WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

Vol. 7
No. 3
Fall
2012

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

Contents

JOURNAL OF APPLIED QUANTITATIVE METHODS

CREDIT ACCELERATOR, CDS RATE AND LONG TERM YIELDS: EMPIRICAL EVIDENCES FROM THE CEE ECONOMIES

Bogdan-Gabriel MOINESCU
PhD, Assoc. Prof.
Bucharest University of Economic Studies
and National Bank of Romania

E-mail: bogdan.moinescu@fin.ase.ro

Abstract:

The study aims to investigate the mechanism by which lending to private sector may induce risks to the long-term interest rates convergence process in the new EU Member States. The added value of this approach consists of three elements. First of all, the analysis provides a quantitative mechanism for assessing the fundamental dependence of the bank portfolio quality to the dynamics of the credit accelerator, econometric results showing that about 30 percent of the squared change in the private sector credit flow is reflected in the jump of the rate of non-performing loans. Secondly, the study shows that sovereign risk premium is dependent on the stability of the banking system, considering that about 20 percent of the changes in the rate of non-performing loans are reflected in the level of the CDS rate. Third, empirical assessment highlights the importance of the sovereign risk premium transmission channel related to long-term interest rate, with approximately two thirds of the CDS rate contributing to the level of government bonds long-term yields. In this context, promoting a mix of macroeconomic policies oriented also to limiting the volatility of credit demand accompanied by poor multiplier effects in the economy becomes a fundamental requirement for ensuring a sustainable cost of financing long term public debt.

Key Words: credit accelerator, nonperforming loans, CDS rate, long term yields, nominal convergence, panel regressions, emerging economies

Introduction

The severity of the recent economic crisis in most new EU Member States shows that economic policies must be cautious in managing the process of economic catching up to euro area, in order to ensure that real convergence takes place while maintaining macroeconomic stability. Promoting pro-cyclical policies, in order to meet the population's excessive expectations related to fast increase of income, fails to yield sustainable results, especially given that the swift rise in living standards is supported by an accelerated indebtedness of real economy. Implementing lax policies during economic boom contributes both to the accumulation of systemic vulnerabilities in the banking sector, by excessively feeding loan demand with modest multiplier effects in the economy, as well as to a considerable reduction in borrowers' repayment capacity during recession, caused by major adjustments of investment budgets and other negative fiscal impulses.

The desire to rapidly advance in increasing living standards by resorting to bank loans has proven to be part of the ingredients for an unsustainable economic growth, in the case of a significant number of CEE countries. Alternatives which are available to banks, given the capital account liberalization, provide macro-prudential policy with limited power compared to fiscal and income policies, in the process of tempering unhealthy credit expansion in the economy. Although on the short term they produced noticeable effects, many prudential measures adopted by CEE countries have lost effectiveness over time, especially in the context of financial integration after entering the EU (**Georgescu**, 2010). In addition, when credit demand is very strong, actions aiming to limit financing supply by using the solvency channel become insufficient, given the rapid growth of profitability, based on the swift increase of business volume.

Bernanke, Gertler and Gilchrist (1999) developed a general dynamic equilibrium model that includes credit market imperfections in the explanatory framework of business cycle evolution. The central piece is represented by the financial accelerator, the framework assuming that financial system is not an independent source of volatility, but acts as an amplifier of exogenous shocks. This concept reflects the role of financial markets in augmenting and spreading macroeconomic shocks (Bernanke, 2007). Furthermore, recent studies showed that one of the mechanisms of global crisis spread in CEE countries is represented by the financial channel (Becker et al., 2010). At the same time, failure to distinguish between temporary and permanent influences on budget revenues has consumed the operating space used by fiscal policy in taking actions towards stabilization (Isarescu, 2011). In addition, looming additional budgetary expenditures in order to maintain financial stability has increased the pressure on the public finance stance, in the context of increased yields required by investors for purchasing government securities. Increase of the sovereign risk premium in conjunction with the dynamic of bank loan portfolio quality has represented, along with the output gap, an important channel for the distribution of second-round effects, while the real and financial economy became more and more interdependent.

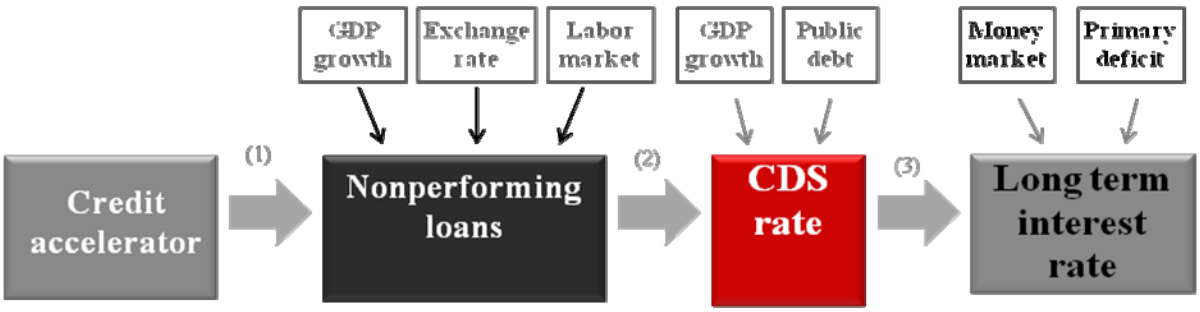
In this context, the study aims to investigate the mechanism by which acceleration of private sector lending may induce risks to long-term interest rates convergence process in the new EU Member States. The operational objective is to build a simplified financial satellite, based on three components, modelling the long-term bond yields dependence on the interaction between sovereign risk premium and the dynamic of non-performing credit loans amid material credit impulses in the economy.

The rest of the paper is structured as follows. The second section presents the methodology underlying the analytical framework for assessing risks induced by the credit accelerator to the evolution of long-term interest rate, emphasizing the main functional forms used. The third section presents the data used in the study and describes in detail both the underlying economic foundation, as well as preliminary statistical results that lead to the selection of explanatory variables. Section four provides an overview of key empirical issues in developing a financial satellite model, which favors estimation of the impact that the private sector credit flow dynamic has on meeting the long-term interest rates convergence criterion.

1. METHODOLOGICAL FRAMEWORK

The analytical framework used for assessing credit effects on long-term interest rates for CEE countries is based on a three components transmission mechanism. The first step is represented by the effect of credit growth on the dynamics of non-performing loans, given that the volatility of private sector credit flow has a direct proportional effect on the quality of bank portfolios. Second step is represented by the deterioration of sovereign risk due to the depreciation of bank loan quality. Step three consists in spreading the CDS rate effects to government bonds long-term yields (see Figure 1).

Figure 1: Transmission mechanism of credit accelerator to long term yields



1.1 Lending impact on the quality of credit portfolio

The dynamics of non-performing loans (NPL) transmission channel is based on the premise that both strong accelerations in lending as well as sudden deceleration feed the increase of credit risk. The harmful effect of contracting financing flow on the repayment of existing loans is similar to inefficient allocation of bank resources (**Jakubik and Moinescu**, 2012), considering the intensification of the struggle for market share and excessive lending, which increases the risk of financing more unfeasible projects amid loose credit conditions.

Thus, the dynamics of non-performing loans (see equation 1) is directly proportional to the squared credit accelerator ($\alpha_1 > 0$), defined as the first order difference of the private sector credit flow, expressed as percentage of GDP.

$$d(NPL_t^i) = \alpha_1 \times (Credit\ Accelerator_t^i)^2 + \beta_1 \times Macro_t^i + \gamma_1 \times Market_t^i + C_1^i \quad (1)$$

The conceptual model for the dynamics of non-performing loans ($d(NPL_t^i)$) assumes a linear relationship, where the set of determinants also includes macroeconomic variables ($Macro_t^i$), such as economic growth, the average gross income and number of employees in the economy.

The mentioned macroeconomic indicators influence in a positive manner the capacity for repayment ($\beta_1 < 0$). The functional form of the explanatory equation also includes financial market variables such as the exchange rate and interbank interest rate, which affects directly and proportional borrowers' financial burden ($\gamma_1 > 0$).

1.2 Loan portfolio quality impact on sovereign risk premium

Increase of the non-performing rate in loan portfolio generates the need of recapitalising banks, which is sometimes covered only from public resources, and also deteriorates investors' perception of sovereign risk, which is followed by significant upward movements of CDS rate ($\alpha_2 > 0$ – see equation 2).

$$CDS_t^i = \alpha_2 \times d(NPL)_t^i + \beta_2 \times Real\ Economy_t^i + \gamma_2 \times Public\ finance_t^i + (\quad (2)$$

The functional form of the sovereign risk premium explanatory equation also includes the inverse relationship with the stance of real economy ($\beta_2 < 0$), expressed by GDP growth and the flow of foreign direct investments. The CDS rate explanatory equation also includes positive dependence on public finance accumulated deficits ($\gamma_2 > 0$), expressed by the share of government debt and budget balance in GDP.

1.3 Sovereign risk premium impact on long-term interest rate

The dynamics of sovereign risk premiums is subsequently reflected in the performance required for the issuance of bonds (see equation 3).

$$LTY_t^i = \alpha_3 \times CDS_t^i + \beta_3 \times MM_t^i + \delta_3 \times Public\ Finance_t^i + C_3^i \quad (3)$$

Along with sovereign risk premium, long-term interest rate explanatory equation also includes the dependence on interbank rates (MM) and on public finances stance, that captures the positive connection between the evolution of financing need and the cost of attracting resources. Structural differences between CEE economies are also captured by fixed effects of panel estimation.

2. DATA

Private sector credit variable is expressed by the indicator *private sector credit flow in % of GDP*, provisioned in the European Commission's macroeconomic imbalance procedure, while the long term interest rate variable is the *yield of long term bonds*, provided by the Maastricht criteria.

