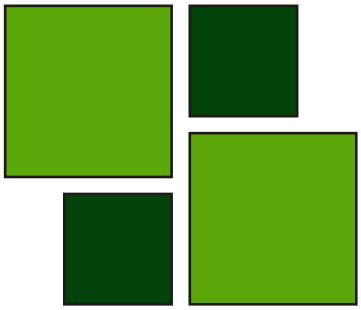

WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

Vol. 10
No. 2
Summer
2015

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

JOURNAL OF APPLIED QUANTITATIVE METHODS

Contents

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

SOCIO-ECONOMIC ANALYSIS OF BURDEN OF PREMATURE DELIVERIES IN ROMANIA[1]

Adrian PANA
MD, PhD Candidate,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: doctorinth@yahoo.com

Adriana CIUVICA
MD, IOMC Alfred Rusescu, Bucharest, Romania

E-mail: adriana.ciuvica@gmail.com

Ionel JIANU
PhD, Assistant Professor,
Department of Accounting and Audit
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: ionel.jianu@cig.ase.ro

Abstract

Premature births represent a major public health problem shared by many stakeholders inside and outside of the health care system. On one side, the direct implied part: child, mother, family, and on the other side, institutions designing and implementing policies towards better health of their populations. The WHO (2014) claims that every year globally, a "estimated number of 15 million of babies are born preterm". In Romania the preterm birth rate was estimated, in 2010, to be 7.3% according to WHO (2010) report. Even if across the world this value may be considered low, in Europe only few countries (Switzerland, Bulgaria and Slovenia) have greater values. In this case, socio-demographic analysis of this problem may offer some hints to different institutions in order to take relevant measures to reduce the impact of this burden.

Key words: premature deliveries, territorial analysis, GIS, quantitative methods

1. Introduction

The preterm births complication represents the most frequent cause of death for the children. Moreover, WHO (2014) report, mention that almost 75% of this kind of deaths can be avoided through cost-effective interventions.

Premature deliveries are associated with several determinants related to mother and the environment she lives. Unfortunately, even with the most recent studies there is no consensus regarding the number and the importance of determinants of premature deliveries. However, psychosocial and or physical stresses, pollution, poor nutrition, drugs consumptions, some infections (Greg, 2007), (Dolatian *et al.*, 2014), education (Lopez and Breart, 2013) etc. are mentioned as direct and causal factors, proven generally by cohort or case-control studies.

Preterm births are associated with different socio-demographics factors like race, education level, urbanization, mother's age, household condition, but also with some intrinsic previous behaviors like smoking, chronic conditions, multiple births, abortions, parity etc. Greg (2007), Florescu *et al.*, (2009), Poalelungi *et al.*, (2013), Dolatian *et al.*, (2014), Lopez and Breart, (2013), Chander *et al.*, (2015).

2. Data and methods

2.1. Data

The first indicator selected in our case is given by the total number of preterm births, defined as gestational period less than 37 weeks, registered in the inpatient care during 2013, and having birthdate registered between 01.01 and 31.12. 2013. The sample of these preterm births was searched within hospitalizations from 2013. The volume of sample represent 15.437 live preterm birth, codified in international classification of diseases, 10th edition (ICD-10) with P07.1x, P07.2x and P07.3x codes, as primary or secondary diagnosis. The sample also retains the locality of the mother (child) at the moment of birth, and the hospital where the newborn was delivered.

Also, a selection of other indicators collected from different sources was used within the statistical analysis. Therefore from Romanian National Institute of Statistics (NIS) were included: (i) population structured by education and occupation (2011, census data); (ii) family physician availability; (iii) live births by age of father and mother (Statistical Research regarding nativity). From the same source (NIS census data) were selected certain data about socio-economic development of communities like: (iv) number of household connected to water pipes or sewerage systems, and (v) share of Roma people within the municipality.

A locality social development index, a measure similar to Human Development Index, developed by Sandu (2010) was also used in the analysis.

2.2. Methods

In order to analyze the impact of several factors considered to be determinants of premature deliveries, two non-linear multiple regression models estimated on an initial sample of 3180 Romanian localities (municipalities) representing the NUTS3/ LAU2 level in the Eurostat classification, NUTS were constructed.

The effect variable used in first model was the number of preterm births in 2013 in each locality.

For the second model we computed the preterm births rate in the locality as the proportion of total number of preterm births on the total number of live births registered in the locality in 2013.

For both models we analyzed multiple possible determinants and or their mix, retaining only the most relevant of them. The general description of them is given below:

- [WAVAGE] - the mean age of females been in the normal age of fertility (15-49 y.o.), assumed as continue variable;
- [RESIDENCE]-a binary variable representing the type of locality with code 1 for urban and 0 for rural;
- [MF] a binary variable having code 1 for localities where a family physician is present;
- [IDSL] a continuous variable representing the level of development of locality as defined by Sandu D (2010);
- [ROMA] –a continuous variable representing the share of Roma people in total population of the locality;
- [SEWAGE*WATER] -a continuous proxy variable for the share of households without sewage nor water system
- [EDU] - women aged over 10y.o. without any formal education as percent of total women aged over 10, in the locality.

3. Results

Some descriptive analysis was performed to facilitate a better understanding of data distribution and behavior. Thus the preterm births recorded are characterized by residence status, and hospital category.

Table 1. Preterm deliveries rates by residence

Residence	Number of live-births	Preterm births	Preterm birth rates (%)
Urban	108,288	7,860	7,26
Rural	89,928	7,577	8,43
Total	**198,216**	**15,437**	**7,78**

In Romania, the obstetrical wards are classified according to their complexity in three categories, 1 being the lowest (usually local city hospitals), and 3 being the highest (referral obstetrical and neonatology centers at regional level), therefore the highest the hospital ranking the more premature births are attended in that facility. The rate of premature deliveries as percentage out of the total deliveries performed in different hospitals according to their complexity level is shown in table 2.

Table 2. Average preterm deliveries rates by obstetrical ward complexity level

Obstetrical ward complexity level	Hospital type	Average premature delivery rate (%)
3	Clinical university hospitals	9.9
2	District hospitals	9.3
1b	Major municipal hospitals	7.4
1a	Minor municipal and local hospitals	5.7

Even if the absolute figures for preterm births in urban and rural are quite similar, the preterm birth rate recorded in rural area is significantly higher. In order to observe if

geographical distribution of preterm births in Romania is uniform across the country we completed the following map:

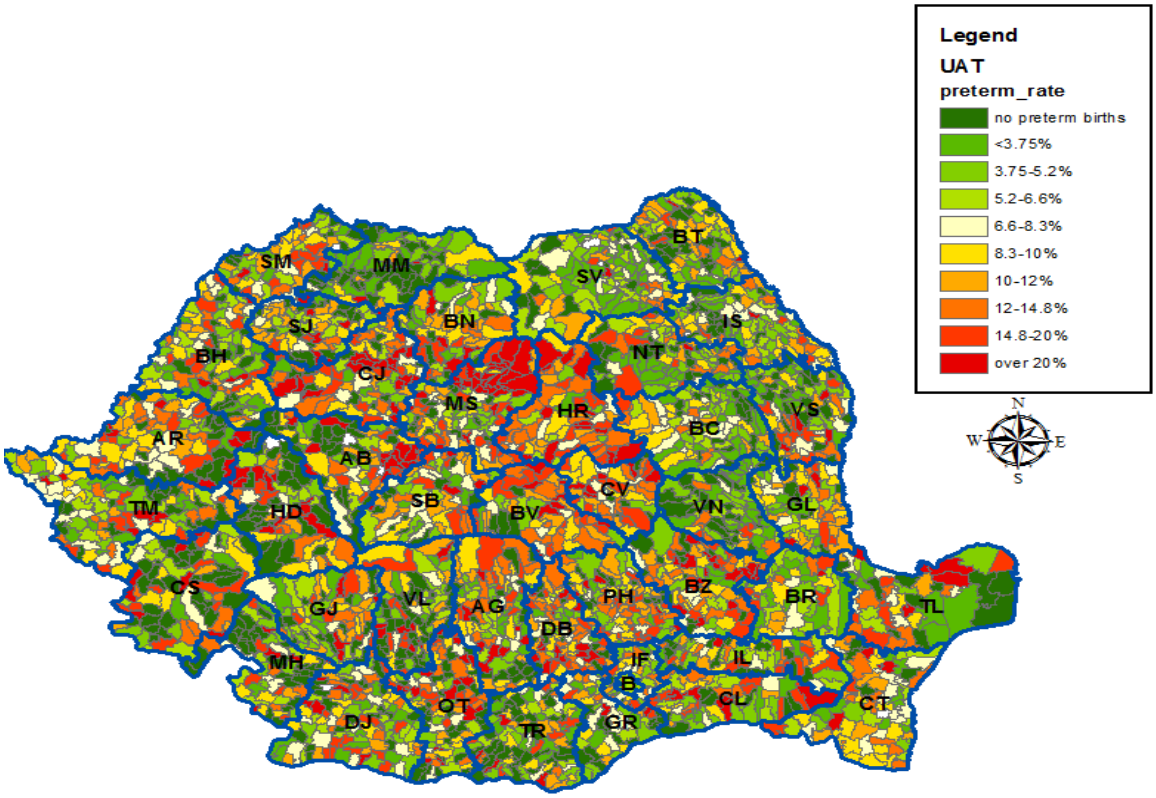

Figure 1. Distribution of preterm delivery rate at NUTS3 level in Romania, 2013, by decile

According to the map, the territorial distribution of preterm deliveries rates is highly heterogeneous. The highest rate appears to be in southeastern Transylvania (a large part of Mures, Harghita and Covasna counties) as well as small parts (non-mountain locations) from Alba and Sibiu counties. On the contrary, Maramures county appear to be the district with the lowest rate of premature deliveries. In the rest of the districts, the results are mixed presenting both very low as well as very high rate of premature deliveries at community level.

In order to analyze the possible link between premature deliveries and a number of socio-economic determinants we employed two models.

The primary model (M1), which has the number of births as count, was treated as Poisson regression. Multiple techniques were tested during estimation process. The best results were achieved when ML estimation was used with Berndt-Hall-Hall-Hausman optimization algorithm and with GLM robust standard errors.

For the second model, due to the fact that some localities have "outliers" regarding the dependent variable or in the covariates values, the classical OLS doesn't perform adequate. As a result we used the Robust Least Square M-estimation with Huber type I SE and covariance.

Both models were run in Eviews 8, and further raw results are listed in the Annex 1 and 2.

Table 3. Analysis of possible factors associated with preterm deliveries in Romania

Model	Factors/Covariates								Adj R^2	Prob(F-stat)/ Prob(LR)
	WAVAGE	RESIDENCE	MF	IDSL	ROMA	SEW*WATER	EDU	C		
M1	0.416***	1.487***	0.709***	0.038***	1.237***	0.251***	-0.028***	-5.34***	0.148	0.00
M2	-0.065***	-0.191***	-0.194***	-0.004***	1.173***	-0.085	-0.021***	-1.34***	0.043	0.00

Note: *** level of significance less than 1%,
** level of significance between 1-5%,
*level of significance between 5-10%

Most of the variables included in the model have statistically significant parameters, showing thus a significant influence on the preterm number of births or rates. When data is somehow normalized (model M2 analyses rates instead of absolute figures) for some dependent variables the coefficients' signs are changed (age, residence, medical doctors' availability, development level, education).

The positive signs of some determinants such as residence and family doctors are not abnormal since the number of preterm births is highly correlated with total number of births (Pearson r=0.98). Measuring as effect the number of preterm births in absolute values they are showing that in better condition we have more births and as a secondary result more preterm births. When rates are computed the signs are generally become as expected.

4. Conclusions

Preterm deliveries in Romania are highly heterogeneous across the country due to several factors such as: social composition (mother's age, mothers' education, share of Roma, access to GP services), as well as community infrastructure (health, sewage, running water). The most relevant factors from those include in the analysis were the one related to infrastructure (access to running water and sewage), followed by the residence status (rural) and percentage of Roma in locality.

One factor that needs to be taken into account as well as a limitation, is the impossibility to verify the accuracy of diagnostic coding knowing that more complex cases are entitled to receive higher budgets. This might explain the unusual high rate of premature delivery rates in some low ranking obstetrical wards, as well as "artificial high" rates of premature deliveries in communities without any unfavorable factors that were included in the analysis.

Lower values for R squared are underlying the presence of some local influences or unmeasured factors (disturbances) that require further investigations.

References

1. Berhman, R.E. and Butler, A.S. (eds.) **Preterm Birth.Causes, Consequences, and Prevention,** Washington (DC): National Academies Press (US); 2007
2. Ceptureanu, S.I. **Knowledge Based Economy in Romania: Comparative Approach,** Journal of Applied Quantitative Methods, Vol. 9, No. 4, 2014, pp. 51-61
3. Chander, P. A., Kacerovsky, M., Zinner, B., Ertl, T., Ceausu, I., Rusnak, I., Shurpyak, S., Sandhu, M., Hobel, C. J., Dumesic, D.A. and Vari, S.G. **Disparities and relative risk ratio of preterm birth in six Central and Eastern European centers,** Croatian Medical Journal, Vol. 56, pp. 119-127, 2015, doi: 10.3325/cmj.2015.56.119
4. Dolatian, M., Mirabzadeh, A., Forouzan, A. S., Sajjadi, H., Majd, H. A., and Moafi, F. **Preterm delivery and psycho-social determinants of health based on World Health**

Organization model in Iran: a narrative review, Global Journal of Health Science,Vol. 5, No. 1, 2012, pp. 52-64

5. Dolatian, M., Mirabzadeh, A., Forouzan, A.S., Sajjadi, H., Alavimajd, H., Mahmoodi, Z. and Moafi, F. **Relationship between Structural and Intermediary Determinants of Health and Preterm Delivery,** Journal of Reproduction & Infertility, Vol. 15, No. 2, 2014, pp. 78-86
6. Florescu, L., Matei, M., Balanica, G. and Azoicai, D. **Cross-sectional study to evaluate the development and risk factors in premature children,** Revista medico-chirurgicala a Societatii de Medici si Naturalisti din Iasi, Vol. 113, No. 4, pp. 1211-1215, 2009
7. Greg, R.A. **Prematurity at Birth: Determinants, Consequences, and Geographic Variation,** in Behrman, R.E. and Butler, A.S. (eds.) "Committee on Understanding, Premature Birth and Assuring Healthy Outcomes, Preterm Birth: Causes, Consequences, and Prevention", The National Academies Press, Washington D.C., 2007
8. Lopez, P.O. and Breart, G. **Sociodemographic characteristics of mother's population and risk of preterm birth in Chile,** Reproductive Health, 2013, pp. 10-26. doi:10.1186/1742-4755-10-26
9. Poalelungi, C., Banateanu, A., Didilescu, A., Hudita, D. and Ceausu, I. **Study of risk factors involved in preterm birth,** Revista Medicala Romana, Vol. LX, No. 3, 2013, pp. 190-193
10. Sandu, D. **Indice al dezvoltarii sociale a localitatilor,** 2010, https://sites.google.com/site/dumitrusandu/bazededate
11. Wooldridge, J.M. **Econometric Analysis of Cross Section and Panel Data, second edition,** MIT Press, 2010
12. * * * **Eurostat NUTS,** http://ec.europa.eu/eurostat/web/nuts/overview
13. * * * **WHO Report Preterm births per 100 births**, http://www.who.int/pmnch/media/news/2012/2010_pretermbirthsper100births.pdf
14. * * * **WHO Report,** 2014, http://www.who.int/mediacentre/factsheets/fs363/en/

Annex 1

Model 1. Determinants of the number of preterm births.
Dependent Variable: NR_PRETERM
Method: ML/QML - Poisson Count (BHHH)
Sample (adjusted): 1 3181
Included observations: 3168 after adjustments
Convergence achieved after 361 iterations
GLM Robust Standard Errors & Covariance
Variance factor estimate = 14.7119128495
Covariance matrix computed using first derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
EDU	-0.028741	0.006255	-4.594942	0.0000
MF	0.709122	0.094345	7.516284	0.0000
ROMA	1.237338	0.213268	5.801812	0.0000
WAVGAGE	0.416986	0.006928	60.18615	0.0000
IDSL	0.038723	0.000539	71.86343	0.0000
RESIDENCE	1.487591	0.034673	42.90305	0.0000
WATER*SEWAGE	0.251872	0.111619	2.256543	0.0240
C	-5.348180	0.129052	-41.44202	0.0000

R-squared	0.150871	Mean dependent var	4.872475
Adjusted R-squared	0.148990	S.D. dependent var	25.46799
S.E. of regression	23.49428	Akaike info criterion	8.854955
Sum squared resid	1744260.	Schwarz criterion	8.870260

Log likelihood	*-14018.25*	*Hannan-Quinn criter.*	*8.860445*
Restr. log likelihood	*-25087.02*	*LR statistic*	*22137.54*
Avg. log likelihood	*-4.424952*	*Prob(LR statistic)*	*0.000000*

Annex 2

Model 2. Determinants of the preterm rate
Dependent Variable: LOG(PRETERM_RATE)
Method: Robust Least Squares
Sample (adjusted): 1 3181
Included observations: 2525 after adjustments
Method: M-estimation
M settings: weight=Bisquare, tuning=4.685, scale=MAD (median centered)
Huber Type I Standard Errors & Covariance

Variable	*Coefficient*	*Std. Error*	*z-Statistic*	*Prob.*
EDU	*-0.021539*	*0.005904*	*-3.648315*	*0.0003*
MF	*-0.194294*	*0.059450*	*-3.268190*	*0.0011*
ROMA	*1.173340*	*0.216816*	*5.411677*	*0.0000*
WAVGAGE	*-0.065214*	*0.019147*	*-3.405910*	*0.0007*
IDSL	*-0.004840*	*0.001424*	*-3.398309*	*0.0007*
RESIDENCE	*-0.190949*	*0.047565*	*-4.014442*	*0.0001*
*WATER*SEWAGE*	*-0.085410*	*0.073669*	*-1.159371*	*0.2463*
C	*-1.343937*	*0.193747*	*-6.936543*	*0.0000*

Robust Statistics

R-squared	*0.030799*	*Adjusted R-squared*	*0.028103*
Rw-squared	*0.043259*	*Adjust Rw-squared*	*0.043259*
Akaike info criterion	*2298.945*	*Schwarz criterion*	*2348.545*
Deviance	*952.6983*	*Scale*	*0.645582*
Rn-squared statistic	*88.58509*	*Prob(Rn-squared stat.)*	*0.000000*

Non-robust Statistics

Mean dependent var	*-2.443715*	*S.D. dependent var*	*0.674859*
S.E. of regression	*0.662779*	*Sum squared resid*	*1105.657*

[1] **Acknowledgements**
This paper was co-financed from the European Social Fund, through the Sectoral 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 authors are grateful to the substantial contribution and guidance provided for this study to the Prof. Gheorghe Peltecu and Lecturer Bogdan Ileanu.

ICT USE BY JUDICIARY SYSTEMS IN EUROPEAN UNION

Carlo CUSATELLI
PhD, Ionian Department, University of Bari, Italy

E-mail: carlo.cusatelli@uniba.it

Massimiliano GIACALONE
School of Economics, Management and Statistics,
University of Bologna, Italy

E-mail: massimilia.giacalone@unibo.it

Abstract
The new technologies have become essential, thanks to the enormous possibilities that they can offer, as the immediate transfer of a document from one continent to another with just a click, saving both time and money and also allowing many simultaneous interpersonal exchanges, that considerably speed up the decision-making processes which involve numerous individuals located in various places. Moreover, we have witnessed over a very short period of time, that most of the human activities which were carried out manually have given space to much more efficient digital implementations. For instance, we can consider the serious problems that the vast documental archives have created in its management, and how centralized computer databases helped to solve most of these problems, speeding up and optimizing all research operations and data mining. This natural easiness of data exchange is still being expanded and facilitated by the development of computer networks, and in particular by the internet.

Keywords: *Information and Communication Technology, ICT, Judiciary systems*

1. Introduction

The need for strong regulatory and valid instruments is clear in contemporary society to resolve disputes in a timely manner and penalize wrongdoing. The judicial function must be measured against this complexity, in order to ensure the safety of citizens, social

cohesion, and economic competitiveness at an international level. The Information and Communication Technology (ICT), that facilitates knowledge and exchange of data and information through an analytical approach to problem solving, has been one of the main levers of change for the judicial authorities. The information systems have entered the offices promising greater operational efficiency, increased transparency, the online exchange of data and documents between offices and litigants. The comparative analysis of experiences in different countries has allowed us to identify both approaches and traps into which some policy makers have fallen (Carnevali, 2010). Understanding the processes of innovation and their peculiarities in highly formalized systems, such as the courts, is therefore a necessary step to be able to support and guide.

ICT in the administration of justice offers possible solutions, improving the administration of justice and helping to streamline procedures and reduce costs. The new concept of electronic justice (e-justice) represents an initial response to the threefold need to improve access to justice, cooperation between legal authorities and the effectiveness of justice. Massive investments in ICT are taking place in all judicial systems to improve the quality of administration of justice and, therefore, better protect the rights and safety of citizens. The ICT projects in the justice sector are quite different and range from the creation of websites for the courts, electronic filing, distance learning, the alternative dispute resolution through web-based technologies. These are projects that are usually accompanied or take place within a broader framework of reforms that affect the entire judicial system, and that often contribute to opening up new possibilities for institutional change, judicial cooperation and even integration between judicial authorities of different countries. It's important to immediately report how the introduction of ICT in judicial systems must necessarily be contextualized within the proper legal and institutional framework that characterizes each country.

2. Introducing technology in Courts

The European Commission, with the Commission Communication COM (2013) 160 final, has developed an evaluation framework of EU justice as a tool to promote effective justice and growth, to ensure a more effective European justice system by identifying reliable and comparable information on the functioning of the judicial systems of the Member States; among the indicators of the Scoreboard, the length of proceedings and the timing of treatment, the turnover rate (defined as the ratio of the completed and introduced judgments) and the number of pending cases, elements have been defined which are absolutely relevant to be considered and optimized in order to increase the quality of the judicial system. Indicators show that the availability of ICT systems for recording and case management and for communication and exchange of information between courts and their context is a determining factor for the effectiveness of justice, for example, electronic forms available on the internet, websites of the courts, follow-ups of suits online, electronic records, electronic processing of small claims and the recovery of uncontested claims, electronic submission of applications and videoconferencing.

The strong drive for innovation of the judicial system is based on three key moments: the introduction of technology, which is able to streamline all the procedures for which discretion of the court is required; a legal system that is capable of facilitating the availability of regulatory instruments which are flexible and capable of adapting to

continuous stress; the internal structure of an organization is able to exploit its possibilities. This has been pursued with approaches related to the idea of cultivation, a method of development of complex systems based on local planning interventions aimed at conversion, adaptation and connection of systems, components and functionality already available, in part, for the purpose of assembling configuration systems that can then be put to the next test. This has allowed, for example, to achieve, in a relatively simple way, applications that would facilitate access to information on the procedures provided by the courts to involved parties or to exchange data necessary for the definition of cases (April, 2011).

Electronic justice initiative at Community level is carried out by the Council of the Union (Justice and Home Affairs) and the European Commission (Justice, Freedom and Security General Directorate) under which a specific Action Plan is defined; such plan includes, among other things, the creation of the "European e-Justice Portal" which supplements the initiatives of computerization of individual Member States and access point privileged access to information, applications and case law on the part of citizens, businesses, professionals and judicial authorities: e-justice can be defined as the use of ICT to improve citizens' access to justice and the effectiveness of the judicial action seen as any kind of activity to resolve a dispute or punishment of criminal behavior. The Commission has always encouraged the use of videoconferencing and the electronic transmission of documents between judicial authorities and actively participated in the project of interconnection of criminal records. The potential scope of e-justice is very broad and likely to evolve in the light of progress within the European Judicial Area and technological developments.

There is, however, another e-justice that, applying the technologies of automatic extraction of information to the court orders, can not only enrich the decision-making process of the actors of the proceedings but may also provide new and interesting perspectives for the management of the courts. Besides the specific scientific techniques related to information extraction, data and text mining, judicial decisions that are so elaborate allow performing synchronic, diachronic and comparative analysis on the functioning of the judicial administration. It is an innovative and original result that clears the way for justice monitoring, without which it is almost unthinkable to identify concrete and effective solutions that improve performances. Online, it is possible to find information on legal systems, legislation and case law; electronic communication systems are developing between the parties and the courts, and in some cases entirely electronic procedures are available (Fabri, 2006). The use of electronic means to record hearings is increasing as well. At a European level, several professional organizations are developing particularly interesting projects for the exchange of information or interconnection, for example, the website of the Association of the Councils of State, the common portal on the jurisprudence of the Supreme Courts or the European register of wills. Several projects in the field of e-justice are currently being developed. In addition to the examples mentioned above, it is appropriate to recall all the projects relating to legal documentation undertaken by the European Union or from institutional and private operators (Seibert-Fohr, 2012).

The Commission supports these projects, but also considers as important factors increasing readability, accessibility and efficiency of EU action in the judiciary field, and to emphasize projects that will truly add value to the European judicial area. Indeed, while the law in the legal field has developed considerably, its impact often remains limited due to the difficulties of transposition (especially in criminal matters) and the operators who often lack

in knowledge.

3. ICT usage in European Courts

The European e-justice wants to target the use and development of ICT at the service of the judicial systems of the Member States, in particular in cross-border situations, so as to enable citizens, businesses and legal practitioners with a greater access to justice and judicial information and to facilitate cooperation between judicial authorities of the Member States. It aims at improving the effectiveness of justice itself, while respecting the independence and diversity of the legal systems of the Member States, as well as of fundamental rights. It is appropriate to ensure that users of the European e-justice system, including citizens, can take advantage quickly of concrete electronic tools.

In Table 1, three ICT partial composite unweighted indicators

$$_{ITC}P_i = \sum_{j=1}^{w_i} r_j / w_i \quad (i = 1, 2, 3;\ w_1=5, w_2=4, w_3=9) \tag{3.1}$$

have been calculated by averaging the usage rates r_j of the w_i computer facilities within each of the three uses of technologies in courts, on the basis of data collected through multiple questions no. 62 (direct assistance to the judges and court clerks), no. 63 (administration and management) and no. 64 (electronic communication and exchange of information) of the ultimate European Commission for the Efficiency of Justice evaluation scheme (CEPEJ, 2014).

The overall synthetic indicator, weighted by respective number of facilities,

$$_{ITC}S = \sum_{i=1}^{3} w_i P_i / 18 \tag{3.2}$$

summarizes the previous partial ones: it shows almost a three-quarter judicial ICT completeness in Europe; Romania, Spain, Germany (representing the median unit with exactly 75% of completeness if UK is considered in its unity) and Netherlands (representing the median unit with precisely the same average 76.7% of completeness considering UK divided in three parts) range around that level, but Greece (the last, with 18.6%), Belgium and Cyprus lie below fifty percent of such computerization, while Austria, Estonia, Malta and Portugal have already completed it.

The results obtained, the limitations encountered and the targets set for the future require a comprehensive European strategy on e-justice to bring the commitment and involvement at a strategic level. The new European e-justice 2014-2018 strategy intends to shift from work already undertaken so that there is a greater use of electronic applications, electronic transmission of documents, video conferencing and the interconnection of registers and administrative records, in order to further reduce the costs of litigation out of court by establishing a mechanism to ensure that future legislation is designed to be used by means of online applications.

Several Member States have already developed and participated in a series of pilot projects in the field of e-justice and an infrastructure for European e-justice is gradually developing: since 2011, a number of Ministries of Justice and central authorities of several European countries have embarked on such a large scale project (http://www.e-codex.eu) in which they have concretely experienced some applications in the field of cross-border proceedings, starting with the European payment order, established by specific Community regulations.

Table 1. Percentage indicators estimated on the basis of 2012 data collected through 5th CEPEJ report

Use of technologies in courts (computer facilities)	Austria	Belgium	Bulgaria	Cyprus	Czech Rep.	Denmark	Estonia	Finland	France	Germany
Direct assistance to the judges and court clerks (word processing, electronic database of case-law, electronic files, email, Internet connection)	100.0	100.0	100.0	80.0	86.0	100.0	100.0	100.0	95.0	81.0
Administration and management (case registration system, court management information system, financial information system, videoconferencing)	100.0	46.3	76.3	50.0	57.5	100.0	100.0	100.0	100.0	76.3
Electronic communication and exchange of information between the courts and their environment (electronic web forms, websites, follow-up of cases online, electronic registers, electronic processing of small claims, electronic processing of undisputed debt recovery, electronic submission of claims, videoconferencing, other)	100.0	10.0	33.9	22.2	92.2	33.3	100.0	88.9	40.0	71.1
Overall ICT (average of the previous three composite indicators weighted by number of computer facilities)	100.0	43.1	61.7	44.4	82.8	66.7	100.0	94.4	68.6	75.0

Use of technologies in courts (computer facilities)	Greece	Hungary	Ireland	Italy	Latvia	Lithuania	Luxembourg	Malta	Netherlands	Poland
Direct assistance to the judges and court clerks (word processing, electronic database of case-law, electronic files, email, Internet connection)	30.0	86.0	100.0	100.0	100.0	100.0	100.0	100.0	81.0	81.0
Administration and management (case registration system, court management information system, financial information system, videoconferencing)	17.5	75.0	58.8	87.5	76.3	87.5	100.0	100.0	93.8	76.3
Electronic communication and exchange of information between the courts and their environment (electronic web forms, websites, follow-up of cases online, electronic registers, electronic processing of small claims, electronic processing of undisputed debt recovery, electronic submission of claims, videoconferencing, other)	12.8	61.1	52.8	77.8	100.0	97.2	44.4	100.0	66.7	39.4
Overall ICT (average of the previous three composite indicators weighted by number of computer facilities)	18.6	71.1	67.2	86.1	94.7	95.8	72.2	100.0	76.7	59.2

Use of technologies in courts (computer facilities)	Portugal	Romania	Slovakia	Slovenia	Spain	Sweden	England and Wales	North. Ireland	Scotland	Average
Direct assistance to the judges and court clerks (word processing, electronic database of case-law, electronic files, email, Internet connection)	100.0	76.0	81.0	95.0	100.0	100.0	100.0	100.0	100.0	*92.1*
Administration and management (case registration system, court management information system, financial information system, videoconferencing)	100.0	87.5	76.3	100.0	93.8	100.0	93.8	93.8	100.0	*83.6*
Electronic communication and exchange of information between the courts and their environment (electronic web forms, websites, follow-up of cases online, electronic registers, electronic processing of small claims, electronic processing of undisputed debt recovery, electronic submission of claims, videoconferencing, other)	100.0	67.8	46.1	75.6	51.7	88.9	72.8	71.1	66.7	*65.0*
Overall ICT (average of the previous three composite indicators weighted by number of computer facilities)	100.0	74.4	62.5	86.4	74.4	94.4	85.0	84.2	83.3	*76.7*

4. Conclusions

The ICT indicators showed that when the process of computerization will be completed, which seems to be quite realistic even in the short term, the next step should include a coordination to ensure, while still respecting the local autonomy and cognitive needs peculiar to different contexts, a common information base homogeneous and shared on a European level.

The problems of data reliability, the provision of appropriate classifications in survey forms and, more generally, the quality of data are attributable, directly or indirectly, to the degree of computerization in statistical-judicial production. In fact, in the presence of a fully computerized detection system (also at the level of case management records) the possibility of transcription errors, manipulation and interpretation of the information required will drastically reduce (due to the non-perfect correspondence between the classification adopted in models of detection and what is recovered in the official records), as well as the time-lag in some cases considerable, between data recording and the actual time/instant of reference; on the other hand, the detailing of the information collected could increase a result of a greater and more appropriate articulation of the detection patterns (certainly not feasible, beyond a given limit, in cases of manual detection) and the activation of an automatic check on the consistency of the data would be possible, not only ex-post, but during the same stage in which information is entered.

The last aspect concerning the quality of production processes regards their duration, which is also highly dependent on the degree of computerization of various systems. In fact, with the completion of the automation, the time of data acquisition will be strongly contracted and the frequency of detections will also increase.

References

1. Aprile, S. **Rapporto ICT Giustizia. Gestione dall'aprile 2009 al novembre 2011,** Direzione Generale Sistemi Informativi Automatizzati, Roma, 2011
2. Carnevali, D. **Soggetti Smarriti. Perché innovazione e giustizia non si incontrano quasi mai,** Franco Angeli, Milano, 2010
3. Fabri, M. **Amministrare la giustizia. Governance, organizzazione, sistemi informativi,** Clueb, Bologna, 2006
4. Seibert-Fohr, A. **Judicial Independence in Transition,** Springer, Heidelberg, 2012
5. * * * **Evaluation report on European judicial systems - 2012 data,** CEPEJ (European Commission for the Efficiency of Justice), Council of Europe Publishing, Strasbourg, 2014

MEASURING SOCIAL RESPONSIBLE BANKS' EFFICIENCY AND PRODUCTIVITY – A NONPARAMETRIC APPROACH[1]

Iustina Alina BOITAN
PhD, Assistant Professor, Faculty of Finance, Insurance, Banking and Stock Exchange, Bucharest University of Economic Studies, Bucharest, Romania

E-mail: iustina.boitan@fin.ase.ro

Abstract

The paper investigates a sample of commercial banks that voluntarily and publicly committed to becoming promoters of sustainability in the financial industry, by applying in their regular banking activity a set of principles related to environmental and societal responsibility.

To assess the individual financial profile of these banks and the similarities in terms of business behavior it has been performed several steps. First, it has been computed the descriptive statistics of key financial indicators, namely the market share, the liquidity position, the financial structure, the operational efficiency, profitability, capital adequacy and the individual contribution of each bank to the domestic financial depth. Secondly, to proxy the performance in achieving their objectives, it has been employed the Data Envelopment Analysis nonparametric technique to estimate the relative efficiency scores. Two models' configurations have been tested, in the assumption of a financial intermediation approach and respectively profit efficiency approach. Then, to gain a comprehensive and dynamic picture, it has been computed the Malmquist productivity index in order to illustrate not only socially responsible banks' productivity's changes over time but also the main sources of changes, in terms of catch-up effect and technological progress.

Keywords: *sustainable bank, social and environmental responsibility, Data Envelopment Analysis, efficiency frontier, Malmquist index*

1. Introduction

Financial industry representatives, practitioners and academia agree that, at present, we are all witnessing momentum in sustainable banking. Mainstream banking business behavior is recording a continuous shift and re-assessment toward increased social and environmental awareness and responsibility. This trend overlaps on the sustainable and inclusive economic growth actively promoted by the European Commission and the process of countries' rethinking of regional development (Davidescu, Strat 2014).

At the origin sustainable banks are traditional, commercial banks that have started to make steps toward implementing several sustainability or social responsibility criteria in their business model, especially in terms of lending decision making. Their common

denominator is the public communication related to their new focus on relationship banking, in which customer centricity is a key attitude.

Sustainable banks' top management has become aware of the direct and tight interdependencies between financial intermediaries' long term business ongoing and viability and its contribution towards economic development, healthier environment and societal wellbeing. Consequently, their stated mission is twofold: maintaining banks financial position, soundness and stability, and at the same time meeting customers' needs and environmental objectives, by adapting banking products and services offering to responsible criteria. It is acknowledged the special role to play both in the global financial industry field, as well as in local, domestic economy and community.

The present paper extends the qualitative research performed in a previous one (Boitan, 2014), that aimed at investigating through an analytical analysis the main international sustainability frameworks and principles that worldwide financial institutions could join, on a voluntary basis. The aim of that paper had been to assess, through a country-by-country survey, where contemporaneous European Union's banking industry stands within the broad sustainability stream and which is the most widespread sustainable standard across the European Union's banks.

The paper found out that 16 countries out of the 28 EU member ones comprise at least one conventional bank that voluntarily adhered to different widespread sustainability frameworks. By restricting the analysis to banks that joined several sustainability frameworks, it has been uncovered a sample of 13 European banks that committed to align their activity at the same three sustainability standards, simultaneously (Equator Principles, the United Nations Global Compact and the United Nations Environment Program Financial Initiative).

The aim of the present paper is to perform an in-depth, bank-level empirical analysis so as to investigate whether these particular 13 sustainable banks depict common features of their business models, as a result of adhering to the same guiding frameworks or on the contrary have strategy specificities.

The paper is structured as follows: part one summarizes the features of the 13 sustainable banks, in terms of their individual financial profile. Part two describes the research hypotheses and employs the Data Envelopment Analysis non-parametric approach to assess the individual degree of efficiency in fulfilling the fundamental financial intermediation role. The third part computes the Malmquist productivity index for each sustainable bank and decomposes it in two components. Last part concludes.

2. Overview of sustainable banks' individual financial profile

The three international sustainability frameworks directly related to financial industry are represented by the United Nations Environment Program – Financial Initiative (UNEP FI) whose aim is to increase banks' understanding and monitoring of environmental issues, as well as to estimate the exposure to environmental risks, the United Nations Global Compact which requires its signatory members to adopt a set of core principles in the areas of human rights, labor standards, the environment and anti-corruption and the Equator Principles, which are applied for investment projects that exceed the value of USD 10 million, are relate to identifying, assessing and managing environmental and social risks.

Their common denominator is the fundamental goal of encouraging financial industry to implement socially responsible behavior, as well as the requirement for greater

transparency and regular reporting related to the progress achieved in implementing these principles. In addition, the frameworks have a complementing nature. Consequently, banks that decide to sign the statements of several sustainability frameworks have to mandatory commit at fulfilling them and aim at transmitting a message of increased responsibility and transparency.

By analyzing the signatory financial institutions of these most widely adopted international sustainability frameworks, it have been identified 13 European commercial banks that committed, on a voluntary basis, to implement in their current activity the standards and principles of increased social and environmental awareness. Figure 1 illustrates the geographical spread of these sustainable banks across EU countries. Banking systems in UK and Netherlands comprise each a number of three sustainable banks, France, Italy and Sweden have 2 sustainable banks while Portugal only one.

