ECONOMETRIC MODELS FOR THE ESTIMATION OF NON-PERFORMING LOANS IN ALBANIA

Phd.Gledjana Zeneli (Foto)¹, Msc Amarilda Kulli², Marsela Xhomaqi³

1) Department of Applied Mathematics, Faculty of Natural Sciences, University of Tirana, Albania
2) Credit Analyst, Credins Bank, Albania,
3) Marsela Xhomaqi, Seas ltd, Albania

ABSTRACT
During the last decade, Albania has experienced fundamental developments, especially in the banking system. From 2008 to 2016, commercial banks faced a continuous increase of non-performing loans. In 2016, the quality of loan portfolio continued its declining performance, due to the extend of non-performing loans versus the total of gross loans.

This paper brings to its attention the peculiar relationship between the non-performing loan level with loan portfolio, through a multiple regression, where loan portfolio is the dependent variable and the types of non-performing loans are the independent variables. To analyse the chosen variables, we use the programmes: SPSS and Microsoft Excel.

Through econometric models, we establish a general approach on the quality of the loan portfolio in Albania over the years and a prognosis of its behaviour via the non-performing loans’ surveys. This indicator, the NPL level provides an useful information for the funding quality of the economy. The Bank of Albania should carefully monitor the rate of NPL, since it could further increase, affecting directly the banking system sustainability and indirectly the economic stability. By the end of 2016, although the NPL level declined, doubtful loans increased. Same conclusions are noted by the analysis we did.

Keywords: banking system, non-performing loans, loan portfolio, econometric models

JEL Classification: C13,E51,G21

Introduction

This study aims to describe the quality progress of the loan portfolio, to prognose its behaviour and the non-performing loan’s number in the near future. The main focus is to predict the NPL’s behaviour and the portfolio quality over the years, by using statistical programmes. Also, we seek to identify the impact of non-performing loans in the loan portfolio in the Albanian banking system.
Since 2008, the year of the beginning of the global financial crisis, the levels of NPL in all over the world, have significantly increased. In Albania, the high level of non-performing loans, from the first quarter of 2010 up to now, has been causing a reduction of new loans.

According to the International Institute of Finance for the approximation of Central Banks of the NPL’s comparison in different countries, are used five categories of loans: standard, loans in pursuit, substandard, doubtful and lost loans. In the majority of countries, as well as in Albania, in non-performing loans are included the last 3 categories (in our analysis are included 4 types, the loans in pursuit, substandard, doubtful and lost loans).

The estimation of non-performing loans through econometric modeling is not a new invention. Many researchers have used different econometric methods of explaining the variation on NPL. Olena Havrylchyk explained the variation of NPL, in South Africa, using a linear regression model with factors such as GDP, inflation and interest rate as independent variables. Gertjan Vlieghe (2001), in his study, analysed the credit risk for England using a autorregresive model. In this article, we will use a multiple regression model to predict the non-performing loans behaviour and the portfolio quality, for the case of Albania.

The conducted model of this study will be as follow: the loan portfolio is the dependent variable and the types of non-performing loans are the independent variables (loans in pursuit, substandard, doubtful and lost loans). Considering the types of NPL as independent variables, our research adds to a growing literature by determining the index through macroeconomic and bank specific factors. More precisely, we exclusively focused on the Albanian banking system for the period 2008-2016, using data from the Bank of Albania and the commercial banks. The focus is on the after-crisis period which is characterized by a declining performance of the loan portfolio, due to the extend of non-performing loans versus the total of gross loans.

Contrary to the vast majority of the existing literature, our research is an empirical study, which presents findings regarding the fragility of the Albanian banking system. In particular, our study contributes to enriching the existing literature by predicting the non-performing loans behaviour and the portfolio quality in Albania. The structure of the paper is organized as follows. The next section presents the existing literature on problem loans, the NPL types, and their impact on loan portfolio. Chapter 3 describes the methodology, chapter 4 presents the results of the econometric analysis and chapter 5 offers concluding remarks and possible topics for further research concerning NPL’s behaviour and the quality of loan portfolio in Albania.

2. Literature Review

The literature review helps in the definition of the problem and the selection of the appropriate methodology of the research development. First section of this chapter, consists in the definition of econometrics, and non-performing loans. Then, we continue with the discussion of several research about NPL and banking system, conducted over the world.
a) Definition of econometrics

Econometrics is a link between the economic theory, mathematical economics, mathematical statistics and economic statistics. The word "econometrics" means economic measure. Although measure is an important phase of econometrics, its purpose is very widespread, as is clear from the following definitions:

- Econometrics, as a result of a review on the role of economics, consists in the application of mathematical statistics on economic data to give an empirical support the models, conducted by math economics in order to obtain numerical results.

