Athens University of Economics and Business
School of Economics
Department of Economics

"Macroeconomic determinants of NPLs for the case of Greece using a VAR model"

Vasiliki Tsourouni

A thesis submitted for the partial fulfillment of the requirements for the Master's Degree

Athens

February 2020

Supervisor

Tzavalis Elias

Athens University of Economics and Business

Examiner

Dendramis Yiannis

Athens University of Economics and Business

Examiner

Pagratis Spyridon

Athens University of Economics and Business

Acknowledgements

At this point I would like to thank my supervisor Mr Elias Tzavalis for providing intellectual stimulation, invaluable guidance, and constructive comments during the research process, writing and review of this thesis. Further, I would like to thank Mr Yiannis Dendramis as well as Mr Spyridon Pagratis not only for the willingness to be the examiners of this thesis but also for the knowledge they offered me during the MSc in Applied Economics and Finance. Also, I greatly feel the need to thank my friend Stamatina Balta. She really deserves a very special word of gratitude not only for her truly support and unconditional help but also for the numerous discussions we have shared regarding a variety of matters. Finally, I would like to thank my family and friends for their love, encouragement and care the whole period of this analysis. All these people motivated me to follow my dreams and achieve my goals in every aspect of my life.

Contents

Abstract

Loans have a vital contribution towards development of economy. It is a fact that its non-repayment leads to great loss on banks and country in general. Studying the determinants of non-performing loans (NPLs) constitute a favorite subject among economists. This study attempts to ascertain the effects of four macroeconomic factors on different categories of non-performing loans using a sample for Greece during the period 2002-2019. These factors consist of unemployment, inflation, economic sentiment index and exchange rate. The aim of this dissertation is to contribute to this field of literature by investigating how the rate of default of business, consumer and mortgage loans in our country is affected by the above factors. In particular, the major purpose of this thesis is to examine these specific determinants and show the importance of credit risk management. A Vector Autoregression (VAR) approach is used for the analysis in order to be able to forecast into the future. Overall, the results indicated that the repayment of loans depends on macroeconomic conditions. We concluded that business loans are actually affected by changes in the aforementioned macroeconomic variables. Also, consumer and mortgage loans seem to be influenced either negatively or positively but we cannot draw a general conclusion.

Keywords

Non-performing loans, Credit risk, Macroeconomic determinants, Time series, Rates of default, VAR (Vector Autoregression)

Chapter 1

Theoretical Analysis

In this chapter the issue of Non-Performing Loans (NPLs) is analyzed in a theoretical basis. The three series of banking regulations (Basel I, II, and III) are also described. Then, there is a reference in the relation of global financial crisis and NPLs, mainly in our country. Continuing our theoretical study we examine the types of risk a bank has to deal with, focusing on credit risk. Finally, a literature review is presented at the end of this chapter.

1.1 Introduction

The past few years presented one of the most turbulent times for international banking system. The global financial crisis was a severe worldwide economic crisis considered by many economists to have been the worst crisis for financial markets since the Great Depression of the 1930s. Despite the fact that massive bail-outs of financial institutions as well as other monetary and fiscal policies were used, a worldwide economic downturn followed. More than one decade has passed since its outbreak which started in the United States in 2007, nevertheless the consequences are still obvious. The Asian markets (China, Hong Kong, Japan, India, etc.) were promptly impacted and volatilized after the U.S. sub-prime crisis. Afterwards, in European Union, it evolved into a sovereign debt crisis, a crisis in the banking system of the European countries using the euro. Economic crises can have significant negative effects on various parts of the economy. A major area that can be affected by economic downturns is banking industry. This is shown by the large stock of non-performing loans (NPL) on balance sheets of many euro area banks. The issue of non-performing loans has expanded in the euro area, and especially in Greece. The situation is clearly illustrated in the figure below.

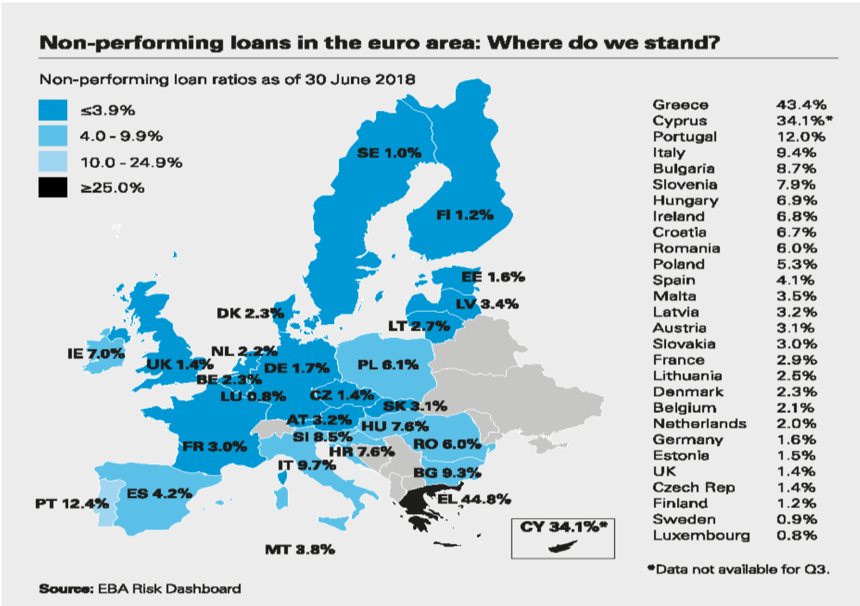

Figure 1.1: This figure presents the non-performing loan ratios in Europe. The statistics refer to the second quarter of 2018 and it is apparent that Greece has the greatest rate compared to the other countries of Euro area. (Source: EBA risk Dashboard)

The deterioration in the quality of loan portfolio had a negative impact even on developed economies. As far as Greece is concerned, the issue of non-performing loans is of increasing interest for more and more researchers, as controlling them is a vital priority for the proper operation of financial institutions. The changing economic environment within which the banks operated, led banks in Greece to adopt a different mode of operation with regards to the ways they handled risk. In order to achieve sufficient levels of profitability and survive in case of more intense competition, the banks were forced to improve the efficiency of their risk management and adopt sophisticated related technology (D. P. Louzis, A. T. Vouldis, V. L. Metaxas, 2010). This dissertation will try to understand the relationship between macroeconomic factors and the rates of default in each loan category (consumer, business and mortgages).

1.2 Non-performing Loans

It is true that financial institutions may face losses due to non-performing loans. The main objective of each bank is to provide loans, as it captures a large part of their profitability. However, the issue occurs when debtors cannot repay their loans or these loans go default. To be more specific, a performing loan will provide a bank with the interest income it needs to make a profit as well as extend new loans as defined by European Central Bank (ECB, 2017). So they make a profit for the banks and apart from that increase their liquidity. These in turn lead to providing new loans. However, there are also cases when a loan is considered as non-performing.

There is not a specific definition for these loans, since its country can set its own terms and thus differentiate it more or less. Nevertheless, according to the European Central Bank, a bank's loan can be classified as non-performing (or "bad loan") when payments of principal and interest are 90 days or more past due, or when it is not expected to receive the future payments in full. A key feature to consider a loan as non-performing is that the 90 day period has passed. Consequently, when clients do not meet their agreed repayment arrangements, the bank must reserve more capital on the assumption that they will not afford the loan. As a result, this puts serious constraints on banks' lending capacity and their ability to build further capital buffers. A high rate of NPL not only influences the stability of the banking system, but also creates systemic risk, which may in turn lead to a run on deposits, significantly reducing the intermediation power of banks. In addition, 'bad loans' are detrimental for the society as a whole, because problems in the banking sector can quickly expand to other parts of the economy, harming the outlook for jobs and growth. Therefore, the ratio of nonperforming loans (NPL) to total gross loans is a measure of the health of the banking system. The general situation of NPLs for the case of Greece is shown in the diagram below (the time period illustrated is from 2008 to 2018).

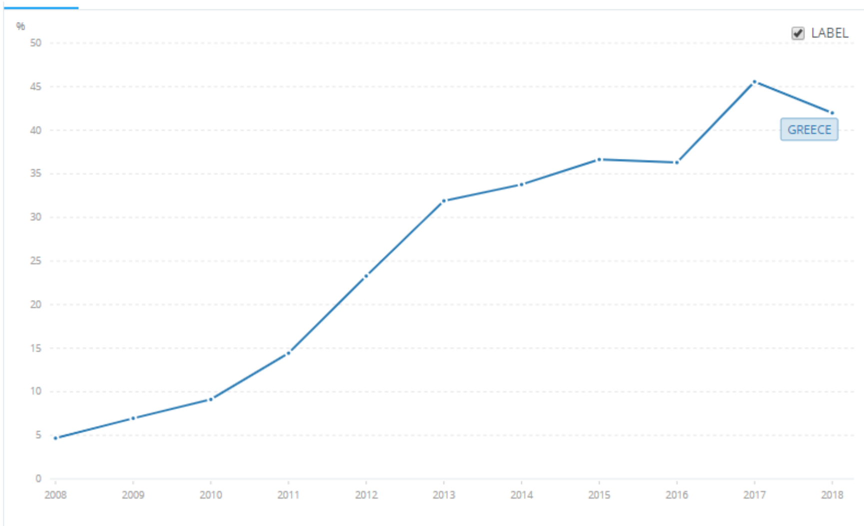

Figure 1.2: This diagram presents the overview of bank non-performing loans to total gross loans (%) in Greece. The horizontal axis represents time on a yearly basis from 2008 to 2018. It is shown that the line gets its maximum value of approximately 45.5% in 2017. (Source: https://www.worldbank.org/)

The existence of non-performing loans is a common phenomenon for banks, especially in our country. Not only persons but also businesses face financial difficulties. In order for banks to be successful in their long-term operation, they must keep the rate of non-performing loans at the lowest possible level. In this way, banks can continue providing loans and still bring profit from them. It is easily understood that the higher the percentage of non-performing loans, the more cautious banks become regarding their lending. Obviously, the 'normal' percentage of default varies to each country, depending on the regulatory environment. However, high rates of NPLs result in the appearance of a slowdown in the economy of the country.

1.3 Basel I, Basel II, Basel III

The necessity for tighter and more robust supervision of the global financial system was apparent long before the recent financial crisis. Bank regulation aims to ensure that a bank keeps enough capital for the risks it faces. Therefore, one of the main purposes of governments is to make the probability of default for any given bank very small.

In view of this, Basel I Accord, developed by the Basel Committee on Banking Supervision, was introduced in 1988 as an internationally agreed set of measures for regulatory capital for banks and it was the first attempt to set international risk-based standards for capital adequacy. Notwithstanding that 1988 BIS Accord paved the way for significant increases in the resources of banks regarding measuring, understanding, and managing risks, the “one size fit all” approach failed. The core drawback of this approach was its simplicity, as a loan to a corporation with a AAA credit rating is treated in the same way as one to a corporation with a B credit rating.