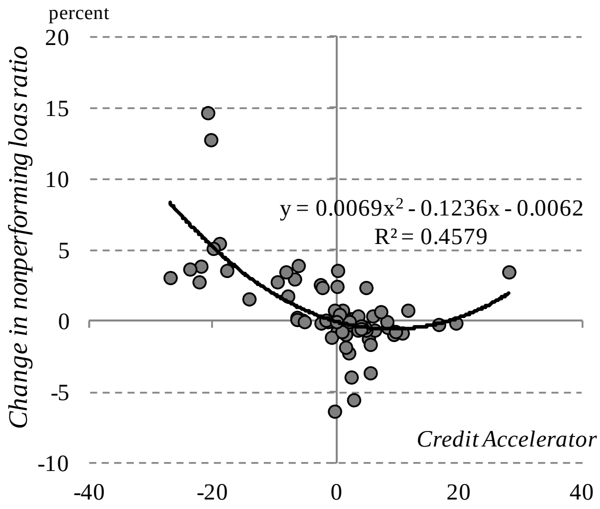
Information underlying the assessment of the impact that credit has on long-term interest rates in the CEE Member States is represented by annual frequency data covering the period 2000 to 2011. The countries under consideration are Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia.

The main source of information is represented by Eurostat, from which were extracted data on credit flow to the private sector, long-term interest rates, economic growth, the average gross income, the number of employees in the economy, foreign direct investment, exchange rate and long term interbank interest rate, the exchange rate, the average inflation rate and the primary deficit of the state budget. The data on non-performing loans rate were extracted from the International Monetary Fund reports on indicators of financial stability and sovereign risk premium was calculated based on daily information extracted from Bloomberg platform.

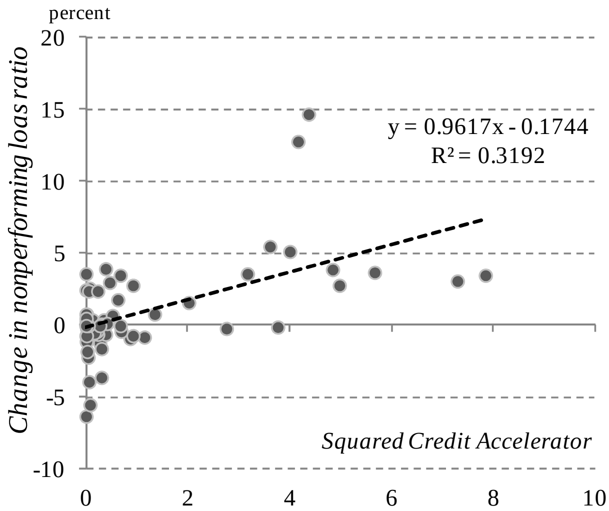
Preliminary empirical analysis shows that the credit acceleration in CEE countries was one of the main factors favoring the accumulation of nonperforming loans (see Chart 1).

Chart no. 1 – Correlation between credit flow and

Chart no. 2 – Correlation between credit the change in non-performing loans



accelerator and the change in non-performing loans

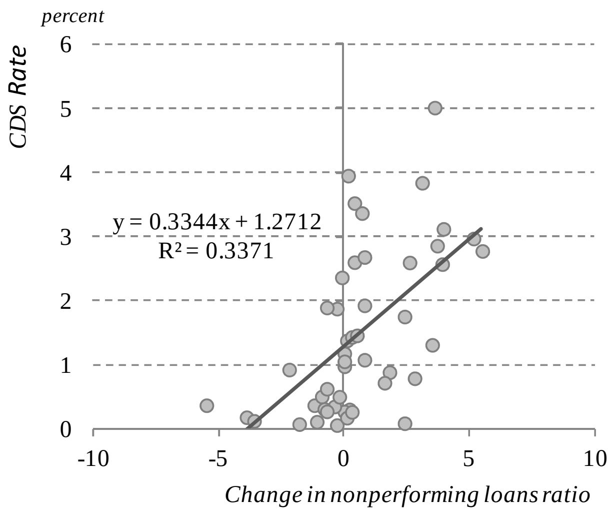


Data source: Eurostat, FMI, own calculations

Univariate tests show a consistent elasticity of the rate of non-performing loans to squared credit accelerator in the area of the new EU member states, given an explanatory power of functional connection of more than 30 percent (see chart 2).

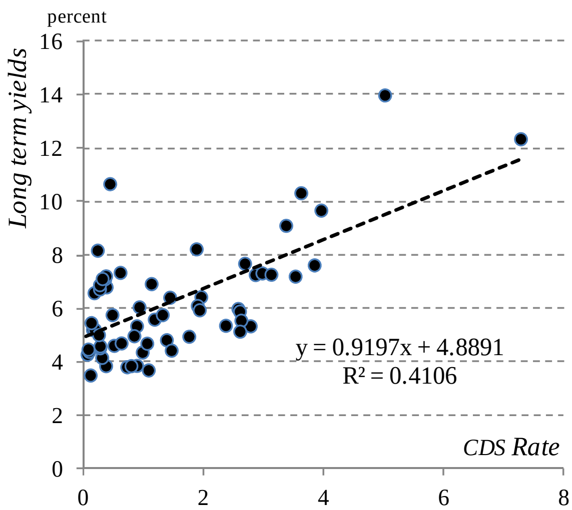
At the same time, empirical evidence in CEE countries shows that deterioration of the credit portfolio quality in the region has increased sovereign risk premium, CDS rate being positively driven by increase in the rate of non-performing loans. Univariate assessment of the sovereign risk premium dependence on the evolution of non-performing loans indicates a significant causal linear form (see Chart 3), both in terms of elasticity levels (about 33 percent) as well as in the degree of determination (approximately 33 percent). The impact occurs instantaneously.

Chart no. 3 – Correlation between the rate of non-performing loans and CDS rates



Data source: FMI, Bloomberg, own calculations

Chart no. 4 – Correlation between CDS rates and long term yields



Data source: Eurostat, Bloomberg, own calculations

Subsequently, the changes in sovereign risk premium propagate almost entirely in long-term interest yields, explaining slightly more than 40 percent of its variance (see chart 3).

The candidate indicators for structuring the models and their expected impact on the dependent variables together with the applied transformation are provided in Table 1.

Table 1. The candidate explanatory variables and the corresponding equations

	Explanatory variables	Expected sign
Equation 1: Nonperforming loans ratio		
1	Squared credit accelerator	+
2	GDP growth	-
3	Earnings (log transformation)	-
4	Employment (log transformation)	-
5	Exchange rate (log transformation)	+
6	Money market interest rate (3M)	+
Equation 2: CDS rate		
1	Nonperforming loans ratio	+
2	GDP growth	-
3	Foreign direct investments (log transformation)	+
4	Private debt (as percent of GDP)	+
5	Public debt (as percent of GDP)	+
6	Primary budgetary balance (as percent of GDP)	-
7	Current account (as percent of GDP)	-
Equation 3: Long term yields		
1	CDS rate	+
2	Inflation	+
3	Money market interest rate (3M)	+
4	Primary budgetary balance (as percent of GDP)	-

Stationarity of the considered indicators was tested. All indicators were I(0) after the appropriate transformation and the first difference. Furthermore, the univariate OLS panel regression was used to make the first selection of variables based on statistical relevance. The applied procedure tested variables on one-by-one basis up to two lags, including the contemporary impact, for each explanatory variable (see Annex 1).

3. EMPIRICAL ANALYSIS

Multivariate empirical assessment is based on a standard backward estimation procedure using macroeconomic factors short-listed in the previous section. The analytical component consists of a set of simplified econometric models, built by panel estimations using annual data, structuring the mechanism by which credit rate affects long-term interest rates.

The first equation of the financial satellite models the dynamics of the rate of non-performing loans. Empirical results confirm that the squared credit accelerator increases credit risk, with a strictly positive coefficient, statistically significant at a probability of 93 percent (see Table 2).

Tabel 2 – Multivariate model estimation output for the non-performing loans ratio

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP growth	-0.326519	0.046012	-7.096344	0.0000
(Credit accelerator) ^2	0.317049	0.170207	1.862720	0.0668
C	1.525334	0.362907	4.203099	0.0001
Fixed Effects (Cross)				
_BG--C	0.594569	_HUN--C		-0.539924
_CZ--C	-0.417393	_POL--C		-1.694735
_EE--C	-0.465320	_RO--C		0.100351
_LET--C	1.015640	_SK--C		0.253345
_LIT--C	1.343094	_SLO--C		-0.189628
Adjusted R-squared	0.597229			
Durbin-Watson stat	1.831699			

The evolution of non-performing loans rate in CEE countries depends, at the same time, on economic growth, each in a ratio of one to three. Thus, in order to prevent increase in the rate of non-performing loans by one percentage point, economic growth of about three percent would be required. These two determinant factors together explain about 55 percents of the variance of the rate of non-performing loans dynamics. Econometric estimations also suggest that there are some structural differences between countries in the sample in terms of loan portfolio quality, with statistically significant fixed effects. However, these structural differences are minor, the model estimated without fixed effects leading to a similar result, only marginally reduced in performance (from 60 to 56 percents).

The estimation result of the CDS rate equation confirms the dependence of sovereign risk premium on the banking system stability, the increase of non-performing loans through a credit impulse being accompanied by an increase in the sovereign risk premium of about 20 percent (see Table 3). The effect occurs in the same year.

Table 3 – Multivariate model estimation output for CDS rate

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP growth	-0.096790	0.025421	-3.807517	0.0003
Change in NPL	0.197709	0.051291	3.854668	0.0003
C	1.478602	0.144212	10.25296	0.0000
Fixed Effects (Cross)				
_BG--C	0.218462	_HUN--C		0.001077
_CZ--C	-0.616806	_POL--C		0.141473
_EE--C	-0.115022	_RO--C		0.614289
_LET--C	0.732251	_SK--C		-0.401846
_LIT--C	0.209642	_SLO--C		-0.831465
Adjusted R-squared	0.704068			
Durbin-Watson stat	1.539265			

At the same time, economic growth acts by reducing sovereign risk premium, with a negative coefficient (about -0.1) and statistically significant for a probability of one percent. The estimated multifactorial functional form manages in capturing slightly more than 70 percent of the CDS rate variance, by also taking into account, through fixed effects, structural differences in sovereign risk. A result which was less expected is the absence of public finances indicators in the final configuration. One possible reason in this respect is expressed by the relatively low level of public debt to GDP ratio in the CEE economies, which probably prompts sovereign risk insurance providers to only marginally include in its price a component related to public finance. At the same time, the basic level of sovereign risk premium is covering enough for insurance providers, with the intercept taking values of around 150 basis points, starting with Slovenia (83 basis points less than the average) and ending with Latvia (73 basis points more than the average). However, the relatively low value of the DW indicator shows the existence of significant autocorrelation between error terms, which indicates the existence of additional determinants. Their nature is most likely of regional contagion, reflecting indirect effects of the risk premium dynamics in countries such as Greece or Austria, provided that the set of macroeconomic variables included only internal sources of risk.