- Econometrics can be defined as a social science, in which the tools of economic theory, mathematics and statistical, are applied in analyzing economic phenomena.

- The art of econometrics consists in finding a set of assumptions that are specific and realistic enough, to provide us the best from the available data.

- The method of econometric research aims to link economic theory and recent measures, by using the theory and the techniques of statistical inferences.

b) Definition of non-performing loans (NPL)

The definition of non-performing loans differs a lot in different countries. Based on the most used definition in the literature or practice, a loan is considered as a non-performing loan, when the borrower is not able to repay within 90 days from the deadline. In general, there are five types of loans: standard, loans in pursuit, substandard, doubtful and lost loans. In the majority of countries, as well as in Albania, in non-performing loans are included substandard, doubtful and lost loans.

c) A brief review of related literature

In recent years, the interest in loan portfolio and non-performing loans has greatly increased, since we encounter more published data at the commercial banks and aggregate banking system level. Many researchers consider NPL as “financial pollution” with harmful effects for both economic development and social welfare (e.g. Brenda Gonzales Hermosillo 1999; Levon Barseghyan 2010; Shihong Zeng 2011). William R. Keeton and Charles Morris (1987) examined a sample of 2,470 insured commercial banks in the United States (US) for the period 1978-1985. Therefore, the banking institutions that undertake greater risk show greater losses. In another study, Boudriga, Taktak and Jellouli (2009b), using aggregate banking, financial, institutional and legal environment data of 59 countries for the period 2002-2006, they examined whether and which factors determine the NPL rate. Their empirical results showed that the NPL is influenced mainly by bank-specific factors, such as capital adequacy, provisions, and bank ownership, while
credit exposure is reduced in countries where legal and institutional conditions are improved. Based on the merits of the aforementioned studies, it is notable that both macroeconomic (e.g. unemployment, house prices indices, inflation, lending rates, credit growth) and bank specific factors (e.g. profitability ratios, capital adequacy, bank size and ownership) seem to determine loan portfolio quality. However, it is obvious that there is a large gap in contemporary literature, regarding the behaviour of NPL and the quality of loan portfolio in Albania, which our empirical investigation hopes to fill in. Therefore, our study extends the existing literature providing safe and valid conclusions for the Albanian’s banking system.

3. Methodology

As mentioned above, this paper brings to its attention the peculiar relationship between the non-performing loan level with the loan portfolio, through a multiple regression, where the loan portfolio is the dependent variable and the types of non-performing loans are the independent variables. To analyse the chosen variables, we use statistical programmes as SPSS, R 3.2.0 and Microsoft Excel. For this, we obtained the data from the BoA and the commercial banks in Albania.

Regression model

The regression model can be used as a tool of the empirical analysis to study the relationship between two variables. Equation (1) represents the functional relationship between $y$ and $x$. If all factors are constant, in order that $u$ is zero, then $x$ has a linear effect on $y$. Before we assume how $x$ and $u$ are related, we should consider always $u$. As long as the assessor is included in the draw, nothing changes from the assumption that the average value of $u$ in the population is 0. Mathematically $E (u) = 0$. 

$$\Delta y = \beta_1 \Delta x, \text{ if } \Delta u = 0$$  \hspace{1cm} (1)

In our case, the regression equation would be:

$$LP_t = B_0 + B_1 X_{1,t} + B_2 X_{2,t} + B_3 X_{3,t} + B_4 X_{4,t} + \varepsilon_{i,t}$$  \hspace{1cm} (2)

where, $LP_t$ is the dependent variable (loan portfolio) and $X_{1,t}, X_{2,t}, X_{3,t}$ and $X_{4,t}$ are the independent variables (NPLs), $X_{1,t}$ (loans in pursuit), $X_{2,t}$ (substandard loans), $X_{3,t}$ (doubtful loans) and $X_{4,t}$ (lost loans).

4. Empirical results

For the maturity 1-12 months, we consider the following database, where, as a dependent variable will be the total number of loans disbursed per year, in a survey conducted from 2011 to
2016. Whereas as independent variables will be the 4 types of loans: in pursuit, substandard, doubtful and lost.