This led to a new and more comprehensive approach, which is known as Basel II. The new rules applied not only to all European banks, but also to securities companies. Essentially, what Basel II does is to adopt a three pillar approach to risk management. In particular, these are minimum capital requirements, supervisory review and market discipline. During the credit crisis, it was realized that some changes were essential for the calculation of capital for market risk. These changes are called Basel II.5.

After 2009, the need for more prudential regulation in the banking sector was imperative. In order to achieve this, in December 2010, Basel committee introduced new regulatory requirement commonly referred as Basel III, which tightened capital requirements and introduced liquidity requirements. The Basel III capital and liquidity standards were adopted by countries around the world. Based on the Basel I and Basel II documents, it aims to improve the banking sector's ability to deal with financial stress, enhance risk management, and strengthen the banks' transparency. Basel III was also adopted in order to reduce the risk of system-wide shocks at the individual bank level. Even though the new framework can play a crucial role in deterring a new financial crisis, it should not be regarded as a panacea for all the inadequacies of the banking sectors. Basel III itself is not the response to problems that were revealed by the global financial crisis.

1.4 Financial Crisis and Non-Performing Loans

At this point we will explain in more detail what economic crisis is. So, economic crisis is the phenomenon in which an economy is characterized by a sustained and noticeable decline in its economic activity. By referring to economic activity we mean all macroeconomic factors of the economy, such as national product, employment (or unemployment), prices, investments, etc. The most important indicator of economic activity is investment because changes of investment influence all the other financial variables.

We could not fail to refer to global financial crisis as a major factor affecting financial institutions and consequently the rates of non-performing loans. It is a fact that the default behavior of borrowers recedes when there is macroeconomic growth and is growing when recession exists as we have mentioned before. Since its establishment, the financial system has faced a great number of crises which are either local - and characterized as small level - or do not have big intensity and duration, are spread through the international operating network of financial lenders and markets, and finally taking a more general approach . Past financial crises have been devastating, affecting economies of both developed and developing countries. In our case, after a period of economic growth, the global economy recovered in the face of the worst financial and economic crisis since the Second World War. This crisis started in the USA financial sector in 2007 focused on the bank debt and in particular the inability to service mortgages loans, since 9.2% of mortgages were overdue or in an auction process. As a result, banks' liquidity was reduced and this had a negative impact on their ability to lend. It soon expanded as an epidemic to developed countries and thereafter around the world resulting in a major global recession. It has been described as the worst crisis since the Wall Street Crash and Great Depression (1920s and 30s) (Ashby, 2010) and even as the greatest crisis in the history of finance capitalism (Turner, 2009). This crisis was perceived in Greece with a delay for two of reasons. First of all, Greek economy is not very open and therefore it is not exposed to international distortions. Secondly, the banking system of our country did not face such big problems compared to the banks of other countries. Nevertheless, more than one decade has passed since the outbreak of the crisis, and the repercussions are still noticeable in the economy of our country.

With the onset of the economic downturn and in general the economic crisis in Greece, the financial system of our country had a direct impact. A major factor that has contributed to the crisis in our country was the uncontrolled lending by banks. This can explain the fact that since the end of 2008, the Greek banking system has imposed high interest rates and apart from that it often refuses to borrow because of liquidity problems. One of the major consequences of the economic crisis is the rise in unemployment, which has reached high rates while it is over 25% among young people. Theories about how financial crises develop and how they could be prevented have been offered by many economists. However, it may be inevitable to avoid them and thus financial crises continue to occur from time to time. Given the current economic crisis, there is an imperative necessity to address various challenges, including how the already high stocks of non-performing loans in a bank’s portfolio could be significantly reduced.

1.5 Types of risks

In an increasingly complex environment of the financial services industry, new complexities arise, requiring an adjustment in risk management systems and procedures. Understanding the risks posed to banks is a key factor, as governments can set better regulations to encourage prudent management and decision-making. Apart from that, investors’ decisions are also affected by the ability of a bank to manage risk. Hence, it is really important to define the categories of risks that financial institutions have to address properly. Many economists argue that the future of banking will undoubtedly rest on risk management dynamics. So, what other types of banking risks exist? As mentioned in several surveys, there are eight main types of risk related to banking industries, which are defined briefly above.

- Market risk

Market risk mostly occurs from a bank's activities in capital markets. Banks not only provide loans, but also hold a significant portion of securities. The unpredictability of equity markets, commodity prices, interest rates, and credit spreads create market risk. Banks are more exposed if they are heavily involved in investing in capital markets or sales and trading. In order to be able to mitigate such risks banks simply use hedging contracts.

- Operational risk

This kind of risk cannot be clearly defined. This type of risk occurs as the result of a failed business processes in the bank's day to day activities. In particular, this includes errors, interruptions, or damages caused by people, systems or processes. Typical examples of operational risk are payments credited to the wrong account or executing an incorrect order while dealing in the markets. It is inevitable for every department in a bank to be protected from operational risks. Main reasons involve hiring the wrong people, fraud, a breakdown of the information technology systems and inappropriate internal controls.

- Moral hazard

Moral hazard is a situation in which one party gets involved in a risky event considering that the other party will incur the cost. Basic requirement is the incomplete information about each party has about the other. In a financial market, moral hazard arises when for example, a borrower has incentives to act in a riskier way knowing that such risk-taking, won't be borne by the person taking that risk. In other words, the borrower might engage this risk in activities that are not desirable from the lender's point of view because they make him less likely to pay back a loan.

- Liquidity risk

Liquidity risk may arise when a prospective investor, business or financial organization fails to meet its short-term liabilities. In other words, it is the type of risk that the bank will not be able to meet its obligations if the depositors come in to withdraw their money. This usually occur when a bank has many short term liabilities (like customer deposits) and not enough short-term assets. The inability to provide cash in a timely manner to customers can result in a snowball effect. However, nowadays in case there is a run on a particular bank, the central bank diverts all its resources to the affected bank.

- Business risk

This can be described as the risk that a given bank may choose the wrong strategy, which can lead banks losing market share over time and being acquired or simply collapse. In contrast to operational risk, business risk is considered to be the risk arising from a bank's long-term business strategy. Suffice it to state the case of banks such as Washington Mutual and Lehman Brothers. Their approach was to be preferred as lender to people who have less than perfect credit scores. However, the whole area of subprime lending failed and taking into account that these banks had heavy exposures to such loans, they led to severe consequences.

- Reputational risk

By referring to Reputational risk we mean the risk arising from negative perception on the part of customers, counterparties, shareholders, investors, debt-holders, market analysts, other relevant parties or regulators that can adversely affect a bank's ability to maintain existing, or establish new, business relationships and continued access to sources of funding (BIS, 2009). Although this type of risk is, obviously, hard to measure, Reputational Risk Management departments of banks try to evaluate potential environmental, social or ethical risks arising from products, transactions and business relations.

- Systemic risk

This type of risk is often triggered by financial institutions. It refers to the risk of the collapse of an entire financial system or market. More specifically, it is the risk that a default by one financial institution will create a ''ripple effect'' that leads to defaults by other financial institutions and threatens the stability of the financial system (Hull). It is worth noting that systemic risk spreads from unhealthy institutions to relatively healthier institutions through a transmission mechanism.

- Credit risk

Credit risk, or default risk, is the risk that a financial loss will be incurred if counterparty to a transaction does not fulfill its financial obligations on time. Alternatively stated, it is the risk of loss that may occur from the failure of any party to abide by the terms and conditions of any financial contract. Hence, banks face credit risks from financial instruments such as acceptances, interbank transactions, trade financing, foreign exchange transactions, futures, swaps, bonds, options, settlement of transactions and others.

1.5.1 Credit Risk

As already mentioned credit risk is the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms. The riskiness is because it affects banks profitability. Banks try to handle large default rates, and thus enhance their performance. Having a solid credit risk management (CRM) is a critical component of a comprehensive approach to risk management and fundamental to the long-term success of any banking organization. Banks practice credit risk management not only to reduce their NPLs, but also succeed over their competition and thus improve return capital. Some studies even attribute bad CRM of commercial banks as one of the determinants for the global financial crisis.

Typically, the main source of credit risk for banks is loans. Huge volumes of loans which are not performing continue to be a great concern and challenge to many banks in Greece. There are two useful ways of analyzing the losses incurred by banks on their loan portfolios: firstly, by looking at the overall portfolio; and secondly, by examining the individual components of the portfolio (ECB, 2007). Considering the former standpoint, looking at the overall portfolios, banks typically expect to lose a certain amount on average, which is known as expected loss (EL). They cover EL by including a risk premium into the interest rate charged to borrowers and using loan impairment charges. But what about losses that transcend expected losses? These are called unexpected losses (UL), although institutions are aware such losses will occur.

The uncertainty derives from the fact that it is unknown when these losses might take place, and which their magnitude be. In order to tackle with unexpected losses, banks have to maintain adequate capital. The amount of capital held depends on bank's management and regulatory requirements, as well as requirements of external parties such as rating agencies, and the investors' view of the bank's risk-return profile. It is obvious that, specifying the optimal level of this capital is crucial as holding in excess entails an opportunity cost, as this money could otherwise be used to finance additional lending. The latter analysis is by looking at its loan portfolio individual components. The expected loss of each loan exposure can be divided, for instance, to the probability of default (PD), the exposure at default, and the loss given default. The first one is the probability of not repaying the loan. The exposure amount (E) is the amount outstanding in the event of the borrower's default. In that case, the loss given default (LGD), i.e. the actual loss faced by the bank, depends on how much of the initial debt can be recovered through a bankruptcy proceeding and the amount of collateral (if available).

It is undeniable that today, credit risk raises concerns of all financial institutions, even if there are not really at risk of lenders being unable to meet their obligations. Credit risk is the most fundamental risk to the operation of banks and financial institutions in general, as it has proven to have destroyed the many market portfolios. In fact there is a number of definitions of credit risk in literatures. According to Basel Committee on Bank Supervision credit risk is the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms (Bank of International Settlements, 2000). Alternatively, credit risk is defined as the risk that unexpected change in a counter-party's creditworthiness may generate a corresponding unexpected change in the market value of the associated credit exposure (Sironi & Resty, 2007). Hence, credit risk induces financial distress on banks and better credit risk management is necessary in order to reduce the number of non-performing loans. The main components of credit risk for its measurement and evaluation are default, probability of default (PD), loss given default (LGD) and finally exposure at default (EAD).

At this point it is worth mentioning what the financial system is. In essence, it is a set of institutions, such as banks, insurance companies, and stock exchanges that permit the exchange of funds. The financial system also includes sets of rules and practices that borrowers and lenders use to decide which projects get financed, who finances projects, and terms of financial deals (Investopedia).

1.5.2 Credit Analysis – Objective view

Credit risk is one of the most important areas of risk management, since it plays a key role especially for banking institutions. This is the reason banks perform their credit risk models - trying to improve them as much as possible - with the aim to ameliorate the quality of each bank portfolio. Credit analysis is thus necessary in order to assess the ability of a customer to pay back the loan or evaluate creditworthiness of new debtors.