The multifactor configuration of the long term interest rate equation confirms its dependence on sovereign risk premium, given that the econometric estimation generated a 67 percent value for the variable coefficient associated to CDS rate (see Table 4).

Tabel 4 – Multivariate model estimation result for long term interest rate

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CDS rate	0.674473	0.104083	6.480127	0.0000
Money market interest rate	0.341499	0.048892	6.984724	0.0000
Primary balance	-0.125469	0.049179	-2.551257	0.0132
C	3.160937	0.300367	10.52357	0.0000
Fixed Effects (Cross)				
_BG--C	-0.325637	_HUN--C		0.455565
_CZ--C	-0.464669	_POL--C		0.263713
_EE--C	0.846987	_RO--C		-0.367869
_LET--C	-0.133740	_SK--C		-0.484542
_LIT--C	0.098006	_SLO--C		-0.149079
Adjusted R-squared	0.755169			
Durbin-Watson stat	2.044825			

Besides CDS rate, the final functional form also includes money market interest rate, which contributes to the level of long-term interest rate in a proportion of 34 percent. The multivariate configuration also captures the impact of financing need on the price asked by investors for buying long term bonds, given that slightly over 12.5 percent of the primary deficit is reflected in the increase of long-term yields. The three determinant factors together explain three quarters of the variance of long term interest rate, taking into account the slight structural differences between CEE economies in this regard. The values of individual constants are relatively low compared to the intercept of the equation (3.16).

FINAL REMARKS

The main contribution of this study is to highlight the ability of credit accelerator theory to explain a significant part of the evolution of long-term interest rates registered in the countries of Central and Eastern Europe.

The added value of this approach consists of three elements. First of all, the analysis provides a quantitative mechanism for assessing the fundamental dependence of the bank portfolio quality to the dynamics of the credit accelerator, econometric results of this study showing that about 30 percent of the squared change in the private sector credit flow is reflected in the jump of the rate of non-performing loans. Secondly, the study shows that sovereign risk premium is dependent on the stability of the banking system, considering that about 20 percent of the changes in the rate of non-performing loans are reflected in the level of the CDS rate. Third, empirical assessment highlights the importance of the sovereign risk premium transmission channel related to long-term interest rate, with approximately two thirds of the CDS rate contributing to the level of government bonds long-term yields.

At the same time, as the credit accelerator theory indicates a significant impact of credit change on economic growth, we advise on a cautious interpretation of the results. During times of financial disintermediation, non-performing loans can record jumps higher than what can be captured by the analytical framework developed in this study and thus, the sovereign risk premium could be higher in reality. In such a challenging context, we can expect even larger deviations of bond yields from convergence tendency, especially when there are signs of consolidating the dependence of CDS rate on the non-performing loans dynamic.

Thus, promoting a mix of macroeconomic policies also oriented to limiting the volatility of credit demand accompanied by poor multiplier effects in the economy becomes a fundamental requirement for ensuring a sustainable cost of financing long term public debt.

AKNOWLEDGEMENTS

This work was supported by the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project POSDRU/89/1.5/S/59184 "Performance and excellence in postdoctoral research in Romanian economics science domain".

REFERENCES

1. Becker T., D. Daianu, Z. Darvas, V. Gligorov, M. Landesmann, P. Petrovic, J. Pisani-Ferry, D. Rosati, A. Sapir and B.Weder di Mauro, **Whither growth in central and eastern Europe?** Policy lessons for an integrated Europe, *BRUEGEL BLUEPRINT SERIES,* Vol. XI, 2010
2. Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, **The Financial Accelerator in a Quantitative Business Cycle Framework**, in handbook of *Macroeconomics*, Volume 1C, Handbooks in Economics, Volume 15, Amerstdam: Elsevier, 1999, pp. 1341–93 (earlier version: NBER Working Paper, 5455.).
3. Bernanke, B, **The Financial Accelerator and the Credit Channel**, 2007 http://www.federalreserve.gov/newsevents/speech/Bernanke20070615a.htm
4. Georgescu Florin, **Contribuţia politicii prudenţiale a BNR în cadrul mix-ului de politici economice, Banca Naţională a României**, 2010, http://www.bnr.ro/PublicationDocuments.aspx?icid=6885
5. Isărescu Mugur, **Măsurarea caracterului anticiclic al politicii fiscale.** Diserţie cu ocazia decernării titlului de Doctor Honoris Causa al Universităţii din Oradea, 2011
6. Jakubik Petr and Bogdan Moinescu, **Assessing optimal credit growth for an emerging banking system**. Paper given at the 10th INFINITI Conference on International Finance, Trinity College and Journal of Banking and Finance, Dublin, Ireland, June 11-12, 2012

Annex 1 – Univariate analysis results

Variable	Coefficient	Std. Error	t-Statistic	Prob.	Adj.-R2	DW
Nonperforming Loans						
(Credit accelerator)^2	**0.010111**	**0.001824**	**5.54222**	**0**	**0.309113**	**1.6935**
(Credit accelerator)^2	0.008329	0.002042	4.078731	0.0001	0.195519	1.940889
(Credit accelerator)^2	0.002016	0.002542	0.79302	0.4305	0.009457	1.621014
GDP growth	**-0.37577**	**0.038325**	**-9.8049**	**0**	**0.582813**	**1.74357**
GDP growth (-1)	-0.138796	0.056338	-2.46361	0.0163	0.082281	1.810661
GDP growth (-2)	0.055424	0.058816	0.942333	0.3493	0.014242	1.44969
Gross earnings	**-0.174591**	**0.036261**	**-4.8148**	**0**	**0.252648**	**1.59397**
Gross earnings (-1)	0.061813	0.040342	1.532233	0.13	0.034411	1.484915
Gross earnings (-2)	0.156779	0.036219	4.328625	0	0.214783	1.760888
Employment	**-0.482558**	**0.070252**	**-6.8689**	**0**	**0.40703**	**1.91391**
Employment (-1)	-0.081103	0.092131	-0.88029	0.3818	0.012645	1.57988
Employment (-2)	0.202004	0.09486	2.129502	0.0368	0.063129	1.528684
Change in money market IR	0.124689	0.131535	0.947954	0.3465	0.014392	1.426249
Change in money market IR (-1)	0.209245	0.124227	1.68437	0.0966	0.040988	1.718849
Change in money market IR (-2)	**0.448351**	**0.11649**	**3.848843**	**0.0003**	**0.178026**	**1.888082**
Change in Exchange Rate	**0.158267**	**0.081383**	**1.94471**	**0.0577**	**0.07901**	**1.55985**
Change in Exchange Rate (-1)	-0.020318	0.073465	-0.27657	0.7833	0.008027	1.493197
Change in Exchange Rate (-2)	-0.087752	0.069544	-1.26181	0.2131	0.038344	1.530851
CDS Rate						
Change in NPL	**0.346704**	**3.675562**	**9.43269**	**0**	**0.636369**	**1.38291**
Change in NPL (-1)	0.20301	0.053715	3.779362	0.0004	0.271692	2.253917
Change in NPL (-2)	0.121572	0.060448	2.011167	0.0498	0.160692	1.629668
GDP growth	**-0.163851**	**0.022393**	**-7.31692**	**0**	**0.51347**	**0.88701**
GDP growth (-1)	-0.097538	0.025155	-3.87742	0.0002	0.291146	1.700063
GDP growth (-2)	-0.041313	0.026472	-1.560645	0.1232	0.160539	1.308615
Foreign direct investment	**-0.052032**	**0.009368**	**-5.55409**	**0**	**0.401207**	**1.70138**
Foreign direct investment (-1)	-0.049995	0.008781	-5.693363	0	0.408689	1.801416
Foreign direct investment (-2)	-0.019761	0.010695	-1.847571	0.0693	0.152551	1.484769
Private debt (%GDP)	0.008911	0.005547	1.606367	0.1125	0.185449	0.977287
Private debt (%GDP) (-1)	0.017548	0.004925	3.56316	0.0007	0.271342	1.296305

Private debt (%GDP) (-2)	**0.024529**	**0.004455**	**5.505905**	**0**	**0.398771**	**1.60089**
Public debt (%GDP)	**0.079844**	**0.011047**	**7.227898**	**0**	**0.508441**	**1.4149**
Public debt (%GDP) (-1)	0.045255	0.014009	3.230296	0.0019	0.251108	1.340338
Public debt (%GDP) (-2)	0.007167	0.016713	0.428835	0.6694	0.133217	1.182794
Primary deficit	**-0.128946**	**0.057541**	**-2.24095**	**0.028**	**0.210937**	**0.94383**
Primary deficit (-1)	-0.145043	0.059536	-2.436231	0.0174	0.207309	1.268233
Primary deficit (-2)	-0.064638	0.065271	-0.990301	0.3255	0.143086	1.245513
Current account	0.110609	0.033369	3.314762	0.0014	0.266986	1.163738
Current account (-1)	-0.040087	0.041404	-0.968182	0.3362	0.152237	1.060357
Current account (-2)	**-0.162442**	**0.04363**	**-3.7232**	**4E-04**	**0.276299**	**1.32155**
Long term yields						
CDS Rate	**0.859373**	**0.127127**	**6.75998**	**0**	**0.557913**	**1.41086**
CDS Rate (-1)	0.151712	0.163338	0.928822	0.3568	0.33775	2.059493
CDS Rate (-2)	-0.319933	0.14312	-2.23542	0.0299	0.484513	2.598009
Inflation	0.020591	0.069357	0.296883	0.7672	0.187995	1.192446
Inflation (-1)	**0.328496**	**0.050811**	**6.46507**	**0**	**0.442556**	**1.32486**
Inflation (-2)	0.237739	0.053031	4.483034	0	0.37402	1.981305
Money market interest rate	**0.43675**	**0.076438**	**5.71377**	**0**	**0.466388**	**1.42123**
Money market interest rate (-1)	0.375622	0.066851	5.618816	0	0.460252	2.161636
Money market interest rate (-2)	0.11795	0.075958	1.552841	0.126	0.246796	1.657688
Primary deficit	**-0.30485**	**0.067166**	**-4.5387**	**0**	**0.335916**	**1.19682**
Primary deficit (-1)	-0.151356	0.074509	-2.0314	0.0451	0.221806	1.37835
Primary deficit (-2)	-0.021028	0.076361	-0.27537	0.7837	0.224949	1.437885

The variables highlighted in bold are those retained for the multivariate analysis, considering their univariate fitting performance.