Table 1. “The number of disbursed loans during 2011-2016”

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>In pursuit</th>
<th>Substandard</th>
<th>Doubtful</th>
<th>Lost</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>7,942</td>
<td>173</td>
<td>121</td>
<td>101</td>
<td>56</td>
</tr>
<tr>
<td>2012</td>
<td>10,195</td>
<td>185</td>
<td>185</td>
<td>171</td>
<td>206</td>
</tr>
<tr>
<td>2013</td>
<td>15,683</td>
<td>229</td>
<td>166</td>
<td>246</td>
<td>462</td>
</tr>
<tr>
<td>2014</td>
<td>17,631</td>
<td>303</td>
<td>166</td>
<td>258</td>
<td>763</td>
</tr>
<tr>
<td>2015</td>
<td>17,455</td>
<td>251</td>
<td>158</td>
<td>197</td>
<td>1,103</td>
</tr>
<tr>
<td>2016</td>
<td>18,664</td>
<td>135</td>
<td>118</td>
<td>176</td>
<td>1,012</td>
</tr>
</tbody>
</table>

Data from commercial banks, 2017

Based on the table 1, since 2009-2014, the number of loans disbursed, has been increasing. The lowest level was in 2011 and the highest in 2016. Also, the number of losses has reached the top, in 2016.

a) Model summary

In order to identify the best econometric model for our study, we developed 4 models and according to their estimations, we will choose the best model. In the following table is presented the summary of two indicators (R-square and Durbin Watson test), that are very important for our analysis. We mentioned, previously, the R-square coefficient. So, 97.9 % of our model is described by the explanatory variables (lost, doubtful, in pursuit and substandard loans). Considering the fact that, including a lot of variables in the model, can cause a illogical and out of control increase of R-square. To avoid that and to raise the quality of our analysis, we use the Adjusted R-square. The Adjusted R-square, too, confirms a good explanation of the model by the independent variables (89.5 %).

Table 2. “A summary for the explanation of the models”

<table>
<thead>
<tr>
<th>Model</th>
<th>R-Square</th>
<th>Adjusted R-Square</th>
<th>Std. Error of the Estimate</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.979</td>
<td>.895</td>
<td>8890.5972</td>
<td>2.506</td>
</tr>
<tr>
<td>2</td>
<td>.967</td>
<td>.918</td>
<td>7860.8141</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.952</td>
<td>.919</td>
<td>7809.6344</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.911</td>
<td>.889</td>
<td>9169.1412</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ projection, 2017

Also, from the table 2, we see that the value of Durbin-Watson (2.5) is bigger than 2, (2.5>2), which means that the model suffers from the autocorrelation.
Table 3. “A summary of the coefficients”

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>-2464.001</td>
<td>11669.146</td>
<td>- .211</td>
<td>1.713</td>
</tr>
<tr>
<td>doubtful</td>
<td>170.543</td>
<td>99.536</td>
<td>.866</td>
<td>-1.751</td>
</tr>
<tr>
<td>lost</td>
<td>-8.029</td>
<td>10.695</td>
<td>-.309</td>
<td>1.751</td>
</tr>
<tr>
<td>in pursuit</td>
<td>99.651</td>
<td>56.924</td>
<td>.614</td>
<td>1.751</td>
</tr>
<tr>
<td>substandard</td>
<td>-54.959</td>
<td>48.334</td>
<td>-.394</td>
<td>-1.137</td>
</tr>
<tr>
<td>(Constant)</td>
<td>517.460</td>
<td>9701.530</td>
<td>.053</td>
<td>1.662</td>
</tr>
<tr>
<td>doubtful</td>
<td>120.868</td>
<td>65.742</td>
<td>.614</td>
<td>1.839</td>
</tr>
<tr>
<td>in pursuit</td>
<td>80.822</td>
<td>45.182</td>
<td>.498</td>
<td>1.789</td>
</tr>
<tr>
<td>substandard</td>
<td>-27.862</td>
<td>28.420</td>
<td>-.180</td>
<td>-1.480</td>
</tr>
<tr>
<td>(Constant)</td>
<td>2.633</td>
<td>9623.944</td>
<td>.405</td>
<td>1.000</td>
</tr>
<tr>
<td>doubtful</td>
<td>79.771</td>
<td>50.312</td>
<td>.603</td>
<td>1.839</td>
</tr>
<tr>
<td>in pursuit</td>
<td>97.862</td>
<td>41.433</td>
<td>.603</td>
<td>2.362</td>
</tr>
<tr>
<td>(Constant)</td>
<td>10534.021</td>
<td>8172.879</td>
<td>1.289</td>
<td>6.402</td>
</tr>
<tr>
<td>in pursuit</td>
<td>154.856</td>
<td>24.191</td>
<td>.954</td>
<td>6.402</td>
</tr>
</tbody>
</table>

Source: Authors’ projection, 2017

From the table 3, with the estimates of the coefficients, we can identify the equation for each of our 4 models.