Credit analysts examine credit risk levels testing not only the subjective but also objective analysis. As far as subjective evaluation is concerned, the five Cs of good credit is usually implied, described by Apostolik, Donohue & Went (2009). A rather interesting issue is whether the ownership status of a bank is related to its profitability, so five Cs concern key features when assessing borrowers. To be more specific, these are the following:

1. *Character*, which suggests the debtor's willingness to repay a loan, his/her reputation for example in the workplace and in general in relationships with other lending institutions. There is no doubt that bank officers will check customer's historic transactions to detect any event related to credit lending.
2. *Capital* denotes the capital structure of the borrower. Credit analysts study the level of leverage of the target firm by assessing the amount of debt as well as equity that is used as source of finance.
3. *Conditions* refer to external factors that might affect the borrower's financial profile and thus enhance or lessen his ability to repay. These factors come from the overall economic environment and related industry.

4. *Capacity* is also an important factor. In specific, banks focus on cash flow reports of customers and since one of their aims is to lend out money to firms, it is significant to ensure predictable, stable cash flow and alternative sources of credit to pay back loans.

5. *Collateral* implies assets of the borrower that are used to securitize loans. In case of default, the bank has the opportunity to sell these assets to compensate for part or all of the loss.

The objective analysis is substantial for Credit Risk Management. It should be underlined that NPLs are among the main reasons of the problems of economic stagnation. Each 'bad' loan raises the possibility to lead company to difficulty and unprofitability. Thus, keeping them in a minimum level is necessary condition in order to improve economic growth. NPLs are likely to hamper economic growth and decrease the economic efficiency (Hou, 2007). In the present study, determinants of non-performing loans for each loan category in Greece, regarding macroeconomic factors, are investigated. The research of exclusively macroeconomic factors lies in the availability of data. For example, bank-specific determinants data (such as loan to deposit ratio, return on assets etc) would not be available. The loan categories are business, mortgage and consumer. It is of urgent importance to understand and identify the factors that affect NPL and provide some guidance both to banks - so that they improve their credit policies - and governments as well as banks' supervisors so that appropriate preventive measures and stress testing models are adopted. The increase in loan defaults underlines the links between macroeconomic and financial shocks and the relationship between the friction in the credit market and the risk of financial instability.

1.6 Brief review of the main Credit Risk Models

Nowadays, more and more researches have been conducted regarding credit risk analysis. The following credit models are widely implemented by banks, which frequently use them to assess their own credit risk.

The most important tool for the assessment of credit risk is credit scoring. Credit Scoring Models use data on observed borrower characteristics in order either to calculate a result (score) which represents the probability of default or to sort borrowers into different default risk classes. Selecting and combining different economic and financial borrower characteristics, can help banks to determine which factors are important in explaining default risk, evaluate the relative degree or importance of these factors, enhance the pricing of default risk, be better able to screen out bad loan applicants and be in a better position to calculate any reserve needed to meet expected future loan losses. A predetermined weight is allocated to each item considered in the model and a credit rate assigned depending on the credit assessment. Several statistical methods are used to develop credit scoring systems, that was introduced in the 1950's, including linear probability models, logit models and linear discriminant models.

- Linear Probability Model (LPM)/ Linear Regression Model

The linear probability model uses past data, such as accounting ratios, as inputs into a model to explain repayment experience on old loans. It assumes that the probability of default varies linearly with these variables. On the other hand, linear regression is related to LPM as it is the process of establishing a relationship between one dependent variable with one independent variable (simple linear regression) or between multiple independent variables (multiple linear regression). These models can be used for assessing the probability of repayment p or the probability of default PD. The model estimated by linear regression is

$$PDi = \Sigma\beta jXij + \text{error}$$

Where i takes different values for consumer, business and mortgage loans, Xj denotes the independent variables and βj is the estimated importance of the jth variable in explaining past repayment experience. If we then take these estimated βjs and multiply them by the observed Xij, we can derive an expected value of PDi for the actual probability of no repayment on the loan which takes values from 0 to 1. In the LPM a slope coefficient βj measures the change in the probability of success (Y=1) due to a change in Xj :

$$bj = [\ \theta \ \Pr(Y = 1\ \mathrm{I}\ X)] / \theta \ Xj$$

Despite the fact that this method has significant advantages, there is one major drawback. The estimated probabilities of default may lie outside the interval [0,1]. In order to address this problem, logit or probit models are used.

- Logit (/Probit) Model/ Logistic regression model

These models utilize more sophisticated regression techniques that constrain estimated default probabilities within a 0-1 range. Logistic regression model is usually used to calculate the probability of default. The logistic model (logistic discriminant analysis) assumes that the default probabilities are given by

$$p\,(X) = \Lambda(b_0 + b_1 X_1 + \ldots + b_p X_p),$$

where Λ is the logistic cumulative distribution function: $\Lambda(z) = e^z / (1 + e^z)$. On the other hand, Probit models assume that default probabilities are normally distributed and the default probabilities are

$$p(X) = \Phi\,(b_0 + b_1 X_1 + \ldots + b_p X_p),$$

where the standard normal CDF $\Phi\,(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}u^2}\, du$.

Due to nonlinear features of these models it is necessary to use maximum likelihood method for parameters estimation. The objective of the logistic model in credit scoring is to assess the conditional probability of a specific observation belonging to a class given the values of the independent variables of the credit applications.

- Linear Discriminant models

These models want to establish a linear classification rule or formula that best distinguishes between particular groups of borrowers. While linear probability and logit models project a value for the expected probability of default if a loan is made, discriminant models divide borrowers into high or default risk classes based on their observed characteristic (X). A basic principal is to maximize the difference between two groups, while the differences among particular members of the same group are minimized. Altman's Z-score is a popular application of multivariate Discriminant analysis in credit risk modeling.

$$Z=1.2\ X_1+ 1.4\ X_2 +3.3\ X_3 + 0.6\ X_4 + 1.0\ X_5$$

Where X_1 = Working capital/total assets, X_2 = Retained earnings/total assets, X_3 = EBIT/total assets, X_4 = Market value equity/ book value LT debt, X_5 = Sales/total assets.

Values less than 1.81 indicate that the borrower has low credit rating, whilst the opposite is true for values more than this critical value. However, a significant drawback of Altman's Z-score is that it only considers two extreme cases (default/ no default) and it ignores to quantify factors including business cycle effects. Apart from that, discriminant models are linear. Non linear models are more able to indicate the credit rating of a borrower.

Apart from the above parametric methods, also non parametric techniques have been established in recent years and they are characterized by increased flexibility owing to the fact that they are not limited to statistical assumption (in contrast with parametric techniques). One of the most important method is:

- Neural networks

Neural networks, which are supervised learning systems, have recently emerged as an effective method for credit scoring. The model consists of multiple layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after crossing the layers multiple times. Each neuron is a node that is connected to other nodes via links. Each connection can transmit a signal to other neurons. Neurons and links typically have a weight that adjusts as learning proceeds in order to determine the type and intensity of the information exchanged. Neural networks perform in the same way as the biological neural. In contrast with the other methods, neural networks can model the relation between a set of inputs and a set of outputs, under the assumption that the relation is nonlinear. The backpropagation method is used to compute the input to a neuron from the outputs of its predecessor neurons and their connections as a weighted sum.

There are several credit risk models except for credit scoring models, such as RAROC models which use Risk adjusted return on capital, which are beyond the scope of the present analysis. Other models also widely used are Value at Risk, CreditMetrics of JP Morgan and CreditRisk+ introduced by Credit Suisse Financial Products. As far as non-parametric techniques, decision trees, Expert Systems (ES), Support Vector Machines (SVM) and Hybrid Models are used as well.

1.7 Literature Review

The assessment of credit risk has always been essential to banks and other financial institutions. Hence, there exists a broad literature on the determinants of NPLs. In general, the surveys conducted in developed economies have confirmed that macroeconomic conditions have a significant effect on credit risk. There exist numerous studies that explore macroeconomic and banking industry specific determinants of NPLs for various countries and regions, and most of them lead to an inverse relationship between economic environment and NPLs. The academic literature provides proof to suggest a robust relationship between the NPLs and plenty of macroeconomic variables, like inflation and unemployment rate. Almost all studies identify the unemployment rate (or gross domestic product growth rate) as well as the inflation as crucial determinants of credit risk.

One of the first and widespread approaches to evaluate credit risk involves Merton's (1974) model. Merton's Model is the theoretical foundation of structural models and it was used with the intention of understanding how capable a company is at meeting financial obligations, servicing its debt, and weighing the general possibility that it will go into credit default. The basic assumption behind this model is that a firm's equity is analogous to a call option on the firm's assets. The key feature of this model is that the underlying state variable that determines a firm's default is the value of its assets. Nevertheless, Merton's model has some important drawbacks. More precisely, the term structure of default free interest rates is assumed to be constant, the assets of the firm do not trade and as a result their prices are not observable, the liability structures change over time and finally in case of default, the absolute priority structures are not adhered to by the bankruptcy courts. It should be mentioned that this model was later extended by Fischer Black and Myron Scholes to develop the Nobel-prize winning Black-Scholes pricing model for options.

Klein (2013) investigated both the macroeconomic and the financial factors that affect NPLs in Central, Eastern and South Eastern Europe (CESEE) for the period 1998-2011. It should be noted that macroeconomic factors were found to have more explanatory power than bank specific ones.

Messai (2013) studied not only macroeconomic but also bank-specific variables on NPL for countries like Greece. The results showed the existence of a significant positive relationship of the unemployment rate with the ratio of NPLs. This can be explained as after the outburst of recent financial crisis higher unemployment rate occurred and thus more people were unable to meet their debt obligations.

The paper of D. P. Louzis, A. T. Vouldis and V. L. Metaxas (2010) utilized dynamic panel data methods using data comprising the 9 largest Greek banks, in order to examine the determinants of non-performing loans, separately for each category of loan (consumer, business and mortgage loans). The macroeconomic factors examined were GDP growth, unemployment and interest rates. Apart from that, real lending rates were taken into account. The dependent variable was the NPL ratio which is defined as the ratio of the NPLs to the value of total loans.

The model performed was the following

$$\Delta NPL_{it}^{h} = a\,\Delta NPL_{it-1}^{h} + \sum_{j=1}^{2} \beta_{1j}^{h}\ \Delta GDP_{t-j} + \sum_{j=1}^{2} \beta_{2j}^{h}\ \Delta UN_{t-j} + \sum_{i=1}^{2} \beta_{3j}^{h}\ \Delta RLR_{it-j}^{h} + \sum_{i=1}^{4} \beta_{4j}^{h}\ X_{it-j}^{h} + \eta_{i}^{h} + \varepsilon_{it}^{h}$$

with a <1, i = 1,...,9 and where X_{it}^{h} denotes bank-specific variables, superscript h denotes the type of loan and Δ means first differences. GMM estimation was used for the analysis. This paper shows that real GDP growth rate, unemployment rate and lending rates have a strong effect on the level of NPLs through the use of dynamic panel data methods. In addition, bank-specific variables, which include performance and efficiency indicators, have also a considerable impact on rates of default when added to the model. Apart from that, different categories of loans respond in a different way to their determinants. For instance, the effect of the GDP growth rate is found to be stronger for business loans compared to the other types of loans.