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

NOWCASTING ECONOMIC TIME SERIES: REAL VERSUS FINANCIAL COMMON FACTORS

Antonio FRENDA[1]
Instituto Nazionale di Statistica, Rome, Italy

E-mail: frenda@istat.it

Sergio SCIPPACERCOLA[2]
Università degli studi di Napoli "Federico II", Napoli, Italy

E-mail: sergio.scippacercola@unina.it

Abstract:

In this paper we want to assess the impact of real and financial variables in nowcasting smoothed GDP. We implement the generalized dynamic factor model, on which Eurocoin indicator is based. We can assess that, during the structural break in 2008, the impact of real variables in estimating smoothed GDP becomes particularly relevant in relation to that concerning financial data as money supply, spreads.

Key words: *Nowcasting; Eurocoin Approach; Medium to' long run component of the growth; Real and Financial Common Factors*

1. INTRODUCTION

Eurocoin, an important application of dynamic factor model, is an indicator of the Euro area economic activity concerning the medium to long-run growth, published monthly by the Bank of Italy and CEPR.

New Eurocoin (NE) has been recently created (Altissimo et al. 2006); it is a timely estimate of the medium to long-run component of euro area GDP (gross domestic product) and it has a measure of performance. NE can be described through the projection of the whole Euro Area bandpassed gross domestic product on a set of regressors – the linear combination of variables contained in the *Thomson Financial Datastream* used by the Bank of

Italy. NE provides an index of the current economic situation in the Euro Area, extracting from the Data Source relevant information which represent the main sources of variation.

We analyze the "medium to long-run component of the growth" (MLRG) that is not precisely the growth-rate cycle or the "business cycle", as in the definition of a cycle even the oscillations of a period longer than 8 years are generally removed (Stock and Watson, 1999). In fact, we are interested in the performance of our indicators with respect to a measure of the "trend-cycle growth" in nowcasting (Banbura et al., 2010) smoothed GDP.

The main aim of this paper is to propose a theoretical framework for implementing Eurocoin methodology by dividing European variables used to build common latent factors, in real and financial variables. We show that a combination between "real MLRG" and "financial MLRG", can be useful to analyze the impact of real and financial variables (e.g. Spread) in estimating smoothed GDP. This subdivisions among real economy and financial economy is substantially confirmed in Forni et al. (2003), where they study the impact of financial variables on real data. Our procedures are based on the Eurocoin methodology in order to obtain smoothing of a stationary time series, therefore avoiding the occurrence of end-of-sample deterioration.

We build real-time monthly estimates of GDP growth purified from seasonal and other short run fluctuations, as well as from errors in the measurement of GDP, and highly reliable at the end of the sample. In fact, Euro Area GDP is a comprehensive measure of economic activity, but:

1. it is released on a quarterly frequency, with a certain delay and may be subject to significant revisions afterwards;
2. GDP growth may be high or low in any quarter depending on seasonal effects and measurement errors.

Removing erratic components can also be done, for example, by applying a band pass filter to the GDP growth series. This technique, however, presents the same problems in terms of frequency and timeliness, producing some estimates that deteriorate at the end of the sample. Previous research has relied on "two-sided filters" to eliminate seasonal and short-run high frequency noise (Baxter and King, 1999; Christiano and Fitzgerald, 2003).

This is the main technical reason why it is worthwhile to develop the disaggregated indicators that we will present in detail in the following sections.

In section 2 we describe the econometric methodology to analyze our data. In section 3, we show that the *real time performance* is a reasonable approach for the examination of estimate accuracy. Not all observations can be used in estimating parameters. The latter sample will be used to build pseudo estimates by a recursive or a rolling window. We test our models using *pseudo real time estimates* at the end of sample. Real time estimates will be compared to bandpassed Euro Area growth components and we will assess if the in-sample results are certifiable.

2. THE GENERALIZED DYNAMIC FACTOR MODEL

The *generalized dynamic factor model*, on which *Eurocoin* indicator is based, must have two characteristics: it must be dynamic, because business cycle questions are typically non-static. Secondly, it must allow for cross-correlation among idiosyncratic components, as orthogonality is an unrealistic assumption for most applications. An important feature of this model is that the common component is allowed to have an infinite moving average (MA) representation, so as to accommodate for both autoregressive (AR) and MA responses to common factors. Dynamic factor model is more general than a static-factor model in which lagged factors are introduced as additional static factors, since AR responses are ruled out in such a model. This model encompasses as a special case the approximate-factor model of Chamberlain (1983) and Chamberlain and Rothschild (1983), that allows for correlated idiosyncratic components but it is static; it generalizes the *exact* factor model of Sargent and Sims (1977) and Geweke (1977), which is dynamic but has orthogonal idiosyncratic components.

In a classic dynamic factor model (Brillinger, 1981), considering the scalar time series variable Y_t to forecast and let X_t be the N-dimensional time series of candidate predictors, it is assumed that (X_t, Y_{t+h}) admits a factor model with r common latent factors F_t:

$$X_t = \Lambda F_t + \varepsilon_t$$

$$Y_{t+h} = \beta_F' F_t + \beta_\omega' \omega_t + \varepsilon_{t+h} \quad (1)$$

where ε_t is an $N \times 1$ vector of idiosyncratic disturbances, h is the forecast horizon, ω_t is an $m \times 1$ vector of observed variables (i.e. lags of Y_t) useful, with F_t, to forecast Y_{t+h}. In the general model, the value of the medium to long run component of the growth c_t, with the coefficients A_i, at the end of the sample is so estimated:

$$\hat{c}_T = A_1 F_{1T} + A_2 F_{2T} + \dots A_m F_{mT} \quad (2)$$

In this paper, we use two groups of common factors on which the GDP is projected: F_i and R_i (*i =1, ..., m*) will be respectively the common factors relevant to the prediction of "real MLRG" and "financial MLRG" (Figure 2), obtained by projecting Euro Area GDP respectively on real and financial variables. The α monthly weights to combine the two smoothed growth indicators will be obtained in real time by the *regression method*.

The methodology that we develop in this section can be so summarized as in:

$$X_t^F = \Lambda F_t + \varepsilon_t^F \quad (3)$$

$$X_t^R = \Lambda R_t + \varepsilon_t^R \quad (4)$$

$$Y_{t+h}^F = \beta_B' F_t + \beta_\omega' \omega_t^F + \varepsilon_{t+h}^F \quad (5)$$

$$Y_{t+h}^R = \beta_R' R_t + \beta_\omega' \omega_t^R + \varepsilon_{t+h}^R \quad (6)$$

The medium to long-run growth (that we name "Combined Eurocoin") will be equal to:

$$\hat{c}_T = \alpha_0 + \alpha_1 (A_{1F} F_{1t} + A_{2F} F_{2t} + ... + A_{mF} F_{mT}) + \alpha_2 (A_{1R} R_{1T} + A_{2R} R_{2T} + ... A_{mR} R_{mT})$$

Or, considering the lags as in (5) and (6), we have:

$$\hat{c}_T = \alpha_0 + \alpha_1 Y^F_{t+h} + \alpha_2 Y^R_{t+h} \qquad (7)$$

The comparison between the medium to long-run components, obtained through the traditional method Eurocoin in (2), and the combination specified above in (7), will offer a more specific knowledge with regard to real and financial economic activities.

In the theoretical case of infinite data series, evaluation of the medium to long-run component can easily be done by applying band-pass filtering. In reality, band-pass filter method provides a good approximation in the middle of the sample, while approximations at its ends are very poor, since they require knowledge of the future values of GDP, which of course we do not have. It is not an appropriate approach for real-time analysis. The idea of Eurocoin approach, that we are proposing in this paper, is based on the assumption that a panel of macroeconomic variables capture some information about future GDP dynamics, to perform equally well within and at the end of the sample.

Each real time indicator will be compared to the target that is a band pass bilateral filter on growth rate. Target value, which is not available at the end-of-sample time T, is available with good accuracy only at time T +h, for a suitable h. As a consequence, disaggregated indicators produced at time T will be compared with the target at T produced at time T + h.

A finite-sample version of the band-pass filter, equation (8), provides a good approximation to the ideal target at time t in the middle of the sample, and it performs badly at the beginning and end of the sample. Precisely, the performance at time t, with $t \le T - 12$, will be measured as the difference between our indicator at time t and the approximate target at t that is obtained using data up to T.

According to Altissimo et al. (2006), within a finite sample the following approximation of the target can be obtained, by augmenting $y_t{}^s$ with its sample mean $\hat{\mu}$ in both infinite directions:

$$c^{*s}_t = \beta(L) y^{*s}_t, \text{ where } y^{*s}_t = \begin{cases} y_t & if \quad 1 \le t \le T \\ \hat{\mu} & if \quad t < 1 \; or \; t > T \end{cases} \qquad (8)$$

Since y_t, the growth rate, is observed only quarterly, while we are interested in a monthly indicator of economic activity, we chose a simple interpolation to calculate the two missing points for each quarter, assuming that y_t is unchanging within a quarter.

It is possible to prove that in a dynamic factor model the principal components (D'Ambra, L., Gallo, M., 2008) of X_t are consistent estimators of the true latent factors.