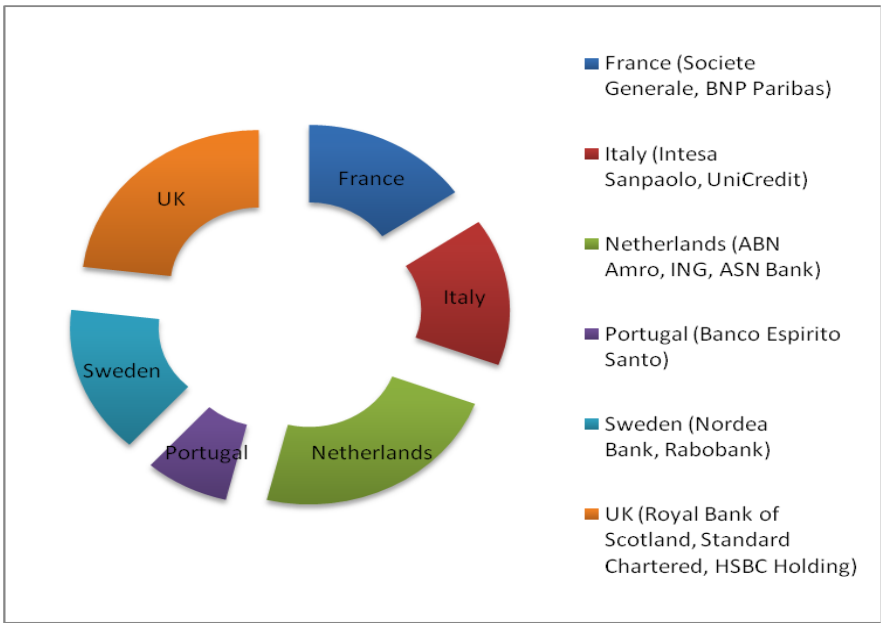

Figure 1. Sustainable banks' country of residence
Source: Boitan (2014)

To gain a comprehensive insight into the financial profile of each bank above mentioned, it have been computed several financial ratios, namely: the market share, the liquidity position, the financial structure, the operational efficiency, profitability, capital adequacy and the individual contribution to the domestic financial depth. The source of data are banks' annual financial statements at end 2013, European Central Bank's Statistical Data Warehouse (the aggregated balance sheet of euro area monetary financial institutions, at December 2013) and Eurostat statistics database (GDP at market prices in millions euro, at end 2013).

The market share hold by a sustainable bank in the banking system of the origin country is represented by the ratio of bank's total assets in domestic banking system's total assets. The liquidity indicator has been proxy by the ratio of loans to customers in total deposits attracted from customers. The financial structure has been assessed by means of two ratios: customer loans to total assets and the share of customers' deposits in total liabilities. Operational efficiency has been computed as a cost to income ratio, capital adequacy is represented by the tier 1 ratio while profitability by two indicators: return on

equity (ROE) and return on assets (ROA). The ratio of a bank's total assets in the GDP of the resident country was computed to measure sustainable banks' individual contribution to the domestic financial depth.

The primary descriptive statistics have been illustrated in the table below.

Table 1. Descriptive statistics

	ROA	ROE	Tier 1	Loans/ Deposits	Loans/ Assets	Deposits/ Assets	Cost/ Income	Market Share	Assets/ GDP
Mean	0.57	5.73	13.28	109.96	53.55	49.77	64.35	18.78	76.22
Maximum	4.1	16.7	21.7	170.59	95.78	94.13	99.04	46.13	179.311
Minimum	0	0	9.4	72.85	27	27.903	49.21	0.46	1.78
Standard deviation	1.09	5.12	3.35	26.00	17.80	16.90	15.47	11.86	49.44
Skewness	2.90	0.52	1.18	0.71	0.81	1.19	1.10	0.71	0.67
Kurtosis	9.99	2.61	4.01	3.43	3.60	4.81	3.12	3.46	2.78

Source: computations using the Eviews software, Boitan (2014)

Standard deviation provides important clues on data features. According to economic theory, it measures the dispersion of a variable's values around its mean. Large values of the standard deviation suggest that data is spread out over a large span of values or the presence of extreme, outlier values. Standard deviation recorded its highest values in the case of total assets to GDP (49.44), followed by loans to deposits ratio (26). Consequently, during 2013 these variables have fluctuated most across the sample of 13 sustainable banks. The four variables that proxy the financial structure, the market share and the operational efficiency show moderate fluctuation between the minimum and maximum values. The smallest variation across banks has been recorded by ROA, with 1.09 followed by Tier 1 with 3.35 and ROE with 5.12. ROA, tier 1 and deposits to assets are leptokurtotic, meaning that the likelihood of an extreme value to occur is higher than if the variable would have had a normal distribution, while all remaining variables depict a kurtosis around 3. The three previously mentioned variables depict an asymmetric distribution with positive skewness, while all other variables record skewness around zero, depicting relatively symmetric distribution.

In respect of the variables that depict sustainable banks' positioning in the domestic banking system and their individual contribution to country's financial depth, raw data show that, irrespective the indicator chosen (market share or total assets in domestic GDP ratio) banks' hierarchy is the same. The highest values are recorded by ING in Netherlands, followed by Nordea Bank and Rabobank, both located in Sweden.

The variables depicting banks' micro features in terms of business strategy and risk profile are relatively heterogeneous. The liquidity indicator and the operational efficiency fluctuate most across the 13 banks, for both of them the minimum values being recorded by HSBC Holding from UK. It is a favorable financial position, as the liquidity indicator is situated at a moderate, comfortable level, the bank not being exposed to liquidity constraints on short to medium term, and the operational costs do not erode the operational income (costs represent almost 50% of operational revenues).

The two complementing indicators related to banks' balance sheet financial structure (loans to assets and deposits to liabilities ratios) record similar values across the sample of banks considered. The lowest levels for both variables (27%, respectively 27.903%) are depicted by Societe Generale in France, while the highest (95.78%, respectively 94.13%) by ASN Bank in Netherlands. This striking difference is a direct consequence of the business model implemented by each of the two banks: the one in Netherlands has as main source of funds the deposits collected from customers and as main use of these funds the loans granted. The one in France has diversified its sources of financing as well as their uses, a fact that explains the low reliance on the two basic banking products (Boitan, 2014).

3. Nonparametric estimation of efficiency

As the sustainable banks in the sample have committed to be more aware and sensitive to local community needs and environmental challenges, the first research direction is to evaluate the manner in which the fundamental financial intermediation role is fulfilled, by assessing their individual degree of efficiency.

It has been employed a nonparametric Data Envelopment Analysis (DEA), to calculate relative efficiency. At the core of this method is the solving of a linear programming problem. The outcome is the computation of efficiency scores, to empirically assess the performance of a given financial unit, and the construction of a best practice or efficiency frontier. In respect of the analytical form of the production function, one of the advantages of DEA is that it does not require an a priori hypothesis; it simply determines the production function by applying minimization or maximization techniques on the available data (Scippacercola, Sepe 2014).

According to the broad literature devoted to DEA, it is argued that the state of inefficiency is caused by the management, but it can be controlled and corrected. Consequently, DEA methodology appears to be related on the microeconomic concept of efficiency and the microeconomic view of production functions (Ferreira 2011, p.7).

The efficiency frontier is composed by all the best-practice input and output combinations. All banks that reached a score of 1 lie on the frontier and are called efficient meanwhile the others are perceived as inefficient. The amount of inefficiency for each bank can be computed by simply subtracting the score obtained from 1.

A peculiarity of DEA is the relative efficiency scores it provides, which means that a bank is qualified as efficient only in relation to those already included in the sample. In other words, it is possible that this fully efficient bank become inefficient when expanding the initial sample. Thus, the results obtained cannot be extrapolated to wider samples. However, in the viewpoint of Repkova (2014), this DEA feature might prove a useful decision-making tool for benchmarking different entities included in a sample.

Our study is not affected by this drawback, as the sample comprises all the sustainable banks. The study relies on an output-oriented model, thus the DEA model's mathematical configuration is :

$$max\ \theta = \alpha + s + e \quad (1)$$

$$\sum_k \mu_k\ y_{ik} = \alpha y_{i0} + s_i\,, \quad i = 1, 2, \ldots . I \quad (2)$$

$$\sum_k \beta_k\ x_{ik} = x_{j0} - e_j\,, \quad j = 1, 2, \ldots . J \quad (3)$$

$$s_i \geq 0, \quad i = 1, 2, \ldots . I \quad (4)$$

$$e_j \geq 0,\ j = 1, 2, \ldots . J \quad (5)$$

$$\beta_k,\ \mu_k \geq 0,\ \ k = 1, 2, \ldots . . n \quad (6)$$

where:

θ = the relative efficiency score of each bank in the sample

n = the number of banks in the sample

I = the number of output variables

J = the number of input variables

μ = the weight of each output variable, belonging to each bank

β = the weight of each input variable, belonging to each bank

y = vector of output variables

x = vector of input variables

α = parameter that shows the amount by which the vector of output variables increases, in the hypothesys of relatively constant inputs

s = parameter depicting deficiences in achieving the output *i*

e = parameter depicting the excessive use of input *j*

It will be tested two DEA output-oriented models, in the assumption of an intermediation approach, respectively of a profit approach. In the intermediation approach banks act as mediators between the demand and the offer of money, between savers and investors.

Due to sample-size constraints, the models to be tested are single input-single output. The selection of variables is grounded on the theory of bank behavior, in terms of a producer of banking products and services or as an intermediary. In respect of the profit efficiency approach, Morita and Avkiran (2009) argue that expert knowledge or generally accepted practices can be useful in selecting the suitable variables.

In the following it has been tested two DEA models, named M1 and M2. Model M1 reflects the intermediation approach and consists of total deposits collected from customers as input variable and total loans provided to customers as output variable. Model M2 corresponds to the profit efficiency approach and comprises interest expenses as input variable and interest income as output variable. Both models have been estimated under the output orientation, in order to highlight banks' potential for increasing or optimizing the amount of outputs given the same level of inputs.

The distance between each bank and the production frontier has been computed through the radial distance or Debreu-Farrell-measure, as its interpretation is intuitive, by depicting the necessary improvements when all relevant variables are improved by the same factor equiproportionally.

For the empirical study it has been employed bank-level data on 13 sustainable banks across Europe, the analysis being run for each year in the time span 2007 – 2013.

Table 1 synthesizes the main conclusions obtained by running the DEA method. M1 represents the first research hypothesis (an output oriented model, financial intermediation approach) and M2 is the output oriented model, in the profit efficiency approach.

Table 1. Summary of results obtained

Year	DEA model	Number of efficient banks	Average efficiency score	Standard deviation of efficiency scores
2007	M1	1	184.26%	1.18
	M2	1	143.07%	0.29
2008	M1	1	200.03%	1.63
	M2	1	138.70%	0.26

Year	DEA model	Number of efficient banks	Average efficiency score	Standard deviation of efficiency scores
2009	M1	1	216.14%	1.75
	M2	1	169.63%	0.44
2010	M1	1	201.69%	2.04
	M2	1	163.97%	0.47
2011	M1	1	216.54%	2.02
	M2	1	166.11%	0.49
2012	M1	1	205.10%	1.87
	M2	1	174.54%	0.53
2013	M1	1	169.94%	0.44
	M2	1	182.13%	0.58

Source: author's computation by means of EMS software

By running both models, in each year during 2007 – 2013 periods, the findings show that a single socially responsible bank had achieved the status of full efficiency. Hence, the production frontier is different for each year considered. On average, the highest inefficiency has been observed for the intermediation approach (model M1), where the presence of outlier scores has been more pronounced. The average inefficiency increased gradually from 2007 to 2009, a time span that overlaps on the financial crisis onset, recorded a small decrease in 2010, then in 2011 arrived at the same level as in 2009. Further it entered on a decreasing path, the lowest inefficiency being recorded in 2013.

Model M2 that illustrates the profit efficiency approach, in other words the manner in which banks succeeded to manage interest expenses so as not to erode interest income and achieve a satisfactory net interest margin, had been relatively more stable than the preceding one. The values of the standard deviation indicate that the efficiency scores estimated are closer to the sample's efficiency mean. The lowest average inefficiency has been recorded in 2007 and 2008, while the highest values belong to 2012 and 2013 years.

The figures 2 and 3 provide a disaggregated picture of the efficiency scores' evolution in each year considered and for each socially responsible bank.

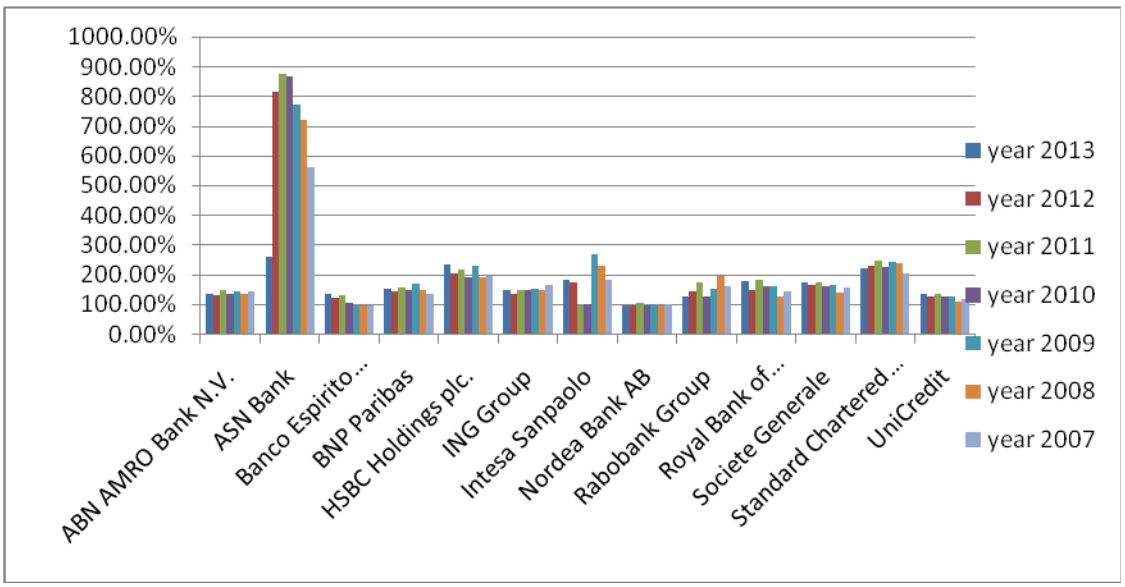

Figure 2. Efficiency trends under model M1

Source: author

A noticeable trend has been recorded by ASN Bank from Netherlands, with huge inefficiency during 2007 – 2012 (varying from a score of 564% to 873%). At end-2013 the inefficiency decreased sharply, to a level of 265%. The explanation of bank's inability in fulfilling its intermediation function lays in its balance sheet peculiarities. The bank depicts a prudent, risk-averse lending strategy, as the deposits collected from customers exceed the amount of loans provided.

HSBC Holdings and Standard Chartered Bank, both from UK, had a relative steady state of inefficiency, of around 200% during the entire period. Most remaining banks recorded lower inefficiency levels that gravitate more closely to the 100% benchmark.

Each year, the efficiency frontier is composed by one bank. In 2007 and 2009, Banco Espirito Santo from Portugal recorded 100% efficiency; in 2011 Intesa Sanpaolo Bank from Italy positioned itself on the frontier. The most efficient bank, from the viewpoint of the financial intermediation role accomplished, is Nordea from Sweden as it positioned most on the best practices frontier (in 2008, 2010, 2012 and 2013). Also, in the other three years considered, it recorded very low, almost negligible levels of inefficiency.

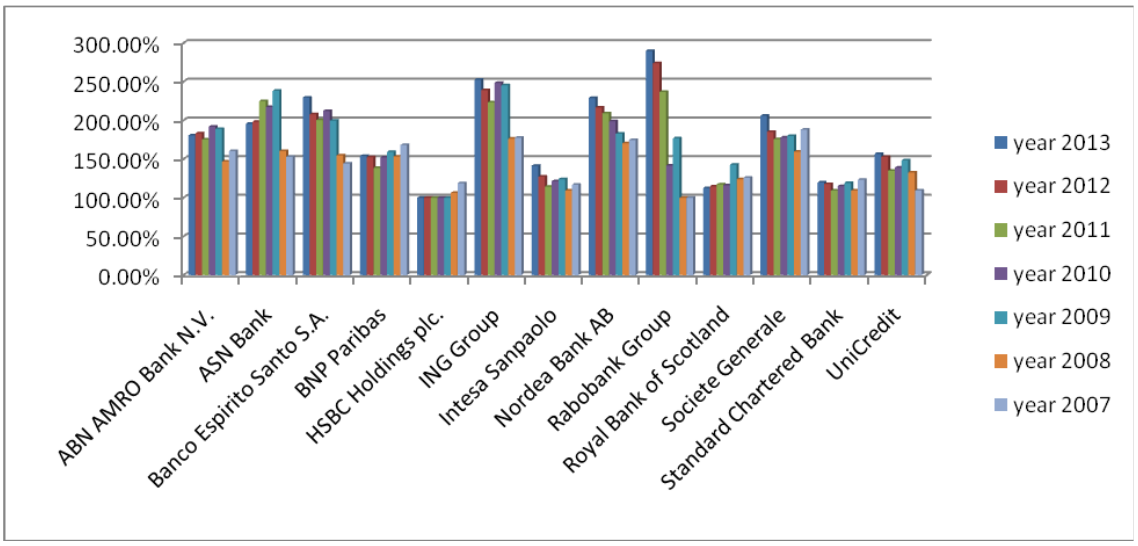

Figure 3. Efficiency trends under model M2
Source: author

From a profit efficiency standpoint, the state of inefficiency is relatively balanced across most banks. Three banks (Intesa Sanpaolo, Royal Bank of Scotland, Standard Chartered Bank) recorded scores in the vicinity of the 100% threshold. The efficiency frontier comprises Rabobank Group from Sweden in 2007 and 2008, while the entire period between 2009 and 2013 is dominated by HSBC Holdings in UK. By looking at its income statements, interest expenses amount to less than half of interest revenues.

4. Productivity assessment

The basic indicators for measuring productivity are usually represented by output/input ratios. However, these indicators provide a static snapshot, at a given moment in time, without providing clues on the leading factors that trigger changes in productivity level. Malmquist productivity index represent a reliable alternative to the traditional

approach, as it depicts not only the productivity's changes over time but also the sources of changes.

An extension of DEA method is to compute Malmquist indices based on bank panel data, to estimate total factor productivity and to decompose it in two components: technical efficiency change (catch up effect) and technological progress, for each bank in the sample. The first step in computing a Malmquist index is the estimation of distance functions, by means of DEA, under different time periods technologies.

In this paper it has been employed the formula proposed by Caves, Christensen, Divert (1982) because it is suited for the output oriented models (one of the assumptions of this study).

$$M_{t,t+1}(y^t, y^{t+1}, x^t, x^{t+1}) = \left[\frac{D^t(y^{t+1}, x^{t+1})}{D^t(y^t, x^t)} \times \frac{D^{t+1}(y^{t+1}, x^{t+1})}{D^{t+1}(y^t, x^t)} \right]^{1/2} \quad (7)$$

where:

$M_{t,t+1}$ = Malmquist productivity index during t and t+1period

y^t, y^{t+1} = output vectors at t, respectively t+1 period

x^t, x^{t+1} = input vectors at t, respectively t+1 period

D^t, D^{t+1} = distance function based on period t, respectively t+1 technology

The first ratio is computed for moment t and measures the productivity change during the t, t+1 period, having as reference point the technology in period t. The second ratio reflects the productivity at t+1 moment, having as benchmark the technology of t+1 period. Values that exceed the threshold 1 indicate productivity growth, meanwhile a value lower than 1 depicts productivity regress. When the Malmquist index equals 1, it is assumed that there is no change in the productivity level.

According to Fare et al. (1989), to better outline the two components of the Malmquist index, the basic formula could be rewritten as follows:

$$M_{t,t+1}(y^t, y^{t+1}, x^t, x^{t+1}) = \frac{D^{t+1}(y^{t+1}, x^{t+1})}{D^t(y^t, x^t)} \times \left[\frac{D^t(y^{t+1}, x^{t+1})}{D^{t+1}(y^{t+1}, x^{t+1})} \times \frac{D^t(y^t, x^t)}{D^{t+1}(y^t, x^t)} \right]^{1/2} \quad (8)$$

where:

$\frac{D^{t+1}(y^{t+1}, x^{t+1})}{D^t(y^t, x^t)}$ represents the technical efficiency change or the catch-up effect. It indicates the extent to which the estimated distance might vary between the observed production frontier and the maximum potential production frontier during the period t, t+1.

$\left[\frac{D^t(y^{t+1}, x^{t+1})}{D^{t+1}(y^{t+1}, x^{t+1})} \times \frac{D^t(y^t, x^t)}{D^{t+1}(y^t, x^t)} \right]^{1/2}$ represents the technological change and indicates the amplitude of the production frontier's shift, as a consequence of technology developments.

Technological efficiency, which is deemed to cause frontier shifts, might be attributed to several developments of banking activity, directly or indirectly driven by information technology, such as the diversification of financial services supplied to customers (electronic payments, internet banking, self banking, e-banking, mobile banking), the improvement of back-office activity, by using economic and statistical models to evaluate the credit, liquidity, market and operational risks and sophisticated, rigorous scoring techniques and discriminant analyses to decide whether to finance or not the credit demands (Dardac, Boitan 2008).

According to Berger (2003), the economic effects of employing financial and software technologies on banking system's productivity consist of improving the quality of banking products and services and extending their range, increasing the processing speed of banking regular operations, and last but not least enhancing the degree of satisfaction felt by clients. On the other hand, costs incurred by the acquisition and use of software techniques, as well as those required by employees' training are a debated issue. Some banks might choose to bear them entirely, while others might be tempted to transfer a fraction of them to customers. However, it is difficult to obtain an accurate, quantitative measure of technological efficiency.

Ho and Mallick (2006) have identified two ways information technology might improve bank performance, namely by diminishing operational cost and by facilitating transactions among customers sharing the same network. By reviewing and analyzing banking developments in the last 25 years, Frame and White (2009) concluded that it have occurred substantial changes in terms of new products or services and new production processes, due to financial innovations driven by technological change (e.g. subprime mortgage loans, online banking, asset securitization, credit scoring, bank risk management through value-at-risk and stress-testing tools). Saeed and Bampton (2013) argue that in developed countries, information and communication technology acts as an engine for challenges in modern banking, its effects being related to lowering costs, providing efficient banking services to customers and enhancing profits.

Figure 4 synthesizes Malmquist productivity evolution over the five years considered, for each bank in the sample.

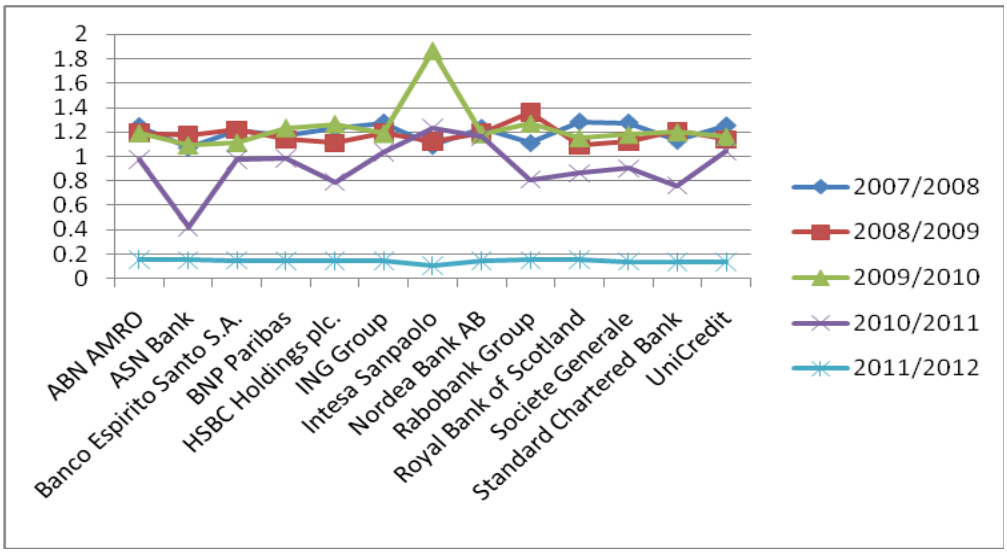

Figure 4. Malmquist productivity change
Source: author

During 2007 – 2010, individual banks' productivity fluctuated in the range 100 – 140%, depicting either no change in productivity level compared to the previous year or productivity gains. Productivity varied most in 2011 compared to 2010, most banks in the sample witnessing productivity declines between 58 percentage points (ASN Bank in Netherlands) and only 2 - 3 percentage points (ABN Amro in Netherlands, Banco Espirito Santo in Portugal and BNP Paribas in France). In 2012 relative to 2011 all socially responsible banks recorded a sharp productivity decline, of around 85 percentage points

(see table 2 for an aggregated picture on the number of banks depicting productivity increase, decrease or stagnation).

Table 2. Features of Malmquist indices – synthesis

	Banks with Malmquist > 1	Banks with Malmquist < 1	Banks with Malmquist = 1
2007 / 2008	13	0	0
2008 / 2009	13	0	0
2009 / 2010	13	0	0
2010 / 2011	4	9	0
2011 / 2012	0	13	0

On average, in 2008 compared with 2007 the productivity increased with 20 percentage points; in 2009 relative to 2008 increased with 17 percentage points; in 2010 the productivity increased most, with 24 percentage points, then it entered on a decreasing path. In 2011 the average productivity compressed with 8 percentage points, while in 2012 it decreased sharply, with 85 percentage points.

To identify the component that contributed most to Malmquist indexes path over time and across banks, it has been have computed the technological and technical efficiency change, the results being summarized in table 3.

Table 3. Sources of total productivity growth

	Catch-up effect (technical efficiency)			Technological efficiency (frontier shift)		
	> 1	< 1	= 1	> 1	< 1	= 1
2007 / 2008	5	7	1	9	4	0
2008 / 2009	11	2	0	7	6	0
2009 / 2010	3	10	0	12	1	0
2010 / 2011	11	2	0	3	10	0
2011 / 2012	1	12	0	0	13	0

Productivity changes recorded by most banks in the sample in 2008 relative to 2007 and 2010 relative to the preceding year are mainly due to a regress of the catch-up effect and increases of technological efficiency. In 2009 and 2011 the productivity levels have been influenced most by increases of the catch-up effect. As expected, the productivity regress in 2013 had been due to decreases in both catch-up effect and technological change.

On average, the catch-up effect has been most prominent in 2011 (increase of around 10 percentage points), followed by 2009 with 8.6 and 2008 with 3.6 percentage points. The frontier shift effect had been very significant in 2010, with an increase of 53 percentage points.

6. Conclusions

The research question of the paper had been targeted toward socially responsible banks and their financial behavior. Keeping in mind that they have joined the same three, complementary sustainability frameworks and hence committed to implement and fulfill all their principles, it is of interest to examine the manner in which their business models have passed through a convergence process, in terms of main financial indicators.

Consequently, it has been performed an in-depth, bank-level empirical analysis during 2007 – 2013, so as to investigate whether these particular 13 sustainable banks

depict common, resembling features of their business models, or on the contrary have strategy specificities.

The descriptive statistics revealed a relative heterogeneity in terms of several sustainable banks' key financial indicators, namely their individual contribution to the domestic GDP, the market share, the liquidity indicator, the operational efficiency and the financial structure of their balance sheet.

In terms of efficiency estimates, both models tested (intermediation approach and profit efficiency approach) show that in each year considered a single socially responsible bank had achieved the status of full efficiency. Therefore, the production frontier looks different for each year considered. On average, the highest inefficiency has been observed in the case of intermediation approach (model M1).

The efficiency estimates provide, however, a static picture for each bank in each year considered. Accordingly, to gain a comprehensive picture it has been computed the Malmquist productivity index in order to illustrate not only the productivity's changes over time but also the sources of changes, in terms of catch-up effect and technological progress. The findings show that in 2008, 2009 and 2010 all banks recorded productivity growth by comparison with the preceding year, but starting with 2011 they entered slowly on a productivity regress trend.

References

1. Berger, A.N. **The economic effects of technological progress: evidence from the banking industry**, Journal of Money, Credit and Banking, vol. 35, no. 2, 2003
2. Boitan, I. **The social responsibility stream in banking activity. A European assessment**, Conference proceedings SGEM: Political sciences, law, finance, economics and tourism, published by STEF92 Technology Ltd., Sofia, Bulgaria, 2014, pp. 793-800.
3. Caves, D. W., Christensen, L. R. and Diewert, W.E. **The economic theory of index numbers and measurement of input, output and productivity**, Econometrica, 50, pp. 1393-1414, 1982.
4. Dardac, N. and Boitan, I. **Assessment of the Romanian banking system performance. A non-parametric approach**, Policies of Economic and Social Development in Europe, din seria Sofia Conferences on Social and Economic Development in Europe, 2010, pp. 89-97
5. Dardac, N. and Boitan, I. **Malmquist index, an alternative technique for measuring credit institutions productivity**, Theoretical and Applied Economics, Vol. 520, no. 3, 2008, pp. 29-36
6. Davidescu A. A and Strat, V. **Coordinates of a New Romanian Regional Policy - Identifying the Development Poles. A Case Study**, Informatica Economica, vol. 18, no. 2, 2014, pp.88-99
7. Fare, R., Grosskopf, S., Lindgren, B. and Proos, P. **Productivity development in Swedish hospitals: a Malmquist output index approach**, Southern Illinois University Carbondale, Department of Economics Discussion paper, 89(3), 1989

8. Ferreira, C. **Efficiency and integration in European banking markets**, School of Economics and Management, Working Paper WP 08/2011/DE/UECE, 2011
9. Frame, W.S. and White L.J. **Technological Change, Financial Innovation, and Diffusion in Banking**, in Berger, A.N., Molyneux, P. and Wilson, J.O.S. (eds.) "The Oxford Handbook of Banking", Oxford University Press, 2009
10. Ho, S.J. and Mallick, S.K. **The Impact of Information Technology on the Banking Industry: Theory and Empirics**, Queen Mary, University of London, UK, 2006
11. Morita, H. and Avkiran N.K. **Selecting Inputs and Outputs in Data Envelopment Analysis by Designing Statistical Experiments**, Journal of the Operations Research, Society of Japan, Vol. 52, No. 2, 2009, pp.163-173
12. Pisano, U., Martinuzzi, A. and Bruckner, B. **The Financial Sector and Sustainable Development Logics**, **principles and actors,** ESDN Quarterly Report 27 – December 2012
13. Saeed, K.A. and Bampton R. **The Impact of Information and Communication Technology on the Performance of Libyan Banks**, Journal of WEI Business and Economics-December, Vol. 2, No. 3, 2013
14. Scippacercola, S. and Sepe, E., **Principal Component Analysis to Ranking Technical Efficiencies through Stochastic Frontier Analysis and DEA**, Journal of Applied Quantitative Methods, Vol. 9, No.4, Winter 2014
15. * * * **A new global partnership: eradicate poverty and transform economies through sustainable development**, United Nations, May 2013
16. * * * **United Nations Environment Program – Financial Initiative,** http://www.unepfi.org
17. * * * **United Nations Global Compact,** http://www.unglobalcompact.org
18. * * * Equator Principles, http://www.equator-principles.com

[1] **Acknowledgement**
This paper was co-financed from the European Social Fund, through the Sectoral 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.

ANALYSIS OF REPORTED CASES OF ROAD TRAFFIC ACCIDENTS IN UMUAHIA METROPOLIS

Emmanuel John EKPENYONG
Lecturer, Department of Statistics,
Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria

E-mail: ekpesstat@yahoo.com

Pius Ida OKO
BSc, Department of Statistics,
Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria

E-mail: okopius87@gmail.com

Abstract

The occurrence of road accidents in Abia State has been of a great concern to the citizens and as a result this research work has examined the Analysis of road traffic accidents' data in Umuahia Metropolis for the period 2010 – 2013 with data collected from the State Traffic Office, Central Police Station, Umuahia. The research identified some of the problems of road accident which includes; the factors that cause the prevailing rate of road accident, the vehicle types involved in such accidents, the days in the week and the year that recorded the highest number of road accidents and inter-dependence of the nature of casualties caused by road accidents and the vehicle type involved. Analyzing and describing the data using descriptive statistics, the Friedman test, the Kruskal Wallis H test and the Chi-square test for goodness of fit and independence, it was discovered that majority of the road accidents are caused by human factors like over speeding, recklessness, drinking and driving, drug abuse, arrogance and illiteracy. Number of reported cases of road accident is not uniformly distributed across the years and vehicle types, and the casualties in road accident are dependent on the vehicle type involved.

Keywords: *Road traffic accidents, Friedman test, Kruskal Wallis test, Chi-square test for goodness of fit and independence, Casualties*

1. Introduction

Over the years, Africans and indeed Nigerians have been deceived to believe or accept that road traffic accident is an integral part of human fate. This belief tends to inspire a considerable indifference to the road traffic accident syndrome, because deaths and injury resulting from road traffic accidents were accepted as part of human destiny.

Back in the 1970s, during the trial of the oil boom, which as a matter of fact reshaped wealth distribution and tests in our society, were many devastating and long lasting side effects among such was the influx or coming in of motor vehicles into a country where only a few number of good road existed.

There is no doubt that this increase in the number of vehicles changed the nations driving style. A large percentage of Nigerian motorists took up the driving business as if they had a set or number of people to kill before sunset; the rate at which accidents happen on our roads has almost turned them into slaughter slabs making road traffic accidents in Nigeria number two life taker after diseases.

No wonder the bank data revealed publication that the ratio of road traffic accidents especially the totality rate in Nigeria was ten times higher than the percentage obtained or recorded in developed countries. The impact of that sudden awareness was translated into the motivation to create a safer road culture in Africa as a whole.

The researchers therefore are much concerned with the problems encountered by citizens with respect to road accidents in Abia State, determining which factors cause higher rates of road traffic accidents, whether road accidents are independent of vehicle types and which vehicle type is mostly involved in road accidents.

Road accidents are defined as unpleasant events in a vehicle that happen unexpectedly and cause injury or damage.

Jorgensen and Kirsten (2011) defined road accident as a result of a chain of events in which something has gone wrong with vehicles on the road, resulting in an undesired outcome. Road traffic accidents according to the Official Statistics of Finland (OFS) (2014) is an event having led to personal injury or damage to property that has taken place in an area intended for public transport or generally used for transport and in which at least one of the involved parties has been a moving vehicle specified in the road traffic Act; such party can also be a tram and in level crossing accidents, a train.

The International Road Assessment Programme (2014) classifies road accidents (collision) into four types namely: head-on collision, single-vehicle collisions, intersection collisions and run-off collisions. (See also Road and Maritime Services (2011)).

Murtala (2014) observed that the causes of road traffic accidents are conventionally classified into Human, Mechanical and Environmental factors. The human factors accounts for up to 90% of accidents, in fact, the mechanical and environmental factors are subsequent to it. Lack of knowledge of road signs and regulations, illiteracy, health problems like poor eye sight, excessive speeding, alcoholism, drug abuse, arrogance, over-confidence are some of the human factors too numerous to mention that cause bad traffic accidents. Mechanical Factors include poor vehicle maintenance, tyre blow outs, poor lights, un-road worthy vehicles, broken down vehicles on the road without adequate warning sign etc. Rainfall, sun

reflection, storm, heavy, wind, pot holes, interred roads are some of the environmental conditions that contribute to road traffic accidents. According to Murtala (2014), Nigerian roads despite the current effort of Road maintenance Agency are still in very bad states especially those leading to rural areas.

The effects or consequences of road accidents are both socially and economically distressing.

According to Ubani (2014), the consequences of road accidents over the years have proved devastating to victims and those affected. It has left a lasting scar on many with bad consequences such as: loss of lives, amputation, regrets, financial deadlock and everlasting pain. (See also Ministry of Road Transport and Highways, Government of India, 2013, Jacqueline *et al*, 2001).

Ajit *et al* (2009) in their research work: 'A statistical Analysis of road traffic Accidents in Dibrugarh city,' noted that the analysis of qualitative data gathered summarizes two principal factors namely: human and environmental, as joint significant contributor to the occurrence of Road Traffic Accidents in Dibrugarh city. Human characteristic (rush and negligence) according to them make the highest contribution (95.38%) to the road traffic accidents in the study area. The environmental factors are related to bad weather and poor road condition. (See also Atubi, 2012, Ohakwe et al, 2011, Eke et al, 2000).

2. Methods of data analysis

For proper analysis of this work, the entire data collected were analyzed using the following:

2.1. The Friedman test

The Friedman test is used to test if the mean effects of the causes of road accidents by different years are equal.

2.2. The Mann – Whitney U – test

The Mann-Whitney U test is used to compare all possible pairs of the causes of road accidents.

2.3. The Chi – square test for goodness of fit

The Chi-square test for goodness of fit is used to test if the number of road accidents is uniformly distributed across the years and among the vehicle types.

2.4. The Kruskal-Wallis H-test

The Kruskal-Wallis H test is used to test if there is a significant difference among the nature of harm caused by road accidents. It is also used to test if the casualties caused by different vehicle types are equal.

2.5. The Chi-square test of independence

The Chi-square test of independence is used to test if the casualties of road accidents are independent of the vehicle type involved.

3. Data presentation and analysis

3.1. Data presentation

Table 1. Yearly Number of Recorded Road Accidents in Umuahia Metropolis (2010 – 2013)

Years	Number of reported cases	Vehicle type				Days of accident							Causes of accident			No. of people	Casualties		
		Heavy duty	Buses	Cars	Motorcycles	Mon	Tue	Wed	Thur	Fri	Sat	Sun	Human	Mechanical	Environmental	Involved	No. killed	No injured	No not injured
2010	53	17	12	17	8	14	7	6	3	8	7	8	45	6	2	123	46	18	59
2011	52	12	14	23	10	11	8	8	6	14	1	4	48	1	3	154	48	46	60
2012	72	12	29	18	14	8	14	13	6	6	10	15	62	5	5	257	80	85	92
2013	68	19	27	25	20	4	8	13	11	10	12	10	60	3	5	363	75	137	151

Source: State Traffic Office, Central Police Station, Umuahia

Figure 1 below shows the number of road accidents in different days for the year 2010 - 2013.