One of the earliest studies tried to understand the major factors concerning NPL in countries like Greece, is that of D. Anastasiou, H. Louri and M. Tsionas (2016). In particular, euro-area core (which consists of Austria, Belgium, France, Germany, Finland, Lithuania, Luxemburg, Netherlands and Slovakia), periphery countries (Greece, Italy, Ireland, Portugal and Spain) as well as the whole euro-area are examined in the analysis. This study tests the existence of a long run effect by both macroeconomic and bank specific determinants using Fully Modified OLS (a non-parametric Approach) and Panel Cointegrated VAR. The dependent variable is again the ratio of non-performing loans to total loans. The econometric models examined are next.

$$\text{Model 1: } \frac{NPLs_{it}}{Total\_Loans_{it}} = \beta_0 + \gamma_i M_t + \text{CRISIS\_DUMMY} + e_{it}$$

$$\text{Model 2: } \frac{NPLs_{it}}{Total\_Loans_{it}} = \beta_0 + \delta_i B_{it} + \text{CRISIS\_DUMMY} + e_{it}$$

$$\text{Model 3: } \frac{NPLs_{it}}{Total\_Loans_{it}} = \beta_0 + \delta_i B_{it} + \gamma_i M_t + \text{CRISIS\_DUMMY} + e_{it}$$

where B is a vector of bank-specific variables, M a vector of macroeconomic factors, and i and t represent euro-area (periphery and core) banks and time respectively. CRISIS is a dummy variable which takes the value of 1when t≥2008Q1, and zero other otherwise. It is included in the model so as to check if the 2008 financial crisis in Europe gave rise to a systemic break in the formation of NPLs. The Augmented Dickey-Fuller (ADF) Fisher type test is used to check the existence of unit root in the panel and then the analysis proceeds to implementation of Fully Modified OLS (FMOLS) method. Moving to the next step, Panel Cointegrated Vector Autoregression is applied. Continuing, test for possible existence of fragmentation between core and periphery banking markets takes place.

The results from FMOLS estimation indicate that unemployment rate, GDP growth rate, output gap, tax on personal income and credit to GDP have a significant influence on NPLs. Also, it was found that crisis has caused a structural break to NPLs by shifting them upwards. Apart from that, it was shown that the NPLs of the periphery react more strongly to the determinants examined than the NPLs of the core. Regarding the results after the panel Cointegrated VAR estimation, these are similar to the previous method. To conclude, this paper proved that NPLs in the euro area have followed an upward shift after 2008 and it is mostly due to worsening macroeconomic conditions especially with respect to unemployment, growth and taxes. Fiscal consolidation and interest rate margins are significant for the euro area periphery (such as Greece) while credit to GDP is significant only for the euro area core. Quality of management (through ROA and ROE) and moral hazard (through the loan to deposits effect) play an important role, while size is negatively significant and exerting a stronger effect in the periphery.

The seemingly unrelated regressions (SUR) framework was used by E. Charalambakis, Y. Dendramis and E. Tzavalis (2017) in order to determine the relationship between these three categories of NPL and lagged values of their determinants. To be more specific, the framework used extends the SUR framework, since it allows for a common break - which can capture the influence of exogenous events such as the deterioration of the economic conditions, sovereign debt crisis, political events etc - in the relationship between NPLs and their determinants. The reduced form model used for the analysis is:

$$\Delta NPL_{it} = (c_i + b_1\, \Delta ROA_{t-1} + b_2 \%EQTY_{t-1} + b_3\, \Delta LTD_{t-1} + \gamma_1\, \Delta UNPL_{t-1} + \gamma_2\, INFL_{t-1}) * DUM_{t-1} + (+\, b_1^*\, \Delta ROA_{t-1} + b_2^*\, \%EQTY_{t-1} + b_3^*\, \Delta LTD_{t-1} + \gamma_1^*\, \Delta UNPL_{t-1} + \gamma_2^*\, INFL_{t-1}) * DUM_{t-1}^* + \rho\, \Delta NPL_{It-1} + u_{it},$$

where Δ denotes first-difference, % denotes percentage change of a variable, i = 1, 2 and 3 denote the three business, mortgages and consumer loans, respectively and DUM_{t-1} is a dummy variable which takes the value of 1 when $t-1 \leq T_0$, when a structural change in model occurs, and zero otherwise. DUM_{t-1}^* is the complementary variable to DUM_{t-1}, and thus it takes the value of 1 when $t-1 > T_0$, and zero otherwise. The bank specific variables included in the model are: ROA defined as earnings before interest and taxes divided by total assets, equity and loan-to-deposit ratio and the macroeconomic factors are the unemployment and inflation rate. T_0 is estimated using an optimization problem. Before proceeding to estimation of the model, all variables become stationary. Then, the lag length is determined based on the Akaike information criterion. Maximum likelihood is used to estimate the model, with and without a structural break in the relationship between NPLs and their determinants.

Regarding the results of this analysis, changes in the unemployment and inflation were found to have a significant impact on NPLs, especially after the first quarter of 2012. On the other hand, bank-specific variables such as changes in equity and loan-to- deposit ratio do not appear to have an important influence on NPLs. However, ROA seems to reflect bank management conditions.

It is important to mention that most of the literature is based on country specific studies and therefore many research papers include a composition of micro and macroeconomic factors. For instance, Louzis et al. (2012) investigated separately business, mortgage and consumer NPLs in Greece in order to test same factors in each category of loans. Based on this paper GDP, unemployment, interest rates, public debt (macroeconomic factors) as well as management quality explain a high rate of NPLs. It is worth mentioning that mortgages were found to be the least responsive to changes in the macroeconomic conditions. This leads to the assumption that for better results each category of loans should be treated separately.

Plenty of papers examine determinants of NPLs at macroeconomic or bank level. Nevertheless, D. Anastasiou, Z. Bragoudakis, I. Malandrakis (2019) present how governance indicators along with some additional macro factors have repercussions on NPLs. Among their findings is that higher levels of governance indicator (which consists of six governance variables: Voice and Accountability, Political Stability, Government Effectiveness, Regulatory Quality, Rule of Law and Control of Corruption) signify a relatively stronger and more stable banking system and hence lower levels of NPLs. The first step of their survey was Principal Component Analysis (also known as PCA), which is actually a dimension reduction method. PCA was useful to aggregate the existing information in the six different Governance Indicators into one governance index. After that, screen plot was used in order to determine which variables must be kept for the investigation. Then, two econometric models are applied, one static and one dynamic. The formulation of these models is next:

$$\text{Static model: } NPLS_t = a + \beta \, \text{GOVERNANCE}_t + \sum_{i=1}^{6} \delta_i \, C_{it} + \varepsilon_t$$

$$\text{Dynamic model: } NPLS_t = a + \beta \, \text{GOVERNANCE}_t + \gamma \, NPLS_{t-1} + \sum_{i=1}^{6} \delta_i \, C_{it} + \varepsilon_t$$

where NPLs, GOVERNANCE and C_i denote non-performing loans to total loans of the Greek banking sector, the first component from the PCA of Worldwide Governance Indicators(WGI) and all the other control variables (DEPOSITS_GDP, GDP_GROWTH, UNEMP, ROA, BANK_CONCENTRATION, CRISIS_DUMMY and SYSTEMIC_LIQ_RISK) respectively. The estimation of the models is based on OLS and using robust standard errors. Apart from that, Augmented Dickey-Fuller and DF-GLS unit-root tests were applied to check for the stationarity condition. The results of the analysis showed that higher levels of WGI – and thus higher levels of the six variables it includes - led to lower levels of default. In addition, it was concluded that the recent financial crisis led to an increase of NPLs in Greece likely due to the fact that after its outbreak, higher unemployment rate occurred and thus more people could not meet their debt obligations. Also, it was shown that systemic liquidity risk has a positive impact on NPLs.

The rest of the dissertation is organized as follows. In the next chapter a presentation of the factors that affect NPLs as well as the categories of loans is conducted. Also, in chapter 2 the explanation of the data and a brief description of the econometric models performed take place. Continuing, the technical framework of this investigation is presented extensively. In section 4 we present the empirical econometric methodologies and the empirical results. Finally, section 5 concludes.

Chapter 2

Data used in the analysis

This chapter describes extensively each of the time series used in the present research. Both the group of macroeconomic determinants (independent variables) and each loan category (dependent variables) are theoretically analyzed. An introduction to the methodology of the model is presented briefly at the end of this chapter.

2.1 Macroeconomic factors considered

Various macroeconomic indicators are used as explanatory variables relating to the indicator of the default rate in the economy. The determinants of NPLs as a proxy for credit risk have been under thorough investigation, especially after the financial crisis. It is of urgent importance to understand and identify the factors that affect NPLs and provide some guidance both to the banks in order to improve their credit policies and to the governments and the banks' supervisors so that appropriate preventive measures and stress testing models are adopted. Based on a variety of thesis studying the phenomenon of non-performing loans, we led to the macroeconomic factors which we think they affect the NPLs. Before we proceed with the statistical analysis, we will explain these macroeconomic factors. In this dissertation, the main explanatory variables under inspection are Economic Sentiment Indicator (ESI), Exchange Rate, Unemployment rate and inflation rate, and they are analyzed below.

2.1.1 Economic Sentiment Indicator

The Economic Sentiment Indicator (ESI), established in 1985, is a composite indicator which consists of five sectoral confidence indicators with different weights (https://ec.europa.eu/eurostat/). To be more specific, it includes:

-industrial confidence indicator (40 %),

-construction confidence indicator (5 %),

-services confidence indicator (30 %),

-consumer confidence indicator (20 %) and

-retail trade confidence indicator (5 %).

But what are confidence indicators? According to Eurostat, confidence indicators are arithmetic means of seasonally adjusted balances of answers to a set of questions closely related to the reference variable they are supposed to track. For example, questions for consumer confidence index may involve financial position and savings of participants. The questions are both backward-looking (over the last 12 months) and forward-looking (over the next 12 months). Surveys are defined within the Joint Harmonised EU Programme of Business and Consumer Surveys. Their aim is to reflect the overall perceptions and expectations at individual sector level in a one dimensional index. Significant benefits of these indicators are that they are both timely and can include information that is known by the survey respondents but not yet reflected in aggregate economic variables, such as consumption expenditures, employment or GDP. These indexes can capture economic developments immediately, since they are available earlier than national accounts or output data and are subject only to limited revision.

Although confidence indicators are not straightforward connected with economic activity, they seem to be more relevant during crisis periods. In particular, many authors underline the special importance of confidence indicators in predicting periods of strong fluctuations in the economy, such as recessions and recoveries, or during periods of major economic or political shocks. Further, surveys conducted have led to the existence of a significant statistical relationship between confidence measures and economic variables, both current and future. During normal times of economic activity, changes in confidence can lead to invalid results. They may either reflect false interpretations about the economic situation or the information content of such indicators may be small. Consequently, sentiment indicators may have poor leading properties. However, during stressed periods, a significant deterioration in confidence can have some predictive power with regard to future economic developments. In such circumstances, confidence indicators can show a significant change in economic agents' behaviour, which is likely, in turn, to have real implications. It is worth noting that the euro area sovereign debt crisis in the summer of 2011 was associated with a large fall in the consumer confidence index.