3. COMBINING REAL AND FINANCIAL VARIABLES: REAL TIME RESULTS

Nowcasting GDP requires to focus on times series data that can provide information on the current state of the economy. There are numerous macroeconomic time series with shorter publication delays than GDP. This is mostly the case for monthly statistics related to employment, industrial production, financial variables or business surveys published by Central Banks or National Statistics Institutes.

At this stage, two main approaches exist in econometric literature to choose the variables useful for the nowcast of growth rate; the first focus on a limited number of series and it consists in selecting a reduced number of variables and tracking their development. The selection criteria are generally based on:

- the ex-post ability of the series to reproduce reference time series movements;
- a priori belief based on economic theory;
- the choice can almost be judged as subjective.

The series can either be individually tracked or aggregated in a synthetic index. The former is a strategy that has been adopted by the Conference Board and the OECD. Since in the present paper we implement Eurocoin indicator, we will be using generalized dynamic factor model (on which Eurocoin is based) to construct some monthly indicators of economic activities in Euro Area, and we assume two different representations for the economic development: the first can be obtained by considering a large dataset of real European variables; the second will consider a smaller dataset of international financial variables. We dispose of a dataset consisting of 157 monthly macroeconomic variables during the period between January 1987 and March 2011. The main blocks of macroeconomic indicators are as follows (table 1):

- Business and consumer confidence indicators – the largest block;
- Industrial production indices;
- OECD Composite Leading Indicator;
- Producer price index for: intermediate and capital goods; energy, industry, investment and intermediate goods; durable and non durable goods;
- Retail Sales;
- Variables describing external transactions: exports and imports of goods and services.
- Financial data: monetary variables, interest rates, effective exchange rates.

Table 1. *Variables used in Estimation by Data Source*

Data Source	***Variables***
Surveys	31
Leading Indicators	6
Demand Indicators	12
Industrial Production	32
Wages Indicators	2
Employment Indicators	5
Producer Price Index	26
Exchange rates	3
Imports-Exports	8
Money Supply	8
STANDARD &POOR'S INDEX	7
(Italy, Germany, USA, UK) SPREAD	10
Benchmark Bond	7
TOTAL	***157***

We focus on medium to long-run components of the growth (MLRG), i.e. the smoothed components of GDP growth rate obtained by removing the fluctuations of period shorter than or equal to one year, and it bears no relationship to any definition of trend. Monthly indicators are commonly used in the prediction of current data on GDP before the data are available. For the Euro area, a flash estimate of GDP is released by Eurostat about six weeks after the end of the reference quarter, and a full set of indicators for the second quarter of the year is not available any earlier than the GDP flash estimate. Also in this section we divide the 157 variables contained in Thomson Financial Datastream in real and financial variables. In (7), we have shown a combination (using regression method to determine the relative weights) of real and financial MLRG; in the *Combined Eurocoin* the regression method is used to determine the relative weights: this combination is useful to analyze the impact of real and financial data in estimating smoothed GDP, that is the main aim of this paper.

Ex post estimate is looked at in this section by analyzing the in-sample 1995-2002; the period 2003-2010 will be analyzed in real time with the end of the sample. Experiments conducted in this paper use 5 generalized principal components, the number estimated over the whole sample period [1 T]. The exercises we develop use the estimates $\hat{c}_t(t+h)$, of each disaggregated indicators at time t using the data from 1 to t + h, h = 0, 1, 2, with t running from January 2003 to December 2010.

Real time estimations are built from 2003 to 2008 and from 2003 to 2010 separately, as in 2008-2010 we observe a strong recession and an high variation in GDP volatility. Analysis of real time performance, in this section will regard:

- ability of real time indicators to approximate the target. It will be measured as the difference between our indicator at time t and the approximate target at t that is obtained using data up to T, by calculating the RMSFE (root mean squared forecast error). Real time error include both uncertainty concerning future values of error term and that arising due to the fact that regression coefficients are estimated (see sub-section 3.1);
- ability of real time indicators to signal the correct sign of target change and in signalling turning points (sub-section 3.2);
- analysis of regression coefficients in equation (7) above described (sub-section 3.3).

Our experiment is useful in the analysis of the impact of real data on estimate smoothed GDP in the different business cycle phases. So, we have outlined two data groups: a first containing "*real economic activity variables*"and a second with "*financial variables*". In our experiments the approximate target is the bandpassed Euro Area growth rate, the same that is generally used to test the performance of the Eurocoin indicator. In the following, the following indicators are compared:

- Eurocoin;
- Financial Eurocoin;
- Real Eurocoin.

3.1 Ability of indicators to approximate the target

Real Time performance is computed by using the following steps and it gives a sense of how well the model has gone at the end of the sample (Figure 1):

1. Select a date near the end of the sample;
2. Estimate model using data up to that date;

3. Use estimated model to produce some forecasts/estimates, by using a recursive window as follows: the initial estimation date is fixed, but additional information is added one at a time to the estimation period.

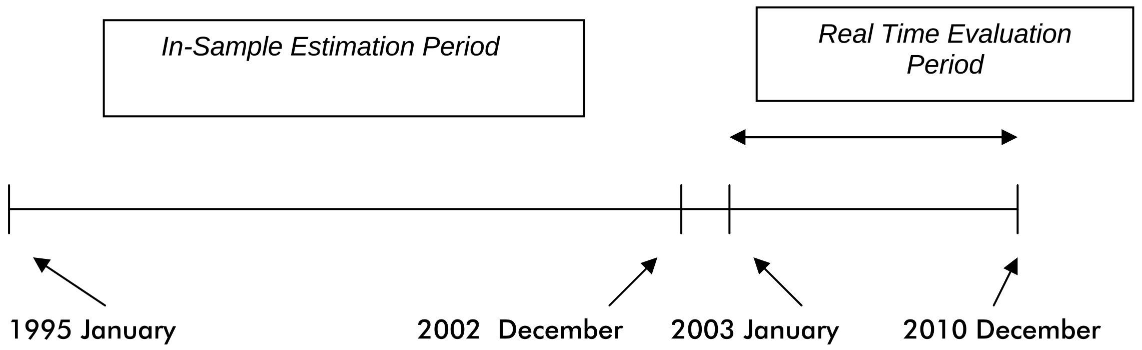


Figure 1. *Use of In Sample and an Real Time Estimation in our research*

In this sub-section we test the capacity of our estimates inside the sample and in real time to approximate the bandpassed target.

Table 2. *RMSFE among Indicators and European bandpassed GDP*

Indicators	1995-2002 (Rmse within the sample)	2003-2008 (Rmse in Real Time)	2003-2010 (Rmse in Real Time)
Real Eurocoin	0.12	0.129	0.44
Financial Eurocoin	0.14	0.161	0.52
Eurocoin	0.12	0.125	0.43

In table 2 and 3 we analyze the performances inside the sample and in real time, since June 1995 to December 2010. We observe, in particular, that performance of Eurocoin Indicator is strongly similar to the one concerning the Real Eurocoin. Our elaboration are based on Thomson Financial Datastream.

Table 3. *Correlation among Indicators and European bandpassed GDP*

Indicators	1995-2002 (Within the sample)	2003-2008 (In Real Time)	2003-2010 (In Real Time)
Real Eurocoin	0.91	0.87	0.88
Financial Eurocoin	0.89	0.77	0.77
Eurocoin	0.92	0.88	0.89

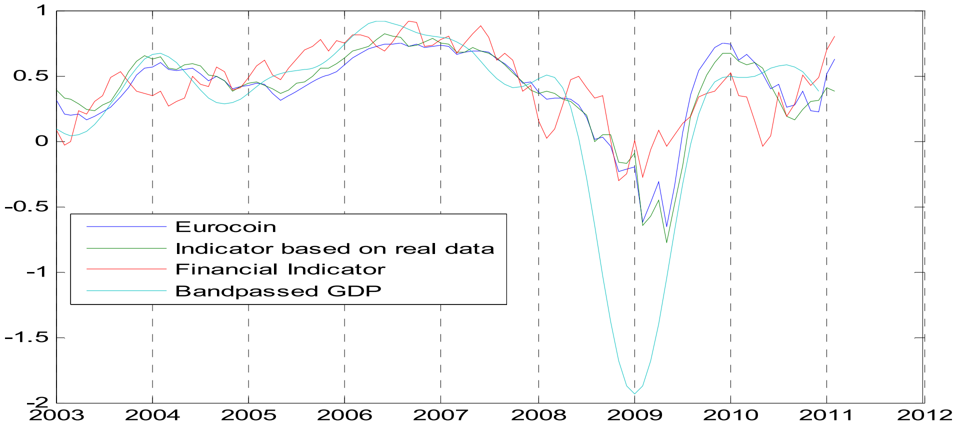


Figure 2. *Pseudo Real Time Estimation*

3.2. Ability to signal the correct sign of target change and turning points

In this sub-section we investigate the capacity of real time indicators to signal the correct sign of target change. To assess the ability of $\Delta \hat{c}_t^{s}(t)$ to signal the correct change of the bandpassed variation $\Delta c_t^{*s}(T)$, we use the statistical test of Pesaran and Timmermann (1992). In synthesis, Pesaran and Timmermann proposed a directional accuracy (DA) test of the hypothesis that there is no relationship between the direction of change predicted by a model and the observed change. Concerning our disaggregated estimates, if P is the proportion of times the sign of the bandpassed Euro Area growth rate (the approximate target) that is correctly predicted by the three indicators in real time, and P_star is the probability of the correct sign being estimated under the assumption that the predictor is independent from the predicted variable, we can shortly highlight, following Pesaran and Timmermann (1992), that

$$S\_n = \frac{(P - P\_star)}{\sqrt{Var(P) - Var(P\_star)}}$$

is approximately normal.

We observe that:

- PT two sided test is above the 99% critical value forEurocoin and Real Eurocoin
- the Real Eurocoin indicator (the one that is based on real variables) strongly rejects the null hypothesis;
- for the Financial indicator we observe a bad performance in terms of correct prediction of sign.