Fistly, before conducting the equations, let’s assume that $Y$ is the dependent variable (loan portfolio), $X$ – the independent variables, $X_1$ – doubtful loans, $X_2$ – lost loans, $X_3$ – loans in pursuit and $X_4$ – substandard loans.

Model 1

$$Y = -2464 + 171 X_1 - 8 X_2 + 99.7 X_3 - 55 X_4$$

Following the regression steps, based on the table 3, we notice that the lowest absolute value of $t$ is for the variable “lost loan” ($|t| = 0.751$), that means that will be “lost loan”, which will be excluded from the model. As we see, the variables “lost loans” and “loan portfolio” are negatively correlated. From that, we can conclude that the level of lost loans has a negative impact in the level of disbursed loans and the beta coefficient confirms that too. Next, we reconduct the model with the remaining variables.
Model 2

\[ Y = 517 + 81 X_1 + 81 X_3 - 28 X_4 \]

After applying the same steps, the equation is represented as above. To conclude, between the substandard loans and the loan portfolio exists a negative relationship. Then we remove from the model the variable with the lowest absolute value of \( t \) (substandard loan) and rewrite the new equation:

Model 3

\[ Y = -3 + 80 X_1 + 98 X_3 \]

From the table 3, we see that R-square is high and its value has not changed a lot because of the variable’s exclusion, confirming that the variable “substandard loans” was not important for the explanation of the model. Same way, the lowest absolute value of \( t \), ( \(|t|\)=0.980), belongs to the variable “doubtful loans”, which is the one that will be removed from the model.

Model 4

\[ Y = 10534 + 155 X_3 \]

The model 4 represent the best model we developed, with the explanation 91.1\% (R-square) and the Adjusted R-square = 88.9 \%. So, it results that the factor which affects the most the loan portfolio, is “loans in persuit”. From this, we can conclude that the Bank of Albania and the commercial banks should be more focused on the loans in persuit, in order to increase the loan portfolio quality.

b) The interpretation of ANOVA table

The SPSS programme provides the ANOVA table, too. Before interpreting its estimates, we formulate the model hypothesis, as follows:

\[ H_0: \beta_i = 0, \quad \forall i = 0, \ldots, p \]

\[ H_a: \beta_i \neq 0 \]
From the ANOVA table, we obtain the F-statistic and the Sig.F. First, we compare the value of F-statistic $= 11.71$ with the critical value of Fisher distribution $\approx 5$. Since $F = 11 > 5$, $H_0$ is rejected and as a result the model is statistically important. The Sig.$F$ confirms that too ($sig = 0.015 < 0.05$).

c) KMO and Barlett’s test

KMO and Barlett’s test is another indicator of the validity and the compatibility of the data analysis. KMO (Kaiser–Meyer–Olkin) estimates the correlation between the variables of the model. If its value is bigger than 0.5, then there is the presence of high correlation between the variables. In our model, the correlation is 0.47 ($\approx 0.05$), so the variables have a strong correlation with one another. While, from the Barlett’s test, we notice that $sig = 0.028$. Knowing that $sig < 0.05$, we can say that there is a good correlation structure between the variables.

d) The Scree Plot analysis (PCA)

The Scree Plot (chart 1) analysis confirms the KMO and Barlett’s results, the graphic presentation of the pair “characteristic roots – dispersions”, is in the same line as in the table 5. The PCA, provided by SPSS programme, is a dimension reduction technique, used to locate the most important components from the components with the biggest characteristic roots ($>1$), helping so in identifying the variables that should be excluded from the model. We can clearly see the translocation, up, on the left, of the two biggest characteristic roots (loan portfolio number, loans in pursuit number). The other components, doubtful, substandard and lost loans...
have the lowest Elgen values (characteristic roots), so they should be removed from the regressive model.