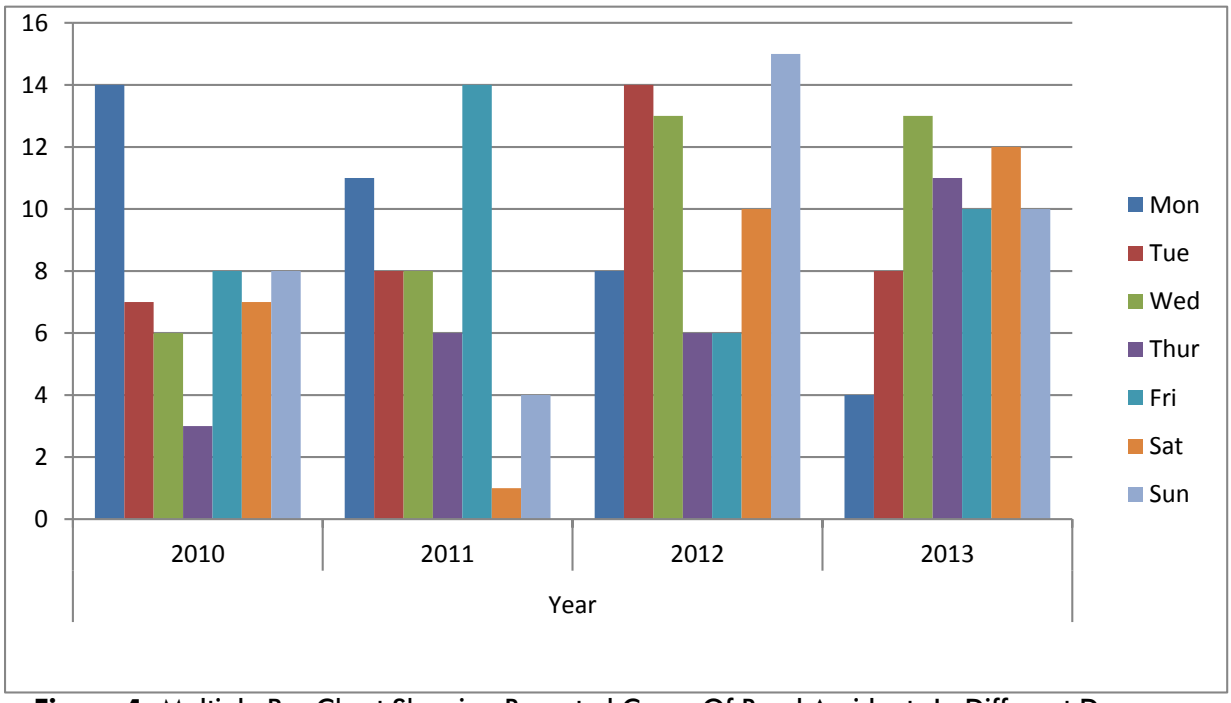

Figure 1. Multiple Bar Chart Showing Reported Cases Of Road Accidents In Different Days For The Year 2010 to 2013

From figure 1and table 1, it can be seen that;

(i) Monday recorded the highest number of reported cases of road accidents of 14 and Thursday has the least number of reported cases of 3 for the year 2010.

(ii) For the year 2011, Friday recorded the highest number of reported cases of 14 while Saturday had the least number of reported cases of 1.

(iii) For the year 2012, it can be seen that Sunday recorded the highest number of reported cases of 15 while Thursday and Friday both have the least number of reported cases of 6 each.

(iv) For the year 2013, Wednesday recorded the highest number of reported cases of 13 road accidents while Monday has the least number of reported cases of 4 accidents.

Generally speaking for the period under study, i.e from 2010 to 2013, Wednesday recorded the highest number of reported cases of 40 road accidents while Thursday had the least number of reported cases of 26 road accidents.

Also, table 1, we can also see that 2012 recorded the highest number of reported cases of 72 road accidents while 2011 recorded the least number of reported cases of 52 road accidents.

3.2. Data analysis

3.2.1. Data analysis using the Friedman test

We test the hypothesis

H_0: The mean effects of the causes of accidents by different years are equal, against

H_1: The mean effects of the causes of accidents by different years are not equal.

source, Sum of the ranks
no r
Environmental 6.5 4
Human 12.0 4
Mechanical 5.5 4
Adjusted for ties
Value: 6.533333
Pvaluechisq: 0.03813333
Alpha : 0.05

Since p (0.03813333) $<$ $\alpha(0.05), we\ reject\ H_0$ and therefore conclude that the mean effect of the causes of road accidents are not all equal. We also conclude that there is a significant difference in the number of accidents occurring yearly.

Since the test above is significant we compare each pair of causes and test if they are significantly different using the Wilcoxon Rank Sum test.

Comparison between Human and mechanical causes

We test the hypothesis:

H_0: The mean effects of human and mechanical causes are equal

H_1: The mean effects of human and mechanical causes are not equal

W = 16, p-value = 0.01519

Since $p\ (0.01519) < \alpha(0.05)$, we reject H_0 and conclude that the mean effects of the Human and mechanical causes of road accidents are not equal.

Comparison between Human and Environmental causes

We test the hypothesis:

H_0: The mean effects of Human and Environmental causes are equal.

H_1: The mean effects of Human and Environmental causes are not equal.

data: x and y

W = 16, p-value = 0.0147

Since $p\,(0.0147) < \alpha\,(0.05)$, we reject H_0 and conclude that the mean effects of the Human and Environmental causes of road accidents are not equal.

Comparing Mechanical and Environmental Causes

We test the null hypothesis

H_0: The mean effect of Mechanical and Environmental causes are equal

H_1: The mean effect of Mechanical and Environmental causes are not equal

data: x and y

W = 8.5, p-value = 0.5

Since $p\,(0.5) > \alpha\,(0.05)$, we accept H_0 and conclude that there is no significant difference between the mean effect of Mechanical and Environmental causes of road accident.

In general, we therefore conclude from the tests conducted above that only the Human factor is significantly different from other causes of road accidents, while the Mechanical and Environmental factors are the same.

3.2.2. Using the Kruskal-Wallis test to test the differences in the nature of harm caused by road accidents

We test the hypothesis:

H_0: There is no significant difference among the nature of harm caused by road accidents, against.

H_1: There is a significant difference among the nature of harm caused by road accidents.

data: x

Kruskal-Wallis chi-squared = 1.457, df = 2, p-value = 0.4826

Since $p\,(0.4826) > \alpha\,(0.05)$, we accept H_0 and conclude that there is no significant difference in the nature of harm caused by road accident.

3.2.3. Using the Chi – square test to test uniformity in the data

Here we test the null hypothesis:

H_0: Number of reported cases road accident is uniformly distributed across the years

H_1: Number of reported cases accident is not uniformly distributed across the years.

χ^2 =26.3469, df = 15, p-value = 0.03453

Since $p\,(0.03453) < \alpha\,(0.05)$, we reject H_0 and conclude therefore that the number of reported cases of road accident is not uniformly distributed across the years 2010 – 2013.

3.2.3.1. Testing uniformity in the number of road accident among . different vehicle types

Here we test the hypothesis,

H_0: Number of road accidents is uniformly distributed among the vehicle type

H_1: Number of road accidents is not uniformly distributed among the vehicle type.

data: x

χ^2 = 10.6101, df = 3, p-value = 0.01403

Since $p\,(0.01403) < \alpha\,(0.05)$, we reject H_0 and conclude therefore that the number of reported case of road accident is not uniformly distributed among vehicle type from 2010 – 2013.

3.2.4. Using the Chi-square test to test independence of casualties of road accidents on the vehicle type

Here, we test the null hypothesis;

H_0: The casualties of road accidents are independent of the vehicle type.

H_1: The casualties of road accidents are dependent on the vehicle type

Table 2. Observed and Expected frequencies of Casualties in Road Traffic Accidents of Vehicle Types

	CASUALTIES			
VEHICLES TYPE	**No. killed**	**No. injured**	**No. not injured**	**Total**
Heavy duty	54 (36.64)	35 (45.09)	43 (53.27)	**132**
Buses	64 (126.03)	163 (144.75)	227 (183.22)	**454**
Cars	85 (51.63)	57 (59.30)	44 (75.06)	**186**
Motorcycles	46 (34.70)	31 (39.86)	48 (50.45)	**125**
Total	**249**	**286**	**362**	**897**

χ^2 = 94.961, df = 6, p-value < 2.2e-16

Since$p\ (2.2e^{-16}) <\ \alpha(0.05)$, we reject H_0 and conclude therefore that the casualties recorded through road accidents are dependent on the vehicle types involved in the road accidents. Also, from table 2, Buses recorded the largest number of casualties and is seconded by Cars.

3.2.5. Using the Kruskal-Wallis H-test to test difference in the casualties caused by different vehicle types

We test the hypothesis

Ho: Casualties caused by different vehicle type are the same

H1: Not all the casualties caused by different vehicle type are the same

data: x

Kruskal-Wallis chi-squared = 7.0513, df = 3, p-value = 0.07028

Since p (0.07028) > $\alpha(0.05)$, we accept H_0 and conclude therefore that there is no significant difference in the nature of casualties by different vehicle types

4. Discussion of results

From the analysis carried out so far, it can be observed that the numbers of road accidents caused by the different factors of human, mechanical and environmental are not all the same. The analysis so far shows a clear difference between the number of road accidents caused by human factors and any other factor causing road traffic accident; but the number of road accident caused by mechanical and environmental factors are approximately equal, this results agrees with that of Ajit et al (2009), Ohakwe et al (2011), Ezenwa (1986), and Odero (1998) that human factors are the lead causes of road traffic accident in Nigeria.

The nature of casualty recorded through road traffic accidents has been shown to be equal; that is, there is no contrast between the number of people killed, number of people injured and number of people not injured. The number of reported cases of road accidents is not uniformly distributed across the years 2010-2013.

The numbers of reported cases of road accidents for the years under study- (2010 – 2013) are not uniformly distributed among the vehicle types involved in the accident, vehicle types which are grouped into heavy duty, buses, cars and motorcycles.

For the year 2010, Monday recorded the highest number of accidents of 14 and Thursday had the least number of reported cases of 3. For 2011, Friday recorded 14 while Saturday had 7. For 2012, Sunday recorded 15 while Thursday and Friday both recorded 6 each. For 2013, Wednesday recorded 13 and Monday recorded 4 accidents.

For the entire period under study (2010-2013), Wednesday recorded the highest number of reported cases of 40 road accidents while Thursday had the least number of reported cases of 26 road accidents. Also 2012 had the highest number of reported cases of 72 road accidents while 2011 recorded the least number of reported cases of 52 road accidents. The increase in

Majority of the road accidents witnessed in Umuahia metropolis for the period under study (2010 – 2013) is caused by human factors which include; lack of knowledge of road signs and regulations, illiteracy, health problems like poor eye sight, excessive speeding, alcoholism, drug abuse, arrogance, over-confidence etc.

The casualties recorded through road accidents depend to a great extent on the vehicle type involved in the road accidents and it is also observed that there is no significant difference between the natures of casualties recorded by the different vehicle types.

5. Conclusion

From the foregoing, it has been inferred that the major factors that cause road traffic accidents are the human factors – lack of knowledge of road signs and regulations, illiteracy, alcoholism, excessive speed, arrogance and other reckless attitudes of the drivers have caused tremendous harm to the lives of commuters in Umuahia Metropolis. Casualties recorded through road traffic accidents depend to a large extent on the vehicle types involved in the road accidents and Buses records the highest number of road accidents within the period. Moreover, for the period under study, Wednesday records the highest number of reported cases of road accidents, while Thursday recorded the least number. It is also found that casualties caused by different vehicle types are the same. Therefore, Drivers of vehicles should be given adequate training from time to time by the Federal Road Safety Commission on road use and traffic regulations. Bus drivers should be given proper orientation on the use of roads as they constitute the majority of commercial drivers.

References

1. Ajit, G. and Ripunjoy, S. **A statistical Analysis of Road Traffic Accidents in Dibrugarh City, Assam, India**, 2009.
2. Atubi, A.O. **A monthly Analysis of Road Traffic Accident in Selected Local Government Areas of Lagos State, Nigeria**, Mediterranean Journal of Social Sciences, Vol. 3, No. 11, 2012, pp 47-62
3. Eke, N., Etebu, E.N. and Nwosu, S.O. **Road traffic Accidents Mortalities in Port Harcourt, Nigeria,** Anil Aggrawal's Internet Journal of Forensic Medicine and Toxicology, Vol. 1, No. 2, 2000

4. Ezenwa A.O. **Trends and Characteristics of Road Traffic Accidents in Nigeria**. *J.R. Soc. Health, 106*, 1986, pp 27 – 29
5. Head on Crashes. International Road Assessment Programme. Retrieved May 12, June 30, 2014
6. Jacqueline B. **Impact of Car Accidents can be long-lasting.** www.webmd.com Aug 20, 2001
7. Jorgensen and Kirsten. **Concept of Accident Analysis**. *Encyclopedia of Occupational Health and Safety*. International Labor Organization, 2011, Geneva.
8. Ministry of Road Transport and Highways, Government of India: **Common Causes of Road Accidents**. Jharkhand Police Portal. www.jhpolice.gov.in/road-safety/common-causes-of-road-accidents. May 18, 2013
9. Murtala M.U. **Road Transport Accidents causes Effects and prevention.** www.gamji.com/article4000/news4881, 2014, referred July 7
10. Odero W. **Alcohol – Related Road Traffic Injuries in Eldoret, Kenya. East Africa**. *Med J.* 75, 1998, 708 – 711.
11. Official Statistics of Finland (OSF): Statistics on Road Traffic Accidents (e – publication). Helsinke Statistics Finland, referred July 2, 2014
12. Ohakwe J., Iwueze I.S. and Chikezie D.C.: **Analysis of Road Traffic Accidents in Nigeria: A case study of Obinze/Nekede Iheagwa Road in Imo State, Southeastern, Nigeria**. *Asian Journal of Applied Sciences*, 4: 2011, 166 – 175.
13. Roads and Maritime Services: **Five Most Common Crash Types,** www.rms.nsw.gov.au. November 3, (2011).
14. Ubani A.T.: **Causes of Road Accidents and Prevention**. *Hubpages*. May 26, (2014)

Appendix 1

Monthly road accident data in Umuahia Metropolis (2010 – 2013)

			Vehicle type				Days of accident							Causes of accident			No. of people	casualties		
Years	Months	Number of reported cases	Heavy duty	Buses	Cars	Motor-cycles	Mon	Tue	Wed	Thur	Fri	Sat	Sun	Human	Mecha-nical	Environ-mental	Involved	No killed	No injured	No not injured
2010	JAN	11	5	2	1	3	0	0	4	0	2	2	3	10	1	0	23	11	10	2
	FEB	3	1	0	2	0	0	0	0	0	1	0	2	3	0	0	6	3	0	3
	MAR	6	2	1	2	1	2	3	0	0	1	0	0	5	0	1	16	6	5	5
	APR	4	1	0	3	0	2	1	0	0	0	1	0	4	0	0	8	4	0	4
	MAY	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	2	1	0	1
	JUN	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	2	1	0	1
	JUL	3	2	1	0	0	0	0	0	0	2	0	1	2	1	0	4	2	0	2
	AUG	6	2	1	2	1	5	0	0	1	0	0	0	4	1	1	22	4	1	17
	SEP	4	0	3	0	1	1	3	0	0	0	0	0	4	0	0	19	3	0	16
	OCT	1	0	1	1	0	1	0	0	0	0	0	0	1	0	0	2	1	0	1
	NOV	4	2	1	1	0	0	0	0	1	0	1	2	4	0	0	9	5	0	4
	DEC	9	2	2	3	2	2	0	2	1	1	3	0	6	3	0	10	5	2	3
2011	JAN	5	2	1	2	2	0	2	2	0	1	0	0	4	0	1	10	5	1	4

	FEB	7	4	3	1	0	3	0	0	0	4	0	0	7	0	0	30	8	16	6
	MAR	2	0	2	0	0	1	0	0	0	0	0	1	1	1	0	16	2	2	12
	APR	4	0	0	3	2	1	0	1	0	2	0	0	2	0	2	10	7	2	1
	MAY	7	3	3	3	1	1	3	0	0	0	0	3	7	0	0	16	7	2	7
	JUN	4	0	0	3	1	2	0	0	2	0	0	0	4	0	0	6	2	1	3
	JUL	8	1	3	4	0	0	2	0	2	4	0	0	8	0	0	42	6	19	17
	AUG	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	2	1	0	1
	SEP	4	1	1	1	1	0	0	3	1	0	0	0	4	0	0	8	3	3	2
	OCT	4	0	0	3	1	0	0	1	0	3	0	0	4	0	0	4	2	0	2
	NOV	3	0	0	1	2	2	0	0	0	0	1	0	3	0	0	4	2	0	2
	DEC	2	1	1	1	0	1	1	0	1	0	0	0	3	0	0	6	3	0	3

Sources: State Traffic Office, Central Police Station, Umuahia

Appendix 2

			Vehicle type				Days of accident							Causes of accident			No. of people	casualties		
Years	Months	Number of reported cases	Heavy duty	Buses	Cars	Motor-cycles	Mon	Tue	Wed	Thur	Fri	Sat	Sun	Human	Mecha-nical	Environ-mental	Involved	No killed	No injured	No not injured
2012	JAN	3	1	2	0	0	0	0	0	2	1	0	0	2	0	1	5	2	0	3
	FEB	7	2	1	3	1	0	2	4	0	0	0	1	6	1	0	12	5	0	7
	MAR	7	0	1	3	3	0	2	2	1	1	0	1	5	1	1	12	6	1	5
	APR	10	2	4	3	1	2	1	1	0	2	3	0	8	2	0	32	8	19	5
	MAY	6	0	3	1	2	0	0	3	1	1	0	1	4	1	1	9	6	1	2
	JUN	8	2	4	1	1	0	0	0	0	1	4	0	7	0	1	38	6	12	20
	JUL	4	1	2	1	0	1	0	1	0	0	1	1	3	0	1	24	6	10	8
	AUG	5	0	3	2	0	0	0	2	2	0	0	1	5	0	0	26	3	0	3
	SEP	5	0	4	1	0	1	3	0	0	0	0	3	5	0	0	29	6	21	2
	OCT	4	0	3	1	0	1	0	0	0	0	0	2	4	0	0	45	4	13	28
	NOV	3	2	1	0	1	1	0	0	0	0	0	0	3	0	0	6	3	1	2
	DEC	10	2	1	2	5	2	0	0	0	0	2	5	10	0	0	39	25	7	7
2013	JAN	5	2	1	2	1	0	3	0	0	0	3	0	4	0	2	10	4	4	2
	FEB	7	4	3	1	1	0	1	1	0	2	1	2	6	1	0	14	6	5	3
	MAR	2	0	2	0	0	0	1	1	0	1	0	2	4	0	0	34	3	6	25
	APR	4	0	0	3	2	0	0	0	2	0	1	0	4	0	0	30	6	5	
	MAY	7	3	3	3	4	0	0	5	2	1	1	0	9	0	1	57	12	28	17
	JUN	4	0	0	3	3	1	1	2	1	2	4	3	13	0	0	68	18	40	10
	JUL	8	1	3	4	0	0	0	1	0	0	0	0	1	0	0	2	1	0	1
	AUG	1	0	0	1	3	2	1	0	2	1	0	1	5	0	2	39	7	20	12
	SEP	4	1	1	1	2	0	0	1	3	1	1	0	5	1	0	46	8	12	26
	OCT	4	0	0	3	3	1	1	0	0	1	1	1	4	1	0	24	25	14	5
	NOV	3	0	0	1	0	0	0	0	1	0	0	0	1	0	0	3	1	0	2
	DEC	2	1	1	1	1	0	0	2	0	1	0	1	4	0	0	36	4	3	29

Sources: State Traffic Office, Central Police Station, Umuahia

INTEGRATION BETWEEN THE ROMANIAN AND THE EURO AREA BANKING MARKETS: AN APPLICATION OF THE JOHANSEN COINTEGRATION TEST TO INTEREST RATES ON LOANS TO NON-FINANCIAL CORPORATIONS[1]

Maricica MOSCALU[2]
PhD, Lecturer, Department of Finance,
Faculty of Finance, Insurance, Banking and Stock Exchange,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: maricica.moscalu@fin.ase.ro

Abstract:

The aim of the present paper is to investigate if there is any level of integration between the Romanian and the euro area banking markets – with focus on lending activities of monetary financial institutions (MFIs) to non-financial corporations (NFCs) – and to assess this level of integration through using both quantity- and especially price-based data. The main empirical instrument used is the Johansen cointegration test applied to pairs of interest rates for euro-denominated loans granted by MFIs located in the two markets to NFCs for different loan maturities and amounts. By employing recent data, the results of the test indicate the existence of a cointegration relationship between the interest rates for loans with floating rate / period of initial rate fixation of up to 1 year and up to and including EUR 1 million euro. These findings suggest that, although Romania is not yet part of the Economic and Monetary Union (EMU), the two markets are not completely disintegrated especially with regard to short-term bank lending operations. Although further investigation is necessary, the findings are relevant from the perspective of Romania entering the EMU and have implications for Romanian NFCs' access to finance.

Key words: banking market integration, interest rate spread, cross-border loans, euro area, Romania, Johansen co-integration test, Granger causality

1. Introduction

Financial integration in general and financial markets integration in particular, has gained considerable and increasing attention considering the recent financial shocks that hit the European Union (EU) member states, particularly the euro area countries. The objective of achieving a high level of financial integration at EU and euro area levels derives from and is consistent with the objective of creating a single market for financial services. The reasons why financial integration counts and should be fostered are related to the issues of (i) monetary policy transmission in the euro area, (ii) financial stability and (iii) functioning of the payment systems, it positively impacting on all of those three areas, in the view of the

European Central Bank (ECB, 2014). De Guevara et al. (2007) synthesize the macroeconomic benefits of financial integration previously highlighted in literature, i.e. an expanded set of financial assets and services, a decrease in their prices and an improvement in portfolio composition. But financial integration is not only important from a macroeconomic and monetary perspective. It has also implications and benefits at microeconomic level. Lucey and Zhang (2011) argue that, based on the expected benefits at country level following financial integration, firms are stimulated to adjust their capital structure by taking advantage of an expanded range of financing resources and a lower cost of capital. Micro-level implications of financial integration can be also traced at household-level (ECB, 2014; Pungulescu, 2013; Maudos, 2013). Maudos (2013), for instance, states that the introduction of euro benefited to firms and households alike by leading to a decrease in their real borrowing costs. According to de Guevara et al. (2007), the awaited benefit of lower financial costs stemming from financial integration, can be viewed as the outcome of tighter competition. In this paper, the implications for NFCs are given priority.

Integration in financial markets can be traced along four main segments – money markets, bond markets, equity markets and banking markets (ECB, 2014). The present paper focuses on banking markets and aims at investigating if there is any level of integration between the Romanian and the euro area banking markets with respect to the MFIs' lending activities to NFCs as this has implications on the access to and the financing conditions for the Romanian NFCs. The decision to focus on this issue given that Romania is not yet part of the euro area can be justified, on one side, by the high presence of foreign banks on the domestic banking market especially banks headquartered in EU and euro area countries. According to the National Bank of Romania (BNR, 2014), 24 out of the 40 credit institutions have a dominant share of foreign capital ownership and hold 80.2% of the aggregate net balance sheet assets. Also, based on the consolidated statistics reported by the Bank of International Settlements (BIS, 2014), it results that the average quarterly level for 2007-2013 of foreign claims by banks located in the euro area (Austria, Belgium, Germany, Greece, Italy, Netherlands, Portugal, Spain) to foreign claims reported by banks located in the EU countries on Romanian counterparts, is 79.8%. These figures are not negligible and underpin the assumption that the two markets are not completely disintegrated much more that some papers suggest that a high presence of foreign-owned banks with facile access to international (euro area) capital (interbank) markets, as is the case of Romania, is indicative of a high level of financial integration (see Owen and Temesvary (2014) and Popov and Ongena (2011)). On the other side, the choice to focus on the banking markets is explained by the fact that, according to the ECB (2014), fragmentation on this market negatively affects the economic recovery in the euro area by restricting the access to finance for firms, especially for small and medium-sized enterprises (SMEs).

The paper uses recent data (2007:1-2014:11) on monthly interest rates on loans to NFCs charged by MFIs located in the euro area and Romania, respectively, which serve as dataset for the empirical tests conducted, as well as monthly / quarterly (transformed to annual) data (2007-2013) on the claims (loans) held by MFIs located in the euro area on Romanian counterparts (volume-based measures) and vice versa which is analyzed in a descriptive manner. The empirical test is the Johansen's test of cointegration applied to the interest rate pair series followed, where co-integration hypothesis is confirmed, by the estimation of the vector error correction model (VECM) and the Granger causality test. The Johansen cointegration test prevailed over the Engle-Granger (1987) procedure given that it

has been recently used in studies of cointegration in financial markets (Boubakri et al., 2012; Demian, 2011). In the main analysis, the overall euro area market is taken as the benchmark market while the German and the Austrian markets are used as benchmarks in the robustness checks section.

The results suggest the existence of a certain level of integration between the Romanian and the euro area banking markets. Specifically, the results showed that the interest rates on new euro-denominated loans with floating interest rate / period of initial interest rate fixation of up to 1 year and of up to and including 1 EUR million euro are co-integrated. The acceptance of the co-integration hypothesis means that there is a long-run equilibrium relationship between the respective interest rates series. However, the estimation of the ECM model showed that the adjustment speed towards the equilibrium is slow especially when the test is conducted for the euro area as a whole (0.03). The convergence speed is higher when the benchmark is considered to be the Austrian market (0.05) and especially the German market (0.08).

The paper is subsequently organized as follows. Section 2 briefly reviews the relevant literature while section 3 presents the data and the method used. Section 4 presents and discusses the results and section 5 briefly concludes the paper.

2. Literature review and research question

2.1. Banking market integration – Euro area and Romania

This section begins with a brief review of the recent evolutions in the euro area banking markets in the light of the recent financial crisis. Maudos (2013) states that the recent financial crisis has cancelled the benefit of lower cost of capital and brought about divergence in the euro area interest rates on loans to firms and to households as well. According to ECB (2014), the fragmentation trend induced by the financial crisis was reverted starting with mid-2012 but, as regards the banking segment, a still large dispersion in the interest rates for loans to NFCs and a reduced level of cross-border lending to the same sector are acknowledged. The widened interest rate differentials on loans to NFCs among the euro area countries are particularly higher for SMEs (ECB, 2014; Maudos, 2013) and this finding should be given special consideration given the SMEs' high reliance on bank financing (ECB, 2014; Blundell-Wignall, 2011). During the sovereign debt crisis, considered one of the causes of the segmentation among the euro area banking markets (ECB, 2014), Maudos (2013) documents that the interest rates on loans to NFCs and households in distressed countries have on average increased while in the other euro area countries they have decreased, especially for smaller loans to NFCs.

Given that Romania is not yet part of the EMU, it is interesting to see whether and how the creation of the EMU impacted on the euro area financial markets integration. Two conflicting opinions can be identified. On one side, some studies support the idea that the EMU contributed to the integration in stock markets in the post-euro period (Gębka and Karoglou, 2013; Bley, 2009; Spiegel, 2009; Kim et al., 2005) as well as to the interest rate convergence on banking markets (Maudos, 2013). Moreover, some of these studies indicate that the positive effect upon integration in stock markets manifested even in the anticipation of the introduction of the single currency (Gębka and Karoglou, 2013; Bley, 2009). This opinion contrasts with that of Bekaert et al. (2013) who document the existence of a positive intrinsic EU membership effect on financial and economic integration, regardless of the

adoption of the euro currency by EU countries. These conflicting results not only stimulate the debates on the topic but also underpin the research question of the present paper by making plausible the hypothesis that a certain level of integration with the euro area might exist even for EU countries not taking yet part in the EMU but aiming to achieve it, as it is the case of Romania. However, this hypothesis was formally tested by Boubakri et al. (2012) who assessed the level of integration between 10 Central and Eastern European EU countries, including Romania, with the euro area and concluded that the integration process is not yet complete – finding consistent with Pungulescu (2013) who additionally found that the process is not yet complete even for the UE15 group – and depends on the institutional monetary arrangements of these countries. The authors also showed that the integration is strongest for countries that are already part of the Exchange Rate Mechanism II (ERM II).

While the literature on financial integration, including banking (credit) markets integration, is quite rich, the references to Romania, especially with regard to banking market, are much less frequent. With respect to integration in credit markets for the 12NMS, Pungulescu (2013) shows improvements in interest rate convergence compared to Germany for three instruments – money market, Treasury Bills and Government bonds. Among the 12NMS, Romania is found to be one of the countries that exhibit the highest speed of convergence in money market interest rates over the period analyzed. With reference to the interbank market integration and its impact on firms' borrowing conditions in 14 old and new EU countries, including Romania, Popov and Ongena (2011) found that Romania was the most non-integrated country prior to its EU accession (1998-2005), finding that authors explain by the changing fiscal and monetary policies leading to high and variable inflation. However, the present study employs data for the period following the Romania' accession to EU.

2.2 Measurement of banking market integration

Regarding the measurement of financial integration, several approaches are possible and have been used in the previous literature. Generally, the literature distinguishes between *de jure* and *de facto* measures of financial integration as well as between *price-based* and *quantity-based* measures (ECB, 2014; Friedrich et al., 2013; Bekaert et al., 2013; Lucey and Zhang, 2011). The use of price-based measures appears natural given that in integrated markets, according to the *law of one price*, identical assets – in terms of risk and maturity – should be identically priced, regardless of the markets they are traded on (Pungulescu, 2013; Yeyati et al., 2009; Bekaert and Harvey, 1995). The opposite of integrated markets is segmented i.e. fragmented markets (Harm, 2001; Bekaert and Harvey, 1995). Regarding the difference between *de jure* and *de facto* measures, Lucey and Zhang (2011) make the following distinction: the first type of measures relate to changes in regulations on cross-border capital flows while the second type reflects the extent to which a country effectively accesses the international capital markets.

For measuring credit (debt) market integration, a price-based measure is the most frequent approach – be it the interest rate or sovereign yields spreads (Bekaert et al., 2013; Pungulescu, 2013). However, some authors warn on the caveats associated with the approach based on interest rate differentials as they can be explained by differences in risk characteristics of the clients serviced by banks located in different markets (Spiegel, 2009) and that their declining trend may be driven by factors other than financial market integration (Pungulescu, 2013). As quantity-based measures of integration in credit markets,

Lucey and Zhang (2011) use the ratios to GDP of outstanding international debt securities and outstanding loans from non-resident banks, respectively. Similarly, Friedrich et al. (2013), considered measures such as: the ratio of the total foreign assets and liabilities over GDP; the ratio of change in net foreign assets over GDP; share of the number of foreign banks in all banks. As it can be seen, quantity–based measures are the results of the effective financial transactions that took place between the markets under analysis. In this paper, I employ several measures of integration that are related to both quantity- and price-based measures of financial integration. They will be described in the data section along with the corresponding data series used to calculate them.

Based on the above findings of the previous literature, the aim of this paper is to investigate, the existence and the level of integration between the Romanian and the euro area banking markets with focus on the assets side i.e. lending activity of the MFIs balance sheets.

3. Methodology

3.1. Data

This section will briefly describe the data series used to investigate the level of integration in banking markets between Romanian and euro area, taken as a whole. Data series come from the European Central Bank (ECB) and the National Bank of Romania (BNR) databases respectively. Data on cross-border loans between the MFIs located in the euro area and counterparties located in Romania and vice-versa comes from the ECB database. The specific data series collected for the 2007-2013 period are as follows: (a) loans by euro area MFIs to non-MFI sector located in Romania; (b) loans by MFIs located in Romania to the MFI and non-MFI sectors located in the euro area; (c) loans by MFIs located in Romania to the MFI and non-MFI sectors in EU countries not belonging to the euro area. By dividing these data series by the Romanian GDP (Eurostat), quantity-based measures of banking market integration are obtained.

The ECB and the BNR databases provided data on interest rates charged by MFIs on loans to NFCs located in the respective areas. Five monthly interest rates (2007:01-2014:11) were used corresponding to the following categories of loans: (a) new loans – total maturity and total amount (ITN_EA, ITN_RO); (b) new loans of up to and including EUR 1 million and with period of initial rate fixation (euro area) or floating rate / period of initial rate fixation (Romania) of up to 1 year (IS1_EA, IS1_RO); (c) new loans of over EUR 1 million and with period of initial rate fixation (euro area) or floating rate / period of initial rate fixation (Romania) of up to 1 year (IS2_EA, IS2_RO); (d) outstanding loans – total maturity and total amount (ITE_EA, ITE_RO); (e) outstanding loans with a maturity of over 5 years and total amount (IL_EA, IL_RO). The series for the euro area are corresponding to a changing composition of the EMU.

3.2. Method

To analyze the data, a descriptive analysis is combined with empirical tests of co-integration and causality. Co-integration was traditionally used as an empirical method for investigating financial integration (see Harm (2001)) and it was recently used, for instance, in Demian (2011) and Boubakri et al. (2012). Even Engle and Granger (1987) mentioned that cointegration analysis could be adequate for investigating the existence of common

trends in e.g. short-term and long-term interest rates or in the price of the same commodity in different markets. Testing for co-integration between the interest rates for Romania and euro area banking markets means testing for the existence of a long-run or equilibrium relationship between the two data series.

This paper employs the Johansen co-integration test implemented in EViews. An alternative earlier procedure is that of Engle and Granger (1987). The idea behind the co-integration tests is to test the nature of the static relationship i.e. the relationship in levels between two (or more) non-stationary time series integrated of the same order (usually 1). If the static relationship is integrated of a lower order (usually 0), then the series are called co-integrated i.e. and the static relationship becomes the cointegration relationship or the long-run equilibrium relationship; otherwise the static relationship is a spurious regression (for a more detailed presentation see Vogelvang (2005) p. 254, p. 266, p. 293).

The Engle-Granger (1987) procedure consists of testing the series for the presence of a unit root in order to establish their degree of integration and, if this is the case, the estimation of the static relationship between them. If the residual series of this equation is stationary then the initial series are called cointegrated (Bourbonnais, 2004 p. 283). The more complex Johansen's test was developed in Johansen (1988), is based on the maximum likelihood theory, and consists of estimating the space of the cointegration vectors in an unrestricted vector autoregressive process (VAR) with Gaussian i.i.d. errors and then test the number of its dimensions using a likelihood ratio test. The work was further extended in papers such as Johansen (1991).

If the series are cointegrated, then a (vector) error correction model (VECM) can be estimated based on the Granger representation theory (Bourbonnais, 2004 p. 283). The error correction model is a dynamic model which contains both the long-run relationship, given by the error correction term, and the short-run deviations from the equilibrium relationship embedded in the variables expressed in first differences (Vogelvang, 2005 p. 254). The coefficient of the error correction term is the adjustment speed towards the equilibrium relationship and is expected to be negative for the (vector) ECM specification to be correct; otherwise, the series depart from the long-run target (Vogelvang, 2005 p. 269; Bourbonnais, 2004 p. 284). The ECM approach was used to assess the level of integration in European interbank markets by Ongena and Popov (2011).

For the case of two (cointegrated) variables, the long-run and the short run (ECM) models can be written as in equations (1) and (2), respectively (Vogelvang, 2005 p. 298), where *s* and *p* denote the number of lags and γ is the coefficient of the error correction term.

$$y_t = \beta_1 + \beta_2 x_t + e_t \quad (1)$$

$$\Delta y_t = \beta_0 + \sum_{i=1}^{s-1} \gamma_i x_{t-i} + \sum_{j=1}^{p-1} \beta_j y_{t-j} + \gamma e_{t-1} + u_t \quad (2)$$

As an additional check of the results of the Johansen's test of cointegration, a Granger causality test was run. Given that the series are supposed to be non-stationary and potentially cointegrated, the Granger test cannot be run in the typical way. If this is the case, the procedure in Toda and Yamamoto (1995) is one of the recommended approaches. It consists of testing linear restrictions on parameters, such as the one implied by the Wald test for Granger non-causality, in an integrated (and cointegrated) VAR by re-estimating the VAR with an additional number of lags given by the maximum number of integration in the process. Subsequently, the coefficients of the additional lagged vectors are ignored and the

linear restrictions can then be tested only for the first coefficients based on the standard asymptotic theory.

4. Results and discussion

4.1. Cross-border loans and interest rates for loans to NFCs

Before presenting the results of the cointegration tests, the evolution of the quantity-based measures since 2007 will be presented.

The most obvious feature in the evolution of series in figure 1 is the enormous difference in magnitude between the levels of cross-border loans from the euro area to the Romanian non-MFI counterpart sector compared to that of cross-border loans from Romania to the euro area non-MFI sector. The general government and the private sector are included in the non-MFI sector. Specifically, the level of cross-border positions initiated by MFIs located in the euro area to GDP reached an average annual level of 8.73% during the period 2007-2013 compared with only 0.21% for the cross-border positions initiated in Romania. The maximum value (11.03%) in the sample (2009) is reached two years after Romania's accession to the EU and is reasonable to suspect that it was stimulated by this event. After that date, the share of cross-border loans from the euro area decreases again without reaching yet the level from the beginning of the period. This average level of 8.73% suggests that there is a certain level of integration between the Romanian and the euro area banking markets following the Romania's entrance to the EU.

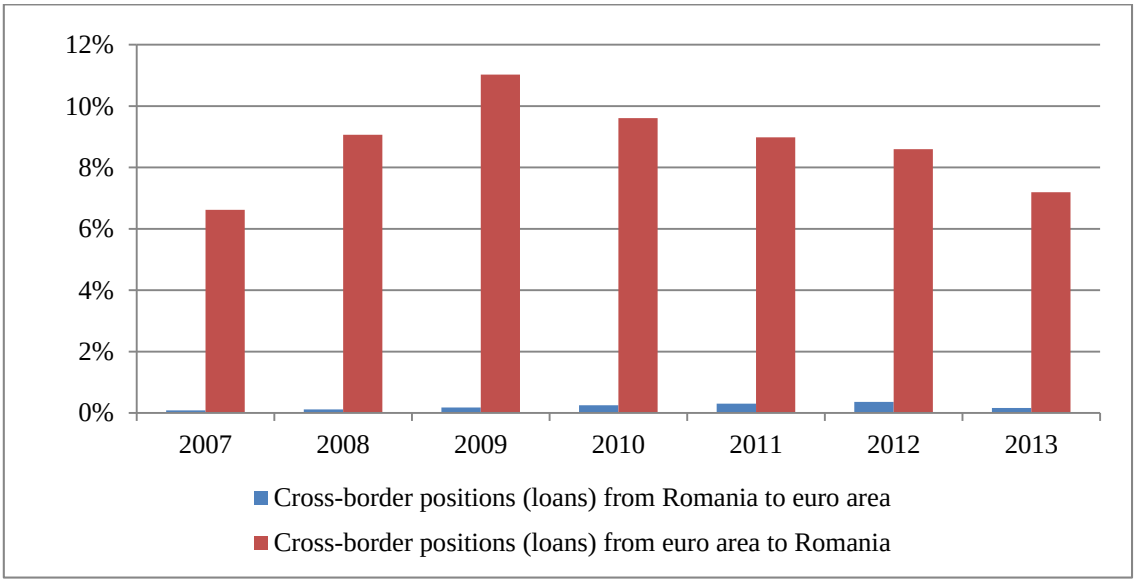

Figure 1. Proportion to GDP of cross-border loans to the non-MFI sector between Romania and the euro area

Source: Own elaboration based on ECB (2014) monthly / quarterly data for cross-border positions (outstanding amounts at the end of the period; the annual value was considered the value for December that year) and on Eurostat data for Romania's GDP (2014).

In order to strengthen the importance of the linkages between the Romanian and the euro area banking markets it is worth to compare the evolution of cross-border loans from Romania to euro area countries and EU non-euro area countries, respectively. This is illustrated in figure 2. It can be noticed that, unlike the positions initiated in the euro area, the cross-border loans initiated in Romania with the MFI sector as counterpart prevail over

those with the non-MFI counterpart, regardless that the euro area or the EU non-euro area is considered. At the euro area level, the share to GDP of loans granted by Romanian MFIs to the corresponding MFI sector is significantly higher than that for the non-MFI sector which can be explained by the high presence on the Romanian banking market of foreign banks headquartered in the euro area.