Table 4. *Non-parametric Statistic of Pesaran - Timmermann (PT)*

Indicators	PT	p-value of the PT test statistic	% Correct prediction of sign of bandpassed Δc*2003-2009
Eurocoin	2.67	0.0075	0.64
Financial Eurocoin	0.15	0.8837	0.51
Real Eurocoin	3.89	0.000	0.69

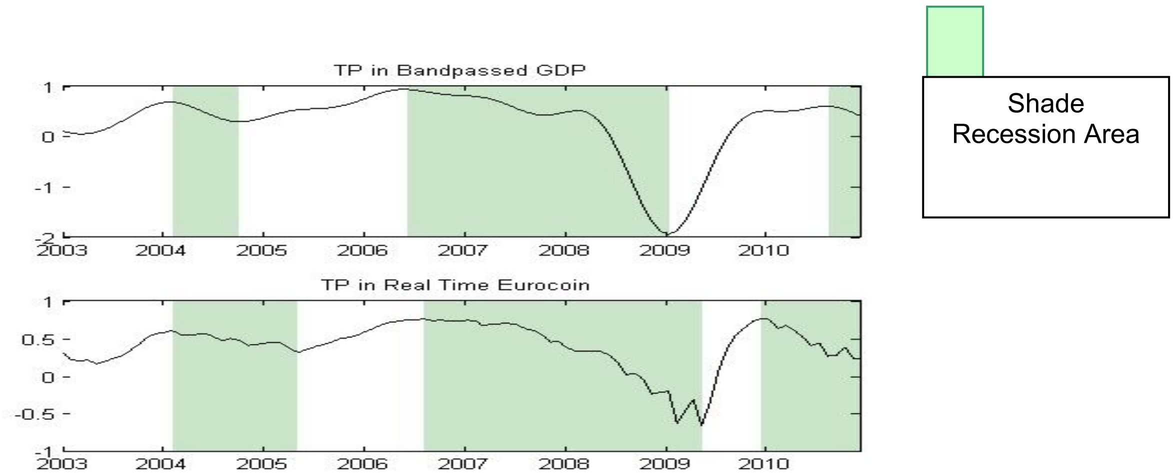
A characteristic of the indicators that we test in this chapter, is the ability to give a correct signal of MLRG turning points in real time. In the simplified *Bry-Boschan* procedure (1971), used in the OECD CLI system for turning point identification, these censor rules guarantee the alternation of peaks and troughs, while ensuring that phases last not less than 9 months and cycles last not less than 2 years". This methodology is based on the concept which focuses on fluctuations in the absolute level of economic activity; however, since this work is based on fluctuations in q-o-q growth rate, we say that an upturn (downturn) signal in $\hat{c}_t^s$ can be predicting or lagging true upturn, tolerating a four-month error.

Table 5. *Number of Turning Points in the Bandpassed Target*

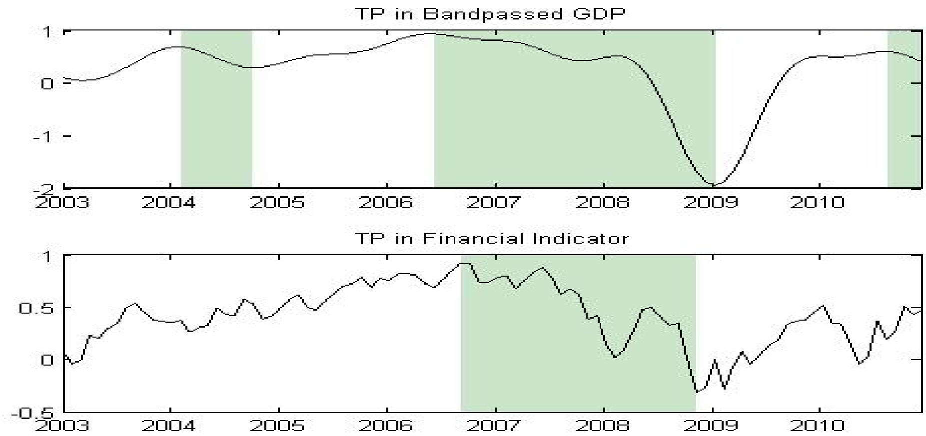
TOTAL TURNING POINTS	DOWNTURNS	UPTURNS
5	3	2

Table 6. *Real time detection of turning points (TP)*

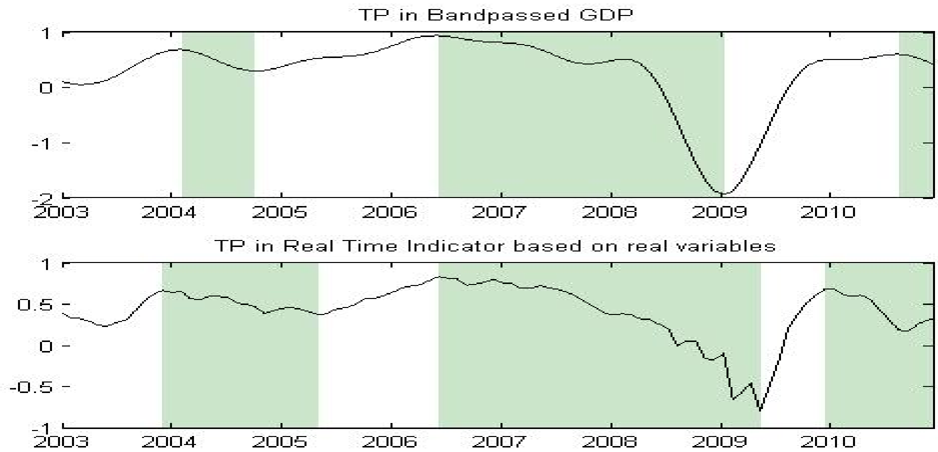
SECTORS	TP Signals	Correct TP	Correct over signalled TP	Missed over all TP
Eurocoin	5	3	3/5	2/5
Indicator based on real variables	5	3	3/5	2/5
Financial Indicator	2	2	2/5	3/5



a) Bandpassed Target versus Eurocoin



b) Bandpassed Target versus Financial Eurocoin



c) Bandpassed Target versus Real Eurocoin

Figure 3. Real Time detection of Turning points

Therefore, we observe that a large dataset of 122 *real variables*, in terms of TP, produces some results similar to Eurocoin (produced with a dataset of 157 variables) in detecting TP in bandpassed target. Differently, a real time indicator based on *financial variables*, and based on a small dataset (35 variables), produces a satisfactory performance in detecting TP when recession lasts for a long period (in our exercise concerning Euro Area, it concerns the 2003 – 2009).

3.3. Forecasting the crisis by the regression coefficients

In the equation (9) the weights α and β are shown updated monthly to underline our regression in real time estimation period (2003-2010)

$$c_t = \alpha_t + \beta_{1t} c_t^R + \beta_{2t} c_t^F \tag{9}$$

in which c_t indicates bandpassed GDP; c_t^R is the "Real Eurocoin" indicator that we calculate only using real variables; c_t^F is the "Financial Eurocoin" indicator that we calculate using financial variables only. The weights are updated every month on the basis of the newly available information. In Figure 4 that follows we show the weights (regression coefficients) calculated in the combination of real and financial indicator: the Combined Eurocoin is calculated by following the equations (7) and (9). In Figure 4, we observe that the relation between the two coefficients is quite stable till 2008; at the beginning of the last recession (during 2008), it changes the impact of real and financial data to estimate smoothed GDP, and in 2009 (at the trough) the distance becomes the minimum; during 2009, when recovery begins, it is shown that the impact of real data to estimate GDP becomes more important than the one concerning financial data. This matter could help the econometrician to forecast economic crisis.

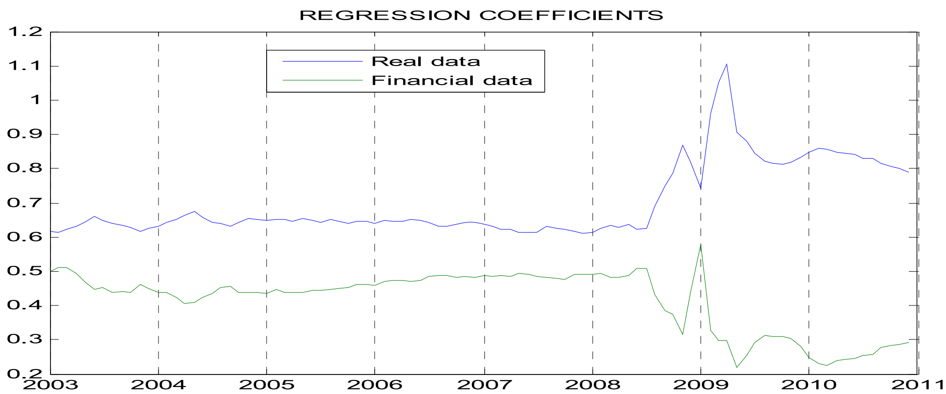


Figure 4. Regression coefficients

4. FINAL REMARKS

In this work we analyze the behaviour concerning a combination of generalized dynamic factor models, compared to a classic Eurocoin indicator (that is produced by using the whole dataset of real and financial variables).

We observe, in particular, that:

- in terms of RMSE and correlation, the performance of Eurocoin Indicator to approximate the bandpassed target is very similar to the one concerning the Real Eurocoin; also concerning the ability in signalling turning points in the target, their performances are quite similar.
- Concerning the Ability of real time indicators to signal the correct sign of target change, the *Real Eurocoin* indicator (the one that is based on real variables) strongly rejects the null hypothesis that there is no relationship between the direction of change predicted and the observed change; also Eurocoin rejects this hypothesis.
- *The financial indicator*, that is based on a small dataset (35 variables), produces a good performance in detecting TP only when recession lasts for a long period.

Finally, we can assess (Figure 4) that the impact of real and financial variables in estimating smoothed GDP, during the structural break in 2008, shows that the role of real

data as industrial production, demand indicators, foreign trade (Import, Export), Employment Indexes, becomes particularly relevant in relation to that concerning financial data as Exchange rates, Money Supply, Spreads. So, one possible explanation could be that interrelations among the recession phase and the variations in production, consumptions and unemployment are highly interrelated.
Concerning further developments of our research, it could be useful to use a larger dataset of historical series.