Chart 1. “The graphic presentation of characteristic values”

In the same database, during 2011-2016, in a monthly period, as a dependent variable we consider the total number of the disbursed loans and as independent variables, the 4 types of loans (doubtful, in pursuit, substandard and lost). Next, after modeling the time series, we use the Expert Modeler to choose the most adequate model.

Table 6. “Model Description”

<table>
<thead>
<tr>
<th>Model ID</th>
<th>No. doubtful</th>
<th>No. lost</th>
<th>No. in pursuit</th>
<th>No. substandard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Type</td>
<td>Model_1</td>
<td>Model_2</td>
<td>Model_3</td>
<td>Model_4</td>
</tr>
<tr>
<td>Winters’ Additive</td>
<td>Winters’ Additive</td>
<td>Winters’ Additive</td>
<td>Simple Seasonal</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ projection, 2017

From the table 6, we see that the doubtful, lost and loans in pursuit are winters’ additive, meanwhile the substandard loans have a simple seasonal behaviour. In our final model, is included only the “loans in pursuit”, so we suggest that in winter season, the banks should pay more attention on the loans in pursuit.

Table 7. “Model Fit”

<table>
<thead>
<tr>
<th>Fit Statistic</th>
<th>Mean</th>
<th>SE</th>
<th>Min</th>
<th>Max</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Stationary R-square</td>
<td>.560</td>
<td>.105</td>
<td>.492</td>
<td>16</td>
<td>.492</td>
</tr>
<tr>
<td>R-square</td>
<td>.673</td>
<td>.257</td>
<td>.391</td>
<td>.391</td>
<td>.391</td>
</tr>
</tbody>
</table>

www.ijsernet.org
Table 7 presents the Model Fit, which tests the fit of the model with the observations and according R-square (almost 100% with the confidence level 95%), we can say that the model best fits with the observed values. This is because of the data we used are real data of the Albanian banking system.

Chart 2. “A graphic presentation of model fit”

Also, from the chart 2, we see that the model fits well with the observed values, which means that this model has good predictive ability. As we can clearly see from the chart 2, from 2009, the total number of disbursed loans has generally increased, with small up and downs, while on the January 2012 has reached the top, compared to the previous years. After that, begins a two-
year period decrease and since January 2014, the total number of disbursed loans has been gradually increasing.

Chart 3. “Forecast from ARIMA (1,0,0)"

Based in the chart 3, provided by the SPSS, after testing and confirmation of the model importance, we can do a forecast for a short period. As we see, for the next 4 years, we expect that the total number of the disbursed loans will arise gradually.

5. Conclusions

In this paper, we used the econometric models to establish a general approach on the progress of the loan portfolio quality in Albania and a prognosis of its behaviour via the NPL’s surveys. We focused on the non-performing loan’s level, because it provides an useful information for the quality of the financing of the economy and the credit risk that Albanian banking system faces nowadays. We studied the Albanian banking system for the period 2008-2016, using data from the Bank of Albania and the commercial banks. The focus is on the after-crisis period which is characterized by a declining performance of the loan portfolio, due to the extend of non-performing loans versus the total of gross loans.

Our research is an empirical study, which presents findings regarding the fragility of the Albanian banking system. In particular, our study contributes to enriching the existing literature by predicting the non-performing loans behaviour and the portfolio quality in Albania. Following the steps of regressive analysis, we developed econometric models with these variables: the loan
portfolio - the dependent variable of the model and the types of NPL - the independent variables (loans in persuit, substandard, doubtful and lost loans).

From the final model developed, with explanation level 91.1%, it resulted that the factor which affects the most the loan portfolio, is “loans in persuit”. From this, we can conclude that the Bank of Albania and the commercial banks should be more focused on the loans in persuit, in order to increase the loan portfolio quality.

Based on the KMO and Barlett’s test, that tests the validity and the compatibility of the data analysis, we can say that there is a good correlation structure between the variables. The Scree Plot analysis, that helps in identifying the variables that should be excluded from the model confirmed the ANOVA and KMO & Barlett’s results.

Next, through the Expert Modeler, we chose the most adequate model, we tested the model fit with the observations and according R-square ( ~100%), we can say that the model best fits with the observed values, which means that this model has good predictive ability. The last step of the regressive analysis step consists in a prediction of the behaviour of the loan portfolio in Albania. For the next four years, we expect the total number of the disbursed loans increase.

6. References


https://www.degruyter.com/view/supplement/s23445416_Instruction_for_Authors.pdf

http://eccsf.ulbsibiu.ro/