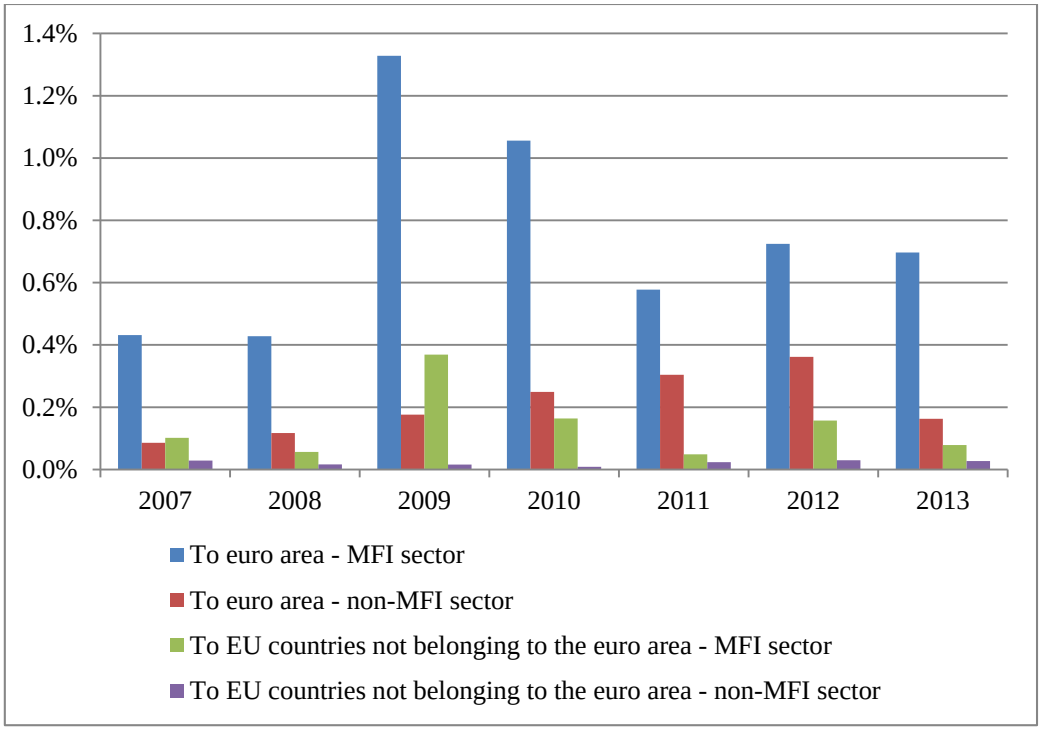

Figure 2. Proportion to GDP of cross-border loans by Romanian MFIs to the MFI and non-MFI sectors from the euro area and EU countries outside the euro area

Source: Own elaboration based on ECB (2014) monthly / quarterly data for cross-border positions (outstanding amounts at the end of the period; the annual value was considered the value for December that year) and on Eurostat (2014) data for Romania's GDP

As a preliminary step before proceeding to the cointegration tests, it is interesting to see if common trends can be grasped from the evolution of the interest rate pairs. Their evolution is presented in figure 3 (the total interest rate series were not presented for reasons of space but their evolution is similar of the others presented). It can be noticed a clear tendency of all the three interest rates to co-move for most of the period. This suggests that a co-integration relationship between the two interest rates for all the three pairs might exist. Next the results of the cointegration tests will be presented.

4.2. Results of the Johansen's test of cointegration in interest rates series

Johansen's test of cointegration can be applied on the necessary condition that the series are non-stationary. Thus the five pair series were tested for the existence of a unit root using the Augmented Dickey-Fuller (ADF) test. As a supplementary check, the Philips-Perron (PP) test was applied. The results are presented in table 1 only for the model with intercept with similar conclusions for the other models, as literature (Vogelvang, 2005 p. 289) considers that this is the specification frequently used for economic variables.

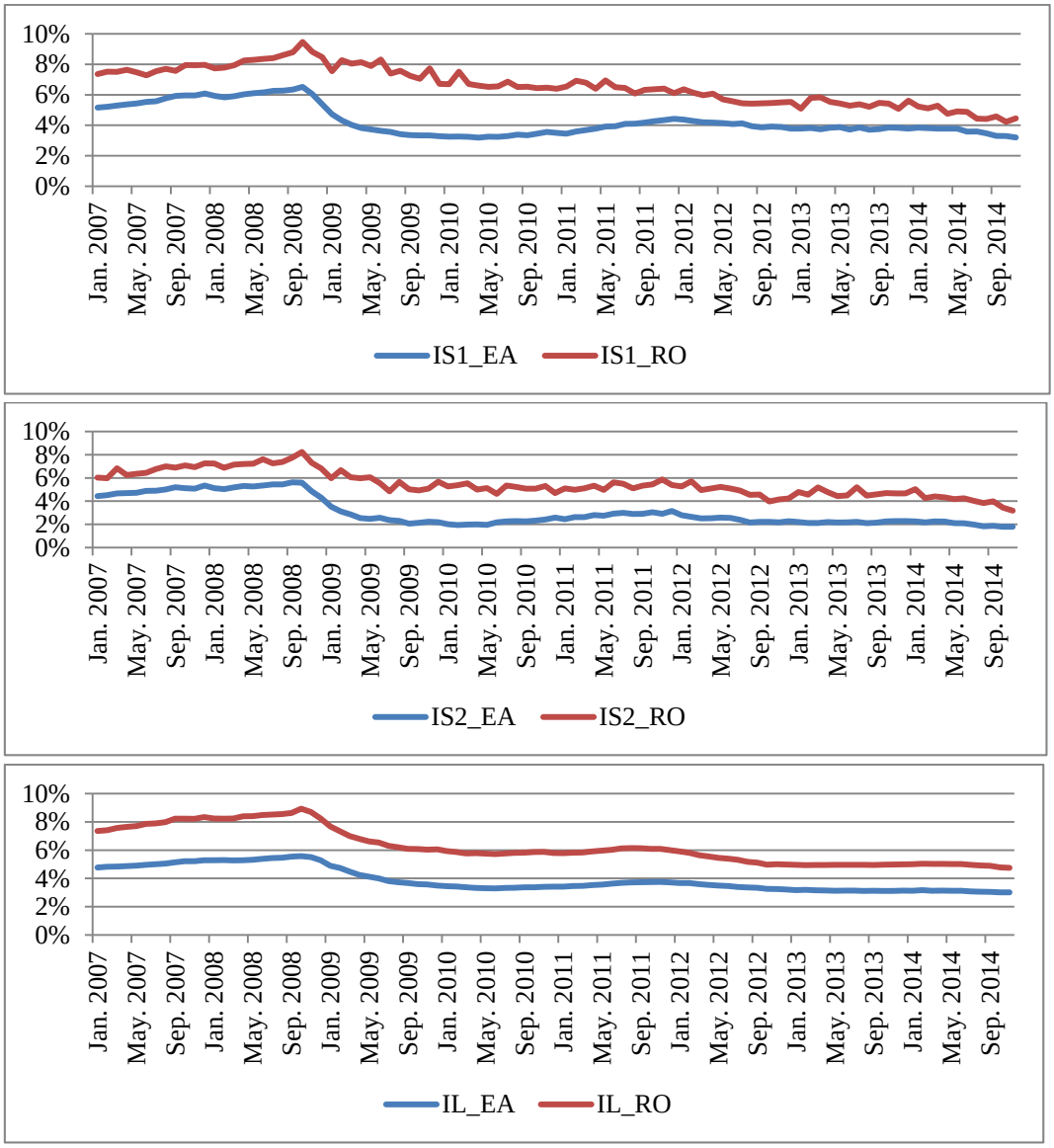

Figure 3. Evolution of MFIs interest rates on loans to NFCs in Romania and euro area between January 2007 and November 2014

Source: Own elaboration based on ECB (2015) and BNR (2015) data for three of the five interest pairs described in the Data subsection

Table 1. Results of the unit root tests

	Levels	First diff.	Levels	First diff.	Levels	First diff.	Levels	First diff.	Levels	First diff.
	ITN_RO		IS1_RO		IS2_RO		ITE_RO		IL_RO	
ADF	0.948	0.000	0.948	0.000	0.904	0.000	0.803	0.000	0.769	0.000
PP	0.927	0.000	0.899	0.000	0.856	0.000	0.835	0.000	0.844	0.000
	ITN_EA		IS1_EA		IS2_EA		ITE_EA		IL_EA	
ADF	0.602	0.000	0.551	0.000	0.656	0.000	0.447	0.024	0.479	0.005
PP	0.676	0.000	0.653	0.000	0.688	0.000	0.751	0.018	0.782	0.005

Source: Own elaboration. The table contains the p-values associated to the null hypothesis that the series are non-stationary

Results in table 1 clearly shows that all the ten series are non-stationary at level but that they become stationary by taking the first difference, according to both ADF and PP tests. Very similar conclusions are obtained based on the results of the KPSS test. Therefore, all the series are integrated of order 1 and the necessary condition is satisfied. To correctly apply the Johansen's test, an unrestricted VAR was previously run in order to determine the optimum number of lags to use in the cointegration test. This was considered to be the one suggested by the majority of the criteria provided that the errors were not autocorrelated in the VAR model; otherwise, the number of lags was increased until the autocorrelation could be rejected using the Breusch-Godfrey LM test of serial autocorrelation.

The results of the cointegration test are given in table 2 for both versions of the Johansen's test i.e. the trace test and the maximum eigenvalue test and for a summary of the models comprising models 2, 3 and 4 as these are considered of practical interest (Sjö, 2008).

Table 2. Results of the Johansen co-integration test – summary of the models 2, 3 and 4

No. of cointegrated relations by model (significant at 5% level)	Model 2	Model 3	Model 4
Group ITN_RO and ITN_EA (no. of lags = 3)			
Trace	0	0	0
Maximum-Eigenvalue	0	0	0
Group IS1_RO and IS1_EA (no. of lags = 4)			
Trace	0	0	1
Maximum-Eigenvalue	1	1	0
Group IS2_RO and IS2_EA (no. of lags = 7)			
Trace	0	0	0
Maximum-Eigenvalue	0	0	0
Group ITE_RO and ITE_EA (no. of lags = 12)			
Trace	0	0	1
Maximum-Eigenvalue	0	0	1
Group IL_RO and IL_EA (no. of lags =7)			
Trace	0	0	0
Maximum-Eigenvalue	0	0	0

Source: Own elaboration; model 2 – intercept (no trend) in the cointegrated equation, no intercept in VAR; model 3 - intercept (no trend) in the cointegrated equation and test VAR; model 4 - intercept and trend in the cointegrated equation, intercept in VAR

Results in table 2 show that there is one cointegrated relation in the case of only 2 group series i.e. short-term interest rates for loans of up to EUR 1 million (IS1_RO and IS1_EA) according to all the three models and the total interest rate for outstanding loans (ITE_RO and ITE_EA) according to the model 4. Consequently, the interest rates on loans granted to NFCs of up to 1 year period on initial rate fixation or floating interest rate and not exceeding EUR 1 million by MFIs from Romania and MFI located in the euro area,

respectively, are co-integrated, meaning that there is a long-term relation between them. The same seem to be true for interest rates on outstanding loans to NFCs as well. The results of the Johansen's test for the specific models for which has been proved that there are cointegrated relations, are given in table 3.

Table 3. Results of the Johansen co-integration test – models 2, 3 and 4

	Unrestricted Cointegration Rank Test (Trace)				Unrestricted Cointegration Rank Test (Maximum Eigenvalue)			
Hypothesized no. of CE(s)	Eigen.	Trace Statistic	0.05 Critical Value	Probab.	Eigen.	Max-Eigen statistic	0.05 Critical Value	Probab.**
Model 2: Group IS1_RO and IS1_EA (no. of lags = 4)								
None	0.170	19.335	20.261	0.067	0.170	16.797	15.892	0.036*
At most 1	0.028	2.538	9.165	0.670	0.028	2.538	9.165	0.67
Model 3: Group IS1_RO and IS1_EA (no. of lags = 4)								
None	0.157	15.351	15.495	0.052	0.157	15.328	14.264	0.034*
At most 1	0.000	0.022	3.841	0.881	0.000	0.022	3.841	0.881
Model 4: Group IS1_RO and IS1_EA (no. of lags = 4)								
None	0.161	27.785	25.872	0.029*	0.161	15.775	19.387	0.155
At most 1	0.125	12.001	12.518	0.061	0.125	12.001	12.518	0.061
Group ITE_RO and ITE_EA (no. of lags =12)								
None	0.212	29.130	25.872	0.019*	0.212	19.560	19.387	0.047*
At most 1	0.110	9.570	12.518	0.148	0.110	9.570	12.518	0.148

Source: Own elaboration. * denotes significance at 5%. ** MacKinnon-Haug-Michelis (1999) values

The analysis proceeds with the estimation of the vector error correction model for the series found to be cointegrated.

4.3. Estimation of the VECM and testing for Granger-causality for the cointegrated series

Estimation of the VECM was conducted using the same lag number found to be appropriate for the cointegration test. The VECM model for the interest rate series IS1_RO and IS1_EA is presented in eq. (3) which is according to the model specified in eq. (2).

$$D(IS1\_RO) = C(1)*(IS1\_RO(-1) - 4.951*S1\_EA(-1) + 0.144) + C(2)*D(IS1\_RO(-1)) + C(3)*D(IS1\_RO(-2)) + C(4)*D(IS1\_RO(-3)) + C(5)*D(IS1\_RO(-4)) + C(6)*D(IS1\_EA(-1)) + C(7)*D(IS1\_EA(-2)) + C(8)*D(IS1\_EA(-3)) + C(9)*D(IS1\_EA(-4)) + C(10) \quad (3)$$

In eq. (3), the expression assigned to C(1) is the cointegrating equation between IS1_RO and IS1_EA. The estimation of the eq. (3) by OLS led to the results presented in table 4.

Table 4. Estimation of VECM for the IS1_RO and IS1_EA series

	C(1)	C(2)	C(3)	C(4)	C(5)	C(6)	C(7)	C(8)	C(9)	C(10)
Coeff.	-0.031	-0.764	-0.558	-0.248	-0.132	0.904	0.335	-0.589	-0.103	-0.000
t-Statistic	-3.376	0.120	-3.863	-1.761	-1.172	3.130	1.056	-1.872	-0.346	-2.409
Probab.	0.001	0.000	0.000	0.082	0.245	0.002	0.294	0.065	0.731	0.0183

Source: own elaboration; Adjusted R-squared = 0.326; probability of F-statistic = 0.000; Durbin-Watson =2.008; the model is robust to: autocorrelation – Breusch-Godfrey serial correlation LM test statistic has a p-value of 0.745 for 12 lags; heteroscedasticity – Breusch-Pagan-Godfrey has a p-value of 0.164 and White test statistic, 0.845; normality: Jarque-Bera test statistic has a p-value of 0.121

The most important result of estimating the VECM relies in the statistics associated with the C(1) coefficient which is the coefficient of the error correction-term. It can be seen that it is statistically significant at 1%. Moreover, it is negative (- 0.031) which, according to the theory (Vogelvang, 2005 p.269; Bourbonnais, 2004 p. 284), mean that the error correction model representation is appropriate i.e. the series converge to the long-run relationship. However, the convergence is slow given that the value is close to zero. Thus, the static relationship between the cointegrated series is not a spurious regression but a long-term equilibrium relation. Its estimation, according to the eq. (1) led to the results given in expression (4) with the main statistics as follows: adjusted R-squared = 0.429, probability of F statistic = 0.000, DW = 0.130. As expected, it presents serial correlation which is corrected in the VECM.

$$IS1\_RO = 0.031 + 0.812*IS1\_EA \quad (4)$$

According to Popov and Ongena (2011), the value of the parameter for the independent variable in eq. (4) is indicative of the degree of integration i.e. if the value is positive and close to 1 it denotes that the two markets are integrated while for disintegrated markets the value is large and positive. In our case, the value is 0.812 which suggests the existence of a high level of integration between the two series.

In the last step of the analysis, the Granger non-causality hypothesis, based on a Wald test, is tested following the procedure recommended in Toda and Yamamoto (1995). The results of the VAR Granger causality test are given in table 5. The additional number of lags (besides the 4 lags previously used which gives the number of d.f.) introduced in the unrestricted VAR as exogenous variables is equal to 1 as the two variables are both integrated of order 1.

Table 5. Results of testing for VAR Granger causality

Excluded	Chi-square	d.f.	Probab.
Dependent variable is: IS1_RO			
IS1_EA	22.152	4	0.000
All	22.152	4	0.000
Dependent variable is: IS1_EA			
IS1_RO	7.029	4	0.134
All	7.029	4	0.134

Source: Own elaboration.

Based on the Granger's test of causality is can be concluded that there is a one-way causality relation running from IS1_EA to IS1_RO which means that the past values of interest rate IS1_EA in the euro area are influencing the current value of the interest rate IS1_RO in Romania. This result strengthens the results of the cointegration analysis. A reverse way causality would not seem plausible in this case as Romania is not part yet of the euro area and thus it cannot be expected to influence the level of interest rate in the EMU which is taken here as a whole.

The results of estimating the VECM for the ITE_RO and ITE_EA series are not presented given that the coefficient of the error correction model is not significant meaning that the error correction model representation is not valid, although the series were found to be cointegrated.

4.4. Robustness checks

The validity of the results obtained so far by considering the whole euro area as reference were checked against the case where a particular euro area market is considered as reference. This was firstly considered to be the German market since it is traditionally used as benchmark in empirical studies concerning financial integration (see Pungulescu (2013) and Ongena and Popov (2011) among others). In a second robustness check, the Austrian market was taken as reference and this is justified by the high presence on the Romanian domestic market of foreign-owned banks headquartered in Austria. The analysis was conducted in the manner previously described and, in order to save space, the results will be presented in short in this section.

Table 6. Main results in the robustness check

Reference market	No. of lags	Models in the Johansen cointegration test	Coefficient for the independent variable in the long-run relation (t-statistic)	Coefficient of the error correction term in VECM (t-statistic)	VAR Granger causality test Chi-square statistic	
					Dependent variable: IS1_RO	Dependent variable: IS1_DE / IS1_AT
Germany	6	Model 4	0.738 (12.180*)	-0.080 (-3.909*)	14.097**	3.298
Austria	4	Model 2, 3,4	0.734 (11.687*)	-0.051 (-3.888*)	10.562**	7.992***

Source: own elaboration; *, **, *** - significant at 1%, 5%, 10% respectively.

Regarding the number of cointegrating relations, it was higher for the Austrian market than for the German market. As in the main analysis, the series IS1_RO was found to be cointegrated with the corresponding series for Germany (model 4) and Austria (models 2, 3 and 4) which confirm the existence of a long-run relation between the MFIs interest rates for new loans to NFCs of up to and including EUR 1 million and with floating rate / period of initial rate fixation of up to 1 year. Additionally, in the case of Austria, two other cases of cointegration were found i.e. for interest rates on existing loans to NFC for total maturity (ITE_RO) and for maturity of over 5 years (IL_RO). However, the corresponding VECM estimated for these two last cases led to a non-significant coefficient for the error-correction term which means that the error correction specification is not appropriate.

For the interest rate series IS1_RO and the corresponding IS1_DE (for Germany) and IS1_AT (Austria), the main results were synthetized in table 6. It can be seen that the coefficient of error correction term in the corresponding VECM models is negative and statistically significant at 1% in both cases. Although still slow, the adjustment speed is higher yet than in the main analysis (0.03) and even higher in the case of Germany (0.08) compared to Austria (0.05). This result confirms not only that the linkages are stronger between the Romanian banking market and the individual markets of Germany and Austria compared to the euro area banking market as a whole, but also the role of the German market as a benchmark for the euro area. The results of the cointegration test are

strengthened again by those of the Granger causality test. In both cases, there is a significant influence (at 5%) exerted by interest rates in Germany (IS1_DE) and Austria (IS1_AT) on the interest rate in Romania (IS1_RO). In the latter case, a reverse causality can also be accepted for a 10% level of significance.

5. Conclusions

The research conducted in this paper aimed at investigating the existence and the level of integration between the Romanian and the euro area banking markets with focus on the lending activities of MFIs located in Romania and the euro area respectively. Using quantity as well as price-based data series, the results suggest that the two markets are integrated to a certain extent. The annual level of cross-border loans from MFIs located in the euro area to Romanian non-MFI sector to GDP was 8.73% between 2007 and 2013. It increased after Romania joined the EU but consistently decreased starting with 2010 suggesting potentially lower financial flows from abroad in the context of the financial and sovereign crises. Regarding the interest rates on loans to NFCs, the cointegration test conducted on five interest rate pair series for new as well as outstanding loans revealed that there is a consistent cointegrating relation between the interest rate series on new loans of up to EUR 1 million and floating rate or period of initial interest rate fixation of up to 1 year for the Romanian and the euro area banking markets respectively regardless which the benchmark for the euro area is considered to be – the overall euro area, Germany or Austria. The convergence towards the long-run relation is however slow in all three cases with the greatest value associated with Germany. These findings confirm that the Romanian and the euro area banking markets are cointegrated with regard to the lending activities to NFCs especially for short-term and smaller amount lending operations. Moreover, the results seem to be consistent with the results of Bekaert et al. (2013) who showed that there is an intrinsic EU membership positive effect on financial (stock) markets integration irrespective that the country has adopted the single currency or not. More and stronger cointegration relations are expected to be found after Romania will join the ERMII as Boubakri et al. (2012) have shown.

The findings of this study are relevant in the perspective that Romania might join the ERMII and eventually the EMU and have implications for the Romanian NFCs' access to and cost of bank financing. Further investigation is, however, necessary as longer data series will become available which is needed to improve the power of the empirical tests. The deposit side of MFIs balance sheets might represent a complementary avenue for further research.

References

1. Bekaert, G., Harvey, C. R., Lundblad, C. T. and Siegel, S. **The European Union, the Euro, and the equity market integration**. Journal of Financial Economics, Vol, 109, 2013, pp. 583-603
2. Bekaert, G. and Harvey, C.R. **Time-Varying World Market Integration**. The Journal of Finance, Vol. L, 2, 1995, pp. 403-444
3. Bley, J. **European stock market integration: Fact or fiction?** Journal of International Financial Markets, Institutions & Money, Vol. 19, 2009, pp. 759-776

4. Blundell-Wignall, A. **Solving the Financial and Sovereign Debt Crisis in Europe**. OECD Journal: Financial Market Trends, Vol. 2, 2011, pp. 201-224
5. Boubakri, S., Couharde, C. and Guillaumin C. **Assessing the financial integration of Central and Eastern European Countries with the euro area: Evidence from panel data and cointegration tests**. International Economics, Vol. 131, 2012, pp. 105-120
6. Bourbonnais, R. **Économétrie**. 5e edition, Dunod, Paris, 2004
7. De Guevara, J. F., Maudos, J. and Pérez, F. **Integration and competition in the European financial markets**. Journal of International Money and Finance, Vol. 26, 2007, pp. 26-45
8. Demian, C.V. **Cointegration in Central and East European markets in light of EU accession**. International Financial Markets, Institutions and Money, Vol. 21, 2011, pp. 144-155
9. Engle, R. F. and Granger, C.W. **Co-Integration and Error Correction: Representation, Estimation, and Testing**, Econometrica, Vol. 55, No. 2, 1987, pp. 251-276
10. Friedrich, C., Schnabel, I. and Zettelmeyer J. **Financial integration and growth – Why is Emerging Europe different?** Journal of International Economics, Vol. 89, 2013, pp. 522-538
11. Gębka, B. and Karoglou, M. **Have the GIPSI settled down? Breaks and multivariate stochastic volatility models for, and not against, the European financial integration**. Journal of Baking and Finance, Vol. 37, 2013, pp. 3639-3653
12. Harm, C. **European financial market integration: the case of private sector bonds and syndicated loans**. Journal of International Financial Markets, Institutions and Money, Vol. 11, 2011, pp. 245-263
13. Johansen, S. **Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector,** Econometrica, Vol. 59, No. 6, 1991, pp. 1551-1580
14. Kim, S. K., Moshirian, F. and Wu, E. **Dynamic stock market integration driven by the European Monetary Union: An empirical analysis**. Journal of Banking and Finance, Vol. 29, 2005, pp. 2475-2502
15. Lucey, B. M. and Zhang, Q. **Financial integration and emerging markets capital structure**, Journal of Banking and Finance, Vol. 35, 2011, pp. 1228-1238
16. Maudos, J. **Fragmentation of the European financial market and the cost of bank financing**. SEFO – Spanish Economic and Financial Outlook, Vol. 2, No. 3, 2013, pp. 5-15
17. Owen, A.L. and Temesvary, J. **Heterogeneity in the growth and finance relationship: How does the impact of bank finance vary by country and type of lending?** International Review of Economics and Finance, Vol. 31, 2014, pp. 275-288
18. Popov, A. and Ongena, S. **Interbank market integration, loan rates, and firm leverage**. Journal of Banking and Finance, vol. 35, 2011, pp. 544-559
19. Pungulescu, C. **European financial market integration in the European Union: EU15 vs. New Member States**. Emerging Markets Review, Vol. 17, 2013, pp. 106-124

20. Sjö, B. **Testing for Unit Roots and Cointegration**. 2008, Accessed in January 2015, https://www.iei.liu.se/nek/ekonometrisk-teori-7-5-hp-730a07/labbar/1.233753/dfdistab7b.pdf
21. Spiegel, M. **Monetary and financial integration: Evidence from the EMU**. Journal of The Japanese and International Economies, Vol. 23, 2009, pp. 114-130
22. Toda, H. Y. and Yamamoto, T. **Statistical inference in vector autoregressions with possibly integrated processes**. Journal of Econometrics, Vol. 66, 1995, pp. 225-250
23. Yeyati, E. L., Schmukler, S.L. and Van Horen, N. **International financial integration through the law of one price: The role of liquidity and capital controls**. Journal of Financial Intermediation, Vol. 18, 2009, pp. 432-463
24. Vogelvang, B. **Econometrics. Theory and applications with Eviews**. Pearson Education, Essex, England, 2005
25. * * * **European Central Bank**. **Financial integration in Europe**, April 2014, Retrieved from http://www.ecb.europa.eu/
26. * * * **European Central Bank Statistical**, Data Warehouse. Accessible online at http://sdw.ecb.europa.eu/browse.do?node=9484266
27. * * * **Banca Nationala a României**. Buletin lunar 6/2014. Retrieved from http://www.bnr.ro/Publicatii-periodice-204.aspx
28. * * * **Banca Nationala a României**, Baza de date interactiva. Accessible online at http://www.bnr.ro/Baza-de-date-interactiva-604.aspx
29. * * * **Bank for International Settlements**, Consolidated Banking Statistics. Accessed in October 2014 http://www.bis.org/statistics/consstats.htm

[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".

[2] **Maricica Moscalu** is a lecturer in the Department of Finance within the Bucharest University of Economic Studies. She is also currently a postdoctoral researcher in the POSDRU/159/1.5/S/134197 project where her research interests are related to the topic of financial markets integration and its effects on firm's access to finance and growth. Other research interests include bankruptcy prediction models and corporate social responsibility.

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

ADAPTIVE NEURO-FUZZY MODEL TUNING FOR EARLY-WARNING OF FINANCIAL CRISES[1]

Erika SZTOJANOV
PhD Student, Bucharest University of Economic Studies, Bucharest, Romania,
European Bank for Reconstruction and Development, London, UK

E-mail:

Grigore STAMATESCU
PhD, Assistant Professor, Department of Automatic Control and Industrial Informatics,
University „Politehnica" of Bucharest, Romania

E-mail: grigore.stamatescu@upb.ro

Abstract:
The paper introduces an early-warning method using multiple financial crises indicators which outputs relevant alerts compared to a precise indication of crisis inception, serving as an effective tool for decision makers. By leveraging fuzzy logic techniques, we design a multi-level fuzzy decision support system based on the evolution of credit growth, housing prices and GDP gap. A neuro-fuzzy approach allows fine tuning of the individual fuzzy sub-systems towards adaptive structures which can follow the particularities of the selected indicators at the individual country level. Simulation results and implementation details are presented, along with conclusion drawn from real economic datasets.

Key words: Early warning indicators, financial crises, fuzzy logic, adaptive neuro-fuzzy inference systems

1. Introduction

Financial crises are extremely costly! The significant losses, which result from them, are linked primarily to the slow-down of the lending process, which leads, as an immediate effect, to a dramatic decrease of the GDP [1], [4], [7]. The same studies reveal that these financial crises are relatively frequent, have long timespans and often represent the underlying factor for other types of crises, such as currency and debt crises, which further amplify their effects. Another aspect to consider is that these crises lead to a lack of trust in that particular economy, resulting in major negative consequences. The above-mentioned factors are the most important effects of banking crises, but besides these, a banking crisis

also has other side effects, which are equally important. The global effect of all of these can be devastating for the economy of the affected countries.

The recovery from such a crisis is a cumbersome and lengthy process! Therefore the stringent need for effective early-warning intelligent systems. Following the unfavourable effects of the previous financial crisis, the international supervisory bodies have established the Basel III and CRD-IV regulatory frameworks, which include the implementation of the Counter-Cyclical Buffers (CCBs) to increase the resilience of the banking sector and its ability to absorb shocks arising from financial and economic stress.

The intention is to set aside a financial reserve in the periods of economic growth and to use it during periods of crises in order to reduce their effects. To establish such a capital buffer, banks require an early warning system for determining the moment of emergence of a future crisis. A warning with one year in advance (Late warning) is not considered to be sufficient for setting up this buffer (CCB) and therefore, it is necessary to have a warning two or three years earlier (Early warning). These warning systems have to use the data and the indicators available at a particular moment, they have to be well-grounded and based on scientific methods, in order to benefit from a good credibility and to be able to indicate the beginning of crises with a sufficiently high degree of probability [3].

2. Early-warning indicators

Predicting financial crises represents a complex process due to the stakeholders involved which account for, aside economic and financial factors, on the local and global political context, social phenomena, etc. These are diffcult to quantify and measure in an objective manner.

A number of studies are dedicated to finding relevant indicators which could be used for an early warning system [1], [9], [2]. A selected indicator has to present an evolution with a peak point which can be easily observed with two or three years before the start of a crisis.

In the study published by Jan Babecky et al. in 2012, a number of 30 local and global indicators have been analysed, which could potentially be used as an early warning signal [1]. From the conclusions of the study, we retained that: "growth of domestic private credit, increasing FDI inflows, rising money market rates as well as increasing world GDP and inflation were common indicators of banking crises". The authors of this article have created and tested models with these single indicators, as well as models with multiple indicators and their conclusion was that using a composite early warning index increases the usefulness of the model when compared to using the best single indicator!

As motivation for crisis indicator selection and taking into account the conclusions stemming from [1] and [2], our multi-level warning system outputs a composite early-warning index (CEWI) by processing three basic indicators as inputs: private credit growth, house price growth and private credit to GDP gap. Figure 1 ilustrates the three components of the basic dataset, using aggregated ECB data on a reference timeframe of +/- 16 quarters relative to the outbreak of the recent financial crisis 2008-2009, whose effects are still active to date.

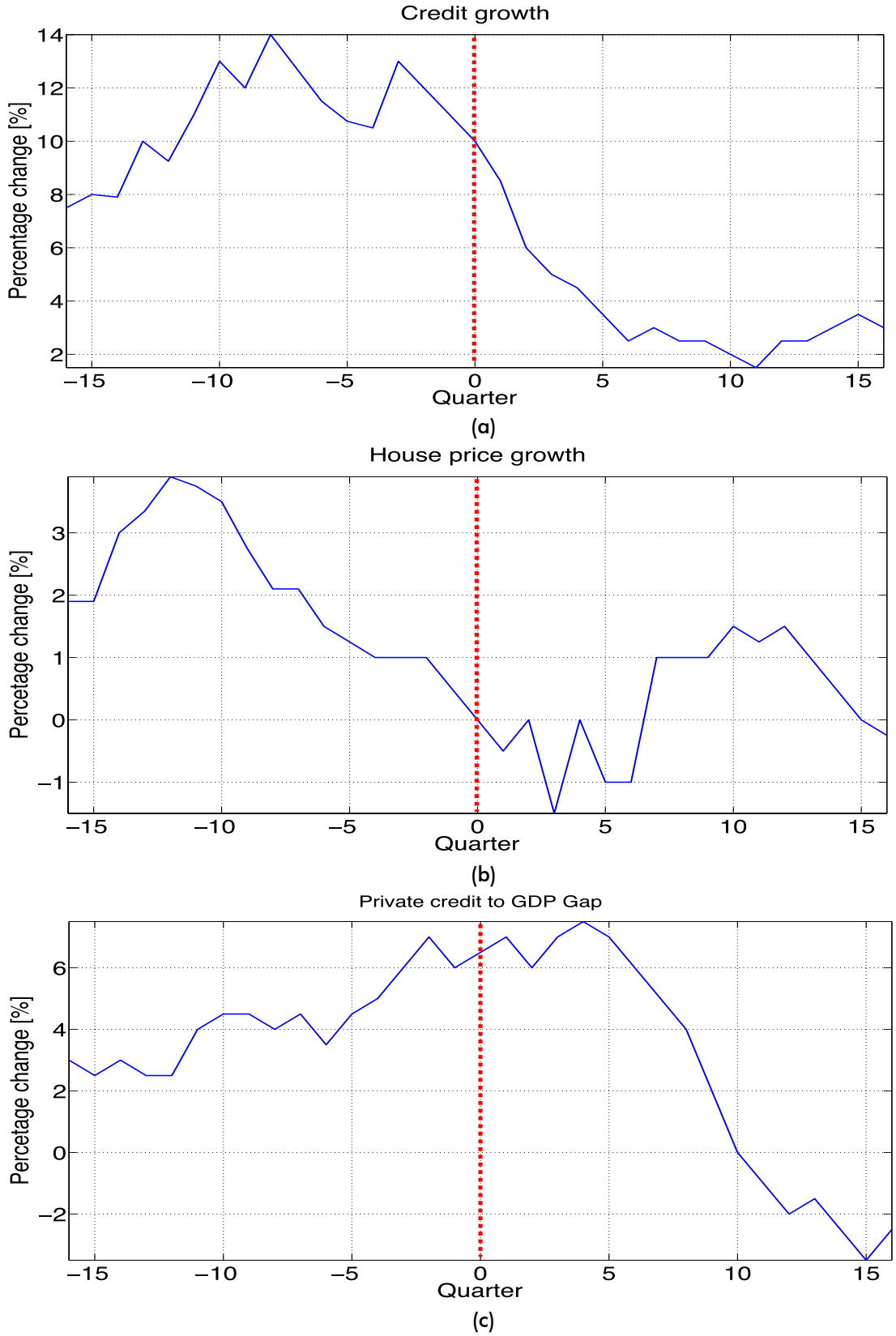

Figure 1. Aggregated reference ECB data: (a) private credit growth – CREDIT, (b) house price growth – HOUSE, (c) private credit to GDP gap - GAP

The private credit to GDP gap is generally acknowledged as one of the most robust crisis indicators [1]. Following its evolution, a value over 2% represents an important warning sign which should not be ignored. Compared to the GAP, which can be seen as a credible crisis indicator, for the other two: CREDIT and HOUSE, extremum points can be detected to be used as early-warning signs of a future crisis. In order to showcase this for the data in Figure 1, the house price growth curve shows a global maximum around three years before the financial crisis of 2008-2009 along with a maximum of the private credit growth indicator around two years before the same event.

3. Multi-level Fuzzy System for Early Warning

Fuzzy logic represents a powerful tool for quantitative method assesment [4-6]. Previous work [10] has dealt with the application of fuzzy logic for early warning of financial crises. As follow up and significant improvement of that approach, one of the main contributions of this work is the design, implementation and evaluation of a multi-level decision support system for early warning, based on fuzzy logic. This has the task of effectively combining the information offered by the three chosen indicators described in the previous section: GAP, CREDIT and HOUSE, by two fuzzy blocks, as first level, generating early-warning signals: EW1 and EW2, which are subsequently interpreted, in the second level, to give the final output. Additionally, using an adaptive neuro-fuzzy inference system (ANFIS) a neural network is trained for each of the individual fuzzy systems, in order to adapt the model to the particularities of a specific country upon departing from the aggregated data used in the design phase. The overall system diagram of the proposed approach is shown in Figure 2. It represents a two level fuzzy system for generation of the early warning signal based on the three input paramenters, evaluated as crisis indicators. The three main blocks are: fuzzy system 1 (FZ1), fuzzy system 2 (FZ2), composing the first processing, input, level and fuzzy system 3 (FZ3) as second processing, output, and level.

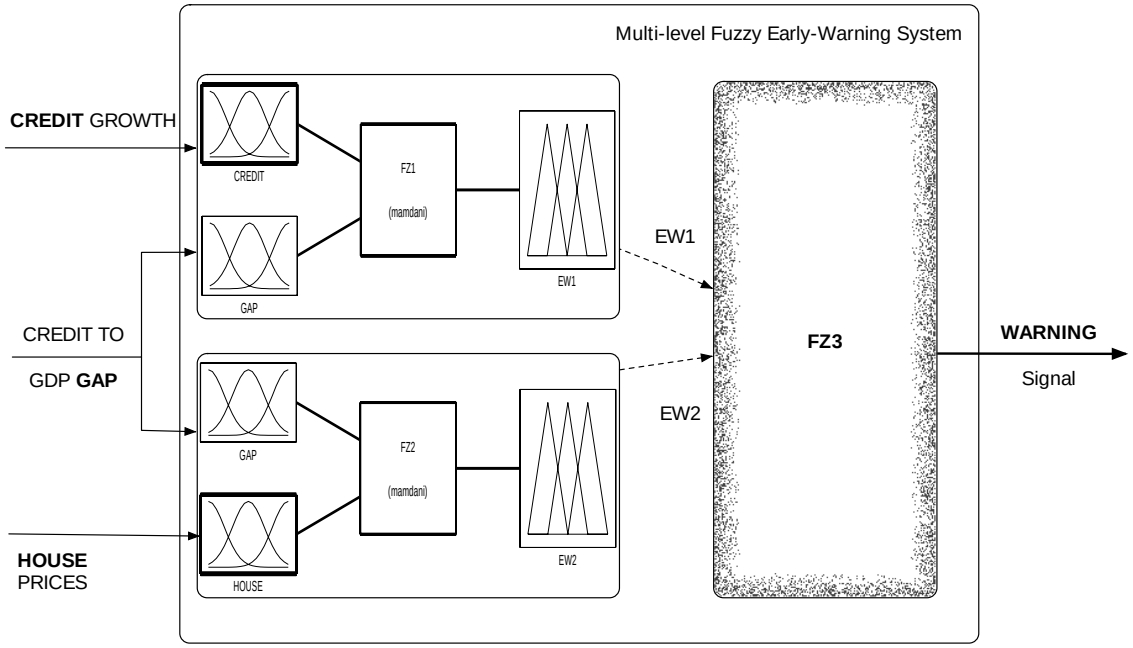

Figure 2. Multi-level Fuzzy Early-Warning System Diagram

The processing block FZ1 takes as input parameters the private credit growths and the private credit to GDP gap and outputs a synthetic, intermediary, warning indicator, denoted EW1. The second processing block of the input level, FZ2, has the private credit to GDP gap and the house price growth. The output signal is denoted EW2.

