Department (31,9%) and Production Department (18,8%). Unfortunately, Management Dept. is ranked 4th.
Perception of organizational changes on the market	Changes results on the market is reflected especially into creation of a product or service (39,2%), use of new resource (e.g. knowledge builders, T managers etc.)-34,3% or using an old idea/product/service into a new manner (13,95)
Determinants of change	The determinants of change- new ideas are represented by higher-level managers (63,61%), changing interests of owners (59,7%), liquidity crises and success crises (58,36%).58,24% of respondents believe that the process of organizational change cannot be controlled completely vs. 41,75% believe that it is possible to direct organizational change.
Areas affected by the change	The areas highly affected by the change are represented by new products/ services (55,31%) human resources (51,52%), organizational structure (49,08%).
The types of change used in companies analysed	Proactive change represent roughly 54,21% from answer opposed to reactive changes-45,78%
The techniques used to implement organizational changes	52% of respondents use techniques such as restructuration in crisis conditions (50,54%), managerial reengineering of BPM instruments (46,15%) and organizational development (28,69%)
The success of	**Negative results during implementation were obtained in roughly**

implemented changes	**62,39% of analysed companies, while only 22,83% of respondents were fully satisfied with the results**
Role of subjects of change	Only 25, 07% of respondents mentioned that mid-level managers played the role of strategists, 57,14% were implementators and 17,79 were passive subjects of change.
Measuring resistance to change the categories of employees	72,28% of mid-level and high level managers have positive reactions to change, the remaining 39,82% saw the change as a threat
Manifestations of resistance to change	Unfortunately 74,72% of employees show an active resistance to change
Frequency of using tactics to reduce resistance to change -actions of senior managers on change	Reducing resistance to change was obtained negotiation with employees reluctant to change (21,5%), Staff training (21,2%), Providing information needed for the adaptation of change (12,85%), Managers personal involvement in change management (18,8%), Stimulation and support in adapting change (14,2%), Rotation posts (6,5%) and Job enrichment (5%)

Source: own research

3. Conclusions

Generally, considering the results, we find out that that:

- Resistance to change was and is a problem that faced all the organizations investigated, and attempts to reduce resistance to change problematic went to all;
- Conduct direct actions change (implementation plan) so was a difficult for domestic enterprises;
- Achieve quick results is only possible if it was developed a good plan of action coupled situational management practices in situations when there were "surprises" that it was not possible to foresee at the planning stage;
- Strengthening the change in corporate culture is an intangible result is sometimes very difficult to get him and requiring time. Respondents recognized that this requirement has been ignored in the past unconscious, lack of knowledge of change management;
- Assess the results of implementing change can be achieved easily by comparing staff to plan, analysing external and internal sources of information taking into account the social implications of changes completed;
- A distinction is made between strategic and operational change;
- Use models to stimulate and clarify thinking about change and impacts;
- Pursue the technical, cultural, etc. - Are interdependent;
- Attention is paid to transition management, and not the final aspects of change;
- Strategies are not filled with procedures, tactics;
- Preparatory measures (changing organizational culture and conducting training with employees) are vital Success is the approach;

References

1. Ansoff, I. and McDonell, E. **Strategic Management,** Palgrave Macmillan Publishing House, 2007
2. Ceptureanu, E.G. and Ceptureanu, S.I. **Model of organizational change by reengineering in Romanian companies,** IBIMA Conference, 2012

3. Ceptureanu, S.I. **Knowledge management model for Romanian companies,** Review of International Comparative Management, Vol.11, No. 1, 2010
4. Ceptureanu, S.I., Ceptureanu, E.G., Tudorache, A. and Zgubea, F. **Knowledge Based Economy Assessment in Romania,** Economia. Seria Management, 2012
5. Clarke, L. **Managementul schimbarii: ghid practic privind producerea, mentinerea si controlul schimbarilor intr-o organizatie,** Bucharest, Teora Publishing House, 2002
6. Ileanu, B. V., Isaic-Maniu, A. and Herteliu, C. **Intellectual capital components as causes of regional disparities. A case study in Romania,** Romanian Journal of Regional Science, Vol. 3, No. 2, 2009, pp. 39-53
7. Isaic-Maniu, A. an Herteliu, C. **Ethnic and religious groups in Romania-Educational (co) incidences,** Journal for the Study of Religions and Ideologies, No. 12, 2005, pp. 68-75
8. Johns, G. **Comportament organizational,** Bucharest, Economica Publishing House, 1998
9. Kotter, J.P. **Leading Change,** Harvard Business Press, 1996
10. Lester M. **Innovation and Knowledge Management, The Long View,** Creativity and Innovation Management, no. 3, 2001
11. Massa, S. and Testa, S. **Innovation and SMEs: Misaligned perspectives and goals among entrepreneurs, academics, and policy makers,** Technovation, Vol. 28, 2008, pp. 93–407
12. Maurer, R. **Beyond the wall of resistance,** Palgrave Macmillan Publishing House, 2010
13. Nicolescu, O. and Nicolescu, C. **Tranzitia organizationala si rezistenta la schimbari,** Revista Economie teoretica si aplicata, no. 7, 2006
14. Prusak L. and Matson E. **Knowledge Management and Organisational Learning,** Oxford University Press, Oxford, 2007
15. Raducanu, A. M., Feraru, V., Herteliu, C. and Anghelescu, R. **Assessment of The Prevalence of Dental Fear and its Causes Among Children and Adolescents Attending a Department of Paediatric Dentistry in Bucharest,** OHDMBSC, Vol. 8, No. 1, 2009, pp. 42-49
16. Sabau, G. **Schimbarile organizationale produse prin reingineria proceselor economice,** Bucharest, ASE Publishing House, 2000
17. Senge, P. **The Dance of Change: The Challenges to Sustaining Momentum in Learning Organizations,** Crown Business, 1999

[1] **Acknowledgements**
This work was cofinanced from the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/142115 „Performance and excellence in doctoral and postdoctoral research in Romanian economics science domain”