The ESI has information content for the GDP growth rate and, thus, it can be used to gauge economic agents' perceptions of future economic activity. In this respect, the domestic ESI may show important information content for the spreads (Rua, 2002). The economic sentiment index aims to capture expectations about immediate economic conditions. It is worth underlying that the Economic Sentiment Indices are forward-looking indices since they are derived by surveys of households and corporations. In general, a sentiment indicator seeks to quantify how current beliefs and positions affect future behavior.

2.1.2 Exchange Rate

Generally speaking, an exchange rate can be defined as the value of one nation's currency versus the currency of another nation or economic zone. In our analysis, Euro to USD rate is selected as this pair has become the most widely-traded pair in the world, since it represents a combination of two of the biggest economies in the world. The main reason for this is the financial zones that these two currencies represent. Euro depicts, in a significant degree, the economy of Europe and on the other hand the US dollar illustrates the financial situation in the USA. It is affected by factors that influence the value of the euro and/or the U.S. dollar in relation to each other and to other currencies. It is important to highlight that the base currency of the pair is fixed and always represents one unit. Consequently, this number solely cannot reflect the source of the strengthening and/or weakening. An increase in the EUR/USD rate can derive from either the euro is getting stronger or the U.S. dollar is getting weaker. Both conditions result in an upward movement in the rate and a corresponding upward movement in a price chart.

It should be pointed out that the risk of sovereign default and exchange rate fluctuations are inextricably linked, considering that the depreciation of a country's currency is often a reflection of poor economic conditions (P. Augustin, M. Chernov, D. Song 2018). Default events tend to be associated with currency devaluations. Such devaluations may either strategically support the competitiveness of the domestic economy, or penalize a country's growth due to increased borrowing costs or reduced access to international capital markets. Also, surveys have shown that exchange rate system establishes a structural frame work for conducting foreign exchange transactions which affect banks profitability. Finally, as it is obvious foreign exchange rate fluctuations cannot make an impact on the NPL rate in the same month, but rather in the periods after that (with delay).

2.1.3 Unemployment rate

There is no doubt that unemployment rates affect consumers' delinquency and bankruptcy behavior. As reported by OECD, unemployment rate is the number of unemployed people as a percentage of the total labor force, where the latter consists of the unemployed plus those in paid or self-employment. Unemployed people are those who report that they are without work, are available for work and have taken active steps to find a job in the last four weeks. Unemployment can be divided into three categories regarding the time period the person remains without employment. These are long-term unemployment, short-term unemployment and seasonal unemployment. This variable is included in our analysis as it captures the business and macroeconomic conditions in Greece. An alternative option would be the real Gross Domestic Product (GDP) growth rate. Nevertheless, based on Monokroussos and Thomakos (2016) choosing one of these two macro variables is sufficient to capture the macroeconomic conditions in the economy. A sharp rise of unemployment in our data is observed after 2011, when it reaches the value of about 25 percent. The unemployment rate is a valuable measure of the mismatches between labor supply and demand. It should be mentioned that it is one of the most extensively produced and used labour market indicators. The high and highly persistent unemployment rates experienced by many countries and regions both in Europe and the US have attracted a significant amount of both theoretical and empirical work. Several surveys indicate this macroeconomic factor as one of the most significant for the rise in rates of default in recent years. In our analysis, unemployment rate is considered as one of the main factors, because we believe that the change in the borrower's financial situation can cause major changes in his/her trading behavior towards banks and its debts. In other words, as unemployment increases, more and more people it is obvious that are unable to meet their debt obligations.

4.1.4 Inflation rate

A simple definition of inflation can be the change in the prices of a basket of goods and services that are typically purchased by specific groups of households. A more sophisticated definition however would be that inflation is a persistent increase in the level of consumer prices or alternatively a persistent decline in the purchasing power of money, due to an increase in available currency and credit beyond the ratio of available goods and services. Inflation rate is calculated based on Consumer Price Index (CPI) and more precisely, it derives from the rate of change in the CPI over a given period. The purpose of the CPI is to measure the general level of the prices of goods and services supplied by the average Greek household. In particular, it is determined by the evolution of certain indicators related to CPI. Each of the these indices is evaluated using a sample of prices for a defined set of goods and services obtained in, or by residents of, a specific region from a given set of outlets or other sources of consumption goods and services. In Greece, the most important categories in the consumer price index are food and non-alcoholic beverages, transport, housing and hotels, cafés and restaurants. Another way for measuring inflation is the Producer Price Indexes (PPI). These indexes measure the change in selling price that a producer is able to get for a good or service.

According to several authors inflation was not considered as a serious threat for economic growth until few decades ago. In contrast, before the beginning of 20th century several countries were experiencing deflation. Nowadays however inflation has a negative impact on the economy as it reduces the purchasing power of incomes and the competitiveness of the economy, encourages imports, reinforces income inequality and minimizes the trend for savings. Hence, this can explain the fact that inflation rate can be a determinant factor for the repayment of all loan categories and thus should be included in our investigation.

To conclude, there are many factors which influence the default rates of loans apart from the above mentioned. To be more specific, the income of the borrower, the status of the client, a variety of ratios such as loan to value ratio can have an important impact on the repayment of loans and consequently affect the corresponding rates of default. However, the availability of these data is not feasible, since these variables are related to specific customer personal data. For this reason our analysis will be limited to the study of the four aforementioned macroeconomic variables.

2.2 Categories of Non-Performing Loans

Starting with the analysis of loans characterized as non-performing, it is reasonable to define and specify precisely their concept. The definition of non-performing loans varies to each country, since some of them have stricter regulations. However, a Non-performing loan (NPL) is considered as loan for which the borrower has not made the scheduled payments (of the total amount or a part of it) of the interests or/and capital for a specified period, or it is under legal action. The exact details of the loan repayment status vary according to the terms of the particular loan.

From a macroeconomic standpoint, Zeng (2012) describes non-performing loans as "Economic pollution" wanting to emphasize in their negative impact on the social prosperity. Keeton and Morris (1987), Meyer and Yeager (2000) and Gambera (2000) were some of the first authors try to investigate the impact of macroeconomics factors on the quality of banks' assets. Through their research it was found a significant relationship between loan quality and the behaviour of certain variables (such as unemployment, income, inflation, etc.). Given this aspect, it is understood that the reduction of the rate of non-performing loans is possible only by a general recovery of the economy. The ratio of non-performing loans to total loans of the Greek banking sector (NPLs) is used in our analysis, since it can be considered as a measure of the aggregate credit risk of the banks in our country.

One would probably wonder which the main categories of these loans are. To answer the question, a banks' portfolio is usually divided into three categories; mortgage, business and consumer loans. These are the categories that will be examined in the present analysis. To be more specific, the rates of defaults of each loan category describe the three variables of loans. Let us now define each loan category.

2.2.1 Consumer loans

To begin with, a consumer loan is a loan that is granted to meet personal needs for the purpose of obtaining products and services and they should not be related to the borrower's potential professional needs. The way they are acquired, their nature and the amount granted, and the purpose of obtaining them varies considerably between financial institutions. In general, however, consumer loan conditions are analyzed below. The amount granted on a consumer loan usually does not exceed EUR 25,000 and may not exceed 65% of the value of the product or service purchased. The difference (at least 35%) is paid by the borrower. Moreover, there are cases where the amount is not given to the customer but is given directly to the merchant from whom the loan was financed and sold. However, there are consumer loans that are not granted for the purpose of purchasing products and therefore the borrower is not required to provide market documents. Consumer loans sometimes include mortgage loans, personal loans or even car loans. That is, the borrower seeks a consumer loan for the purpose of buying or repairing a home, buying a car or purchasing goods and services.

Consumer loans can usually be obtained by all persons, whether employed or self-employed, who file a tax return. However, the necessary documents and the conditions for obtaining a consumer loan may differ between the different financial institutions.

2.2.2 Business loans

Business loans are defined as those that are granted by a financial institution to a client, usually a large business, or a small and medium-sized or even freelance professional. Every financial institution checks whether a loan can be given to a company, since the terms and conditions of the loan usually differ. A business loan is given in order to meet certain specific needs depending on the type of borrower. There are many purposes for which a business loan can be obtained, as for example the purchase of land area, purchase of ready or under construction business premises, completion of a business premises, business renovation (upgrades, additions, repairs), purchase of business goods, payment of suppliers. Business loans are a type of financing a business. Generally, a company in order to meet its specific needs chooses through a wide variety of financing methods provided by financial institutions. A loan is one of these ways.

2.2.3 Mortgage loans

A mortgage loan is a very common type of securities used by many people to buy a house. In this type of loan the borrower utilizes the money to buy a real estate. The financial institution, however, ensures a security until the mortgage is totally paid. Should the loan fail to repay, the bank will have the legal right to seize the home and sell it in order to recover the amounts owed. Although the mortgage loan seems to be a secured loan because of the bank's safeguards, however, the value of the real estate may be decreased and thus even after a possible foreclosure the bank cannot cover the amount of customer's debts. This was the case in the United States of America, where a crisis in the mortgage industry came up. The mainly reason for this was that mortgage prices had fallen far below total customer debt, and thus repaying loans was not beneficial for borrowers. In addition to that, from the bank's perspective, the value of the properties was not enough to cover the loan amount. Consequently, seizing is not advantageous even for banks. This can be explained not only for the above mentioned reason, but also because extra cost is needed in order to lead to the sale of the property, since banks do not benefit from acquiring real estates. The basic reasons for which one gets a mortgage loan can be the purchase, erection, completion, extension, improvement, repair, maintenance of a home or business (for owning or hire), purchase of land intended for residential or commercial use. Apart from that, a common purpose of mortgages is refinancing of mortgage loans of other banks.

It is observed from our dataset that consumer and business loans have the maximum rate of default during 2016 and in particular the first (2016 Q1) and third quarter (2016 Q3) respectively. Consumer loans reached the highest value of 63.7% of the total rates of default, and corresponding business percentage was 50.5%. As far as mortgages, these seem to have the largest rate (44.7%) in the first quarter of 2019 (2019 Q1). In addition, we can conclude that consumer loans have the highest average value compared to the others. Both business and mortgages have approximately the same mean value, which is significantly lower than the one of consumers.

2.3 Brief Model methodology

The main question this dissertation is going to answer is how the aforementioned macroeconomic factors influence the rates of default for the three loan categories. As far as the model is concerned, the estimation will be made through the Eviews software package and will be based on Vector Autoregression (VAR) model. VAR model is an extension of univariate autoregression model to multivariate time series data and represents the relationships among a set of variables. It is often used when two or more time series influence each other in order to analyze certain aspects of the relationships between the variables of interest. That means, the basic requirements in order to use VAR are at least two time series (variables) and time series should influence each other. Before applying the model, tests for stationarity of the data as well as serial correlation of standard errors are performed. Continuing, impulse response functions are used in order to interpret the results. Next, a short analysis of cointegration testing is performed.