The two intermediary outputs are further processed at the second level through the FZ3 block. Establishing the processing rules requires expert knowledge of the underlying process and we base our design on some assumptions listed below:

The exact date of the outbreak of a financial crisis cannot be reasonably established in practice due to the lack of an universally accepted notion for how it is defined, with a diversity of opinions among economists. Therefore we found it more suitable that instead of outputing a crisp value in years for the crisis start estimation, to use a warning singal instead from fuzzy processing. Three levels have been defined as outputs of the FZ3 block: NORMAL, DANGER and CRITICAL. The NORMAL signal is generated when the economic indicators taken into consideration follow a predictable evolution, between adequate limits leading to the assumption that there is no impending danger of a financial crisis. The DANGER signal is associated to an early warning prediction. It will be basically generated when yearly credit growth surpasses 15-20% along with credit gap values over 2% or a housing price growth of over 5-6% associated to a similar increase in the GAP value. In this context the early warning given by the DANGER indicator signals a high probability of a financial crisis in the next 2 to 3 years. The third, CRITICAL, signal points to a crisis situation where already the economic situation is out of control. We associate this to a late warning signal. Tables 1-3 synthesize the definition of the membership functions for the input and output variables of the fuzzy systems.

Table 1. CREDIT input fuzzy membership functions

Slow credit	SL	0-6%
Moderate credit	MO	5% -10%
Medium credit	ME	8%-12%
Big credit	MR	11%-15%
Alarming credit	AL	14%-20%

Table 2. HOUSE input fuzzy membership function

Decrease	DC	-4%-0%
Stationary	ST	-1%-1%
Moderate growth	MG	0%-2%
Growth	GR	1.5%-4%
High growth	HG	3.5%-6%

Table 3. GAP input fuzzy membership function

Normal	NO	0%-1.8%
Dangerous	DS	1.5%-3%
Critical	CS	2%-8%

To exemplify the implementation of the system, carried out by means of the MATLAB Fuzzy Logic Toolbox [11], with triangular membership functions. The definition of the CREDIT input variable is illustrated in Figure 3.

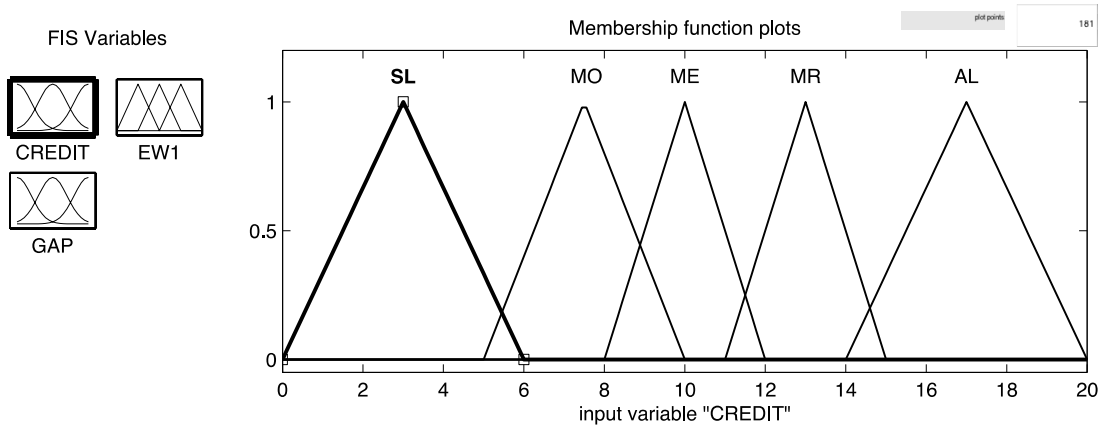

Figure 3. CREDIT membership function definition for FZ1

The results obtained by applying this approach on aggregated ECB data are shown in Figure 4. It can be seen how the average of the two indicators EW1 and EW2 correctly converges upon the starting point of the financial crisis of 2008-2009.

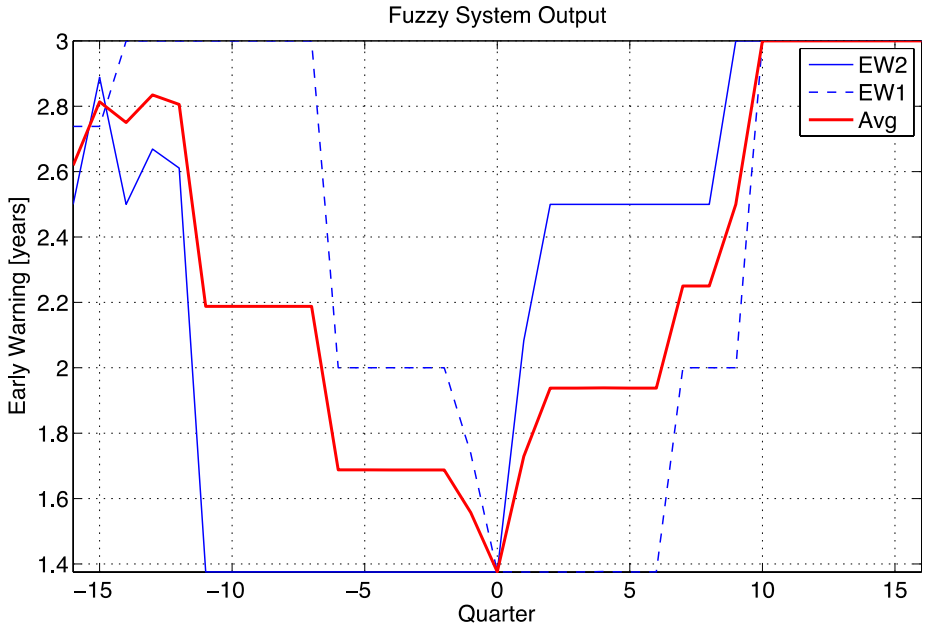

Figure 4. Synthetic output warning signals for FZ1 and FZ2

At the output level, the warning signal of the system is computed based on the intermediary warning signals EW1, based on CREDIT and GAP, and EW2, based on HOUS and GAP. The classification of the two signals is carried out according to Table 4.

Table 4. Output signal fuzzy classification

	Crises starts in	EW1	EW2
Iminent crises	1-1.75 years	IM	IK
Crises in two years	1.5-2.5 years	C2	K2
Crises between 2 and 3 years	2-3 years	2P	23
Crises in three years	2.5-3.5 years	C3	K3
Evolution without crises	3-5 years	FC	FK

An option for the implementation of the output decision block FZ3 is based on the combination of EW1/EW2 by means of an OR operator as described in Table 5. The output variables representing direct early warning signals: Late Warning – LW, Early Warning – EW, and No Crisis – NO, are classified by means of an intensity map as in Figure 5.

Table 5. FZ3 decision block output

	IM	C2	2P	C3	FC
IK	**LW**	**LW**	**LW**	**LW**	**LW**
K2	**LW**	**EW**	**EW**	**EW**	**NO**
23	**LW**	**EW**	**EW**	**EW**	**NO**
K3	**LW**	**EW**	**EW**	**EW**	**NO**
FK	**LW**	**NO**	**NO**	**NO**	**NO**

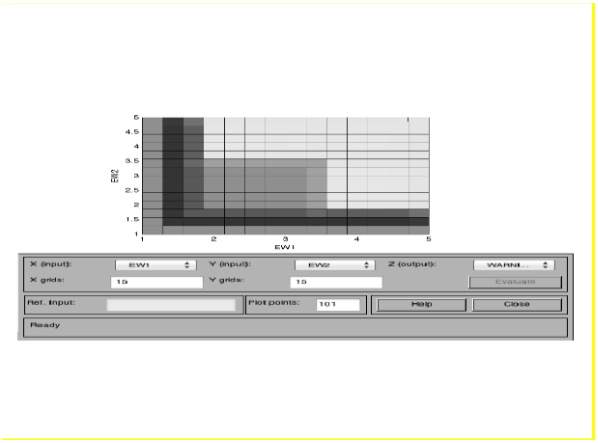

Figure 5. Warning signal of the multi-level structure

4. Adaptive neuro-fuzzy model tuning approach

In order to fine-tune our approach and provide for a more flexible way of adapting the fuzzy-inference systems to country specific variations, we build an adaptive neuro-fuzzy model. ANFIS systems were initially proposed by Jang [8] and offer neural-network based training for fuzzy logic membership function intervals. This requires an initial data set based on which the modelling of the neural network is carried out using a hybrid: backpropagation and linear lest squares, learning method.

Main results were obtained macro-economic data represented in percentage variation quarterly series for Hungary over 42 quarters starting in 1997. Figure 6 shows the input data for the three variables of choice: credit, house and gap. It can be seen how the need to adjust the fuzzy models designed initially for the aggregated data, appears due to significant country specific variations.

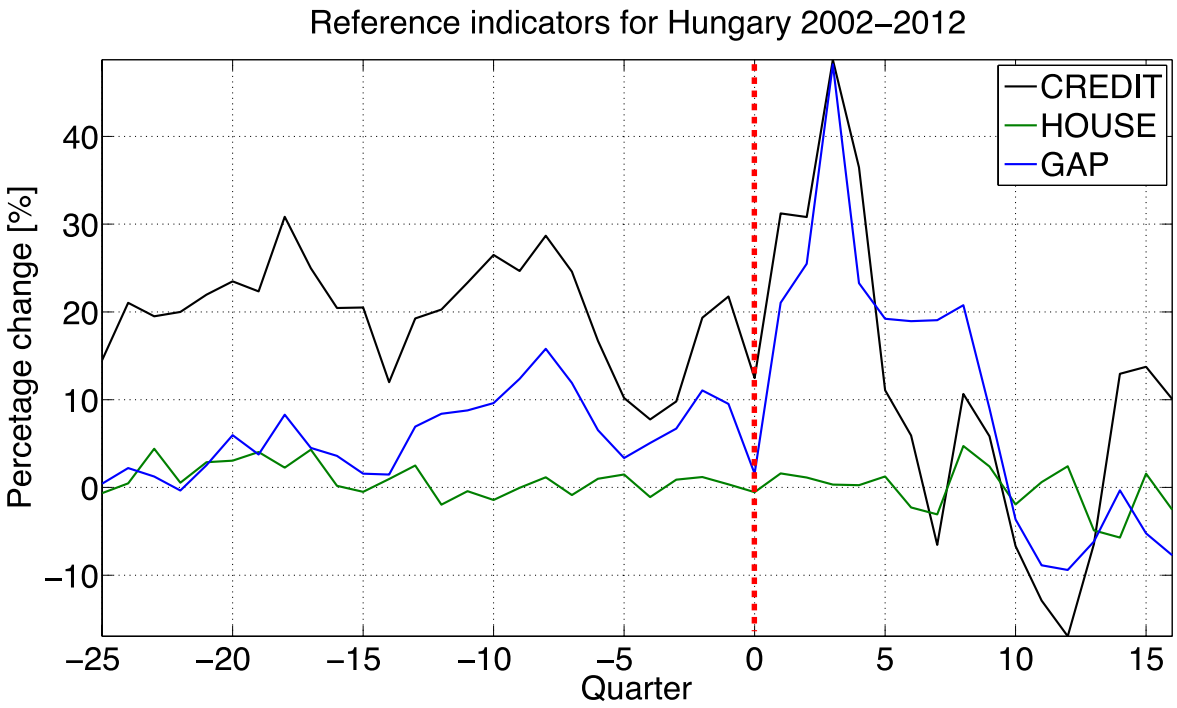

Figure 6. Input variables for Hungary 2002-2012

The neural network used for training of the fuzzy logic blocks is shown in Figure 7. Figure 8 lists the numeric results of the ANFIS training stage.

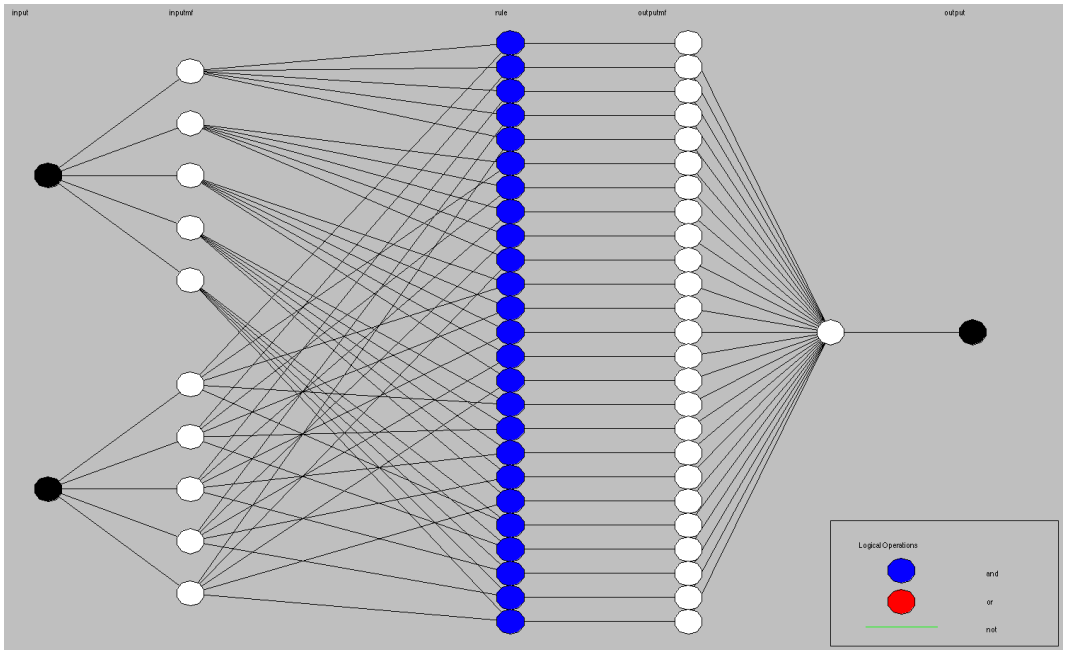

Figure 7. Neural-network based training of fuzzy systems

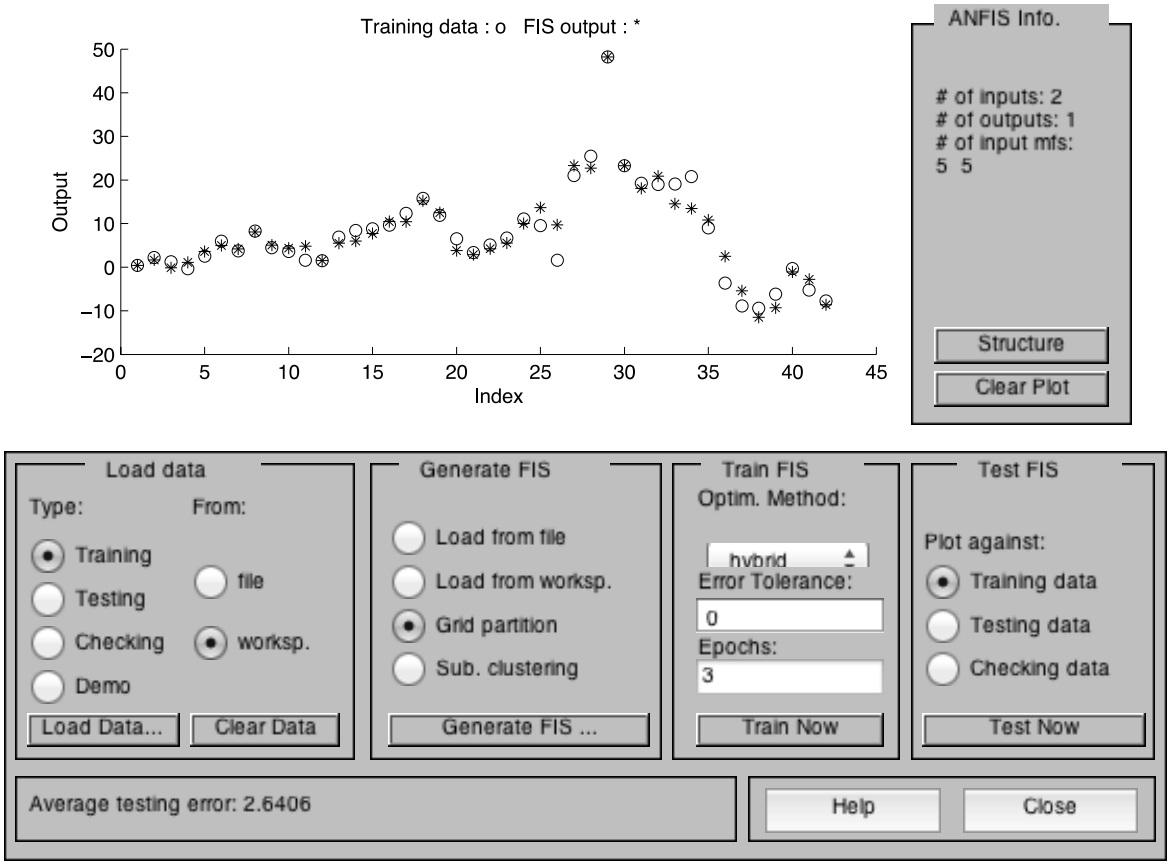

Figure 8. ANFIS training results

The final result is shown in Figure 9. It illustrates the surface view on an equivalent FZ1 fuzzy system with CREDIT and GAP as input1 and input2 respectively. An important observation is that the ANFIS modelling process generates Sugeno-type fuzzy systems which generate the output as a weighted average of the individual rule evaluation.

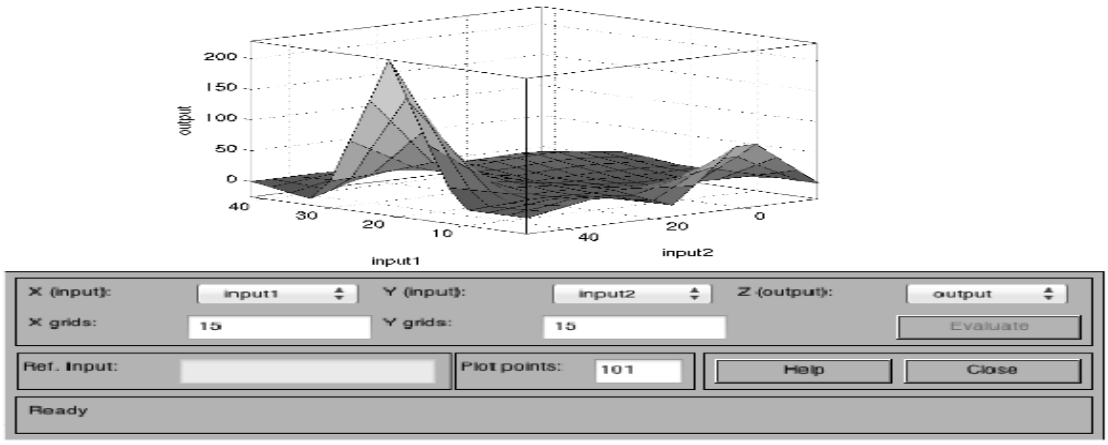

Figure 9. Surface view of Sugeno-type fuzzy system output

5. Conclusions

The paper discussed the application of fuzzy logic structures for decision support in early warning of financial crises scenarios. A multi-level fuzzy structure has been designed, implemented and validated using aggregated ECB macroeconomic indicators. Leveraging ANFIS modeling methods for fuzzy logic membership tuning has shown promising results in the scenario of early warning systems, adapted to specific socio-economic conditions of individual countries.

Main findings resulted in defining fuzzy variables and rule bases for a two level system able to offer reliable indices for financial crises forecasting. This allows incorporating expert knowledge into a flexible prediction framework which can be adjusted depending on country-level characteristics as well as the global economic environment e.g. by selecting different indicators as inputs.

Future work will be focused on developing the global model and application to Romania, subject to availability of extensive datasets to use as input variables.

References

1. Babecky, J., Havranek, T., Makju, J., Rusnak, M. and Smidkova, K., **Banking and currency crises Early Warning Indicators for Developed Countries**, European Central Bank, Working paper series, Nr. 1485, October, 2012
2. Behn, M., Detken, C., Peltonen, T. and Schudel, W., **Setting Countercyclical Capital Buffers based on early warning models would it works?**, ECB Working Paper Series, No. 1604, November, 2013
3. Belingher, D. and Bodislav, D.-A., **The Optimal Interval for the Taylor rule Appliance in Romania**, International Journal of Mathematical Models and Methods in Applied Sciences, Issue 2, Volume 5, 2012, pp. 231-238
4. Costea, A. **Applying Fuzzy C-Means and Artificial Neural Networks for Analyzing the Non-Banking Financial Institutions' Sector in Romania**, Journal of Applied Quantitative Methods, Vol. 7, No. 3, 2012, pp. 26-32
5. Dostal, P., **The Use of Soft Computing Methods for Forecasting in Business, Their Applications in Practice**, Advances in Intelligent Systems and Computing, 2012, vol. 192, no. 1, pp. 49-60
6. Dostal, P., **Forecasting of Time Series with Fuzzy Logic**, Advances in Intelligent Systems and Computing, 2013, vol. 210, no. 1, p. 155-162
7. Georgescu, F. **România – mediul international, provocari actuale si solutii**, Dissertation on the occasion of the decerning of the title of Doctor Honoris Causa of the National School for Political and Administrative Studies, Bucharest, November 21st 2013
8. Jang, J.-S.R. **ANFIS: adaptive-network-based fuzzy inference system**, Systems, Man and Cybernetics, IEEE Transactions on , vol.23, no.3, pp.665,685, May/Jun 1993
9. Peltonen, T. and Schudel, W., **Predicting Financial Vulnerabilities to guide the Set – Up of Counter – Cyclical Capital Buffers**, ECB Financial Stability Review, November, 2013, pp.118 – 128

10. Sztojanov, E., Stamatescu, G. and Sztojanov, I., **Early - warning of Financial Crisis based on Fuzzy Logic**, Proceedings of the 6th International Workshop on Soft Computing Applications – SOFA, Timisoara, 24-26 July, 2014
11. * * * **Fuzzy Logic Toolbox: User's Guide**, The MathWorks Inc., R2014b, 2014

[1] **Acknowledgements**
The work has been partially funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398.

ENDOGENOUS REGIONAL DEVELOPMENT IN ROMANIA. A KNOWLEDGE PRODUCTION FUNCTION MODEL

Zizi GOSCHIN
Professor, Department of Statistics and Econometrics,
Bucharest University of Economic Studies, Romania
Senior Scientific Researcher, Institute of National Economy

E-mail: zizigoschin@csie.ase.ro

Abstract:

Results from research - development and innovation sector, embodied in capital, are an undisputed factor of economic growth, included in most macroeconomic models. Drawing on the New Growth Theory that states the importance of R&D in all economic and social domains, as well as its key role in endogenous development, this paper is aiming to assess the nature and the impact of technological progress on the development of Romanian regions in recent years. We try to capture R&D's influence on regional economic growth by means of a knowledge production function model that employs county level data for the period 2001 to 2011. Our main finding is the positive and significant, although relatively small, contribution of technical progress (as captured by R&D expenditures) to regional GDP growth in Romania. This calls for improved regional research and development strategy, able to stimulate balanced territorial distribution of R&D and innovation activities, as well as a closer link with the business sector, in order to take advantage of the economic growth potential of regional R&D activity.

Key words: endogenous growth, R&D, Cobb-Douglas production function, region, Romania

1. Introduction

Research and development (R&D) activities are nowadays largely acknowledged as a main driver of economic growth and are routinely included in the macroeconomic models. Modern research in macroeconomic growth started from the neo-classical models, which considered that long-run growth was based on external sources and consequently viewed population, capital accumulation and technological change as exogenous factors of economic growth (e.g. Swan, 1956; Solow, 1957; Barro, 1997). In opposition to the neoclassical models, the New Growth Theory introduced the concept of endogenous growth and brought theoretical and empirical evidence in favour of human capital and innovation as factors of growth originating inside the economic system.

The delimitation between exogenous and endogenous factors of growth is relevant at regional and local levels as well. Endogenous growth originates inside the regional economy, being created by domestic private or public enterprise, while exogenous growth has external sources, outside the region. One of the main endogenous resources for regional economic growth is technical progress emerging from R&D activities. Recent

European empirical research, such as Drivera (2008) and Buesa (2010) confirmed that regional innovation is crucial for economic growth. In Romania, studies relying on Cobb-Douglas production functions, such as Zaman and Goschin (2007a), Sandu and Modoran (2008) and Zaman and Goschin (2010), revealed the positive impact of R&D expenditures on economic growth at national level, while Silaghi and Medeşfălean (2014) found an unexpected negative coefficient on patents (as proxy for innovation), possibly due to inefficiency in patenting activity. At regional level, Goschin (2014), using a panel data model, reported significant positive impact of R&D expenditures on the regional economic growth process in Romania over 1995-2010. In the same register, Nae (2013), employing Enterprise Survey data, revealed significant influence of endogenous factors like innovation on regional economic growth in Romania, while R&D is found to have only indirect impact, through its effects on patenting activity.

Drawing on the New Growth Theory that suggests the need to increase the role of R&D in all economic and social domains, as the direct source of technological progress and an important resource of economic growth, we aim to assess the nature and the impact of technological change in the development of Romanian regions. The issue is of interest for both central and local public authorities, as they should design economic policies in support of endogenous regional development. Therefore, we intend to test the theory of endogenous economic growth fuelled by innovation in Romania, using data at county level (NUTS3). To this aim we are going to employ the knowledge production function model in order to capture potential R&D influence on regional economic growth.

The remainder of this paper proceeds as follows. Next section briefly explains how exogenous and endogenous technical progress might be modelled using Cobb-Douglas production function framework. Section 3 describes the model to be employed for our county level analysis, alongside variables and data. Section 4 discusses the results and section 5 concludes.

2. The knowledge production function model

The production functions were first introduced by Cobb and Douglas (1928), who used them to test economic hypotheses related to marginal productivity and competitiveness. Solow (1957) further defined the aggregate production function including exogenous technical progress captured by the variable time, as follows:

$$Y_t = A_t K_t^{\alpha} L_t^{\beta} \ , \qquad (1)$$

where *Y* denotes the output, while A_t is a function of time which allows for neutral technical progress and *K* and *L* represent capital and labour, respectively. Differentiating the previous relation with respect to time and dividing it by Y results:

$$\frac{\dot{Y}}{Y} = \frac{\dot{A}}{A} + \alpha \frac{\dot{K}}{K} + \beta \frac{\dot{L}}{L} \qquad (2)$$

where α and β represent the share of capital and labour in the output and $\frac{\dot{A}}{A}$ is the technical progress determined as a residual.

Further developments of Solow's model allowed for more complex analyses of the effects of technical progress by including into the equation factors such as human capital, technological improvements embodied in capital, multiple sectors and so on. As a direct consequence of increasing the number of explanatory variables in the economic growth models, the share of technical progress in economic growth declined from 87.5% in Solow (1957) to about a third in more recent empirical research (Jorgenson, 1990; Denison, 1985; Matthews et al., 1982).

A new hypothesis, stating the endogenous nature of technical change, emerged in the papers of the advocates of the New Growth Theory: Lucas (1988), Romer (1990), Grossman and Helpman (1991), Aghion and Howitt (1992). In their view, growth is endogenously generated by innovations triggered by investments in research and development activity and others types of knowledge, such as human capital. Consequently, R&D was introduced in the standard Cobb-Douglas production function (e.g. Griliches, 1980; Mansfield, 1980; Scherer, 1982; Griliches and Lichtenberg, 1984) resulting the following knowledge production function model:

$$Y_t = AD_t^{\beta} K_t^{\alpha_1} L_t^{\alpha_2} e^{\lambda t} \qquad (3)$$

where Y_t is output, D_t is the stock of knowledge, L_t is the labour input, K_t is capital input, A is a constant and λ is a trend variable which catches other influences. An important result of applying the knowledge production function model is the opportunity to single out the output elasticity depending on knowledge (parameter β), which might be considered, in a broader view, a measure of social efficiency of scientific knowledge.

One difficult problem related to such models is how to separate knowledge from other production factors. Supporters of New Growth Theory explicitly modelled knowledge as an output of R&D activity and the stock of knowledge D_t was measured either as accumulated capital of R&D, as R&D flow (of expenditures, personnel, etc.) or as R&D intensity (e.g. R&D expenditures relative to turnover at microeconomic level, or relative to GDP at macroeconomic level). Based on data availability and accuracy, R&D expenditures are the most common choice.

The New Growth Theory analyses technological change in the context of economic processes (as knowledge creation is part of the current economic activity), indicating that knowledge and technology are the key factors of increasing returns and therefore the main driving forces of economic growth. The stock of knowledge generated by R&D activity is increasing marginal productivity, thus offsetting the diminishing returns of the other inputs.

Exponents of New Growth Theory also entered the human capital as a new factor of production and explained its potential for increasing returns to all factors of production (Romer, 1986; Lucas, 1988). For instance, the endogenous economic growth model of Romer (1990) is focused on four production factors: capital, labour, human capital and technology, all depending on the technological level of production. Technology is represented by a stock of manufacturing industrial models (designs) of goods, which are accumulated in time, as result of research efforts. Aghion and Howitt (1998) explained growth on the long-run in relation to constant technological progress embodied in new goods, markets and processes.

The New Growth Theory is helping to understand the ongoing change from resource–based economy to a knowledge-based economy, which has major implications for economic theory and practice.

3. Model, variables and data

We start from the New Growth Theory approach on technical progress as endogenously generated by research and development activities. Considering the advantages of Cobb-Douglas model, that made it a common choice in empirical economic growth research, we are going to employ it in order to assess the relevance of technical progress as a factor of endogenous regional development in Romania.

In our model GDP is used as the most appropriate measure of the economic development of the Romanian counties (NUTS 3 level), capital K and labour L enter the model as the traditional production factors, and R&D expenditures are added as a proxy for the endogenous growth potential of the counties (Table 1). Foreign direct investments had been used as a proxy for the production factor capital. Even if FDI data do not reflect entirely the production factor capital, they represent currently the best available information at county level.

Total expenditures are used in this model as a measure of total investments (material and intangible) in the R&D sector. The construction of the R&D data series is usually the key issue for this type of analysis. In many studies the R&D stock is calculated as the accumulated value of R&D expenditure after depreciation, procedure which implies the assumption that all research-development expenditure is accumulated with 100 percent certainty and that the R&D stock depreciates at a certain fixed rate. Since long time-series data, essential for building long time series of flow data for research and development, are rarely available, other studies assume that the growth rate of R&D flow is equal to that of R&D stock (which implies that the ratio of expenditure to stock is stable). We chose to use data on R&D expenditures instead of R&D stock, which brings about the advantage that there is no need for strong assumptions on research and development activity, such as a fixed rate of depreciation and the linear and certain accumulation of knowledge.

Table 1. Variables for the knowledge production function model

Variable	**Description**	**Data source**
GDP	Gross domestic product at county level (RON)	National Institute of Statistics (NIS) database
Capital	Foreign direct investments at county level (RON)	Romanian National Trade Register Office
Labour	Civil employment in the county economy (persons)	NIS database
R&D	County's total expenditures for research and development (RON)	NIS database

We are further going to apply the model of aggregate production functions of Cobb-Douglas type, including R&D expenditures, in the form of the standard knowledge production function model:

$$GDP_i = AK_i^{\alpha} L_i^{\beta} R_i^{\delta} \quad (4)$$

where GDP is the output (Y), α and β stand for the elasticity of output with respect to capital K and labor L, respectively (α, β> 0), A is a constant, and R represent the R&D expenditures. R&D is the variable of interest, as it captures the endogenous technological change that might impact regional economic development.

In order to estimate the model, we are going to use logarithms of the variables, as follows:

$$\ln GDP_i = \ln A + \alpha \ln K_i + \beta \ln L_i + \delta \ln R_i + \varepsilon_i \quad (5)$$

We are going to estimate the parameters of the production function, annually, for the period 2001-2011, using county level (NUTS 3) data from the National Institute of Statistics and from the Romanian National Trade Register Office. Time and space datasets have been built for GDP, foreign direct investments, employed population, total research and development expenditures, for the period 2001 to 2011 and the 42 counties of Romania. Lacking county data on capital, we used foreign direct investments as proxy.

4. Results and discussion

Results of annual parameter estimation of knowledge Cobb-Douglas production function (Table 2) clearly indicate that endogenous technical progress has had a positive and statistically significant contribution to regional economic growth in Romania, in every year of the period under consideration.

Table 2. Annual parameter estimates for knowledge Cobb-Douglas production function, 2001 to 2011

Variable	2001		2002		2003	
	Coefficient	Probability	Coefficient	Probability	Coefficient	Probability
Capital	0.070155	0.1144	0.068029	0.3084	0.021117	0.7110
Labour	0.978998	0.0000	1.014530	0.0000	1.019004	0.0000
R&D	0.056351	0.0252	0.048223	0.0695	0.083588	0.0026
Constant	1.249613	0.0167	1.439624	0.0010	1.823966	0.0000

Variable	2004		2005		2006	
	Coefficient	Probability	Coefficient	Probability	Coefficient	Probability
Capital	0.071142	0.1825	0.067390	0.2195	0.047798	0.3724
Labour	0.932758	0.0000	1.004414	0.0000	1.028033	0.0000
R&D	0.055551	0.0482	0.064180	0.0193	0.063213	0.0078
Constant	2.242087	0.0000	2.041086	0.0000	2.314369	0.0000

Variable	2007		2008		2009	
	Coefficient	Probability	Coefficient	Probability	Coefficient	Probability
Capital	0.027819	0.6371	0.089645	0.0099	0.125711	0.0000
Labour	1.093265	0.0000	0.984025	0.0000	0.945053	0.0000
R&D	0.055576	0.0164	0.038414	0.0058	0.030038	0.0336
Constant	2.402909	0.0000	2.575139	0.0000	2.412848	0.0000

Variable	2010		2011	
	Coefficient	Probability	Coefficient	Probability
Capital	0.119264	0.0194	0.105073	0.0291
Labour	0.927935	0.0000	0.935280	0.0000
R&D	0.025739	0.0683	0.035665	0.0362
Constant	2.658761	0.0000	2.748403	0.0000

The results in Table 2 show that labour had the expected positive influence on the county output and was statistically significant for all years, but the capital (proxied by FDIs) had been insignificant between 2001 and 2007 and became statistically significant since 2008. It is likely that FDI (that we only used in absence of other statistical data on capital at the county level) may not be a suitable option for capturing the production factor capital.

Our results on low but positive impact of R&D on the economic growth in Romania are in accordance with similar findings in Zaman and Goschin (2007b), Silaghi and Medeşfălean (2014), and Goschin (2014).

Of special economic interest is the analysis of the parameters of the production function, as well as the economic policy conclusions arising therefrom. Thus, the estimated parameters allow measuring the contribution of each input (K, L and R) in creating the output Y with the following relations:

- capital's contribution to growth: $\frac{\alpha}{\alpha+\beta+\delta}$,

- labour's contribution: $\frac{\beta}{\alpha+\beta+\delta}$,

- R&D's contribution: $\frac{\delta}{\alpha+\beta+\delta}$.

Based on the previous formulae, we used the estimated parameters to calculate the average contribution of each production factor to regional GDP, over the period 2001 to 2011, obtaining the following results:

- the production factor labour contributed on average by 90% to GDP creation;
- R&D expenditures explain on average 4.5% of regional GDP;
- the capital (using FDIs as proxy) had a contribution of only 5.5%, which suggests that FDIs have relatively small effects on regional economic growth in Romania.

The standard statistical tests carried out have validated the model, which has a high explanatory power (approx. 90%). The high heterogeneity of territorial distribution of the variables used in the model, especially in the case of FDIs, raised estimation problems. To fix the problem, we used White Heteroskedasticity-Consistent Standard Errors & covariance while estimating the annual models (Annexes).

In conclusion, the main result from the annual estimations of the knowledge production function model is the positive and significant, but relatively small, contribution of technical progress (as captured by R&D expenditures) to regional GDP growth in Romania. This should be a concern and alert decision makers at national and local level on economic and social policy mix needed to increase the contribution of technological progress, especially considering the current international trend towards knowledge society. R&D driven technological progress - the main factor of modern economic growth - as demonstrated by the experience of developed countries - should act more strongly in the future regional development of the Romanian economy.

5. Conclusion

Economic theory states the possibility to increase the competitiveness of regional economies and to fuel economic growth by capitalizing on local technological potential which might impact upon businesses.

As the origin of innovations and technological change, research and development is a main source of endogenous growth. We tested this hypothesis for Romanian counties and found positive and significant, although relatively small, contribution of R&D expenditures to regional GDP growth. This calls for improved regional research and development strategy, able to stimulate balanced territorial distribution of R&D and innovation activities, as well as a closer link with the business sector, in order to take advantage of the economic growth potential of regional R&D.

Post-crisis regional programs for development should target diversification of local economies by boosting private investment in R&D, adequate specialization and performance of local research, development and innovation systems, stimulation of innovative activities and technology transfer from universities and research centers to production sector, according to the business needs of local communities, assistance for the development of innovative SMEs, financial support for companies so that they can acquire advanced technologies and improve their production activity.