References

1. ALTISSIMO, F., CRISTADORO, R., FORNI, M., LIPPI, M., VERONESE, G. **New Eurocoin: Tracking Economic Growth in Real Time**, The Review of Economics and Statistics, MIT Press, vol. 92(4), 2010, pp. 1024-1034

2. BAI, J., NG, S. **Determining the number of factors in approximate factor models**, *Econometrica* 70, 2002, pp. 191-221.

3. BANBURA, M., GIANNONE, D., REICHLIN, L. **Nowcasting**, ECB Working Paper Series, No. 1275, 2010

4. BAXTER, A., KING, R.G. **Measuring Business Cycle Approximate Band-Pass filters for Economic Time Series,** The Review of Economics and Statistics 81, 1999, pp. 575-93.

5. BRILLINGER, D. **Time series – Data Analysis and Theory**, expanded ed., San Francisco, 1981.

6. CHRISTIANO, L.J., FITZGERALD, T.J. **The Band-Pass Filter**, International Economic Review 84, 2003, pp. 435-65.

7. BRY, G., BOSCHAN, C. **Cyclical Analysis of Time Series: Selected Procedures and Computer Programs**, National Bureau of Economic Research, New York, 1971

8. CHAMBERLAIN, G. **Funds, factors, and diversification in arbitrage pricing models,** Econometrica 51, 1983, pp. 1281-1304.

9. CHAMBERLAIN, G., ROTHSCHILD, M. **Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets**, Econometrica, 70, 1983, pp. 191–221.

10. CRISTADORO, R., FORNI, M., REICHLIN, L., VERONESE, G. **A core inflation indicator for the euro area**, Journal of Money Credit and Banking 37(3), 2005, pp. 539-560.

11. D'AMBRA, L., GALLO, M. **Nonlinear constrained principal component analysis for the multivariate process control..** In:Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R., Data Analysis, Machine Learning and Applications. Springer, Berlino, 2008, pp.193- 200

12. FORNI, M., HALLIN, M., LIPPI, M., REICHLIN, L. **The generalized dynamic factor model: consistency and rates**, Journal of Econometrics 119, 2004, pp. 231-255.

13. GEWEKE, J. **The Dynamic Factor Analysis of Economic Time Series Models**, in Latent Variables in Socioeconomic Models, North Holland, 1977, pp. 365-383.

14. GIANNONE, D., LENZA, M., REICHLIN, L. **Business cycles in the euro area,** European Central Bank Working Paper Series, 2009

15. HALLIN, M., LISKA, R. **Determining the Number of Factors in the Generalized Factor Model**, Journal of the American Statistical Association, 102, No. 478, 2007, pp. 603-617

16. PESARAN, M.H., TIMMERMANN, A. **Testing dependence among serially correlated multi-category variables**, CESifo Working Papers, 1770, 2006.

17. SARGENT, T. J., SIMS, C.A. **Business Cycle Modeling Without Pretending to have too Much a priori Economic Theory**, New Methods in Business Cycles Research, Federal Reserve Bank of Minneapolis, 1977, pp. 45-110.

18. STOCK, J.H., WATSON, M.W. **Forecasting using principal components from a large number of predictors**, *Journal of the American Statistical Association* 97, 2002, pp. 1167-79.

---

[1] Researcher at the Italian National Statistical Institute (ISTAT) since 2000. He has participated in Eurostat task force on Statistical Business Registers and collaborated with Eurostaf – Group Les Echos (Paris). Lecturer at the University of Naples Federico II, Faculty of Economy. Member of Italian Statistical Society (SIS) and Euro Area Business Cycle Network (EABCN).

[2] Associate Professor, Computer Engineering Department, University of Naples "Federico II", Italy. Member of the college of the Ph.D. on Statistics of University of Naples. Membership of the Italian Statistical Society. Member of Scientific Committees. Author of many books and journal articles in the field of the Multidimensional Data Analysis, Numerical Algorithms, Cluster Analysis.

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

APPLYING FUZZY C-MEANS AND ARTIFICIAL NEURAL NETWORKS FOR ANALYZING THE NON-BANKING FINANCIAL INSTITUTIONS' SECTOR IN ROMANIA

Adrian COSTEA
Ph.D., Academy of Economic Studies, Statistics and Econometrics Department

E-mail: adrian.costea@csie.ase.ro

ABSTRACT:

In this paper we apply a neural approach to develop classification models in order to assess the performance of non-banking financial institutions (NFIs) in Romania. Our objective is twofold: to empirically validate our methodology and understand how different financial factors can and do contribute to the NFIs' movements from one performance class to another.

Keywords: non-banking financial institutions, performance evaluation, neural networks, class prediction

1. INTRODUCTION

In this paper we analyze comparatively the financial performance of non-banking financial institutions (NFIs) in Romania using Data Mining (DM) techniques. The NFIs' financial performance is measured according to our previous work (Costea, 2011a) in terms of financial ratios that define three performance dimensions: capital adequacy, assets' quality and profitability. Our methodology consists of two stages. In the first stage we apply a DM clustering techniques, namely Fuzzy C-Means (FCM) algorithm, in order to find performance clusters within the data. According to Costea (2011b), FCM algorithm performed better than other clustering algorithms in terms of the formed clusters when applied on the same NFIs' performance dataset. At this stage we characterize each cluster in terms of average characteristics of the observations that are allocated in that particular cluster and attach to each observation a label (performance class variable) that identifies the observation as belonging to the cluster. In the second stage, we apply feed-forward artificial neural networks (ANNs) algorithms in order to map the input space to the newly created performance class variable so that we might be able to predict the performance of different NFIs as data become available. For a detailed technical explanation of the FCM algorithm we refer the reader to Costea (2011b). In the next Section, we present a brief explanation of the ANNs for classification.

2. ARTIFICIAL NEURAL NETWORK CLASSIFICATION ALGORITHMS

The generic classification model based on neural approaches is depicted in Figure 1 (adapted from Costea, 2005). As it can be seen from Figure 1, when building classification models, firstly, we perform some preliminary steps: we separate the data into training (*TR*) and test (*TS*) sets, we construct the performance class variable by applying a clustering method (in our case – FCM algorithm). Then, we determine the proper ANN architecture. This step consists of determining the proper number of hidden layers, and the appropriate number of neurons in each hidden layer. Also, we decide how to code the class variable: using one neuron or as many neurons as the number of performance classes (clusters). Finally, we train and test the ANN using different values for the parameters involved in training.

According to Basheer & Hajmeer (2000, p. 22), the choice of the number of hidden layers and the number of neurons in each hidden layer depend on "input/output vector sizes, size of training and test subsets, and, more importantly, the problem of non-linearity". Basheer & Hajmeer (2000) presents a list of papers that provide different rules of thumb regarding the correspondence between the number of hidden neurons (*NH*) and the number of input (*NI*) and output (*NO*) neurons or the number of training samples (N_{TRN}). For example, Lachtermacher & Fuller (1995) links the number of input and hidden neurons *(NI, NH)* for one output *ANN* with the number of training samples N_{TRN} with the formula: $0.11N_{TRN} < NH(NI+1) < 0.30N_{TRN}$. Upadhyaya & Eryurek (1992) connect the total number of weights N_w with the number of training samples with the formula: $N_w = N_{TRN}\log_2(N_{TRN})$. Masters (1994) calculates the number of hidden neurons in the hidden layer around the geometric mean of the number of inputs (*NI*) and of outputs (*NO*). Choosing these parameters is more art than science. We base our decision on the advice given by Basheer & Hajmeer's (2000, p. 23): "the most popular approach to finding the optimal number of hidden nodes is by trial and error with one of the above rules". For example, in this study we chose the Lachtermarcher & Fuller (1995) rule and varied *NH* depending on the size of the training set. Concerning the number of hidden layers, we performed in each case a number of experiments for ANN architectures with one and two hidden layers to see what the appropriate number of hidden layers is. Depending on the dataset used, an ANN with one or two hidden layers performed better in terms of the training mean square error. The three hidden layer cases was discarded to avoid the increase in network complexity given we obtain high training accuracy rates for less complex ones.

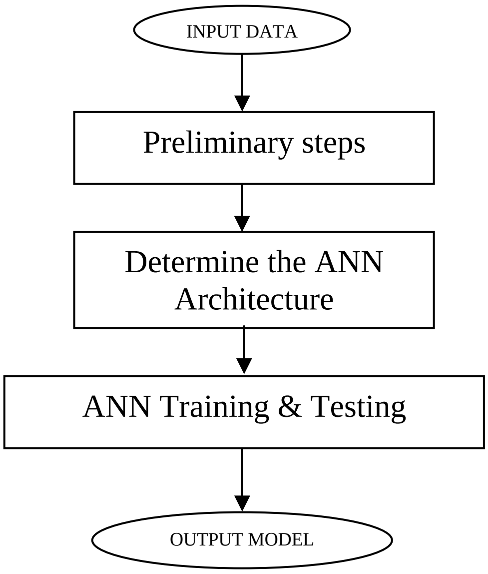


Figure 1. ANN generic classification model (Source: Costea, 2005)

We have used the sigmoid and linear activation functions for the hidden and output layers respectively, as this combination of activation functions provided the best results in our experiments. We base our decision in choosing the best training algorithm on the studies that have been written were different algorithms were compared in order to find the best algorithm for a particular problem (Demuth & Beale, 2001; Nastac & Koskivaara, 2003; Costea, 2003). In Costea (2003) we compared four training algorithms in terms of error rates and convergence speed. Our findings suggest that there is a negative correlation between error rates and the convergence speed. Therefore, in choosing the training algorithm, one should seek a compromise between these two factors. We observed that the Scaled Conjugate Gradient (SCG) algorithm (Moller, 1993) performs well over a wide variety of problems. The SCG is not the fastest algorithm, but it does not require large computational memory and it has a good convergence. Furthermore, in order to avoid the network over-fitting the training samples, we apply the validation stop method: we separate the training data in effective training (*TRe*) and validation (*VAL*) datasets and the training process stops when the difference between the effective training error and the validation error is greater than a small value given as a parameter. Moreover, it is well known that for validation stop, one must be careful not to use an algorithm that converges too rapidly (Hagan *et al*., 1996; Demuth & Beale, 2001). The SCG is well suited for the validation stop method.