Chapter 3

Technical framework

In this chapter we describe the methodology we have followed and present each test separately. Initially, the significance of stationarity condition in the dataset is described as well as the corresponding methodology to ensure there are no unit roots. Then, the general formulation of Vector Autoregressive Model and the reasons we opted for this model are presented. In continue, tests for serial correlation along with lag length criteria are described. An overview of impulse responses and variance decomposition follows. Furthermore, Granger-causality test is theoretically illustrated and finally the procedure of cointegration is presented.

3.1 Stationarity

3.1.1 Introduction to Stationarity

It is essential for our econometric analysis that the time series under consideration is stationary. The absence of stationarity can cause different problems in our investigation. In particular, non-stationary data can lead us to misleading conclusions, since there seems to be a really strong relationship between two variables, whereas there is just a 'spurious relationship'. In fact there is evidence of contemporaneous correlations rather than meaningful causal relations. It is really often that non-stationary time series may produce a spurious regression with high invalid r-squared

(R^2), which is the estimator of our model. In addition, another issue can be the errant behavior of the variables used since the analysis is based on invalid assumptions, such as t-ratios will not follow a t-distribution.

A stationary time series can be defined as one whose statistical properties such as mean and variance are constant over time. More precisely, if y_t is a stationary time series, then for all s (either a positive or negative number), the distribution of $(y_t,\ldots,y_{t+s})$ does not depend on t. The strict definition of stationarity refers to all the properties of one stochastic process, so when only the above conditions are satisfied, the stochastic process is characterized as weakly stationary. For our further analysis it will be sufficient for a time series to be weakly stationary. It does not mean that the series does not change over time, it just mean that the way it changes does not itself change over time. That is, the following conditions are true:

- $E(y_t) = \mu$, independent of t
- $V(y_t) = \sigma^2$, independent of t
- $Cov(y_t, y_{t+s}) = Cov(y_{t+m}, y_{t+m+s}) = \gamma_s$, independent of t

All variables to be included in the VAR are required to be stationary in order to carry out joint significance tests on the lags of the variables and reach accurate conclusions. To check stationarity condition for our dataset we use unit root test for each variable. The main intuition behind this test is that if the process has no unit root, meaning it is stationary, it then exhibits reversion to the mean. As a result, the lagged values will provide relevant information in forecasting the change of the series. To understand the interpretation of unit root test, let the VAR(p) system in a matrix form, be defined as

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + E_t ,$$

where q is the number of equations and thus the number of our variables, Y_t, Y_{t-i} $(i=1,\ldots,p)$ are $q\times1$ vectors of the variables of interest and their lagged values respectively, Φ_0 is the vector of the constant terms, Φ_i indicate the matrices of the unknown coefficients and E_t indicates the vector of the white noise process. We can now examine the tests for stationarity condition from a more mathematical perpective. More specifically, a VAR(p) model is characterized as stationary as long as all solutions (roots) of the following characteristic equation

$|Iq - \Phi1\, z - \Phi2\, z^2 - \ldots - \Phi p\, z^p| = 0$ lie outside the unit circle, that is $|z| > 1$.

Therefore, if |z| > 1, Y_t is said to be stationary or in other words I(0). This mathematic explanation can justify the fact that these tests are called unit root. "Order of integration" is used to describe a unit root process in time series analysis. Particularly, it demonstrates the minimum number of differences needed to obtain a stationary series. There are several ways to check the stationarity condition of a time series. One possible way is by studying the graphic illustration of the series. Another way can be the implementation of the autocorrelation function and its corresponding correlogram. One can also perform statistical tests for the coefficient of correlation (Q statistic) or apply unit root tests.

3.1.2 Tests for stationarity condition

The most commonly used unit root tests are Augmented Dickey-Fuller (ADF), Pillips-Perron (PP) and KPSS. The KPSS test for stationarity examines the following cases:

Ho: no root unit (time series is stationary)

Ha: unit root (time series is non-stationary)

In our analysis Augmented Dickey-Fuller test, which has the advantage of handling bigger and more complex models, is implemented. A reasonable question would be what the theory of ADF is based on. To reach the answer let $Y_{i,t}$, $X_{j,t}$ present rates of default for different loan categories and the macroeconomic variables respectively. Dickey and Fuller test the null hypothesis, which is ρ=1, against the alternative ρ≠1 (unit root test) considering the following equation:

$$Z_t = \gamma + \delta t + \rho Z_{t-1} + e_t ,$$

where $Z_t = (Y_t, X_t)$, meaning the dependent and independent time series. In other words, the ADF test examines the following cases:

Ho: unit root (time series is non-stationary)

Ha: no root unit (time series is stationary)

The ADF test differs from the KPSS test since zero hypothesis and alternative hypothesis have opposite roles in the two statistics. Consequently, in order for all time series to be I(0), i.e. stationary, the null hypothesis must be rejected. The ADF test returns a negative value. The more negative this value is, the higher the probability that the null hypothesis can be rejected. If not, differences are applied to each variable until the series become stationary. That means that the given dataset can be adjusted to first or even second differences. Using leveled variables (which are stationary) in VAR models can result in spurious regression. But, differenced variables will remedy the problem.

3.2 VAR model

3.2.1 An overview of VAR model

As far as the model is concerned, the estimation will be based on Vector Autoregression (VAR). VAR modelling is a major area of interest in multivariate time series analysis. VAR models were initially used in finance in order to forecast, draw statistical conclusions, describe the procedures that data follow and help the economic policy. Before the use of VAR different models and simultaneous equation models failed to predict and interpret the economic variables correctly. Increasing the number of variables and equations as well, did not lead to better conclusions since they could not take into account the interactions among variables of the system. Nevertheless, Sims (1980) suggested a new model - the well known VAR – as a different version to standard econometric models that were used until then and which were characterized by their dubious exclusion restrictions. It is worth mentioning that one of his first publications continues to be very significant today.

In VAR models all variables are considered as endogenous and they are interpreted by their own lagged values as well as the lagged values of the other variables. This allows affecting each other and studying their relationships. To be more precise, VAR models represent the correlations among a set of variables and thus are often used to analyze certain aspects of the relationships between the variables of interest. These models have plenty of benefits. For instance, they are useful for modelling multivariate time series. In fact, they apply liner interactions between multiple time series. VAR is a natural generalization of autoregressive models (AR) allowing for more than one dependent variable. It is in a sense a systems regression model. Another advantage of these models is that we have the opportunity to use Ordinary Least Squares (OLS) separately on each equation so as to estimate the parameters, provided that there are no contemporaneous terms in the right part of the equations in a VAR model. The estimation gives consistent and effective estimators of the system coefficients. In addition, they allow the values of the dependent variable to depend not only on its own lags but also on the other variables, hence it is more general than VARMA models. Other than that, the predictions with the VAR models are better than those obtained from the equation systems. The basic condition to apply a VAR model is a list of variables that can be assumed to influence each other over time. The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting.

3.2.2 Model description – VAR

As mentioned above, there is no need to specify which variables are endogenous or exogenous since all variables in a VAR model are treated as endogenous (dependent). One more advantage of VAR modeling is that each variable is affected not only by its own lags but also by the current and past values of other variables. The number of lags used in the model is defined based on the data and data frequency. The system order indicates the lag number of each series. The simplest case is a bivariate VAR. An example of a VAR(q) model (meaning the maximum lag of the variables is ‘q’) with two variables is following.

$$y_{1t} = c_1 + \alpha_{11} y_{1t-1} + \dots + \alpha_{1q} y_{1t-q} + \beta_{11} y_{2t-1} + \dots + \beta_{1q} y_{2t-q} + u_{1t} \quad (3.1)$$

$$y_{2t} = c_2 + \alpha_{21} y_{1t-1} + \dots + \alpha_{2q} y_{1t-q} + \beta_{21} y_{2t-1} + \dots + \beta_{2q} y_{2t-q} + u_{2t} \quad (3.2)$$

Where y_{it} are the endogenous variables (i=1,2), c_i represent the constant and u_{it} is a white noise process. Stationarity hypothesis for y_t variables is a prerequisite to apply VAR. That means that each y_t has a constant mean, constant variance and the covariance matrices between y_t and y_{t+s} depend only on s and not on t. Continuing, a white noise process is one with no discernible structure. In particular, we mean an independent and identically distributed (iid) disturbance term with zero mean (so $E(u_t)=0 \ \forall t$), and no correlation between its values at different times, meaning $E(u_t u_s)=0$ for $t\neq s$ and $E(u_t u_s) = \Omega$ for t=s. Ω is the variance - covariance matrix for the bivariate VAR and it is defined as next:

$$\Omega = E\,(u_t\ u_{t=s}) = \begin{pmatrix} var(u_{1t}) & cov(u_{1t}u_{2t}) \\ cov(u_{1t}u_{2t}) & var(u_{2t}) \end{pmatrix} \quad (3.3)$$

It should be pointed out that the error term of each equation should behave like a random noise process, with zero mean and constant variances in the main diagonal of the Ω matrix. In addition, it should be highlighted that an equation's error term might be correlated with another equation's error term during the study period. In other words, the covariances in the non-diagonal elements of the matrix might not be zero. Apart from that, in order to fit a model like the above one we should assume (or better check) that the error terms u_t are not correlated with any other variable in the right hand side of the equation, videlicet they should not be correlated with the lagged values of y_t. An also fundamental hypothesis when applying VAR is that all variables included in the model should be stationary. All the aforementioned conditions are extensively explained in the next sections, and they are tested in practice in the next chapter.

It is easily observed from the system of (3.1) and (3.2) equations that there is one equation for each variable as dependent variable. Each equation explains the evolution of the left-hand side variable including lagged values of all dependent variables in the right-hand side of the system (but no contemporaneous variables exist in the right-hand side) and an error term.

The system of (3.1) and (3.2) equations can be presented in a more compact form as next,

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11} & \beta_{11} \\ a_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \dots + \begin{pmatrix} a_{1q} & \beta_{1q} \\ a_{2q} & \beta_{2q} \end{pmatrix} \begin{pmatrix} y_{1t-q} \\ y_{2t-q} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (3.4)$$

Thus, a VAR model of order q, denoted as VAR(q), is a system in which each outcome variable is regressed on a constant and p of its own lagged values as well as p lagged values of each of the other existing variables. The analysis can obviously be extended to a VAR model so that there are p variables and p equations (where p>2), but the implementation of that model is left up to the reader.

One important feature of VAR models is the compactness with which they can be written. For instance, in case of q=2, a VAR(2) model is defined as

$$y_{1t} = c_1 + \alpha_{11}\, y_{1t-1} + \alpha_{12}\, y_{2t-1} + \beta_{11}\, y_{1t-2} + \beta_{12}\, y_{2t-2} + u_{1t} \quad (3.4)$$

$$y_{2t} = c_2 + \alpha_{21}\, y_{1t-1} + \alpha_{22}\, y_{2t-1} + \beta_{21}\, y_{1t-2} + \beta_{22}\, y_{2t-2} + u_{2t} \quad (3.5)$$

Or $\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (3.6)$

Or even more compactly in tables form

$$Y_t = C_0 + A\, Y_{t-1} + B\, Y_{t-2} + U_t \quad (3.7)$$

where $Y_t = (Y_{1t}\ Y_{2t})'$ is a 2×1 vector and contains two variables, C_0 is the constant 2×1 dimension, A, B are 2×2 matrices, Y_{t-1}, Y_{t-2} (2×1 dimension) present the lagged values of Y_t and U_t (2×1) is assumed to be a white noise vector.