References

1. Aghion, P. and Howitt, P. **A Model of Growth through Creative Destruction**, Econometrica, 1992, pp .323-351
2. Aghion, P. and Howitt, P. **Endogenous Growth Theory**, Cambridge: MIT Press, 1998
3. Arrow, K. **The economic implications of learning by doing**, Review of Economic Studies, vol. 29, 1962, pp. 155-173
4. Buesa, M. **The determinants of regional innovation in Europe: A combined factorial and regression knowledge production function approach**, Research Policy, Vol.39, Iss. 6; 2010, p.722
5. Cameron, G. and Muellbauer, J. **Knowledge, Increasing Returns, and the UK Production Function**, in Mayes, D. (ed.) "Sources of Productivity Growth in the 1980s", Cambridge: Cambridge University Press, 1990
6. Cobb, C.W. and Douglas, P.H. **A Theory of Production**, American Economic Review, Vol. 18, 1928, pp.139-165
7. Drivera, C. and Oughtonb, C. **Dynamic models of regional innovation: Explorations with British time-series data**, Cambridge Journal of Regions, Economy and Society, Vol.1, Iss. 2; 2008, p. 205
8. Goschin, Z. **R&D as an Engine of Regional Economic Growth in Romania**, Romanian Journal of Regional Science, vol. 81, 2014, pp. 24-37
9. Griliches, Z. **Issues in Assessing the Contribution of Research and Development to Productivity Growth**, Bell Journal of Economics, no. 10, 1979, pp. 92–116
10. Griliches, Z. **Returns to R&D Expenditures in the Private Sector**, in Kendrick, K.W. and Vaccara, B. (eds.) "New Developments in Productivity Measurement", Chicago: University Press, 1980

11. Griliches, Z. and Lichtenberg, F. **R&D and Productivity Growth at the Industry Level: Is There Still a Relationship?**, in Griliches, Z. (ed.) "R&D, Patents and Productivity", Chicago:University of Chicago Press, 1984
12. Griliches, Z. and Mairesse, J. **Production functions: The search for identification** Econometrics and economic theory in the 20th century: the Ragnar Frisch Centennial Symposium, Cambridge University Press, 1998, pp. 169-203
13. Grossman, T. and Helpman, E. **Innovation and Growth in the Global Economy**, MIT Press, 1991
14. Hicks, J.R. **The Theory of Wages**, Second Edition 1963, St Martin's Press, New York, 1932
15. Kaldor, N. and Mirrlees, J. **A New Model of Economic Growth**, Review of Economic Studies, 1962, pp. 174-245
16. Kiefer, D.M. **Winds of Change in Industrial Chemical Research**, Chemical and Engineering News, vol. 42, 1964, pp. 88-109
17. Lichtenberg, F., R. **The Output Contributions of Computer Equipment and Personal: A Firm-Level Analysis**. Economics of Innovation and New Technology, Vol. 3, 1995, pp. 201–217
18. Lucas, R. **On the Mechanics of Economic Development**, Journal of Monetary Economics, Vol. 22, No. 1, 1988, pp. 3-42
19. Mairesse, J. **New estimates of embodied and disembodied technical progress**, Annales de l'INSEE, 30-31, 1978, pp. 681-720
20. Mansfield, E. **Social and Private Rates of Return from Industrial Innovations**, Quarterly Journal of Economics, vol.91, 1971
21. Nae, G.G. and Sima, C., **Economic Growth at Regional Level and Innovation: Is There Any Link?**, Annals of the University of Petroşani, Economics, 131, 2013, 149-156
22. Romer, P. M. **Increasing Returns and Long-run Growth**, Journal of Political Economy, University of Chicago Press, vol. 945, 1986, pp. 1002-1037
23. Romer, P. M. **Endogenous Technological Change**, Journal of Political Economy, vol. 98, 1990, pp. S71-S102
24. Solow, R. **Technical Change and the Aggregate Production Function**, Review of Economics and Statistics, vol. 39, 1957, pp. 312-20.
25. Silaghi, M. and Medesfalean, R. **Some insights about determinants of economic growth in Romania. An empirical exercise,** Theoretical and Applied Economics, No. 6, 2014, pp. 23-36
26. Stiglitz, J. **Endogenous Growth and Cycles**, Stanford University Working Paper, 1992
27. Swan, T. **Economic Growth and Capital Accumulation**, Economic Record, vol. 32, 1956, pp. 343-361
28. Zaman, G. and Goschin, Z. **Analysis of Macroeconomic Production Functions for Romania. Part one- the time-series approach**, Economic Computation and Economic Cybernetics Studies and Research, no. 1-2, vol.41, 2007, pp. 31-46
29. Zaman, G. and Goschin, Z. **Analysis of Macroeconomic Production Functions for Romania. Part two- the cross-section approach**, Economic Computation and Economic Cybernetics Studies and Research, no 3-4, vol. 42, 2007, pp. 23-32

30. Zaman, G. and Goschin, Z., **Technical Change as Exogenous or Endogenous Factor in the Production Function Models. Empirical Evidence from Romania**, Romanian Journal of Economic Forecasting, vol. 13, iss. 2, 2010, pp. 29-45
31. * * * **Database TEMPO** - time series, National Institute of Statistics 2013, https://statistici.insse.ro/shop/
32. * * * **Companies by FDI,** Romanian National Trade Register Office, Statistical Synthesis of the National Trade Register's Data, www.onrc.ro/

Annexes

Estimations from Cobb-Douglas production function including R&D, annually, 2001-2011

2001

Dependent Variable: LOG(GDP_1)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_1)	0.070155	0.043418	1.615806	0.1144
LOG(PO_1)	0.978998	0.105023	9.321703	0.0000
LOG(RD_1)	0.056351	0.024188	2.329733	0.0252
C	1.249613	0.499265	2.502905	0.0167

R-squared	0.896303	Mean dependent var	7.700897
Adjusted R-squared	0.888116	S.D. dependent var	0.585227
S.E. of regression	0.195753	Akaike info criterion	-0.333537
Sum squared resid	1.456127	Schwarz criterion	-0.168044
Log likelihood	11.00427	F-statistic	109.4840
Durbin-Watson stat	1.823299	Prob(F-statistic)	0.000000

2002

Dependent Variable: LOG(GDP_2)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_2)	0.068029	0.065889	1.032480	0.3084
LOG(PO_2)	1.014530	0.123799	8.194967	0.0000
LOG(RD_2)	0.048223	0.025819	1.867722	0.0695
C	1.439624	0.404581	3.558305	0.0010

R-squared	0.904738	Mean dependent var	7.939739
Adjusted R-squared	0.897217	S.D. dependent var	0.608348
S.E. of regression	0.195035	Akaike info criterion	-0.340880
Sum squared resid	1.445472	Schwarz criterion	-0.175388
Log likelihood	11.15849	F-statistic	120.2995
Durbin-Watson stat	1.736207	Prob(F-statistic)	0.000000

2003

Dependent Variable: LOG(GDP_3)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_3)	0.021117	0.056574	0.373261	0.7110
LOG(PO_3)	1.019004	0.121554	8.383132	0.0000
LOG(RD_3)	0.083588	0.025963	3.219476	0.0026
C	1.823966	0.369009	4.942876	0.0000

R-squared	0.920624	Mean dependent var	8.208860
Adjusted R-squared	0.914357	S.D. dependent var	0.598785
S.E. of regression	0.175233	Akaike info criterion	-0.555006
Sum squared resid	1.166853	Schwarz criterion	-0.389513
Log likelihood	15.65512	F-statistic	146.9106
Durbin-Watson stat	2.034236	Prob(F-statistic)	0.000000

2004

Dependent Variable: LOG(GDP_4)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_4)	0.071142	0.052391	1.357897	0.1825
LOG(PO_4)	0.932758	0.088584	10.52968	0.0000
LOG(RD_4)	0.055551	0.027212	2.041412	0.0482
C	2.242087	0.327385	6.848469	0.0000

R-squared	0.934848	Mean dependent var	8.439265
Adjusted R-squared	0.929704	S.D. dependent var	0.591650
S.E. of regression	0.156867	Akaike info criterion	-0.776450
Sum squared resid	0.935070	Schwarz criterion	-0.610958
Log likelihood	20.30545	F-statistic	181.7495
Durbin-Watson stat	1.944210	Prob(F-statistic)	0.000000

2005

Dependent Variable: LOG(GDP_5)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_5)	0.067390	0.053983	1.248365	0.2195
LOG(PO_5)	1.004414	0.095893	10.47434	0.0000
LOG(RD_5)	0.064180	0.026275	2.442665	0.0193
C	2.041086	0.339542	6.011286	0.0000

R-squared	0.927854	Mean dependent var	8.548619
Adjusted R-squared	0.922158	S.D. dependent var	0.640156
S.E. of regression	0.178604	Akaike info criterion	-0.516895
Sum squared resid	1.212182	Schwarz criterion	-0.351402
Log likelihood	14.85479	F-statistic	162.9033
Durbin-Watson stat	1.836456	Prob(F-statistic)	0.000000

2006

Dependent Variable: LOG(GDP_6)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_6)	0.047798	0.052955	0.902621	0.3724
LOG(PO_6)	1.028033	0.103005	9.980379	0.0000
LOG(RD_6)	0.063213	0.022483	2.811533	0.0078
C	2.314369	0.317979	7.278366	0.0000

R-squared	0.934495	Mean dependent var	8.733498
Adjusted R-squared	0.929323	S.D. dependent var	0.636223
S.E. of regression	0.169141	Akaike info criterion	-0.625780
Sum squared resid	1.087125	Schwarz criterion	-0.460287
Log likelihood	17.14137	F-statistic	180.7017
Durbin-Watson stat	1.954964	Prob(F-statistic)	0.000000

2007

Dependent Variable: LOG(GDP_7)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_7)	0.027819	0.058503	0.475520	0.6371
LOG(PO_7)	1.093265	0.109485	9.985532	0.0000
LOG(RD_7)	0.055576	0.022138	2.510424	0.0164
C	2.402909	0.346065	6.943527	0.0000

R-squared	0.936400	Mean dependent var	8.909312
Adjusted R-squared	0.931379	S.D. dependent var	0.650643
S.E. of regression	0.170440	Akaike info criterion	-0.610470
Sum squared resid	1.103897	Schwarz criterion	-0.444978
Log likelihood	16.81987	F-statistic	186.4938
Durbin-Watson stat	1.837007	Prob(F-statistic)	0.000000

2008

Dependent Variable: LOG(GDP_8)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_8)	0.089645	0.032999	2.716616	0.0099
LOG(PO_8)	0.984025	0.073496	13.38889	0.0000
LOG(RD_8)	0.038414	0.013142	2.923048	0.0058
C	2.575139	0.337103	7.639030	0.0000

R-squared	0.957576	Mean dependent var	9.105453
Adjusted R-squared	0.954227	S.D. dependent var	0.645535
S.E. of regression	0.138110	Akaike info criterion	-1.031136
Sum squared resid	0.724829	Schwarz criterion	-0.865644
Log likelihood	25.65386	F-statistic	285.9064
Durbin-Watson stat	2.059713	Prob(F-statistic)	0.000000

2009

Dependent Variable: LOG(GDP_9)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_9)	0.125711	0.024451	5.141311	0.0000
LOG(PO_9)	0.945053	0.072679	13.00319	0.0000
LOG(RD_9)	0.030038	0.013624	2.204771	0.0336
C	2.412848	0.314898	7.662311	0.0000

R-squared	0.952710	Mean dependent var	9.090880
Adjusted R-squared	0.948976	S.D. dependent var	0.642325
S.E. of regression	0.145091	Akaike info criterion	-0.932518
Sum squared resid	0.799953	Schwarz criterion	-0.767026
Log likelihood	23.58288	F-statistic	255.1828
Durbin-Watson stat	1.892736	Prob(F-statistic)	0.000000

2010

Dependent Variable: LOG(GDP_10)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_10)	0.119264	0.048840	2.441908	0.0194
LOG(PO_10)	0.927935	0.091306	10.16288	0.0000
LOG(RD_10)	0.025739	0.013719	1.876198	0.0683
C	2.658761	0.410253	6.480785	0.0000

R-squared	0.924287	Mean dependent var	9.130794
Adjusted R-squared	0.918310	S.D. dependent var	0.640845
S.E. of regression	0.183163	Akaike info criterion	-0.466485
Sum squared resid	1.274853	Schwarz criterion	-0.300993
Log likelihood	13.79619	F-statistic	154.6317
Durbin-Watson stat	1.612135	Prob(F-statistic)	0.000000

2011

Dependent Variable: LOG(GDP_11)
Included observations: 42
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(ISD_11)	0.105073	0.046332	2.267846	0.0291
LOG(PO_11)	0.935280	0.096466	9.695392	0.0000
LOG(RD_11)	0.035665	0.016428	2.170961	0.0362
C	2.748403	0.482757	5.693135	0.0000

R-squared	0.919890	Mean dependent var	9.178704
Adjusted R-squared	0.913565	S.D. dependent var	0.644287
S.E. of regression	0.189419	Akaike info criterion	-0.399319
Sum squared resid	1.363421	Schwarz criterion	-0.233827
Log likelihood	12.38570	F-statistic	145.4490
Durbin-Watson stat	1.543762	Prob(F-statistic)	0.000000

IDENTIFYING HIGH POTENTIAL BIOMASS-HOME-HEATING CUSTOMERS: A BAYESIAN CLASSIFICATION[1]

Adee ATHIYAMAN
Illinois Institute for Rural Affairs,
Western Illinois University, Macomb, USA

E-mail: A-Athiyaman@wiu.edu

Abstract:

This paper presents a Bayes' classification computer application that would help biomass managers optimize their marketing decision making. Programmed in Mathematica, this decision tool would help managers understand the size of the high potential market at the US county level: number of households that would be receptive to a telemarketing / direct marketing campaign about pellet heating appliances, for example.

Key words: Market Segmentation, Biomass, Residential Heating, Interactive App, Decision Theory

Introduction

A recent McKinsey Global Survey of marketing executives suggests that the most pressing competitive challenge for the executives is gaining customer insights to drive sales (Davis and Freundt, 2011). This is not surprising given that the present-day business environment consists of:

- ✓ multiple customer segments – our multicultural society and the more divided income groups have created multiple customer segments;
- ✓ multitude sub-brands and line extensions that target these segments;
- ✓ multiple media (for example, web banners, magazine ads, and Facebook pages), and
- ✓ multiple distribution touch points (Internet, product re-sellers, big-box retailers, third-party tele-sales providers, etc.).

This complex environment calls for new marketing capabilities such as data-management and analytics - skills that marketing executives say that their companies do not have (Breuer et al, 2013). While it is not uncommon for companies to outsource analysis, it is now well known that managers seldom act on numbers that they don't fully understand (Wierenga, 2002).

In this paper, we outline a relatively straight-forward computer program that would help biomass businesses make better decisions about customers. We believe that the biomass industry needs such "technical" assistance since only a mere 10% of the industry's market potential has been tapped as at date (Athiyaman, 2014).

Classification Algorithms

Linear discriminant function (LDA)

LDA is often employed to discriminate among groups of objects such as customers and products (Sharma, 1996). It is similar to a regression function with a nominal dependent variable:

$y = \lambda_1 x_1 + \cdots + \lambda_n x_n$, where, the λs are weights to be applied to the *x* properties of the objects.

For two groups, the λs are found by maximizing an objective function such as the following:

$$MaxG = \frac{(\overline{y_1} - \overline{y_2})^2}{\sum_{\substack{i=1,2 \\ j=1,..,n}} (y_{ij} - \overline{y_i})^2}$$

The index number *y* is used to allocate a new object to be classified to the class whose value of the function is the closer.

Problems with the use of LDA stem from its stringent measurement requirements - continuous or ratio scale property of the object being measured is presupposed (Lilian and Rangaswamy, 2004). In marketing, information on customer characteristics (for example, home ownership, gender, etc.) is essentially nominal; discrete measures in statistical terminology (Maddala, 1986). Hence, it is our contention that the Bayesian classification procedure should be employed for classification tasks in marketing.

Classification based on Probability Axioms

We illustrate the procedure using data derived from a survey of 25,000 households conducted during late 2014 (see Athiyaman (2015b) for details about the survey). Our interest is in building a customer segmentation scheme for the biomass residential heating industry. Specifically, we want to segment households in the US into "high" versus "low" potential customers.

The survey data suggests the following attributes as important determinants of purchase:

Dichotomous Attribute	Correlation with Purchase Intention
x_1 –Attitude	0.66 (Negative = 0; Positive = 1)
X_2 – Biomass heating is low cost	0.44 (No = 0; Yes = 1)
X_3 – Education of the head of household	0.41 (High School = 0; College = 1)
X_4 – Year Home was Built	-0.38 (1979 or older = 0; 1980 and newer = 1)

For a dichotomous purchase intention measure (high likelihood of purchase (H) versus low likelihood (L)), the probabilities of the two groups having or not having each of the above attributes are shown in Table 1. To elaborate, in Table 1 the attribute pattern *X* = 1 includes all four dichotomous attributes each with level = 1. The reverse is true for attribute pattern *X* = 16: each of the four attributes has level = 0.

Table 1. Attribute Possession and Conditional Probabilities: Estimates from Survey Responses

	Attribute Patterns (levels)				Conditional Probability	
X	x_1	x_2	x_3	x_4	***P*(*X*|High)**	***P*(*X*|Low)**
1	1	1	1	1	0.28	0.15
2	1	1	1	0	0.42	0.33
3	1	1	0	1	0.06	0.01
4	1	1	0	0	0.11	0.05
5	1	0	1	1	0.04	0.06
6	1	0	1	0	0.05	0.05
7	1	0	0	1	0.01	0.01
8	1	0	0	0	0.02	0.02
9	0	1	1	1	0.00	0.07
10	0	1	1	0	0.01	0.04
11	0	1	0	1	0.00	0.02
12	0	1	0	0	0.00	0.01
13	0	0	1	1	0.00	0.06
14	0	0	1	0	0.00	0.11
15	0	0	0	1	0.00	0.01
16	0	0	0	0	0.00	0.01

Note: Total number of responses = 2717

Table 1 suggests that 99% of all customers classified as "high potential" possess positive attitude towards biomass heating. Furthermore, 42% of all "high potential" customers are educated beyond high school, live in older homes, and believe that biomass heating is low cost. In fact, the age of home matters as much as 14 points in predicting customer's purchase likelihood; older the home, higher is the purchase probability. How could a biomass company make use of this information to fine-tune its marketing?

Decision Aid

Assume that a biomass company is planning to advertise nationally. To minimize opportunity loss, the firm employs decision theory. Specifically, based on US Census data (American Community Survey (ACS) Selected Housing Characteristics) the firm estimates the prior probability of a randomly chosen household being classified as high sales potential, and applies this number to opportunity loss estimates to gain insights into expected payoffs (Table 2).

Table 2. Opportunity Losses: Household Level Illustration

Decision	State of Household		Opportunity Loss
	High Potential	**Low Potential**	
D_1 – Advertise	\$0	\$42	$E(Loss) = \$0 \times .1 + \$42 \times .9 = \$37.8$
D_2 – Do not advertise	\$1500	\$0	$E(Loss) = \$1500 \times .1 + \$0 \times .9 = \$150$

Note:

(i) ACS estimates of priors: proportion of households using biomass for heating = 0.1;

(ii) Advertising to a low potential household will cost the company \$42. This estimate was derived from averaging national telemarketing costs and national direct marketing costs per household (source: WebpageFX);

(iii) Not advertising to high potential is assumed to cost the company, actually the industry, sale of a pellet stove which is estimated at \$1500; "top-line" or revenue loss for the company from lost sale for a stove is \$1500 (US Environmental Protection Agency).

If the company were to choose an act or decision on the basis of information given in Table 2, then action "D_1 – Advertise" will be chosen (the company is minimizing the

maximum loss). However, if households could be screened for possession of attribute patterns given in Table 1, then a much more precise decision could be made about advertisements.

To elaborate, the joint probabilities of each of the 16 attribute patterns given in Table 1 can be computed by weighting the prior probabilities of **H**igh/**L**ow potential (p=0.1 / p=0.9) with the conditional probabilities given in Table 1. Since "**H**" and "**L**" constitute a partition over the set of households, the marginal probabilities represent the sums of the joint probabilities (Green, 1964). Finally, the posterior probabilities are the results of the application of Bayes' theorem:

$$p(H|X) = \frac{p(H)\times p(X|H)}{p(H)\times p(X|H)+p(L)\times p(X|L)}$$

Based on the posterior analysis, the company can determine which segments of households should be targeted by advertisements. The critical posterior probability that informs us of the relevant attribute pattern(s) to target is:

$$C^* = \frac{O\text{pportunity Loss } (\mathrm{D}_1)}{Opportunity\ Loss\ (D_1) + Opportunity\ Loss\ (D_2)}$$

In other words, the decision rule for advertising is:

$$If\ p(H|X) > C^*, take\ action\ D_{1;}$$
$$If\ p(H|X) < C^*, take\ action\ D_{2;}$$
$$If\ p(H|X) = C^*, take\ either\ action\ D_1\ or\ D_2$$

Interactive Application

Figure 1 shows the interactive web application that implements the marketing decision model. It is available online at www.instituteintelliegence.com. To illustrate, suppose that a biomass company wants to see the number of households in McDonough County, Illinois, that would be receptive to a telemarketing / direct marketing campaign about pellet heating appliances. In the starting page, the analyst / decision maker will specify the county that is of interest – McDonough in this instance. This will result in a listing of number of owner-occupied households in the county. Then, the analyst will specify the attribute pattern that is of interest: assume that it is one of the first six rows of Table 1: "1, 0, 1, 0", for example. This specification would result in a listing of the proportion / number of households in the county that are associated with the given pattern (the home owner has positive attitude towards biomass heating, believes that it is inexpensive, has a high school education, and lives in a home that is classified as "new"). In addition, the app will also specify the number of households that satisfy the decision criterion,

$$If\ p(H|X) > C^*, take\ action\ D_1\ (advertise)$$

Appendix 1 shows that that this classification procedure or decision model yields 93% savings to the company than the "generic" decision theory framework given in Table 2. Although privacy concerns prevent us from specifying the addresses of these households, the

app is being developed to "map" the geographical location of majority of these households (for example, one or more of the 10 census tracts in McDonough County).

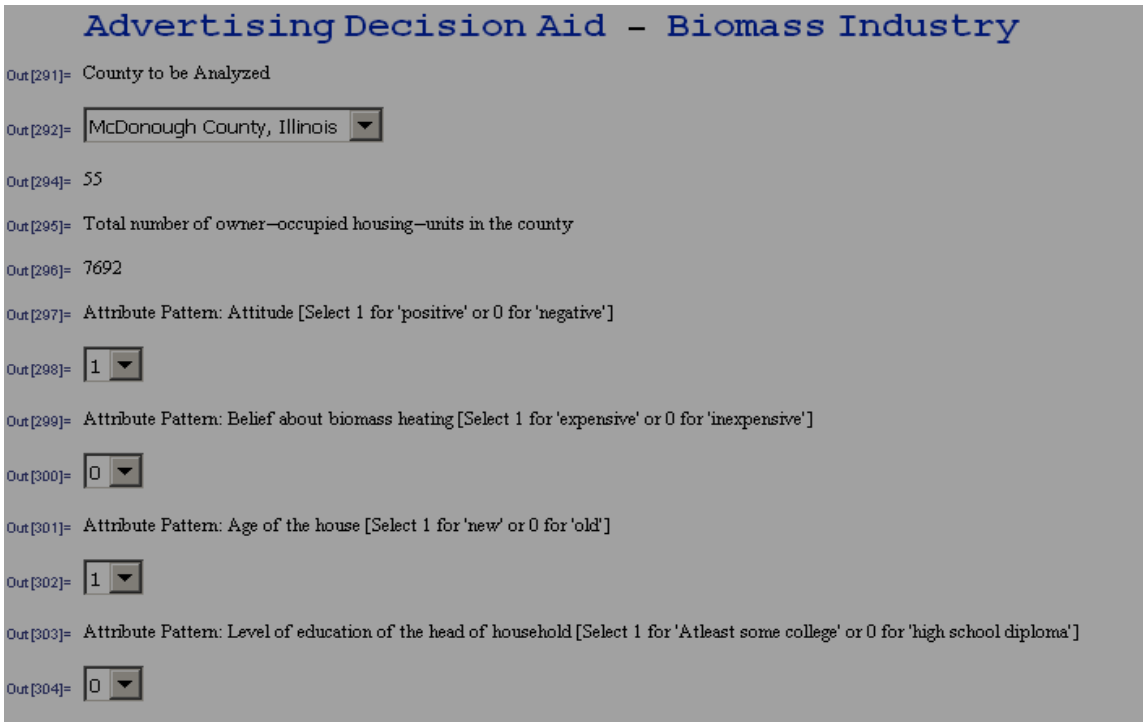

Figure 1. The Classification Application

Summary and Conclusion

Marketing in this age of micro segmentation requires skills such as knowledge of distributed computing and machine-learning techniques. Most companies do not possess these skills hence the demand for "canned" software that would aid marketing decision making. This paper presents a Bayes' classification computer application that would help biomass managers optimize their marketing decision making. This decision tool would help managers understand the size of the high potential market at the county level: number of households that would be receptive to a telemarketing / direct marketing campaign about pellet heating appliances, for example.

Earlier, Athiyaman (2015a) highlighted that company induced "push" marketing influences product "purchase" or "closure", albeit at the final stages of customer decision sequence. Given this finding, it is essential that the biomass industry utilizes analytical tools such as the one presented in this paper for marketing-program optimization. As aptly observed by Aron and van den Direst (2014), use of data-driven decision tools is a necessary condition for being successful in todays' marketplace.

References

1. Arons, M.D.S., van den Driest, F. and Weed, K. **The ultimate marketing machine,** Harvard Business Review, Vol. 92, July/August 2014, pp. 54-63
2. Athiyaman, A. **Tapping Unstructured Data to Gain Insights into Consumer Purchases of Biomass Residential Heating Appliances: An Exercise in**

Big Data Analytics, Biomass Research Report 3, Working Paper, Macomb, IL: Institute Intelligence, 2015a

3. Athiyaman, A. **Analysis of Consumer Actions: A Descriptive Study of the Biomass Residential Heating Industry,** Biomass Research Report 5, Working Paper, Macomb, IL: Institute Intelligence, 2015b
4. Athiyaman, A. **Market Potential for Residential Biomass Heating Equipment: Stochastic and Econometric Assessments,** Biomass Research Report 1, Working Paper, Macomb, IL: Institute Intelligence, 2014
5. Breuer, P., Forina, L. and Moulton, J. **Beyond the Hype: Capturing Value from Big Data and Advanced Analytics,** McKinsey & Company, 2013
6. Bogdanović, J., Vladana Lilić, A. and Gijić, N. **Implementation Of Advanced Analytics In Consumer-facing Companies**, International Journal of Economics & Law, Vol. 10, 2014, pp. 35-42
7. Davis, C. and Freundt, T. **What Marketers Say About Working Online, McKinsey Global Survey Results,** McKinsey Quarterly, November, 2011
8. Green, P. E. **Bayesian Classification Procedures in Analyzing Customer Charcateristics,** Journal of Marketing Research, May 1964, pp. 44-50
9. Lilien, G.L. and Rangaswamy, A. **Marketing Engineering: Computer-assisted Marketing Analysis and Planning,** DecisionPro, 2004
10. Lilien, G.L., Rangaswamy, A., Van Bruggen, G.H. and Wierenga, B. **Bridging the Marketing Theory–Practice Gap with Marketing Engineering,** Journal of Business Research, Vol. 55, 2002, pp. 111-121
11. Maddala, G.S. **Limited-dependent and Qualitative Variables in Econometrics,** No. 3, Oxford: Cambridge University Press, 1986
12. Sharma, S. **Applied Multivariate Techniques,** New York: John Wiley & Sons, 1996
13. Wierenga, B. **On Academic Marketing Knowledge and Marketing Knowledge that Marketing Managers Use for Decision-making,** Marketing Theory, Vol. 2, No. 4, 2002, pp. 355-362
14. * * * **DP04: Selected Housing Characteristics"American Community Survey,** Accessed February 6, 2015. http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_5YR_DP04&prodType=table
15. * * * **Pellet Stove Fact Sheet,** US Environmental Protection Agency, Accessed February 6, 2015, http://www.epa.gov/burnwise/pdfs/PelletStoveFS08-04-11.pdf
16. * * * **The Cost of Advertising Nationally Broken Down by Medium,** WebPageFX, Accessed February 6, 2015, http://www.webpagefx.com/blog/business-advice/the-cost-of-advertising-nationally-broken-down-by-medium/

Appendix 1.

The Rationale for Using Customer Attribute Patterns in Marketing Decisions

In this section, we empirically demonstrate that taking action based on a posterior analysis for each subgroup of sales prospects is optimal than deciding on a prior analysis alone (Table A1).

Table A1. Utility or Benefits of Analyzing Customer Attribute Patterns

Attribute Pattern "X" (See Table 1 in Text)	Posterior: P(H|X)	P(L|X)	Marginal P(X)	Expected Opportunity Loss (D2: Do not advertise)	Expected Opportunity Loss (D1: Advertise)	Optimal Decision
1	0.173484	0.826516	0.16	41.86046512	5.584091	D1
2	0.125185	0.874815	0.34	63.6627907	12.45682	D1
3	0.362438	0.637562	0.02	8.720930233	0.429545	D1
4	0.193563	0.806437	0.06	16.56976744	1.932955	D1
5	0.06747	0.93253	0.06	6.104651163	2.3625	D1
6	0.102088	0.897912	0.05	7.848837209	1.932955	D1
7	0.102088	0.897912	0.01	0.872093023	0.214773	D1
8	0.131638	0.868362	0.02	3.488372093	0.644318	D1
9	0.00	1	0.06	0	2.577273	D2
10	0.015983	0.984017	0.04	0.872093023	1.503409	D2
11	0	1	0.02	0	0.644318	D2
12	0	1	0.01	0	0.214773	D2
13	0	1	0.05	0	2.147727	D2
14	0	1	0.10	0	4.295455	D2
15	0	1	0.01	0	0.429545	D2
16	0	1	0.01	0	0.429545	D2

The opportunity loss of $339 is 45% less than the decision based on a prior analysis alone (expected opportunity loss for a prior-only decision is $618).

[1] **Acknowledgement**
State funds for this project were matched with Federal funds under the Federal-State Marketing Improvement Program of the Agricultural Marketing Service, U.S. Department of Agriculture.

A PROPOSED CONTEXTUAL EVALUATION OF REFERENDUM QUORUM USING FUZZY LOGICS[1]

Ovidiu TURCOANE[2]
PhD Candidate, Department of Economic Informatics,
Bucharest University of Economic Studies, Bucharest, Romania

E-mail: turcoaneovidiu06@stud.ase.ro; ovidiuturcoane@yahoo.com

Abstract:

This conceptual article proposes a new approach for referendum quorum size calculation, based on criteria that do not take into account voters' preferences or levels of information. Although some political commissions and plenty of past researches support the absence of a quorum in a referendum, the herein model relies on voting exogenous variables, referring some objective criteria of quorum calculation. A Mamdani fuzzy inference system is used to build a controller that yields the value of the output based on three inputs: Type, Discrepancy and Age. The results obtained through elementary simulations are of a wide range, from facilitating to obstructing the usage of a referendum depending on the context.

Key words: quorum; referendum; fuzzy logic; fuzzy sets; Mamdani fuzzy inference system

1. Introduction

The etymology of word *referendum* leads to the gerund of Latin verb *referre*, thus implying an action of restoring, of turning back to the people's opinion. An exact definition is that of direct voting that establishes the community, region or country electorate's will in a rather contextual matter. Examples are the constitution modification or the sovereignty debate (when a referendum becomes a plebiscite) that are forms of participative or direct democracy (DD).

DD becomes more prominent nowadays, especially from the perspective of using the institution of referendum. Although skeptics are afraid of voter's lack of information, high costs or the cleavage between mass and elites, the economic perspectives are encouraging for a functional DD (Matsusaka 2005). Low turnout in case of a referendum is sometimes an important issue, especially after excessive use of this instrument, but not as important as the problem of informed citizens in Switzerland of 1981-1999, a state with great appetite for referendums, where some tedious manifestations appeared because of the lack of information (Lutz 2007). In the same state and similar period, there were identified some aspects of DD that have an economic positive impact: preliminary debates lead to better information, the citizens take greater responsibility and accept a relaxation of their principles, and, as a direct impact, tax evasion is lower while some public services are significantly improved (Feld and Kirchgassner 2000). In an inter-country DD analysis, the

impact over fiscal policies and government efficiency is reconfirmed, but not the over productivity and individual happiness self-perception; in exchange it is emphasized the need for guiding referendums towards contextual problems (Blume, Müller and Voigt 2009).

The issue of the referendum quorum (RQ), meaning establishing a relative (percentage) or absolute (number) level for validating a referendum by representative participation, is an important topic of the recent or old DD debates. Subject of specific political decisions paradoxes, belonging to the class of composition and decomposition (Nurmi 1998), RQ is not recommended for two reasons (Venice Commission 2007): i) "a turn-out quorum (threshold, minimum percentage), because it assimilates voters who abstain to those who vote no" and ii) "an approval quorum (approval by a minimum percentage of registered voters), since it risks involving a difficult political situation if the draft is adopted by a simple majority lower than the necessary threshold". To prevent the bias towards acceptance, the result of the non-valid referendum should take rejection into account, while the settling of RQ should encompass the collective referendums' memory and relate to the paradox that the turn-out may exceed a non-imposed quorum. Moreover, a non-optimal vote relying on tendencies and preferences of citizens is preferred when targeting the result instead of the turn-out (Zwart 2010). Reclaiming the idea of quorum paradox, Herrera and Matozzi (2010) propose settling it to a level that is half of the desired turn-out target. If there is an established RQ to assuring legitimacy for the referendum's political decision, then it has negative effects on turn-out and representation rather than promoting absenteeism; it opposes status-quo, hardly disadvantaging minorities, and it facilitates pressure over voters in an undemocratic way (Aguiar-Conraria and Magalhães 2010).

With all these many scientific and socio-politic (e.g. Venice Commission) advocacies for not imposing an RQ, the commitment to support the contrary ought to be questioned. There are three reasons to defend an RQ: i) legitimacy and preventing the tyranny of minority, the need for representation; ii) context and iii) objective criterion, detached from voting preference, intention or turn-out. Proposals for abandoning RQ are based on citizen's inclination to participate or not and on voter's preference for yes or no. On the contrary, RQ should be established on objective and exogenous criteria, aiming to prevent referendum's trifling or abuse in non-consolidated states subjects to the rule of law, see pre-Nazis Germany (Zurcher 1935) or, in a more recent time, see post-communist contemporaneous Romania (Scheppele 2012).

RQ is more important in the context of E-democracy, where online petitions are solved using E-referendum. Discussions on this issue and a conceptual model have been presented in a previous paper (Turcoane 2014b).

This paper is structured as follows: section 2 briefly presents the Mamdani fuzzy inference system (MFIS) relying on a human-like decision making controller; the herein introduced model for evaluation of RQ is discussed in section 3; results and examples are in section 4 and the last section presents the conclusions of this conceptual exploration.

2. Fuzzy logic and MFIS

Fuzzy logic (FL) is an extension of the binary logic, which, from a mathematical point of view, works with two values: 0 and 1. FL is a multi-value logic that uses the whole range between 0 (nothing, absence) and 1 (absolute, certain thing) to negotiate with problems that naturally deal, under the constraint of imprecision and well-defined criteria,

with classes of values and possibility rather than with random variables (Zadeh 1965). FL have been already used in social (Montero 2008) or political (Nurmi and Kacprzyk 2007) sciences. The role of uncertainty is discussed using propositional logic from political perspective (Sen 2009) or even under fuzzy aspects in social systems (Treadwell 1995). In order to build a decisional model (MFQ) for settling RQ based on FL, it will be used an MFIS that usually guides technological processes formalized in an imprecise linguistic-human manner. MFIS gives remarkable results in the case of vaguely defined problems, as the controversial issue of RQ. There are also examples of using an MFIS in supply chain management (Ayadi, Cheikhrouhou and Masmoudi 2013) or in web shopping analysis (Liu, Geng and Zhang 2005).

Figure 1 illustrates the steps on an MFIS (Mamdani and Assilian 1975, MathWorks 2015), which is based on the conversion of input ordinary crisp values to fuzzy values that are part of fuzzy sets and which, filtered through a knowledge base given by fuzzy rules (FR), are transformed and aggregated in a final fuzzy set that is in the end subject to re-conversion to a crisp result.

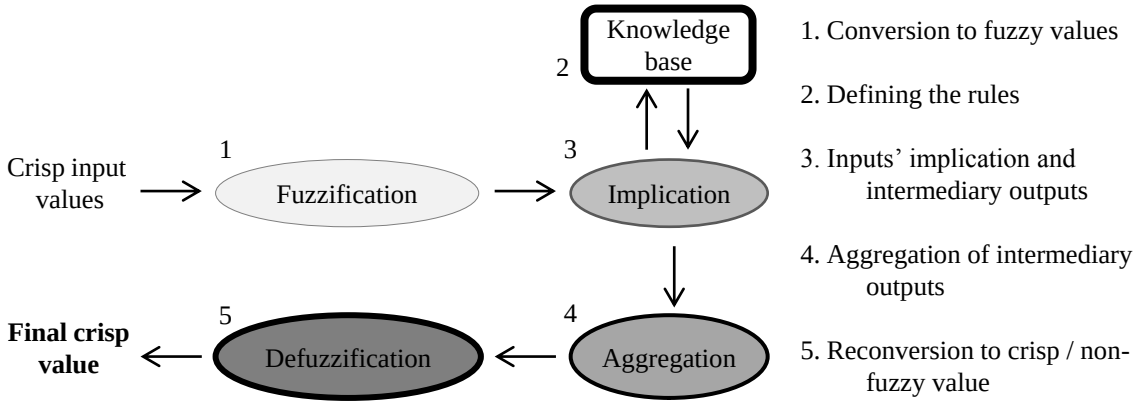

Figure 1. MFIS Without insisting on further details of MFIS formalization, see (MathWorks 2015), section 3 will build and explain MFQ in a practical manner, thus exemplifying the inferential system briefly and visually introduced in Figure 1.

3. MFQ

Prior to build an MFIS, there must be identified the inputs and output, and there must be created their fuzzy sets (FS). FS are an extension of the classic logic sets and they are described through membership functions (usually denoted by μ) that take values in the range of [0;1] or [0%;100%]. Any of the MFQ inputs or output will be built as a set made out of other fuzzy (sub)sets that define the variables of the model using a linguistic and qualitative-quantitative approach. While the output is easily identifiable as the (dimension of) Quorum (i.e. RQ), the inputs are far more difficult to be established. Apart from previous researches, see section 1, not the pattern of the voter is the key to determining the inputs, but some electorate exogenous variables that are subject to a flexible approach depending on the context.

This paper proposes three inputs for MFQ: Type, Discrepancy and Age. While the inputs may not be found reasonable and viable to dictate the value of RQ, although they may be found in different forms in Venice Commission's code for referendums (2007), the author believes that at least the model may give incentives to others for further explorations

in this area. Let us explain each of the inputs, using examples to better understand MFQ.

Type represents the class, kind or category of law, decision, decree (generic denoted as draft) that is subject of the referendum and it has two subtypes: proposal and abrogation. The former materializes a pure original political statement that is supposed to be enacted; the latter embodies a radical contestation of a functioning / active draft. It is true that labeling one of the two subtypes in a crisp (i.e. binary) way may not be difficult using classification algorithms or plain human approach (e.g. by an appointed organism). However, the classification of the draft should be based on some computational approaches when seeking for accuracy and reliability, especially for the fuzzy way (i.e. the draft is neither proposal nor abrogation, but somewhere in between). There already exists enough literature in this area, from binary classification based on prior knowledge and support vector machines (Lauer and Bloch 2008) or based on knowledge sets (Orchel 2015) to fuzzy concepts in text classification (Li and Tsai 2013). The Venice Commission (2007) also discusses the problem of presenting texts in various forms, but valid from unity point of view; these make easier the draft's classification.