3. THE NFIS FINANCIAL PERFORMANCE DATASET

In this application we use three performance dimensions to evaluate a NFI: capital adequacy (C), assets' quality (A) and profitability (P). We select different indicators for each dimension based on the analysis of the periodic financial statements of the NFIs. In the following table we present the indicators for each performance dimension.

Table 1. The performance dimension and the corresponding financial ratios

Dimension	Indicators
Capital adequacy	1. Equity ratio (Leverage) = own capital / total assets (net value) 2. Own capital / equity 3. Indebtedness sources = borrowings / own capital
Assets' quality	1. Loans granted to clients (net value) / total assets (net value) 2. Loan granted to clients (net value) / total borrowings 3. Past due and doubtful loans (net value) / total loans portfolio (net value) 4. Past due and doubtful claims (net value) / total assets (net value) 5. Past due and doubtful claims (net value) / own capital
Profitability	1. Return on assets (ROA) = net income / total assets (net value) 2. Return on equity (ROE) = net profit / own capital 3. The rate of profit = gross profit / total revenues 4. Activity cost = total costs / total revenues

The next step of the analysis is to choose the best set of indicators for each dimension and collect the data necessary to calculate these indicators. We have changed indicators number 3 for the "degree of capitalization" dimension and number 3 for the "profitability" dimension by replacing the denominator with Total Assets (net value). We have done this in order to be able to interpret the indicators since the former denominator (own capital) could take negative values. At the same time we have eliminated the indicator 5 for the "assets' quality" dimension for the same reason. Finally, we have 11 indicators: 3 for the degree of capitalization, 4 for assets' quality and 4 for profitability. The data were collected quarterly from 2007 to 2012 for the NFIs registered in the Special Register that have been active since the introduction of the regulatory framework for these institutions in Romania. In total there were 68 NFIs that met the above criteria and 990 observations. Out of these 990 observations, 5 observations were discarded due to lack of data for certain financial indicators.

4. EXPERIMENT

In this experiment we try to evaluate comparatively the performance of 68 Romanian NFIs registered in the Special Register that have been active since 2006, the first year when this sector has been regulated in Romania. This analysis can help the Supervision Department of the National Bank of Romania to allocate more efficiently its resources. Identifying poorly performing NFIs would support supervisors to concentrate on a smaller number of NFIs that face difficulties. Other authors have studied the sectoral dynamics of non-performing loans (e.g.: Moinescu & Codirlasu, 2012) having similar research goals.

As the Figure 1 shows the first step of the methodology consists of some preliminary steps. Our dataset that consist of 11x985 observations has been transformed by levelling the extreme values for each variables in the [-20, 20] interval. We have done this in order to avoid the algorithms' results being affected by these extreme values.

In the next step, we apply FCM algorithm in order to build cluster with similar performance. We chose 4 clusters as we have done with a version of the same dataset in our previous work (Costea, 201x). The other parameters of FCM were as follows: m = 1.5, no_of_iterations = 10000, the limit for the stopping criterion = 0.00001. After we run the FCM algorithm on the 11x985 dataset we obtained the following structure of the clusters: cluster 1 (95 observations), cluster 2 (770 observations), cluster 3 (59 observations), and cluster 4 (61 observations). Based on the clusterization we have constructed the class variable by associating to each observation the number of the cluster that the observation belongs to.

In order to have an uniform number of observations in each cluster to train the classification model we selected 59 observations (the number of observations in the smallest cluster) from each cluster, totalling 236 observations. Also, at this stage, we have split the data in training (*TR*) and testing (*TS*) sets by selecting one testing instance for every nine training instances. Thus, we obtained randomly 212 observations for training and the rest for testing (24 observations).

The next step of the methodology was to determine the proper architecture for the ANN-based classification model that maps the 11-dimensional input space to the newly constructed performance class variable. In our experiments regarding the application of ANNs for classification (performed using Matlab's Neural Networks toolbox) we have kept all parameters of the ANNs constant (the learning algorithm - SCG, the performance goal of the classifier, the maximum number of epochs), except the number of neurons in the hidden layers (*NH* when we had one hidden layer and NH_1, NH_2 when we had two hidden layers).

Next, we present the empirical procedure to determine the architecture for an ANN with two hidden layers. Firstly, we performed three trainings in order to find the best ANN architecture. For each training we have split further the training set (*TR*) in the effective training set (*TRe*) and the validation set (*VAL*), obtaining each time approximately 186 observations for effective training and 26 observations for validation (we have used validation stop method as stopping criterion). We followed the Lachtermarcher & Fuller (1995) rule and varied NH_1 and NH_2 from 5 to 8 and trained the network for each ANN architecture based on the effective training dataset. We saved the best ANN architecture in terms of mean squared error for the effective training dataset (MSE_{TRe}) and if the mean squared error based on the validation set (MSE_{VAL}) is less than $6/5*MSE_{TRe}$. This condition has been imposed in order to avoid saving ANN architectures for which the effective training and validation mean squared error are too far from each other. The final ANN architecture consisted of 8 neurons on the first hidden layer and 5 neurons on the second hidden layer.

Finally, at the last methodological step, we have trained the obtained ANN with the same-way generated effective training, validation and testing datasets and obtained the following accuracy rates: effective training dataset accuracy rate (ACR_{TRe}) = 100 percent, validation dataset accuracy rate (ACR_{VAL}) = 100 percent, total training dataset accuracy rate (ACR_{TR}) = 100 percent and testing dataset accuracy rate (ACR_{TS}) = 95.83 percent. The high values for the accuracy rates and the small difference between testing and training accuracy rates show that we obtained a very good classification model. Moreover the empirical procedure to find the best ANN architecture has been validated by the same high accuracy rates. Based on the chosen architecture we can test different values for the other ANN parameters and further improve the performance of the ANN-based classifiers.

5. CONCLUSIONS

In this study we have applied Data Mining to formalize the process of assessing comparatively the performance of non-banking financial institutions in Romania. We addressed this research problem by associating two Data Mining tasks: a clustering task by which we followed a description strategy showing what is the current situation of the NFIs' sector and a classification task used for creating a mapping between the performance class variable and the multidimensional input space.

For the clustering phase we employed a fuzzy logic algorithm called Fuzzy C-Means algorithm and identified four performance clusters. Based on the average characteristics of the input variables we characterized each individual cluster. For the classification phase we selected an even number of observation in each cluster to allow the classifier to learn the characteristics of each cluster. As classification technique we used feed-forward neural networks trained using variants of backpropagation algorithm (e.g.: the Scaled Conjugate Gradient algorithm).

A secondary goal of this study was to find a procedure to determine the proper neural network architecture for our particular research problem. We obtained very high training and testing accuracy rates and small differences between these rates. Compared with other classification models applied on the same dataset in our previous work, the neural network-based model is the best in terms of training and testing accuracy. However, the explanatory capabilities of the decision trees have to be taken into account in the process of choosing the best model.

ACKNOWLEDGEMENTS:

This work was supported from the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/89/1.5/S/59184 „Performance and excellence in postdoctoral research in Romanian economics science domain”.

References

1. Basheer IA, Hajmeer M., **Artificial neural networks: fundamentals, computing, design, and application**. *Journal of Microbiological Methods* 43, 2000, 3-31.

2. Costea A., **Economic Performance Classification Using Neural Networks**. *Proceedings of the Sixth International Symposium on Economic Informatics*, I. Ivan. and I. Rosca (eds), Academy of Economic Studies Press, Bucharest, Romania, 2003.

3 .Costea A., **Computational Intelligence Methods for Quantitative Data Mining**, Turku Centre for Computer Science (TUCS), PhD thesis no. 67.,2005

4. Costea A., **Prudential supervision of non-banking financial institutions: developments and prospects**. *Romanian Statistical Review* 7, 2011, 81-100.

5. Costea A., **A fuzzy-logic approach to assessing the performance of non-banking financial institutions in Romania**, *Proceedings of First International*

Conference on Financial Modeling – FM 2011, organizers: Society for Financial and Actuarial Mathematics, Sofia, Bulgaria, editor: P. Stoynov, publisher: PERUN-SPRINT Ltd., November 19-20, 2011, pp. 48-54, ISSN: 1314-5916.

6. Costea A., **Performance benchmarking of non-banking financial institutions by means of Self-Organising Map algorithm**, accepted for publication in *East-West Journal of Economics and Business*, Volume XVI, Number 1, 2013, ISSN 1108-2992. The paper has been presented to the 2nd international conference International Conference on International Business (ICIB 2011), organized by the Department of International and European Studies, University of Macedonia, Thessaloniki, Greece, European Center for the Development of Vocational Training, Thessaloniki, Greece, Faculty of Political Science, University of Messina, Messina, Italy and CRIISEA, University of Picardie, Amiens, France, 19-21 May 2011, Thessaloniki, Greece.

7. Demuth H, Beale M., **Neural Network Toolbox**. The MathWorks Inc, Natick Press: MA, USA, 2001.

8. Hagan MT, Demuth HB, Beale M., **Neural Network Design**. PWS Publishing Company, Boston, USA, 1996,

9. Lachtermacher G, Fuller JD., **Backpropagation in time series forecasting**. *Journal of Forecasting* 14, 1995, 381–393.

10. Masters T., ***Practical Neural Network Recipes in C++.*** Academic Press, Boston, MA, 1994,

11. Moinescu B, Codirlasu A., **Assessing the Sectoral Dynamics of Non-performing Loans: Signs from Financial and Real Economy**, *Theoretical and Applied Economics*, Volume XIX (2012), No. 2(567), pp. 69-80.

12. Moller MF., **A scaled conjugate gradient algorithm for fast supervised learning**. *Neural Networks* 6, 1993, 525-533.

13. Nastac I, Koskivaara E., **A Neural Network Model for Prediction: Architecture and Training Analysis**. *TUCS Technical Report* 521, April 2003. [Available at: http://www.tucs.fi/publications/insight.php?id=tNaKo03a&table=techreport]. (Accessed on: 21.11.2012).

14. Upadhyaya BR, Eryurek E., **Application of neural network for sensory validation and plant monitoring**. *Neural Technology* 97: 170–176.