In addition to the stationarity condition, a VAR should also meet other conditions so as to provide reliable results. Hence, these prerequisites should be taken into account for our investigation. A theoretical analysis follows in the next sections.

3.3 Serial Correlation

A basic assumption of the linear regression model, and thus for VAR models, is that the error terms, i.e. u_{1t}, u_{2t},…, u_{nt} are not correlated. Many methods have been established to check properly the presence of correlations in the error terms in time series data. But what do we mean with the term of serial correlation (also known as autocorrelation)? In short, when error terms from different periods are correlated, we say that the error terms are serially correlated. Mathematically autocorrelation can be stated as $Cov(u_t,u_s) = 0$, for $t \neq s$. An unsophisticated way to think of serial correlation follows, considering an AR model for simplicity. Let us for example, assume that y_t is correlated with x_{1t} and x_{2t}, but x_{2t} is mistakenly not taken into account in the model. It is apparent that the effect of x_{2t} will be included in the disturbing term u_t. If the variable x_{2t} shows a trend over time, as most financial series do so, then x_{2t} will depend on its previous values, meaning x_{2t-1}, x_{2t-2} etc. Similarly the disturbing term u_t will depend on u_{t-1}, u_{t-2} and so on.

Serial correlation 'arises' in time-series studies when the errors associated with a given time period, continue over future time periods. This mainly happens in time series data because the observations are obtained at discrete points in time and thus they may have positive correlated errors. However, this can cause several problems. First of all, although the estimators are still unbiased and consistent with the presence of autocorrelation, they will not be efficient anymore. There can be either positive or negative autocorrelation. In case of positive serial correlation, the issue of smaller estimates of the standard errors comes up. The underestimation of true standard errors can lead us thinking that predictors are statistically significant when they are actually not. For instance, a 95% confidence interval may have a much lower probability than 0.95 of containing the true value of the parameter. As a result, the parameter estimates seem to be more precise than they really are and there will be a tendency to reject the null hypothesis when it should not be rejected, since probability values will be lower than they should be and t-statistics tend to be higher. Further, in most of the cases the r squared would be overestimated, which suggests exaggerated goodness of fit.

For all these reasons the assumption of uncorrelated errors is extremely important. A simple way to recognize the existence of serial correlation is through plots of residuals as a function of time. It can be understood that the errors are uncorrelated if they have no discernible pattern in the graph. More professional methods include a variety of tests for detecting the presence of autocorellation. We can briefly mention some of them, such as Durbin-Watson, Durbin h and Breusch-Godfrey LM test, that have been performed for this purpose. Regarding the Durbin-Watson (DW) statistic, it is given by:

$$DW = \frac{\sum_2^T (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_2^T \varepsilon_t^2}$$

where ε_t denotes the residuals of the model.

This test measures the linear association between adjacent residuals from a regression model. However, there are some limitations when using this test. For instance, it is restricted to detecting only first-order autocorrelation. In addition, if there are lagged dependent variables on the right-hand side of the regression, the DW test is no longer valid.

To overcome the above difficulties another test for autocorrelation in the residuals is used. This is Breusch-Godfrey (BG) test and it will be implemented in our analysis. This test is more general than the Durbin–Watson statistic. This is in fact a general test for autocorrelation of any order. More specifically, consider a VAR(q) model. Since the residuals are given by the following equation:

$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \ldots + \rho_q u_{t-q} + e$, in order to check for the existence of serial correlation, a test with the null hypothesis H_0: $\rho_1 = \rho_2 = \rho_3 = \ldots = \rho_q = 0$ can be implemented. In the methodology of Breusch-Godfrey test an auxiliary regression of residuals is used, nevertheless it is beyond the scope of this analysis, so it will not be explained more thoroughly.

3.4 Lag length criteria

Once the approach of VAR modelling is adopted for our data set, choosing the appropriate VAR model for best modelling would become the next important step of this research. To begin with, as it is already said *Y* responds to *X* with a lapse of time. This lapse of time is called lag. If the chosen – or the assumed – lag order is unnecessarily large, the Augmented Dickey-Fuller test, which was previously mentioned, faces many difficulties and erroneously results will be concluded. The presence of more lags than those that are necessary reduces the power of the test to reject the null hypothesis of unit root. It has also been found that when the number of lags is large enough the forecast precision of the corresponding VAR model will be reduced.

Apart from that, extended lag structures require a sufficient data set (regarding its size) which is not possible most of the time. Consequently, reducing the sample size often leads to the reduction in the number of degrees of freedom in the estimation procedures. So, there are several possibilities in order to choose the appropriate number of lags. Statistical tests can be applied considering the hypothesis that a certain lag equals zero. Another way is to use the well known information criteria (IC) for choosing the proper lag order. The basic principle of information criteria is the minimization of the Mean Squared Error (MSE). There is a variety of existing information criteria in bibliography. In essence, an information criterion is a measure of the quality of a statistical model taking into account how well the model fits the data and the complexity of the model. Information criteria help us compare alternative models fitted to the same data set. The most commonly used criteria are Akaike (1974), Schwarz (1978) and Hannan-Quinn. Each of these criteria is characterized by its own specific properties, but we will not give more emphasis for now. The performance of these information criteria is examined in order to choose the optimal lag length in our vector autoregressive (VAR) model. All of these measures are simple to use and compute.

Akaike, Schwarz and Hannan-Quinn information are given by:

$$AIC = \ln |\Omega| + 2k / T$$

$$\mathrm{BIC} = \ln |\Omega| + \frac{k}{T} \ln(T)$$

$$\mathrm{HQIC} = \ln |\Omega| + \frac{2k}{T} \ln(\ln(T))$$

where Ω is the variance – covariance matrix of the residuals and k is the total number of regressors in all equations (this will be equal to $g^2k + g$ for g equations, each with k lags of the g variables, plus a constant term in each equation).

It can be easily understood that since $\ln n > 2$ for $n > 7$, the BIC statistic generally places a heavier penalty on models with many variables compared to AIC. But how can we compare the results of these models? The model with a lower value is superior to a model with a higher one and can be considered as the best estimation of the unknown true model, considering that all else being equal.

3.5 Impulse Responses and Variance Decompositions

From an economical point of view it is difficult to interpret coefficients of the model. We can of course study the results of F-tests and suggest which of the variables in the model have statistically significant impacts on the future values of each of the variables in the system. However, F-test cannot explain if the changes in the value of a given variable have a positive or negative impact on other variables in the system, (in other words the sign of the relationship) or how long these effects require to take place. These are the reasons we use the so-called impulse response functions (IRFs) as well as variance decompositions (VD). Both of them are analyzed in the next sections.

3.5.1 Impulse Responses

Impulse response analysis is an important step in econometric analysis which employs vector autoregressive models. Their main objective is to explain the evolution of a model's variables in reaction to a shock in one or more parameters. More specifically, Impulse Response Functions (IRFs) help us identify how present and future values of each variable (response) react to an increase to the standard deviation of a shock to the other variable (impulse). The variable that experiences the shock is called 'Impulse' and the time series which (possibly) responds to the shock is the 'response' variable. This can help us draw significant conclusions about the transmission of a single shock within an otherwise noisy system of equations. As a result, this feature makes IRFs very useful tools in the assessment of economic policies.

To explain the theory behind the IRFs, when there is a shock to a variable, it does not only immediately affect this variable but is also is transmitted to all of the other endogenous variables through the dynamic (lag) structure of the VAR. To be more specific, given that there is no serial correlation between the error terms, the interpretation of IRFs is quite naive. However, the error terms are actually correlated. This may happen because there is a common factor that affects all variables and it is not included in the model. An impulse response function traces the effect of a one-time shock to one of the innovations on present and future values of the endogenous variables. Usually, this shock is expressed in terms of standard deviations of the error terms. Thus, for each variable from each equation separately, a unit shock is applied to the error, and the effects upon the VAR system over time are noted. If there are for example g variables in a system, the number of impulse responses that could be generated is g^2.

In practice, it is based on the fact that a VAR model can be expressed as a vector moving average (VMA). Let us use a bivariate VAR(1) to explain it more thoroughly. The VAR(1) is given by: $y_t = A_1 y_{t-1} + u_t$,

where $y_t = (y_{1t}, y_{2t})'$, A_1 denotes a 2×2 matrix and $u_t = (u_{1t}, u_{2t})'$. Consider now the effect at time t = 0, 1, … , of a unit shock to y_{1t} at time t = 0.

$$y_0 = \begin{pmatrix} u_{10} \\ u_{20} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$y_1 = A_1 y_0 , \ y_2 = A_1 y_1 \text{ and so on.}$$

When A_1 is given, it is quite simple to see what the effects of shocks to the variables will be in such a VAR. However, when the VAR contains more equations it is much more difficult to observe what the interactions between the equations are.

Using the lag operator we can show the MA(∞) representation for the VAR(1) is

$$y_t = u_t + \phi\, u_{t-1} + \phi^2\, u_{t-2} + \ldots + \phi^{j}\, u_{t-j} + \ldots$$

Thus, the coefficient in the MA representation measures the impulse response is

$\phi^j = dy_t/du_t$ (where ϕ^j is a 2×2 matrix for a bivariate system).

Let now Ω be variance covariance matrix for the error vector. In general, error terms are contemporaneously correlated (not-orthogonal), which means that the non-diagonal elements of matrix Ω are non-zero. Therefore we cannot hold the one error term (i.e. u_1) constant and let only the other one (u_2) vary. By applying various mathematical techniques the error terms u_t of the VAR system can be transformed into a vector which elements are not correlated. The most common of these methods is the Cholesky decomposition which finds a lower triangular matrix A so that $\Omega = AA'$ (Cholesky Decomposition). Then we can define the new error vector $\widetilde{u_t}$ as $\widetilde{u_t}$ = A^{-1} u_t (meaning a linear transformation of old error vector u_t). It is clear that the new error is orthogonal since its variance-covariance matrix is given by:

$$\mathrm{var}(\widetilde{u_t}) = A^{-1}\, \mathrm{var}(\widetilde{u_t})\, A^{-1\,\prime} = A^{-1}\, \Omega\, A^{-1\,\prime} = A^{-1}AA'\, A^{-1\,\prime} = I$$

In other words, the variance-covariance matrix is diagonal. In fact, Cholesky decomposition imposes an ordering of the variables in the VAR. Is should be noted that responses can change if we change the ordering of the variables.

Of course, impulse response functions would be of little interest if they had not been useful for explaining changes on the structure of the economy. They have actually played a crucial role as inputs into a variety of important substantive debates. It is worth mentioning that Impulse Response analysis changed the way Economists think. In fact, Economists started thinking differently about the aggregate economy because of what they have learned using VARs.