From this perspective, Figure 2 illustrates the FS of Type, which allows a binary and as well a fuzzy representation of the draft. If proposal is 0 then abrogation is 1 and vice versa; both are half present at the set interval's middle, i.e. μproposal(0.5) = μabrogation(0.5) = 0.5. The scale suggested for Type is 0–1, but it could be 0–100 or 0%–100%. Actually, the scale for the inputs or output should be intuitive and simple and this affects the model only relatively.

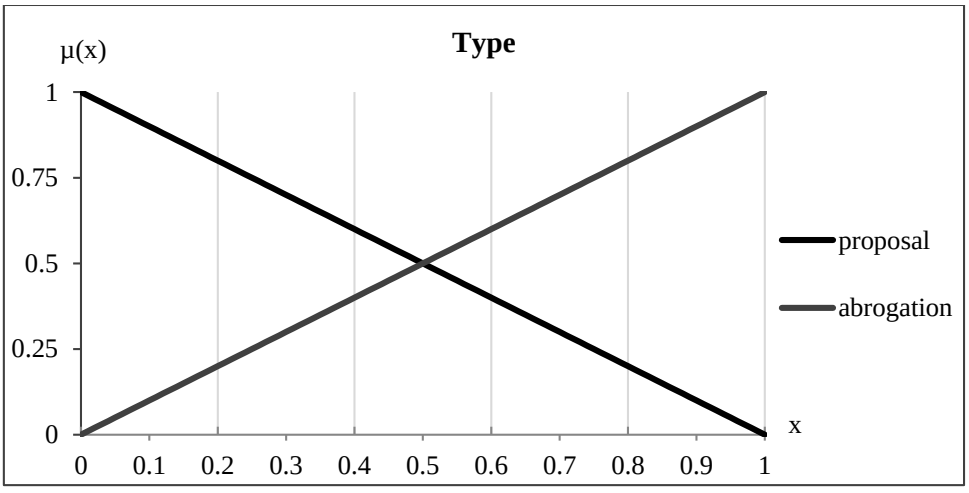

Figure 2. FS of *Type*

An example would be the intention of the European Commission or some other supranational body to legislate a *draft*, the newly digital decentralized currencies (e.g. Bitcoin). Although there is some guidance provided in US by the treasury or other exchange commissions, legislation in this area would practically be classified as a pure proposal. However, there are some laws and rules regarding digital economy and they may be somehow or partially abrogated by the new *draft*. The author doubts that a pure proposal exists, given the fundamentals of any society, global or not. An abrogation example is the impeachment and dismiss of a representative such as a mayor or a president.

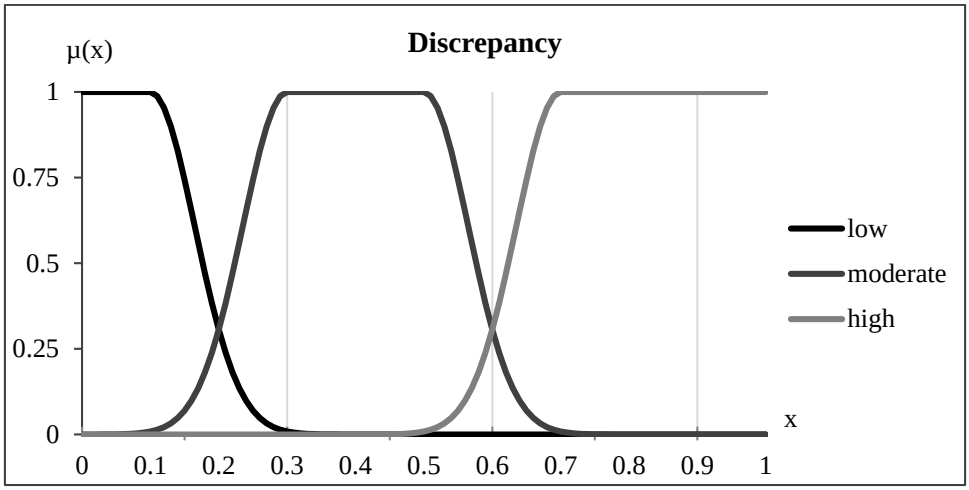

Figure 3. FS of *Discrepancy*

The second input is Discrepancy and it refers the inconsistency of the draft with other laws and rules. The Venice Commission also points in this direction, specifying that both referendum and "texts put to a referendum must not be contrary to international law or to the Council of Europe's statutory principles (democracy, human rights and the rule of law)". Inconsistency would never be eradicated in a system that is based on self-optimization and perpetual search for better solutions under the constraint of epistemic uncertainty, as it might be E-democracy (Turcoane 2014b). But even in nowadays representative democracy there are plenty of judicial inconsistencies (Niblett 2013, Fischman 2014) and examples of law promulgations that contradict the constitution in young states subject to the rule of law (APADOR-CH 2015). The last reference proves that this kind of discrepancy is observable in some cases using a simple human approach, but ways to measure it by mathematical (Doder, et al. 2010) and computational (Olson and Fusco 2012, McAreavey, Liu and Miller 2014) methods have already been proposed.

Figure 3 illustrates Discrepancy in a fuzzy approach with three FS that are not symmetric as in the case of Type and that are also based on author's perception (using the same scale), assuming that low and moderate classes of values are not as important as high level of inconsistency that may drastically affect the system output, see also the knowledge base of MFQ.

Resuming the example with digital currency, inconsistency at high level would occur if a draft enacted the new currency as the only one sanctioned by governmental bodies. Moderate discrepancy is when digital currencies are not allowed on the market as this contradicts the economic freedom (but only in virtual environment); low level appears, in author's opinion, when any digital currency should rely on centralized bodies' decisions.

The third and last input of MFQ is Age and it takes into account the draft's duration from its beginning to any given time, being inspired by the concept of product lifecycle found in engineering, software development, marketing etc. This paper sticks to the classical approach with four stages (Productlifecyclestages 2015): introduction, growth, maturity and decline. The first stage, of introduction, includes research and innovation, which are similar in politics to deliberation and decision; the correspondent for it in MFQ is the fuzzy subset new, which is described by a descendent slope immediately after its inception. Mature is the second subset of Age and it encompasses growth and maturity of a product, while old represents the fourth stage, decline, which is adjusted here based on the idea of a lifespan

extension of the product that still proves its usefulness (Bakker, et al. 2014).

Figure 4 illustrates Age and its fuzzy subsets (i.e. new, mature and old) using a time scale that has its units expressed in years (different approach from the other two inputs). In author's opinion, a draft loses its newness immediately after its promulgation, becoming mature between its age of three to five years and entering the old / obsolete stage after six or seven years. This way of building the third input fits to an accelerated rate of development of new products and of society itself, subject to rapid technological transformations. It is a common thing to acknowledge the fact that predictability must play an important role in legislation (with consequences in economics, research etc.), but Age is developed so as to also respect the idea of innovation.

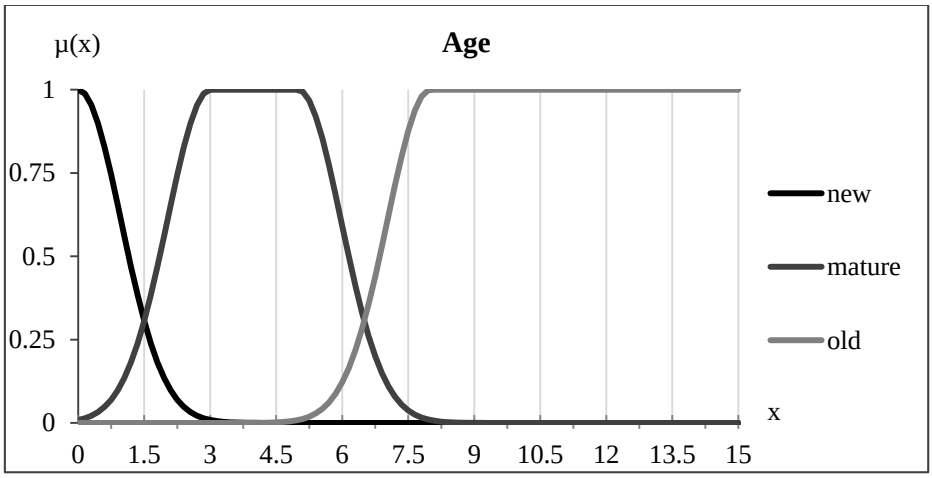

Figure 4. Fs of *Age*

A final example regarding the inputs and digital currency will put the latter in the origin point of the Age's scale, i.e. μnew(0) = 1; μmature(0) = 0 and μold(0) = 0. On the contrary, if the draft is about actual monetary legislation then it will be located somewhere on the scale that defines the FS of old.

The final variable of MFQ to be discussed is Quorum, which represents the value of RQ, the expected result. Subjectively chosen, the three subsets of the output are yet common and intuitive: low, medium, high. Figure 5 illustrates the output and its FS, using the same scale as for the first two inputs, i.e. Type and Discrepancy.

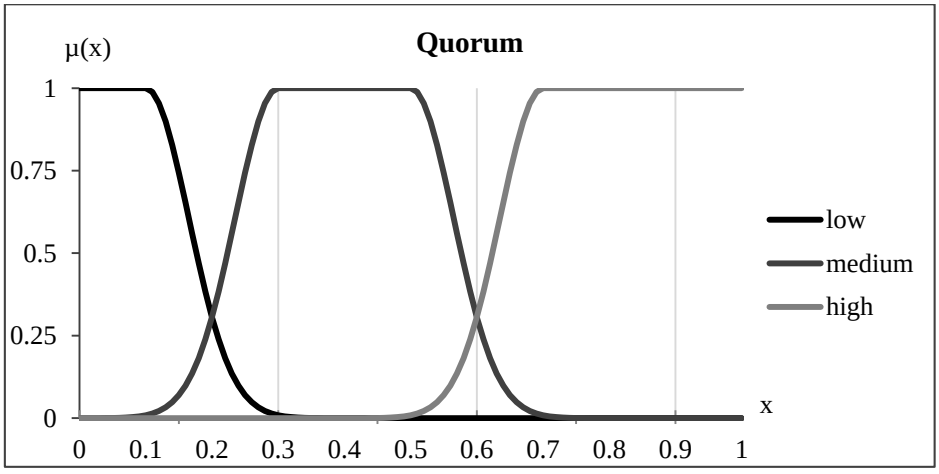

Figure 5. FS of *Quorum*

Table 1 presents the values of parameters of Matlab's functions (MathWorks 2015) used to build the FS of the inputs and the output of MFQ. There are two main types of membership functions in Table 1: triangular-shape and Gaussian-shape. In order to build a Gaussian curve one needs a standard deviation or sigma value and a mean value. The mean value is any bound of the crisp interval of FS of the inputs and the output. For a 0–1 scale, the sigma value is 0.065 (see Figure 2, Figure 3 and Figure 5) and it is based on a proxy determined in some other extended exploration (Turcoane 2014a). Extrapolating this value to a 0–15 scale, the sigma value becomes 0.975 as it is shown in Figure 4.

Table 1. FS of the inputs and the output of MFQ

Input	Fuzzy subsets	Matlab function
Type	proposal	trimf(0 0 1)
	abrogation	trimf(0 1 1)
Discrepancy	low	gauss2mf(0.065 0.0 0.065 0.1)
	moderate	gauss2mf(0.065 0.3 0.065 0.5)
	high	gauss2mf(0.065 0.8 0.065 1.0)
Age	new	gaussmf(0.975 0.0)
	mature	gauss2mf(0.975 3.0 0.975 5.0)
	old	gauss2mf(0.975 8.0 0.975 15.0)
Quorum	low	gauss2mf(0.065 0.0 0.065 0.1)
	medium	gauss2mf(0.065 0.3 0.065 0.5)
	high	gauss2mf(0.065 0.7 0.065 1.0)

The inputs and the output proposed in this paper are built based more on the author's educated guess rather than on self-evidence truth. However, these variables of MFQ should be subject to deliberation and practical experience refinement (e.g. using artificial intelligence to identify real life patterns), not to a proposal from an individual or a group (even one of certified scholars). Using the same principle, the FR of MFQ would derive from true life experiments, while this article only proposes a starting point in defining the knowledge base. There are ten FR that follow the *if–then* statement paradigm and each of them is briefly explained.

FR 1) If (Type is proposal) and (Discrepancy is low) then (Quorum is low). Any proposal that does not contradict any other rule or law requires a low representation and participation to be enacted; it will bring nothing but added value.

FR 2) If (Type is proposal) and (Discrepancy is not low) then (Quorum is high). Any proposal that, although innovative, contradicts the already implemented system demands a high level of approval from citizens; the system should not easily become unstable.

FR 3) If (Type is abrogation) and (Discrepancy is low) and (Age is new) then (Quorum is high). Any *draft* that does not contradict the system cannot be dismissed without high representation if the *draft* is not old / obsolete or if it is not mature; otherwise it will bring instability to the system.

FR 4) If (Type is abrogation) and (Discrepancy is low) and (Age is mature) then (Quorum is medium). A medium RQ is necessary to abrogate a mature *draft*, if the *draft* does not contradict the already implemented system, thus proving its utility.

FR 5) If (Type is abrogation) and (Discrepancy is low) and (Age is old) then (Quorum is low). An old *draft* may be abrogated as any system needs a refresh; having a low inconsistency and being old, the *draft* requires a low RQ to be dismissed.

FR 6) If (Type is abrogation) and (Discrepancy is moderate) and (Age is new) then (Quorum is high). Although inconsistent with the system, if a *draft* is new but there is need to reshape the system then it demands a high representation for abrogation (in order to support stability).

FR 7) If (Type is abrogation) and (Discrepancy is moderate) and (Age is mature) then (Quorum is medium). A *draft* requires medium RQ to be dismissed, if it is mature (granting stability) but it has an evident discrepancy.

FR 8) If (Type is abrogation) and (Discrepancy is moderate) and (Age is old) then (Quorum is low). An inconsistent and old *draft* should be easily abrogated, thus improving the system.

FR 9) If (Type is abrogation) and (Discrepancy is high) and (Age is new) then (Quorum is medium). A brand new *draft*, even with a high discrepancy, needs a medium representation to be abrogated, in order to support the idea of predictability and stability.

FR 10) If (Type is abrogation) and (Discrepancy is high) and (Age is not new) then (Quorum is low). Inconsistent with the system and already mature and old, a *draft* should be easily dismissed.

As can be seen in the ten FR, this research emphasizes the need for a stable system, which is also subject to flexibility and innovation. This section has tried to describe a model that is not indebted to citizen's preference or voter's behavior, but it is objectively bound to the context of the matter deliberated in a referendum.

4. Results of simulations and discussions

Using the Matlab's Fuzzy Toolbox (MathWorks 2015) it is easy to simulate the behavior of MFQ. For minimization and maximization, i.e. minimum and maximum of RQ, a fuzzy optimization algorithm was applied (Turcoane 2014a).

Table 2 presents different combinations of the inputs and the outputs yielded by MFQ.

Table 2. Output values

Simulations	Type	Discrepancy	Age	Quorum
1. Maximization	0	0.3624	0.5	**0.80941042**
	0	0.85	0.5	**0.80941042**
2. Minimization	0	0	0.5	**0.09311753**
	0	0.05	15	**0.09311753**
	1	0.87	11.5	**0.09311753**
3. Other high values	1	0	3.57	0.80935687
	1	0.05	0	0.7975
4. Average values	0.5	0.5	7.5	0.575
	0.5	0.5	1	0.501
	1	0	8	0.4
	0.9	0.5	7.5	0.227
	1	0.25	3	0.227
5. Other low values	1	0.85	4.02	0.09323834
	1	0.05	4	0.09324849

First of all, one notices very easily that values of minimum and maximum RQ are both yielded by different input combinations. Thus, the maximum value of 0.80941042 is determined using significantly different values of Discrepancy. The same thing also applies

for minimization, but more keenly, where the value of 0.09311753 is yielded by significantly disparate value of all the three inputs.

The optimum values of RQ empirically prove the nonlinearity of an MFIS and of the herein proposed MFQ. An MFIS is subject to a human approach that deals with gradual truth and partial belief as components of the uncertainty of the model. The if-then rules give an easy way for non-academics to control the model that is rather suitable for a non-classical mathematical approach. A standard mathematical or computational procedure would hardly fit the prototype described by MFQ.

Secondly, insignificant changes of the output minimum and maximum are identified by MFQ in Table 2. This proves that MFQ provides accurate results when needed. Moreover, the values of inputs could be provided to MFQ with a desired precision that affects the output accuracy; the latter may be adjusted to a round number or should provide a precise RQ, depending on practical decisions.

Thirdly, average values are also yielded in a large range with different input combinations, proving that not only extreme values may occur.

From sociopolitical point of view, MFQ is capable to provide a large range of results based on the contextual problems which need to be identified by FR. This means that RQ is a flexible concept, not a fix point given by a pure political deliberation that can be subject to non-democratic diversion. On one hand, RQ provides incentives to using a referendum in contextual matters (see the low outputs). On the other hand, the institution of referendum could not be used as a non-democratic instrument and a bagatelle mastered by some politicians (see the high outputs).

5. Conclusions

This conceptual paper has introduced a flexible quantitative-qualitative model of evaluating the size of a referendum quorum based on the contextual drafted matter, arguing against models relying on voters' preferences or behavior. Moreover, the quorum should not be determined by simple political debate or quarrel that seeks for individual or group favoritism and it definitely should not be a fix percentage or number. The quorum must give incentives for participation when the system requires enhancement or rejuvenating and it must stop abusing the institution of referendum when stability is needed. This exploration has proposed three variables that define the draft to evaluate the context of the referendum: type, inconsistency / discrepancy and age. Using these three variables, an objective evaluation of the quorum size is achievable depending on the context. While there are still many issues to address regarding the proposed model (e.g. deliberation over the knowledge base, setting up the inputs and the output), the author believes that this research will give incentives to further investigations from others.

References

1. Aguiar-Conraria, L. and Magalhães, P.C. **How quorum rules distort referendum outcomes: Evidence from a pivotal voter model,** European Journal of Political Economy, Vol. 26, No. 4, 2010, pp. 541–557
2. Ayadi, O., Cheikhrouhou, N. and Masmoudi, F. **A decision support system assessing the trust level in supply chain based on information sharing**

dimensions, Computers & Industrial Engineering, Vol. 66, No. 2, 2013, pp. 242-257

3. Bakker, C., Wang, F., Huisman, J. and den Hollander, M. **Products that go round: exploring product life extension through design,** Journal of Cleaner Production, vol. 69, 2014, pp. 10-16
4. Blume, L., Müller, J. and Voigt, S. **The economic effects of direct democracy - a first global assessment,** Public Choice, vol. 140, no. 3-4, September 2009, pp. 431-461
5. Doder, D., Raškovic, M., Markovic, Z. and Ognjanovic, Z. **Measures of inconsistency and defaults,** International Journal of Approximate Reasoning, vol. 51, 2010, pp. 832–845
6. Feld, L.P. and Kirchgassner, G. **Direct democracy, political culture, and the outcome of economic policy: a report on the Swiss experience,** European Journal of Political Economy, vol 16, no. 2, 2000, pp. 287-306
7. Fischman, J.B. **Measuring Inconsistency, Indeterminacy, and Error in Adjudication,** American Law and Economics Review, vol. 16, no. 1, 2014, pp. 40-85
8. Herrera, H. and Mattozzi, A. **Quorum and Turnout in Referenda,** Journal of the European Economic Association, vol. 8, no. 4, June 2010, pp. 838–871
9. Lauer, F. and Bloch, G. **Incorporating prior knowledge in support vector machines for classification: A review,** Neurocomputing, vol. 71, no. 7–9, 2008, pp. 1578-1594
10. Li, S.-T. and Tsai, F.-C. **A fuzzy conceptualization model for text mining with application in opinion polarity classification,** Knowledge-Based Systems, vol. 39, 2013, pp. 23–33
11. Liu, F., Geng, H. and Zhang, Y.Q. **Interactive fuzzy interval reasoning for smart web shopping,** Applied Soft Computing, vol 5, no. 4, 2005, pp. 433-439
12. Lutz, G. **Low turnout in direct democracy,** Electoral Studies, vol. 26, no. 3, 2007, pp. 624-632
13. Mamdani, E.H. and Assilian, S. **An experiment in linguistic synthesis with a fuzzy logic controller,** International Journal of Man-Machine Studies, vol. 7, no. 1, 1975, pp. 1-13
14. Matsusaka, J.G. **Direct Democracy Works,** The Journal of Economic Perspectives, vol. 19, no. 2, Spring 2005, pp. 185-206
15. McAreavey, K., Liu, W. and Miller, P. **Computational approaches to finding and measuring inconsistency in arbitrary knowledge bases,** International Journal of Approximate Reasoning, vol. 55, no. 8, 2014, pp. 1659–1693
16. Montero, J. **The impact of fuzziness in social choice paradoxes,** Soft Computing, vol. 12, no. 2, 2008, pp. 177-182
17. Niblett, A. **Tracking inconsistent judicial behaviour,** International Review of Law and Economics, vol. 34, 2013, pp. 9–20
18. Nurmi, H. and Kacprzyk, J. **Fuzzy Sets In Political Science: An Overview,** New Mathematics and Natural Computation, vol. 3, no. 3, 2007, pp. 281-299
19. Nurmi, H. **Voting paradoxes and referenda,** Social Choice and Welfare, vol. 15, no. 3, 1998, pp. 333-350

20. Olson, D. and Fusco, S. **Rules versus Standards: Competing Notions of Inconsistency Robustness in Patent Law,** Alabama Law Review, vol. 64, no. 3, 2012, pp. 647-696
21. Orchel, M. **Solving classification problems by knowledge sets,** Neurocomputing, vol. 149, 2015, pp. 1109–1124
22. Scheppele, K.L. **Romania Unravels the Rule of Law,** The Conscience of a Liberal, Krugman, P. (ed.) "Opinion Pages", 5 July, 2012. http://krugman.blogs.nytimes.com/2012/07/05/guest-post-romania-unravels-the-rule-of-law/, accessed May 1, 2013
23. Sen, A. **The idea of justice,** Cambridge: The Belknap Press of Harvard University, 2009
24. Treadwell, W.A. **Fuzzy Set Theory Movement in the Social Sciences,** Public Administration Review, vol. 55, no. 1, 1995, pp. 91-98
25. Turcoane, O. **Digital Democracy in Knowledge Society: a proposed architecture based on Cloud and complementary technologies,** Informatica Economica, vol. 18, no. 4, December 2014b, pp. 1-18
26. Zadeh, L.A. **Fuzzy sets,** Information and Control, vol. 8, no. 3, June 1965, pp. 338-353
27. Zurcher, A.J. **Hitler Referenda,** The American Political Science Review, vol. 29, no. 1, January 1935, pp. 91-99
28. Zwart, S. **Ensuring a representative referendum outcome: the daunting task of setting the quorum right,** Social Choice and Welfare, vol. 34, no. 4, 2010, pp. 643-677
29. * * * **Amicus Curiae,** apador.org. APADOR-CH, April 7, 2015. http://www.apador.org/wp-content/uploads/2015/04/Amicus-curiae-Legea-96.pdf, accessed April 20, 2015
30. * * * **Code of Good Practice on Referendums. Study,** Venice Commission, European Commission for Democracy through Law, Venice: European Council, 2007
31. * * * **MathWorks. Matlab Fuzzy Toolbox Documentation Center,** 2015, http://www.mathworks.com/help/fuzzy/index.html/, accessed 3 Feb. 2015
32. * * * **Product Life Cycle Stages,** Product Life Cycle Stages, April 30, 2015. http://productlifecyclestages.com/, accessed April 30, 2015
33. * * * **E-Democracy's outputs in a Mamdani fuzzy inference,** turcoane.ro. 2014a. http://turcoane.ro/cercetare/articole/EDemocracy%20outputs%20in%20MFIS.pdf, accessed December 3, 2014

[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".

[2]**Ovidiu Turcoane** graduated the Faculty of Cybernetics, Statistics and Economic Informat-ics in 2009. He holds a Masters Degree in Computer Science and a second specialization de-gree in Finances from 2011. He is currently studying, as a PhD Candidate of Doctoral School of Academy of Economic Studies, digital democracy in knowledge society. His fields of in-terest are methods and techniques of optimization in economics and social choice problems. He also studies fuzzy logics as an alternative to bivalent logic, especially in the field of de-velopment of a model of E-democracy that better fits inclusion, participation and delibera-tion.

USING PCA IN FINANCIAL MARKETS FIELD[1]

Stefan-Alexandru IONESCU
PhD, Assistant Professor, Romanian-American University, Bucharest, Romania
Postdoctoral Fellow, The Romanian Academy - POSDRU/159/1.5/S/137926

E-mail: ionescusa@gmail.com

Abstract:
Principal Component Analysis is a multivariate technique which aims the reduction of the initially causal space, with a minimal loss of the variance. Using this technique I identified a very interesting aspect, subsequently confirmed too by specific methods of discriminant analysis. If before the financial crisis, companies differentiate themselves primarily by rates of profitability, now it's obvious the increasing discriminatory power of liquidity and efficiency rates. We can also see that most companies register values under the standard parameters considered, even for a period of crisis.

Key words: PCA, multivariate, financial market, quantitative

1. Introduction

In the economic-social field, any problem for analysis and prediction entails, as a necessary step, the detailed investigation of the functional ties between explicative variables (independent variables). These represent symbols that express various quantitative or qualitative aspects of phenomena that constitute *influencing factors* or *causes* for other phenomena and processes.

Hence, the sum of all explicative variables involved in a certain multidimensional analysis defines a certain numerical space, called ***initial causal space***. The number of causal variables involved in the analysis defines the dimension of the initial causal space. Every causal variable is represented on one axis of this space, and a certain point on a certain axis represents a possible value that the variable associated with the respective axis can take. The points of the causal space are represented by the objects under investigation and the projections of these objects on the axes of the space are the values registered by these objects in relation with the characteristics associated with the axes.

The most important *characteristic* for all the techniques of multivariate analysis of data revealed by the *causal space* is the *variability* of this space. The variability contained in the causal space determined by explicative variables may be expressed under various more or less efficient forms. From this vantage point, the principal component analysis can be regarded as a *technique to decompose the total variability of the initial causal space* into *a smaller number of components* and without any *informational redundancies*.

In some analyses, finding out the principal components represents a goal in itself, everything being finalized with the actual interpretation of data. However, in other cases, the

analysis of the principal components is just an intermediary step, the principal components constituting the input for other analyses. Rancher (1998) shows that a special case is represented by regression analysis. If the number of initial variables is relatively great in comparison with the number of observations, or if the initial variables are strongly correlated among themselves, the testing can be inefficient and the regression coefficients can be unstable.

2. The Principal Components

The principal component analysis technique consists in the calculation of the projections of every point from the initial space, determined by the original variables under analysis, onto the axes of a new space, the dimension of which is significantly reduced. This method of multivariate analysis has as a purpose the determination of new variables called ***principal components***. These are in fact abstract vector variables, in the form of linear combinations of maximum variance of the original variables, and practically define a new space of objects. Ruxanda (2009) shows that the determination of linear combinations used in the construction of principal components takes into consideration the following aspects:

- The number of principal components is equal to the number of original variables.
- The sum of the variances of the principal components coincides with the sum of the variances of the original variables, so that the principal components take the variability contained in the original variables.
- *The first principal component* is a normalized linear combination, the *variance* of which *is maximum*, and the *second principal component* is a linear combination that is *uncorrelated* with the first principal component, having a variance that is as great as possible, but smaller than that of the first component etc. The principal components are scaled according to the magnitude of its variance, the first having the maximum variance and the last the minimum variance.
- The axes of the new space are orthogonal two by two and define the new variables, that is the *principal components* that are obviously uncorrelated two by two.
- The sum of the squares of the coefficients that define the linear combination that corresponds to a principal component equals one;

The coordinates of objects in the new space, i.e. the projections of the objects onto its axes, are evaluations of objects in relation with the new variables and are called scores of the principal components or ***principal scores***.

2.1. The mathematical model of principal components

We consider that the initial causal space under investigation is determined by an *n* number of explicative variables, noted $x_1, x_2, \dots, x_n$, which symbolize characteristics of the objects under analysis. What results is that every object is characterized by *n* variables.
The determination of the principal components is described through a transformation of the following type:

$$\Psi: \mathfrak{R}^n \to \mathfrak{R}^k \qquad (1)$$

where $\mathfrak{R}^n$ and $\mathfrak{R}^k$ are two real vector spaces, and the dimension of the second space is smaller than the dimension of the first space, respectively $k < n$.

Through the Ψ transformation, an object ***x,*** belonging to the *n*-dimensional $\mathfrak{R}^n$ space, is transformed into an object ***w*** belonging to the *k*-dimensional $\mathfrak{R}^k$ space. Through this transformation the coordinates of the object change and the number of coordinates is reduced.

If $\boldsymbol{x} \in \mathfrak{R}^n$ and $\boldsymbol{w} \in \mathfrak{R}^k$, then the transformation Ψ is a linear application of the following type:

$$\boldsymbol{w} = A^t \cdot \boldsymbol{x} \tag{2}$$

where A is a matrix of real numbers, of a $n \times k$ dimension.

The solution to this problem consists in the determination of matrix A, so that an object ***w*** should constitute the best possible representation for object ***x***.

The *n* principal components, corresponding to the analyzed causal space, present themselves in the form of an *n*-dimensional vector, noted ***w***:

$$\boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \tag{3}$$

Every w_i component of this vector represents a principal component defined in relation with the original variables, through the following linear combination:

$$w_i = \alpha_1^{(i)} \cdot x_1 + \alpha_2^{(i)} \cdot x_2 + \cdots + \alpha_n^{(i)} \cdot x_n \qquad i = 1,2,\dots,n \tag{4}$$

In order to determine the w_i principal component, the determination of the $\alpha_j^{(i)}$ coefficients which define the corresponding linear combination of this principal component is necessary. The $\alpha_j^{(i)}$ coefficients are precisely the coordinates of the eigenvectors corresponding to the covariance matrix of the original variables $x_1, x_2, \dots, x_n$, and the variances of the principal components are precisely the eigenvalues of this matrix. According to the definition of the principal components, the determination of these coefficients is performed so that the w_i principal component should have the maximum variance.

If we consider that the *n* coefficients $\alpha_1^{(i)}, \alpha_2^{(i)}, \dots, \alpha_n^{(i)}$ of the above-mentioned linear combination are the coordinates of the *n-dimensional* $\propto^{(i)}$ vector, namely:

$$\boldsymbol{\alpha^{(i)}} = \begin{pmatrix} \alpha_1^{(i)} \\ \alpha_2^{(i)} \\ \vdots \\ \alpha_n^{(i)} \end{pmatrix} \qquad i = 1,2,\dots,n \tag{5}$$

we can define the principal component w_i under the form:

$$w_i = \left(\boldsymbol{\alpha^{(i)}}\right)^t \cdot \boldsymbol{x} \qquad i = 1,2,\dots,n \tag{6}$$

where the coordinates of the $\boldsymbol{\alpha}^{(i)}$ vector are chosen so that the maximization of the variance of the principal component w_i be ensured.

Since the *i* principal component, w_i is a linear transformation of the elements of vector ***x***, supposedly normally distributed, of a $\boldsymbol{\mu}$ average and covariance matrix Σ, what results is that this principal component is a random, normally distributed variable. Based on the relationship in (6), the average and the variance of this principal component can be inferred, as follows:

$$E(w_i) = E\left[\left(\boldsymbol{\alpha^{(i)}}\right)^t \cdot \boldsymbol{x}\right] = \left(\boldsymbol{\alpha^{(i)}}\right)^t \cdot \boldsymbol{\mu} \tag{7}$$

$$Var(w_i) = \left(\alpha^{(i)}\right)^t \cdot \Sigma \cdot \alpha^{(i)} \quad (8)$$

The result is:

$$w_i \sim N\left[\left(\alpha^{(i)}\right)^t \cdot \mu, \left(\alpha^{(i)}\right)^t \cdot \Sigma \cdot \alpha^{(i)}\right] \qquad i = 1,2,\dots,n \quad (9)$$

The solution to this principal component analysis is equivalent to the solution of the following extreme value problem:

$$\begin{cases} \underset{A \in M_{n \times k}}{\text{opt}} \ \Phi(\boldsymbol{x}, \boldsymbol{w}) \\ SR: \quad \boldsymbol{w} = A^t \cdot \boldsymbol{x} \end{cases} \quad (10)$$

where the optimum criterion can be either maximum or minimum, according to the nature of the Φ function:

- If the Φ function is a distance function, then the optimum criterion will be represented by the minimization of the Φ function.
- If the Φ function is a measure of the quantity of information brought by the new means of object representation, the optimum criterion will be represented by the maximization of the Φ function.[2]

In order to define the mathematical model of the principal component analysis, we will consider that the $\boldsymbol{\alpha}^{(i)}$ vectors represent the columns of an A matrix of $n \times n$ dimension of the form:

$$A = \begin{pmatrix} \alpha_1^{(1)} & \alpha_1^{(2)} & \dots & \alpha_1^{(n)} \\ \alpha_2^{(1)} & \alpha_2^{(2)} & \dots & \alpha_2^{(n)} \\ \vdots & \vdots & \dots & \vdots \\ \alpha_n^{(1)} & \alpha_n^{(2)} & & \alpha_n^{(n)} \end{pmatrix} \quad (11)$$

We assume that $\boldsymbol{x}$ is the vector whose coordinates are the original variables $x_1, x_2, \dots, x_n$ and that $\boldsymbol{w}$ is the vector whose coordinates are the principal components $w_1, w_2, \dots, w_n$. Taking into account relationship number (6) as well, the linear combinations that define the principal components can be written under the following matrix form:

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} \alpha_1^{(1)} & \alpha_2^{(1)} & \dots & \alpha_n^{(1)} \\ \alpha_2^{(1)} & \alpha_2^{(2)} & \dots & \alpha_n^{(2)} \\ \vdots & \vdots & \dots & \vdots \\ \alpha_1^{(n)} & \alpha_2^{(n)} & & \alpha_n^{(n)} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (12)$$

The mathematical model of the principal component analysis is defined as follows:

$$\begin{cases} \underset{A \in M_{n \times n}}{\max} \ Var(\boldsymbol{w}) \\ \boldsymbol{w} = A^t \cdot \boldsymbol{x} \end{cases} \quad (13)$$

The *n* columns of the A matrix represent in fact the normalized eigenvectors of the Σ covariance matrix, and the variance of every w_i principal component, which is a maximum variance in relation with the variances of the previous principal components, is represented precisely by the λ_i eigenvalue of the same covariance matrix. This method to determine the elements of the A matrix is equivalent to the calculation of the object projections of the $x \in \mathfrak{R}^n$ type onto the linear subspace generated by the vectors of the A matrix columns.

3. Data used in the analysis

For this analysis we have selected 101 Romanian firms of various sizes which have existed for more than one year on the market. All these firms have unfolded their activity

and have submitted their balance sheets for the 31st of December to the Registry of Commerce.

An analysis of the size of these firms can be made according to the total assets. The smallest value is of 15,821 lei, and the biggest is over 30 million lei. More than half of these firms are medium-sized, the small and big ones being in similar proportions.

For all the enterprises we have included, data from both the yearly balance sheet and the balance sheets on December has been collected. The most important values extracted have taken into account:

- assets, as well as their classification;
- debts, divided as well into categories, including those to banks and leasing companies;
- capitals and equity capitals;
- data connected with the turnover, profit, taxes and duties.

Although the information contained in the financial statements of the companies offers important elements referring to company performance, it is advisable that the statements should be accompanied by an analysis of financial rates as well. The objective is precisely to analyze the financial situation focusing on four aspects: liquidity, solvability (risk), activity and profitability.

In this analysis we refer to the following eight rates:

- Profitability Ratios*:* Return on Assets ***(ROA),*** Return on Equity ***(ROE),*** Return on Capital Employed ***(ROCE).***
- Efficiency ratios*:* Total Assets Turnover ***(RAT).***
- Liquidity Ratios*:* Current ratio ***(CR),*** Quick ratio ***(QR),*** Cash Ratio ***(CashR).***
- Solvability Ratios*:* General Solvability ***(SP).***

Data processing has been done with the *STATISTICA 8.0* program package.

4. Results.

We have shown above how the dimensionality of initial causal space can be reduced while preserving the least informational loss. Obviously, informational redundancy from the initial observations will be eliminated. We have performed the principal component analysis on the covariance matrix.

Table 1. Eigenvalues of the covariance matrix

	Eigenvalue	% - Total variance	Cumulative Eigenvalue	Cumulative %
W1	1.681320	49.22606	1.681320	49.2261
W2	1.015914	29.74415	2.697234	78.9702
W3	0.258015	7.55421	2.955248	86.5244
W4	0.197061	5.76958	3.152309	92.2940
W5	0.129885	10281	3.282194	96.0968
W6	0.065222	1.90958	3.347416	98.0064
W7	0.059987	1.75631	3.407403	99.7627
W8	0.008105	0.23731	3.415508	100.0000

In the second column we have the eigenvalues of the covariance matrix of the new factors that have been extracted successively. In the third column is presented the variance values in percentages. As we can see, the contribution of the first factor is approximately half

of the total variance, whereas the second contributes with approximately 30% to the total variance. The sum of the 8 eigenvalues, i.e. the variance of the 8 principal components, equals the sum of the variances of the original variables. The third column contains the cumulative variances, the last line showing the actual total variance. The last column reveals the cumulative variances in percentages.

As we can see, the first two principal components account for approximately 79% of the variance of the 8 initial variables, and, if we were to take into account the third component as well, 86.5% of the initial variance would be represented.

Although the number of factors that can be taken into account is generally an arbitrary decision, there are nevertheless certain criteria that practitioners use. The first was proposed by Kaiser (1960)[3] and referred to keeping those factors the eigenvalues of which are bigger than 1.

The second criterion is called the "scree test" and was proposed by Cattell (1966)[4]. The eigenvalues from the second column of table 1 are represented and the points are united. In figure 1 we can see that, starting with the third variable, the chart slope is close to zero.

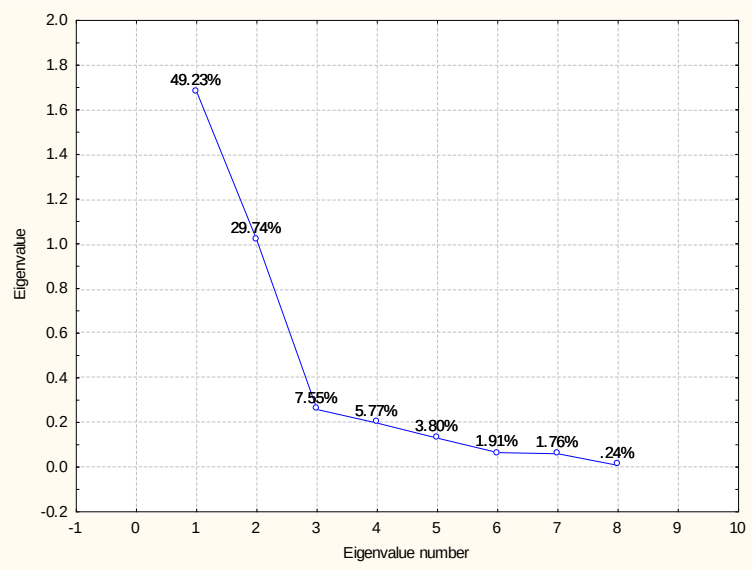

Figure 1. Eigenvalues

Applying both criteria, keeping the first two factors seems the most plausible. Also, keeping more than two factors can make the drawing of future charts more difficult. Hence, the dimension of the new representation space for objects will be 2.

In table 2 we have the eigenvectors of the covariance matrix:

Table 2. Eigenvectors of the covariance matrix

	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8
ROA	-0.040117	-0.037710	-0.026787	-0.073412	-0.219236	-0.082868	-0.046354	**0.966324**
ROE	-0.054749	-0.072769	**-0.961826**	-0.014836	0.115778	-0.141111	-0.179759	-0.027359
ROCE	-0.067856	-0.077984	-0.204885	-0.158220	**-0.781037**	0.305138	0.442325	-0.153373
RAT	-0.223403	**-0.960808**	0.096669	-0.039278	0.115039	-0.026270	0.040940	-0.021263
CR	**-0.660927**	0.148632	-0.022416	0.636065	0.051507	-0.131084	0.338144	0.042728
QR	**-0.599534**	0.105551	0.071886	-0.215882	-0.109665	0.421674	**-0.620428**	-0.053660
CashR	-0.369520	0.175291	0.016442	**-0.718330**	0.280798	-0.269826	0.405954	-0.002575
SP	-0.089828	-0.009507	0.130041	-0.025133	-0.470905	**-0.782064**	-0.322708	-0.191789

Every principal component can be defined in relation with the original variables with the help of linear combinations. The coefficients are practically the coordinates of the eigenvectors corresponding to the covariance matrix of the original variables, and the variances of the principal components are precisely the eigenvalues of this matrix and have been presented in table 1. We may notice the relatively small importance of the ROA, SP and ROCE indicators which are represented mainly through the last 4 components. As we can see from table 1, these components preserve less than 8 percent from the total variance. The principal components of the causal space will be determined by the following equations:[5]

$$\begin{cases} W_1 = -.04 \cdot ROA - .05 \cdot ROE - .07 \cdot ROCE - .22 \cdot RAT - .66 \cdot CR - .60 \cdot QR - .37 \cdot CashR - .09 \cdot SP \\ W_2 = -.04 \cdot ROA - .07 \cdot ROE - .08 \cdot ROCE - .96 \cdot RAT + .15 \cdot CR + .11 \cdot QR + .18 \cdot CashR - .01 \cdot SP \\ W_3 = -.04 \cdot ROA - .96 \cdot ROE - .20 \cdot ROCE + .10 \cdot RAT - .02 \cdot CR + .07 \cdot QR + .02 \cdot CashR + .13 \cdot SP \\ W_4 = -.07 \cdot ROA - .01 \cdot ROE - .16 \cdot ROCE - .04 \cdot RAT + .64 \cdot CR - .22 \cdot QR - .72 \cdot CashR - .03 \cdot SP \\ W_5 = -.22 \cdot ROA + .12 \cdot ROE - .78 \cdot ROCE + .12 \cdot RAT + .05 \cdot CR - .11 \cdot QR + .28 \cdot CashR - .47 \cdot SP \\ W_6 = -.08 \cdot ROA - .14 \cdot ROE + .31 \cdot ROCE - .03 \cdot RAT - .13 \cdot CR + .42 \cdot QR - .27 \cdot CashR - .78 \cdot SP \\ W_7 = -.05 \cdot ROA - .18 \cdot ROE + .44 \cdot ROCE + .04 \cdot RAT + .34 \cdot CR - .62 \cdot QR + .41 \cdot CashR - .32 \cdot SP \\ W_8 = +.97 \cdot ROA - .03 \cdot ROE - .15 \cdot ROCE - .02 \cdot RAT + .04 \cdot CR - .05 \cdot QR - .003 \cdot CashR - .19 \cdot SP \end{cases} \quad (14)$$

As we have shown previously, the first two principal components ensure the preservation of 78.9702% of the variance contained in the initial causal space, determined by the 8 variables. Within the range of a loss of 21.0298%, we can express the 8 initial variables through the intermediary of the first two principal components. Table 3 represents the first two columns of the factor matrix, with the help of which the correlations between the variables and every factor can be interpreted.

Table 3. Matrix of the intensity of factors for the first two principal components

	W 1	W 2
ROA	-0.052018	-0.038009
ROE	-0.070991	-0.073346
ROCE	-0.087986	-0.078602
RAT	-0.289677	**-0.968422**
CR	**-0.856995**	0.149810
QR	**-0.777390**	0.106387
CashR	**-0.479141**	0.176681
SP	-0.116476	-0.009583

Every element may be interpreted as a measure in which the original variable participates in the formation of principal components. The first component describes very well the liquidity variables CR, QR and CashR. To a smaller extent, the variable of efficiency in the use of RAT assets and the variable of patrimonial solvability SP contribute to the significance of this principal component. The second principal component is strongly and in a negative sense correlated with the RAT efficiency variable, displaying weak correlations with the liquidity indicators. All the other correlations, among the first two principal components and the rest of the variables have values of less than 10 percent. The profitability indicators display correlations of very similar intensities with the first two principal components. RAT is the only indicator from the composition of the *DuPont model* that displays important correlations with the two principal components.

Generally speaking, we can say that the first factor, the one that covers 49.23% of the variance, reflects ***liquidity and solvability***, whereas the second factor, which covers 29.74% of the variance, synthesizes, from an informational point of view, company ***efficiency***. Thus, we have an indicator of general solvability, ***W1,*** and an indicator of general efficiency, ***W2***.

Each line in table 3 defines the coordinates of an original variable, seen as a point in the bi-dimensional reduced space. Figure 2 presents in an extremely suggestive manner the interpretation of these factors. What needs to be mentioned is that the 8 variables appear only in quadrants II and III, which shows correlations that are rather negative.

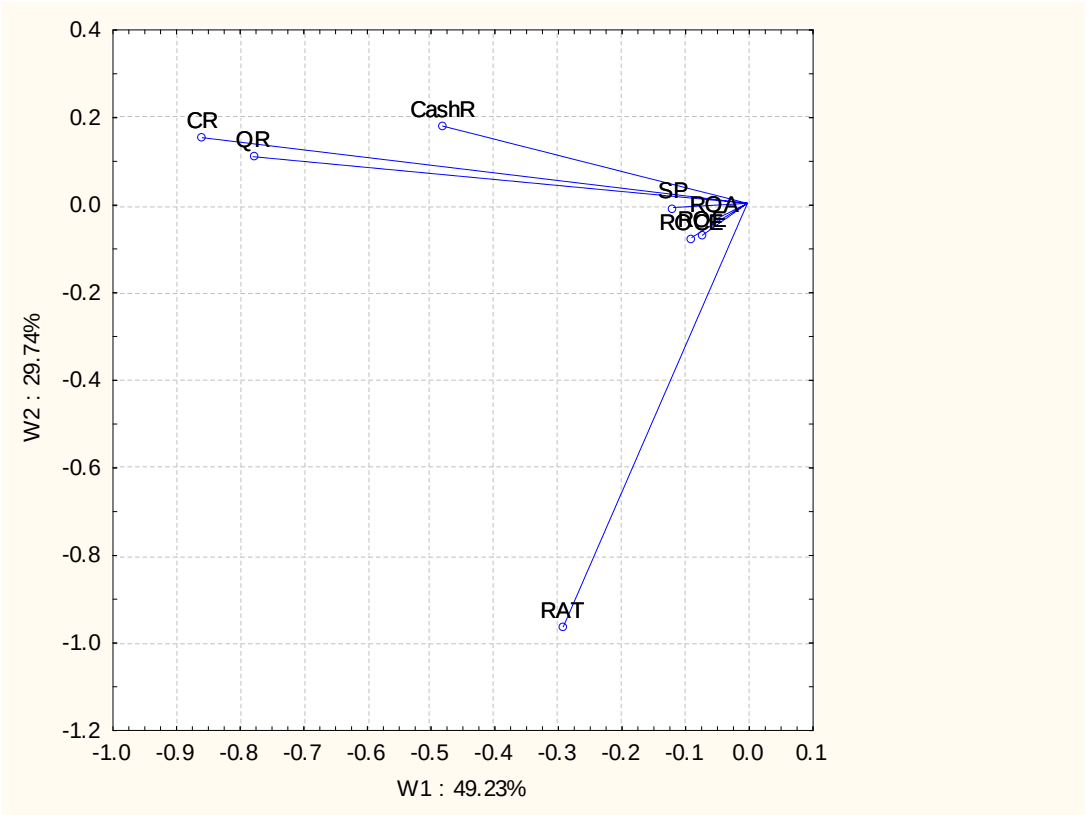

Figure 2. Projection of variables in a Cartesian system of W1 and W2

The variables that are strongly correlated with the first factor, namely liquidity rates, are farther from the *Oy* axis and appear represented in quadrant II. The RAT variable, being correlated in the first place with the second factor, appears even farther apart from the *Ox* axis and is represented in quadrant III.

Table 4 contains the coordinates of each and every one of the 101 firms under analysis, for the bi-dimensional finite space defined by the eigenvectors of the two principal components that we have kept.

Object coordinates in the reduced space are also called the principal scores of objects. Principal scores can be used in the analysis as a substitute of original observations, thus simplifying the initial information base. Ruxanda (2009) considers that principal scores are better suited to be used in analyses, as they are less likely to be affected by errors in comparison with original measurements. The fact that principal scores are more robust in relation with perturbations introduced by errors, that they have a certain invariance in relation with errors, makes them more salient from an informational point of view than original observations.

Due to the fact that the new reduced space has only two axes, the 10 objects can be represented graphically in this space. The graphical representation from figure 3 shows the positioning of the 101 firms in relation with the axes of the new space.

Table 4. Object coordinates in the bi-dimensional principal space

Obj.	W 1	W 2	Obj.	W 1	W 2	Obj.	W 1	W 2
1	0.92246	0.17083	34	0.11369	-0.75858	68	-2.25986	-1.83606
2	-2.30047	1.62656	35	-0.25561	-0.54107	69	1.11784	1.02697
3	0.54505	0.58010	36	1.44912	0.39254	70	1.16220	0.98494
4	0.94290	0.34056	37	-1.01628	-2.55144	71	1.31252	0.29129
5	-0.41592	-1.50400	38	-0.31971	-0.68276	72	0.09453	-1.71136
6	-0.92440	1.15265	39	1.06464	0.40374	73	0.75357	0.79035
7	-2.27346	1.75198	40	0.29465	-1.24653	74	-1.83639	1.59073
8	0.30099	-0.31820	41	0.60322	0.07182	75	1.53481	0.63960
9	-0.16085	1.18733	42	0.75149	-0.31289	76	0.30684	-0.18174
10	-1.15324	-2.15315	43	-0.10130	0.80636	77	0.07209	0.71098
11	1.18478	0.05475	44	1.67841	0.76009	78	0.59218	-0.03836
12	0.26761	-0.02226	45	1.55056	-1.51286	79	0.00806	0.47547
13	-1.35034	-4.05489	46	1.64630	0.72357	80	-2.03139	1.54840
14	-0.26368	0.69912	47	1.73311	0.45565	81	0.21322	-0.08724
15	-1.48029	0.56571	48	1.53607	0.82753	82	1.01396	-0.54419
16	0.20445	0.54344	49	-0.06948	-1.00137	83	-0.04110	-0.80137
17	0.11657	-1.01666	50	-5.06088	0.88351	84	-4.99077	1.16111
18	-1.77799	-0.43436	51	1.44254	0.79235	85	0.58313	-0.62409
19	-1.38050	0.34749	52	-0.85130	1.04399	86	0.51682	-0.12252
20	-2.35150	0.23165	53	-1.52406	0.62633	87	0.30317	-0.42902
21	-0.26505	0.34568	54	0.57404	0.37003	88	-0.41400	-0.29846
22	0.13788	0.91824	55	0.07909	-1.19050	89	-2.03296	-0.87358
23	-0.33733	-2.48703	56	0.27728	-0.36044	90	-0.57988	-0.62456
24	0.94363	0.54725	57	0.30565	-1.21259	91	0.19400	-1.16720
25	0.94021	-0.70235	58	0.23057	-0.43096	92	0.55482	-0.53463
26	-1.50101	1.24276	59	-0.34549	0.46831	93	1.69031	0.51705
27	1.68698	0.67703	60	0.37427	-0.24178	94	0.48793	-0.44790
28	1.78411	0.59879	61	1.42302	0.55391	95	-0.95689	-1.43593
29	0.98932	0.21634	62	-0.40780	-0.92208	96	0.26765	-1.03124
30	0.15451	0.94375	63	1.20108	-0.73368	97	-2.47667	-0.31404
31	1.54794	0.83926	64	0.98130	0.01579	98	-0.52374	1.20294
32	1.09693	0.56957	65	-0.59865	-0.72998	99	-0.10314	0.51226
33	0.57210	0.66365	66	1.23019	0.14112	100	0.37494	0.58453
			67	-1.51410	1.63574	101	0.21414	0.40241

The positioning of objects in relation with the two axes renders a first image regarding the similarities or dissimilarities among the firms. The best distinction seems to be offered by the first factor. We notice a more numerous group of objects gathered on the right side of the *Oy* axis, on the one hand, and a more rarefied, less numerous group on the left side of the *Oy* axis. Analyzing the composition of the first factor, we can say that the firms in quadrants II and III are characterized by liquidity indicators bearing reassuring values. Practically speaking, the farther on the left side of the *Oy* axis is an object placed, the better indicators for liquidity it will have. The second factor separates objects from under the *Ox* axis, characterized primarily by a more intense asset turnover than those above the *Ox* axis which do not use their resources in a satisfactory manner. We notice in the first place the big group in quadrant I. These are firms that register mediocre values for all the indicators and, automatically, for the two principal components as well.

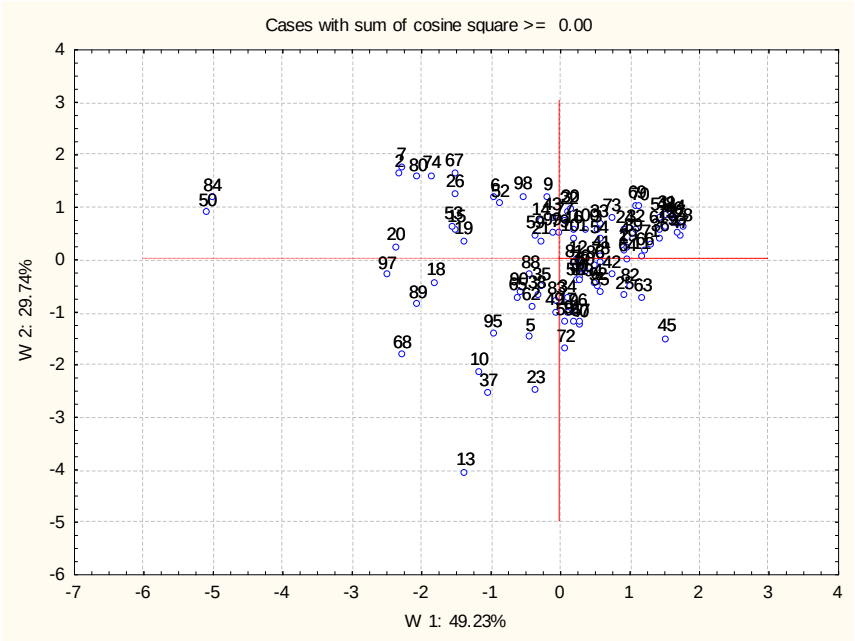

Figure 3. Object representation in the bi-dimensional principal space

5. Conclusions

To conclude, we may assert that principal component analysis is useful in solving two categories of problems: the ***simplification of the structure of causal dependency*** and the ***reduction of space dimensionality***.

In the context of principal component analysis, the values of characteristics are coordinates of the points that define the elements of the analyzed population. The most advantageous means of representation from an informational point of view is obtained by considering a new *space of representation*, which defines through its axes new characteristics of objects. These new characteristics are the ***principal components***, and the values registered by objects to these new characteristics are called ***scores***. Representation in a reduced space is known as *dimensionality reduction* and the principal component analysis is known as the *dimensionality reduction technique*.

The representation of objects in the new principal space is simplified and non-redundant. Every firm appears as a bi-dimensional vector, the components of which are represented by the corresponding values of the two principal components. These values are called principal scores. Having in mind the values from table 2, the two principal components can be expressed in the following manner:

$$\begin{cases} W_1 = -.04 \cdot ROA - .05 \cdot ROE - .07 \cdot ROCE - .22 \cdot RAT - .66 \cdot CR - .60 \cdot QR - .37 \cdot CashR - .09 \cdot SP \\ W_2 = -.04 \cdot ROA - .07 \cdot ROE - .08 \cdot ROCE - .96 \cdot RAT + .15 \cdot CR + .11 \cdot QR + .18 \cdot CashR - .01 \cdot SP \end{cases} \quad (15)$$

In broad lines, we may say that the first component, which covers 49.23% of the variance, reflects liquidity and solvability, whereas the second component, which covers 29.74% of the variance, synthesizes firm efficiency in an informational manner.

The advantage of using principal components is given by the simplification of calculations, and the downside consists in a more difficult interpretation of results, if the analysis is corroborated with other techniques of multidimensional analysis.

Bibliography

1. Abdi, H. and Williams, L.J. **Principal Component Analysis,** John Wiley & Sons, Inc. WIREs Comp Stat, vol. 2, 2010, pp. 433–459
2. Back, A.D. and Weigend, A.S. **Discovering Structure in Finance Using Independent Component Analysis,** Advances in Computational Management Science, Vol. 2, 1998, pp 309-322
3. Cattell, R. **The Scree Test for the Number of Factors. Multivariate Behavioral Research,** vol. 1, 1966, pp. 245-276
4. Chen, K.H. and Shimerda, T.A. **An Empirical Analysis of Useful Financial Ratios,** Financial Management, Vol. 10, No. 1, Spring 1981, pp. 51-60
5. Dieckmann, S. and Plank, T. **Default Risk of Advanced Economies: An Empirical Analysis of Credit Default Swaps during the Financial Crisis,** Review of Finance, vol. 16, no. 4, 2012, pp. 903-934
6. Geambasu, C., Jianu, I., Herteliu, C. and Geambasu L. **Macroeconomic Influence on Shares' Return. Study case: Arbitrage Pricing Theory (APT) Applied on Bucharest Stock Exchange,** Economic Computation and Economic Cybernetics Studies and Research, Vol 48, no. 2, 2014, pp 133-150
7. Kaiser, H. **The Application of Electronic Computers to Factor Analysis,** Educational and Psychological Measurement, 1960, pp. 141-151
8. Leon, C. and Murcia, A. **Systemic Importance Index for Financial Institutions: A Principal Component Analysis Approach,** ODEON, no. 7, 2013.
9. Rencher, A. **Methods of Multivariate Analysis,** New York, USA: John Wiley & Sons, 2002
10. Ruxanda, G. **Analiza Datelor,** Bucharest, ASE, 2001
11. Ruxanda, G. **Construirea, estimarea și implantarea software a metodelor matematice,** Cercetarea stiintifica in ASE, 2010

[1] **Acknowledgements**
This work was financially supported through the project "Routes of academic excellence in doctoral and post-doctoral research - READ" co-financed through the European Social Fund, by Sectoral Operational Programme Human Resources Development 2007-2013, contract no POSDRU/159/1.5/S/137926.

[2] This situation is specific for the standard variance used to solve the principal component problem, in which the objective is to maximize the variance of the principal components, as a measure for the quantity of information expressed by each of these. Thus, in what follows we will particularize matters for this case.

[3] It is the criterion that bears his name: *the Kaiser criterion*.

[4] "Scree" means rock debris, it is a term used in geology and refers to the bits of rock resulted from mountain fragmentation that gather at the base of mountain slopes. What is important is to find the point where the slope becomes visibly less abrupt, namely the base of the slope.

[5] It is about the equations described in (12).

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

MEANS OF AUTOMATING THE PROCESS OF EVALUATING THE QUALITY OF CITIZEN ORIENTED SOFTWARE SOLUTIONS USING SOFTWARE METRICS

Emanuel Eduard HERŢELIU
PhD Candidate, Doctoral School of Economical Informatics
The Bucharest University of Economic Studies, Romania

E-mail: emanuel.herteliu@gmail.com

Abstract:
This paper presents means of evaluating the quality of citizen oriented software solutions using software metrics. Highlights about citizen oriented software solutions are presented. The need of automation is presented when dealing with a big number of informatics applications. The importance of choosing the right metrics that lead to meaningful results when dealing with citizen oriented informatics applications is presented. A sample of web based software solutions is used and the modality of getting and parsing the source code using automated tools is highlighted.

Key words: quality, citizen oriented, software metrics, stability, evaluation

1. Introduction

Software solutions provided by local government authorities are used on a daily basis in order to help citizens solve their administrative tasks. The quality of this type of informatics applications is very important because the computer skills of the users are unknown. Thus efforts must be made whilst developing citizens oriented solutions in order to assure a high level of quality complying to international software quality standards and providing a high utilizability level in the same time.

The variety of automated services provided to citizens is different from a government authority to another. A large number of web based solutions has been analyzed from the prospective of the quality of the services they provide for the citizens and the results show that there are improvements to be made in terms of what services they provide and how good the tools to implement the services are.

Because of the large number of solutions the need of automating the quality evaluation process had risen, this paper presents an approach towards automating this process.

2. Software metrics for citizen oriented web based solutions

As the analyzed software solutions are citizen oriented choosing the right metrics is very important for obtaining meaningful quality evaluating results.

Two prospectives of evaluating the quality are approached:

- The quality of the service provided;
- The quality of the implemented software solutions which provide these services;

A sample of localities from Romania have been chosen and the websites provided by the local government authorities were analyzed in order to emphasize their quality level. The analisys showed that not all of the localities proved online services to citizens and this brings the need of extra governmental funding to correct this matter. The websites of the localities that provide online services were used in an automated quality evaluation process. This process speeds up the evaluation by crawling each individual website, parsing and analyzing the HTML code using software metrics. The metrics have been chosen to show weather a specific service exists or not on the analyzed web based solution. The services provided to citizens are:

- Tax paying;
- Administrative forms available for downloads;
- Informative services;
- Contact information on individual departments of the local government authorities;
- New events calendar schedule;
- Media content provided to citizens.

The metrics are presented in table 1.

Table 1 – *Web page metric for citizen oriented computer applications* [1]

Metric	Applications	Significance regarding to the service provided
Total number of pages per application	Summing up all pages crawled for an app	The complexity of the problem solved by the application
The total number of nodes per page/app	Summing up all nodes parsed per page/app	The structure of page/application
Form nodes present on page/app	Searching for nodes of type *form*	Application is interactive
The weight of *input* nodes in the total number of app/page nodes	Dividing the total number of *input* nodes to the total number of nodes found on page/app	Application's extent of interactivity
Input nodes of type *file* present on page/app	Summing up the input nodes of type *file*	Application allows file uploads
Total number of files available for download per page/app	Summing up all files available for input per page/app	Application allows file downloads
Secure protocol present	Checking the protocol of app's URL	Application allows secure transactions

The metrics presented in table 1 are calculated based on real data crawled using the main URL of the websites corresponding to the localities included in the sample. Then each individual page is crawled and based on the number of the URLs crawled on the same domains the number of the pages for each website is calculated. Each page is searched for the specific HTML nodes presented in the first column of table one.

3. The automated quality analyzing process

The applications sample is built based on the homepage URL of the websites provided by the local government authorities of each locality. The information is stored in a database and one of the tables stores all of the analyzed URLs. Each individual URL is used during an iterative process. The homepage is crawled then out of the links found on the page the ones corresponding to the same domain are crawled as pages of the same website without including the links to pdf, document, image files and media content which are used to calculate the amount of the downloadable content provided to citizens.

The structure of saved information about the applications analyzed is presented in figure one.

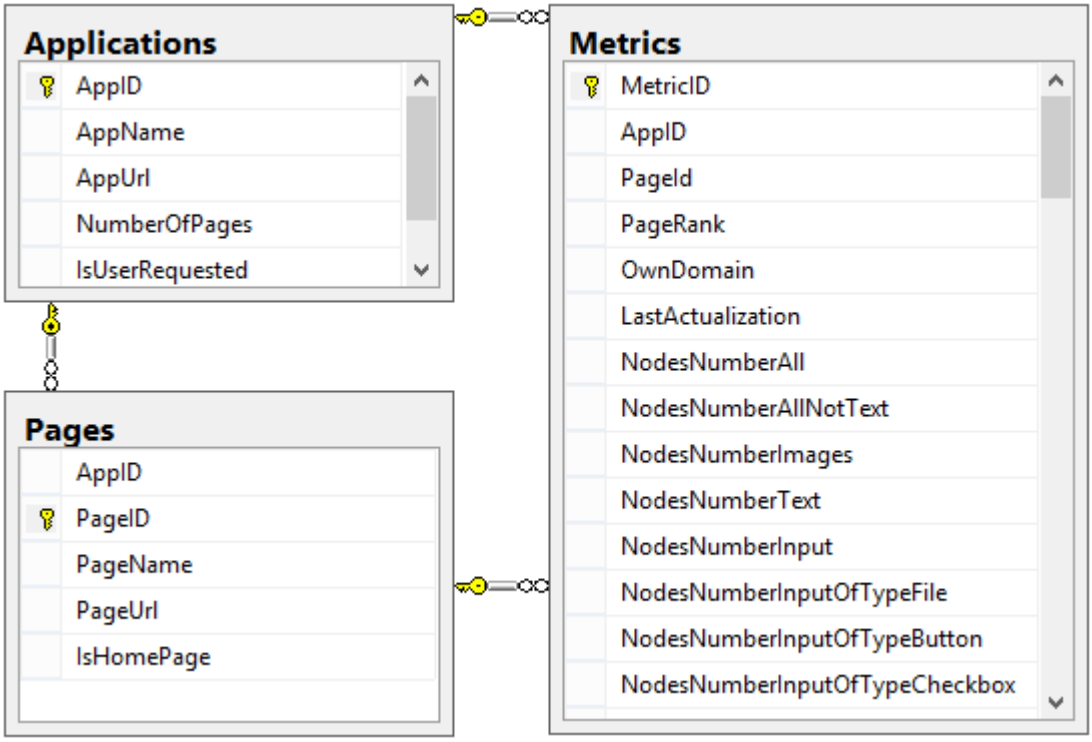

Figure 1 – *Persistent information aout the analyzed applications*

The list of URLs is passed in an iterative process. The number of steps is equal to the number of applications analyzed. At the beginning about the app the only known information is the URL. A new application entity gets created as a pre phase of the process and an identification token is generated for it. This information is saved in the database inside Applications table. App name will be filled in after parsing the title of the homepage. When all pages of the website are completely crawled the number of pages is filled in. The pre phase of the process is finished as long as the application entity is created and successfully saved into the database. The table of applications has the purpose of storing generic information about the app as url, number of pages and name.

The first step of the automated quality analyzing process is crawling the homepage. Crawling involves a standardized procedure for all pages so at this step the app url previously saved in the database is used to create a page entity who's info gets stored in Pages table. In pages table the only difference between the homepage and other pages on

the same app is a boolean value which tells weather a specific page is the entry point of crawling process or not. Page name is filled in as soon as the title of the page is parsed.

At the second step of the process the crawler gets the html content of the page as a crawled page entity. This entity is not saved in the database but is used as an entry point for the third step of the process which involves parsing the content of the page and calculating counting metrics on the go while analyzing specific html nodes.

The third step is where parsing the content of the page and feeding metrics calculation modules with information takes place. The content is analyzed at a node level and specific nodes are counted. The nodes to be counted are chosen based on the importance to citizen oriented applications as stated in table one. The work of the third step is done by the parser module of the tool that was built particularly for this process. The parser consists of two specialized submodules: a links parser and a html nodes parser.

The first submodule is the links parser that transforms all the crawled urls from string structures to uniform resource identifier structures using Microsoft Visual Studio specific libraries. Further, specific url parts are analyzed in order to get link metrics based on the URL.

The metrics analyzed using the links parser are:

- a boolean metric that specifies if the used protocol is secure; this metric is meaningful for citizen oriented applications because it tells about how secure user data is and it tells about how risky is to use data of high importance on the specific application; when dealing with administrative software solutions all data filled in the forms, all the documents and media files need to be safe so using secure protocol is recommended [2];
- counting metrics that specify if the analyzed url points to a certain type of file; on citizen oriented informatics applications users should be able to download administrative forms in order to prefill them, informative and media files; the file extensions analyzed are categorized; documents: .pdf, .doc, .docx, .rtf, .xls, .xlsx; pictures and media files: .gif, .jpg, .png, .flv; presentations: .ppt, .pptx; the number of files of each type is counted as parsing all links found on the page;
- specific domain names that tell if the link found on the page points to a certain website; sites checked: Facebook, Twitter, YouTube, PayPal; the social network links found on a citizen oriented page tells about how easy is the information reachable; in the automated quality check the number of links social links is a metric that tells how easy people can find useful administrative information; the existence of paypal links is a way to tell weather the application supports paypal payments so users can pay for certain services online; the links are analyzed as being internal or external; the internal links are the ones that point to a page of the app and the external links point to external sources.

The second submodule of the parser module is the HTML nodes parser. The crawled content is divided into nodes using HtmlAgillityPack library and xpath expressions. This is done in order to count the number of specific nodes on the page which tells about the nature of the application.

The metrics analyzed with the nodes parser are usually counting metrics calculated as searching for specific node types and for certain nodes an attribute level counting is done. All the metrics are presented in table 2.

4. Calculating the metrics

The presented steps of the automated quality evaluation process have been applied on a large number of apps. Table two presents raw results for a number of five apps out of them together with a total an average and a page level result.

Table 2 – *Web page metrics calculated for a set of citizen oriented informatics applications*

Metric	Ap 1	Ap 2	Ap 3	Ap 4	Ap 5	Total	Average	Per Page
PagesNumber	106	21	37	81	4	**249**	**50**	
NodesNumberAll	54292	6696	17429	13353	1423	**93193**	**18639**	**374.269**
NodesNumberAllNotText	26513	2511	7739	5832	812	**43407**	**8681**	**174.325**
NodesNumberImages	570	100	9	45	7	**731**	**146**	**2.936**
NodesNumberText	27779	4185	9690	7521	611	**49786**	**9957**	**199.944**
NodesNumberInput	302	3	158	3	0	**466**	**93**	**1.871**
NodesNumberInputOfTypeFile	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeText	104	1	42	3	0	**150**	**30**	**0.602**
NodesNumberInputOfTypeButton	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypePassword	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeSubmit	98	0	38	0	0	**136**	**27**	**0.546**
NodesNumberInputOfTypeRadio	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeCheckbox	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeNumber	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeEmail	0	1	0	0	0	**1**	**0**	**0.004**
NodesNumberInputOfTypeTime	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeDateTime	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeSearch	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeDate	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberInputOfTypeUrl	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberCanvas	0	0	0	0	0	**0**	**0**	**0.000**
NodesNumberIframe	0	1	0	0	0	**1**	**0**	**0.004**
NodesNumberForm	98	1	38	1	0	**138**	**28**	**0.554**
NodesNumberScript	532	109	259	287	0	**1187**	**237**	**4.767**
NodesNumberStyle	2	0	36	0	4	**42**	**8**	**0.169**
NodesNumberLink	285	80	218	369	5	**957**	**191**	**3.843**
NodesNumberDiv	8493	431	626	708	10	**10268**	**2054**	**41.237**
NodesNumberSpan	2052	14	712	47	153	**2978**	**596**	**11.960**
NodesNumberLabel	125	4	8	5	0	**142**	**28**	**0.570**
NodesNumberUl	1048	60	409	246	0	**1763**	**353**	**7.080**
NodesNumberLi	3909	481	1701	1353	0	**7444**	**1489**	**29.896**
NodesNumberComment	888	179	474	574	43	**2158**	**432**	**8.667**
NodesNumberProcessingInstruction	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberAll	4994	301	1444	1303	27	**8069**	**1614**	**32.406**
LinksNumberDoc	10	0	6	0	0	**16**	**3**	**0.064**

Metric	Ap 1	Ap 2	Ap 3	Ap 4	Ap 5	Total	Average	Per Page
LinksNumberDocx	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberGif	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberJpg	0	0	1	0	1	**2**	**0**	**0.008**
LinksNumberPng	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberPdf	585	8	32	0	21	**646**	**129**	**2.594**
LinksNumberRtf	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberFlv	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberPpt	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberPptx	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberXls	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberXlsx	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberFiles	595	8	39	0	22	**664**	**133**	**2.667**
LinksNumberDocs	10	0	6	0	0	**16**	**3**	**0.064**
LinksNumberPhotos	0	0	1	0	1	**2**	**0**	**0.008**
LinksNumberPresentations	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberExcel	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberInternal	54	15	37	29	4	**139**	**28**	**0.558**
LinksNumberExternal	4940	286	1407	1274	23	**7930**	**1586**	**31.847**
LinksNumberYouTube	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberFacebook	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberTwitter	0	0	0	0	0	**0**	**0**	**0.000**
LinksNumberPaypal	0	0	0	0	0	**0**	**0**	**0.000**
IsSecure	0	0	0	0	0	**0**	**0**	**0.000**

The table presents the results of crawling and parsing concreted in counting metrics for a set of five citizen oriented web based software solutions. The results have been calculated for all individual pages of the five apps and each "app" column presents the already aggregated results.

The first column holds the metric name. They are basically divided in two categories depending on the submodule used to parse them:

- links metrics: the links pointing to downloadable files categorized as: media files, documents and presentations; links pointing to social networking websites, links to paypal; the total number of links: external and internal links;
- html nodes metrics: the total number of nodes; the total number of text nodes; the total number of input nodes which tells about the level of interactivity; the input nodes are analyzed based on their "*type*" attribute as well as their number.

Following the results presented in table two on average each analyzed app has a number of fifty pages with a minimum of four pages for app five and a maximum of 106 pages for app one. This average value is quite high because the results for the entire sample shows that the administrative web based software solutions have a fewer number of pages.

On average each application has a number of 18639 html nodes with a corresponding nodes per page ratio of 374.26 nodes. A meaningful result is that out of the 374.26 nodes found per page 174.32 nodes are text nodes which corresponds to 46.57%.

This tells that almost half of the nodes on a page are text nodes. This value shows a high diversity of nodes for these citizen oriented applications as long as usually the text nodes rate is higher. Although text content is important, a big ratio of non-text nodes, in this case almost 50%, tells that user experience is enhanced by providing extra features other than just plain text.

It is considered that for a citizen oriented solution providing user with the possibility to input data and presenting him results based on his data brings flexibility and dynamism to the human-computer interaction process [3][4]. Out of all the apps on app number five no input nodes were found which makes it categorized as a static app. Within the citizen oriented metrics calculated the number of downloadable files provided to citizen counts when analyzing the interactivity level of the page but for app number five the downloadable files number is zero. At the other end app number one has an absolute value of 302 input nodes but although app number three has an absolute value of 158 input nodes at a page level the ratio is better for app three (4.27 input nodes per page) then app one (2.84 input nodes per page). On average on the five apps have 1.87 input nodes per page so the assumption can be made that results are displayed based on citizen requests. This is because none of the apps have input nodes of type password which means that users cannot create accounts on the websites so the almost 2 input nodes per page are for different kind of interaction than logging in.

Analyzing the metrics calculated using the link parser module the last row of table two highlights that none of the apps has secure protocol pages. The same value was obtained for PayPal links so most probably none of the apps supports online payments.

On average 28 internal links per app were found and 1586 links pointing to external sources per app. This leads to 0.5 internal links per page. The internal links number tell about how easy the user can navigate from a section of the website to another and the higher the number the quicker the access to other pages. The value being under one means that on average users can navigate to maximum one different page on the same app which decreases application's usability level [5]. The external links number have a better ratio at a page level with 31.8 links per page pointing to an external resource.

Informatics applications provided by administrative authorities should provide downloadable useful content to citizens. On average each one of the analyzed apps provides to citizens 133 files for download. This average value is very high because of app number one which provides a total of 595 files for download but app number four provides no files and the rest of them under fifty. Out of the total of 664 files provided by all five apps 646 are .pdf files 4 photos and 16 .doc files.

None of the applications provides the feature of user file upload as long as the count for input nodes of type file is 0. Although the downloadable content exists at a good rate the upload feature doesn't. This highlights the fact that no services that require citizen scanned documents or forms are implemented.

5. Conclusions

This paper presents a way to evaluate the quality of the services provided by a big number of citizen oriented web based software solutions using automated tools. The analyze is done based on the content the local governance authorities expose to the public via their websites. The chosen metrics are meaningful for evaluating how good the services provided

to citizens are. Further work is to be done in order to be able to get information about how users actually interact with the applications and correlate it with the results obtained using the metrics. This implies using analytics services at owners initiative and shared information about what the services measure.

Bibliography

1. Emanuel Eduard HERȚELIU, **Quality Metrics for Evaluating Interactivity Level of Citizen Oriented Web Based Software Solutions**, The 14th International Conference on Informatics in Economy, Research & Business Technologies, Bucharest Academy of Economic Studies, 30 April, 2015, Bucharest, Romania
2. Sai Peck LEE, Malik HNEIF, **Using Guidelines to Improve Quality in Software Nonfunctional Attributes, IEEE Software**, Vol. 28, Issue 6, November/December 2011, pp. 72-77
3. Ion IVAN, Bogdan VINTILĂ, Dragoș PALAGHIȚĂ, **Types Of Citizen Orientated Informatics Applications**, Open Education Journal, Russia, No.6, 2009
4. Stephen KAN, **Metrics and Models in Software Quality Engineering**, Publisher: Addison Wesley, Boston, 2002, pp. 560
5. Robert BAGGEN, José Pedro CORREIA, Katrin SCHILL, Joost VISSER, **Standardized code quality benchmarking for improving software maintainability**, Software Quality Journal, Vol. 20, Issue 2, June 2012, pp. 287-307