3.5.2 Variance Decomposition

Another way to interpret a VAR model is using Variance Decomposition (VD). As it was mentioned in the previous section, impulse response functions essentially trace the effects of a shock to one endogenous variable on to the other variables included in the model. From another point of view, variance decomposition splits the variation in an endogenous variable into the component shocks to the VAR. Thus, the variance decomposition gives us information about the relative importance of each random innovation in affecting the variables in the VAR model. In particular, the VD method indicates (in percentage) how much of a variable's change is due to the shock itself and how much is due to the disturbances in the other variables. In other words, it determines how much of the variation in the error of each variable can be explained by exogenous disturbances in the other variables. Most commonly, most of the variation is due to shocks of the variable itself. However, as the effect of lagged variables appears the rate of the effect of the other shocks increases as time passes.

3.6 Granger- Causality

3.6.1 The purpose of Granger-Causality test

Although there is no direct connection between Impulse Response Functions and Granger causality tests, one can describe both of them as techniques for explaining the behavior of different variables. Similarly with IRFs, Granger causality test is concerned with short-run relationships between variables, however, correlation does not necessarily imply causation in any meaningful sense of that word. The basis behind VAR is that each of the time series in the system influences each other. That means that we can predict the series with past values of itself along with other series in the system after building the appropriate model. The advantage of Granger's causality test is that it is possible to test this relationship before even create the model. Granger causality (or "G-causality") was proposed in 1969 and has been widely used in economics since the 1960s.

The basic core of Granger causality test is that it attempts to determine the direction of a relationship between two variables. This can shed light to the recognition of the short term behavior of these variables as well as the estimation of the short term predictions based on the systematic part of a regression equation, in contrast to IRFs, which define what will happen to a variable if the random part of the equation (error term) changes. One thing we should emphasize is that the statement 'x Granger causes y' does not imply that the variable y is the effect or the result of the variable x. The basic "Granger Causality" definition is quite simple. To be more specific, Granger's method, examines not only how a variable can be explained by its own past values, but also how the lags of another variable can improve the interpretation of the first one.

A variable X_t is said to 'Granger-cause' (or 'G-cause') another Y_t variable if the prediction of Y_t's values (meaning Y_{t+1}) improves when we use past values of X_t, given that all other relevant information is taken into account. The definition leans heavily on the idea that the cause takes place before the effect, which is after all the basis of most causality definitions. It should be pointed out that short run relationships could be unidirectional, bidirectional or neither, where the last one means the variables are independent of each other. As it is easily understood, the study of the relationship of more than a pair of variables through this method is complicated enough.

G-causality is normally tested in the context of linear regression models. More complex extensions to nonlinear cases exist, however these extensions often face more difficulties when applying in practice. Let us consider a linear autoregressive model with two variables X_1 and X_2 for simplicity. The mathematical formulation of the model is illustrated below:

$$X_{1t} = \sum_{j=1}^{p} A_{11,j}\ X_{1(t-j)} + \sum_{j=1}^{p} A_{12,j}\ X_{2(t-j)} + E_{1t}$$

$$X_{2t} = \sum_{j=1}^{p} A_{21,j}\ X_{1(t-j)} + \sum_{j=1}^{p} A_{22,j}\ X_{2(t-j)} + E_{2t}$$

where p denotes the maximum number of lagged observations included in the model, the matrix A contains the coefficients of the model, E_1 and E_2 vectors are the residuals (prediction errors) for each time series.

If the variance, for example, of E_1 is reduced by the inclusion of the X_2 terms in the first equation, then it is said that X_2 'Granger-causes' X_1. In other words, the term of X_2 is significant, which means that the presence of its lagged values improves the model (in the first equation). As far as the matrix of coefficients is concerned, X_2 G-causes X_1 if the coefficients in A_{12} are jointly significantly different from zero. Thus, we should check the statistical significance of these coefficients. Given assumptions of covariance stationarity on X_1 and X_2 and no residuals correlation, this can be tested by performing an F-test, since the notation is in matrix form and thus a joint probability test is needed. The hypotheses of the test are the following:

$$H_0 : A_{12} = 0 \ \ (X_2 \text{ does not Granger-causes } X_1),$$

$$H_\alpha : A_{12} \neq 0 \ (X_2 \text{ Granger-causes } X_1).$$

If F > $F_{Critical}$ in a certain chosen level of significance, then the null hypothesis H_0 is rejected. Thus, the X_2 lags have an important effect on the X_1 forecasting ability.

In order to determine if this effect is two sided we can apply the same procedure, meaning an F-test with the following hypotheses:

$$H_0 : A_{21} = 0 \ (X_1 \text{ does not Granger-causes } X_2),$$
$$H_\alpha : A_{21} \neq 0 \ (X_1 \text{ Granger-causes } X_2).$$

Note also that BIC or AIC can be used to determine the appropriate model order p. The above test can be performed for testing the statistical significance of A_{21} in the second equation and thus check if X_1 G-causes X_2.

3.6.2 Granger-Causality for multiple variables

Repeated pairwise analyses among multiple variables can sometimes lead us to misleading conclusions. For this reason it is crucial to be able to extend the analysis for more than two variables. As previously mentioned, G-causality can be extended to the n variable case, where n>2, by estimating a VAR model with n variables. However, in this case the procedure requires more attention, since it examines Granger's causality of two variables concerning the existence of additional variables which also influence the result. To be more specific, we can say that X_2 Granger-causes X_1 if lagged observations of X_2 help predict X_1 (the same as in the bivariate case), when lagged observations of all other variables $X_3 \ldots X_N$ are also taken into account. This generalization of Granger-causality in multivariate extension is also referred as "Conditional G-causality". In order to better understand it, assume a VAR model with three time series as represented in matrix form here.

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} a_{11,i} & a_{12,i} & a_{13,i} \\ a_{21,i} & a_{22,i} & a_{23,i} \\ a_{31,i} & a_{32,i} & a_{33,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \\ y_{3,t-i} \end{bmatrix} + u_t \qquad (3.6.1)$$

Let us now check if y_2 G-causes y_1 variable. The hypothesis that should be tested is:

$$a_{12,i} = 0 \text{ , i=1,2,\ldots } (3.6.2)$$

It is important to underline that in a system like this the constraints (3.6.2) are equivalent to one step prediction, and for our example this is $y_{1,t+1}$. Essentially, the information included in lagged values of y_{2t} may be beneficial for predicting y_1 in later time horizon, even if (*) is true (Lütkepohl 1993). For instance, observations of y_2 up to time t, may be useful to predict the value of $y_{1,t+2}$, despite the fact that they may not be useful for the prediction of $y_{1,t+1}$. This is possible to happen when for example observations of y_{1t} could improve the prediction of y_3 (i.e. $y_{3,t+1}$), which in turn could help in the prediction of y_2 in a following time period. In general, Conditional G-causality can measure the effect of one time series on another time series in the presence of a third.

3.7 Cointegration

3.7.1 Introduction to Cointegration

A common issue in time series analysis as that they are characterized as non-stationary, because of various shocks and cyclic fluctuations. Non stationarity can be found using a variety of tests , but the most common one is probably the Augmented Dickey-Fuller test which was discussed earlier. It has already been mentioned that non stationary time series suffer from several problems. Simple O.L.S. (Ordinary Least Squares) regressions do not capture shocks and cyclic events. Consequently, the results from any hypothesis test will be biased (spurious regression) or misleading. In order to tackle these issues, different methods have been proposed for the analysis of these variables. Cointegration, which is one of these methods, has become an important property in contemporary time series analysis. In particular, cointegration tests are used to analyze non stationary time series which have variances and means that vary over time. The benefit of this the method is that it allows you to estimate the long-run relationships in systems with unit root variables (Rao, 2007).

However, we should not confuse cointegration with correlation. Correlation is simply a measure of the degree of mutual association between two or more variables. When two variables move in the same direction, they are positively correlated and if they move in opposing directions, the correlation is said to be negative. On the other hand, cointegration helps identify the degree to which two variables are sensitive to the same average price over a specific time period. In simple terms, cointegration does not reflect whether the pairs would move in the same or opposite direction, but it can help us whether the distance between them remains the same over time. For this reason, it is certainly possible for two time series to be correlated but not cointegrated, cointegrated but not correlated, both or none. Cointegration might provide a more robust measure of the linkage between two financial quantities than correlation which is very unstable in practice. Cointegration is based on the possibility of non-stationary time series to have a common stochastic trend. This fact may not allow them to drift very far from one another in their long run relationship.

But what do we mean by referring to cointegration in practice? Two sets of variables (or even more) are said to be cointegrated if a linear combination of those variables has a lower order of integration. For example, we can think of individual time series that are first-order integrated, denoted as I(1), and they can be modeled via a (cointegrating) vector of coefficients so as to form a stationary (I(0)) linear combination of them. The order of integration, for instance I(1) indicates that a single set of differences can transform the non-stationary time series to stationarity. To be more specific, let u_t be the random walk:

$$u_t = u_{t-1} + \varepsilon_t$$

where $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma^2$, i.e. ε_t is stationary.

Now let

$$X_t = \alpha\, u_t + \nu_t$$

and

$$Y_t = \beta\, u_t + \eta_t$$

where ν_t and η_t are both stationary processes similar to ε_t.

Then both X_t and Y_t are non stationary because they are linear functions of the non-stationary (stochastic trend) variable u_t.

However

$$\beta\, X_t - \alpha\, Y_t = \beta\, \nu_t - \alpha_t\, \eta_t$$

is a linear combination of the stationary disturbances and is therefore stationary. When this happens X_t and Y_t are said to be cointegrated.

Let us now move on to the case of *n*-dimensional. An *n*-dimensional time series $y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is cointegrated if there is some linear combination of the component variables, such as $\beta_1 y_{1t} + \ldots + \beta_n y_{nt}$, that is stationary. The combination is called a cointegrating relation, and the coefficients $\beta = (\beta_1, \ldots, \beta_n)'$ compose a cointegrating vector. It should be pointed out that the number of cointegrating vectors shall be one less than the number of variables. For example, if we have two variables, there should be up to one cointegrating vector.

It is obvious that the idea of cointegration can be also applied in a more general case. Particularly, systems of higher-order variables can be cointegrated if a linear combination reduces their common order of integration.

3.7.2 Methods for testing for cointegration

There are some methods for testing for cointegration used to identify the long-term relationships between two or more sets of variables. In the literature there are two prominent cointegration tests that can be performed. These are the Engle-Granger cointegration test and Johansen test. The procedure that Engle-Granger test follows is firstly creating residuals based on the static regression and then testing if they have unit roots (Augmented Dickey-Fuller test or other tests can be used for shocking the stationarity condition). If the series is cointegrated then the Engle-Granger test will show the stationarity of the residuals. If you have two single time series variables, then Engle-Granger is just the appropriate test for analysis. Nevertheless, this test has an important drawback since it is a single equation model. In other words, in the presence of more than two variables, the test may show more than two cointegrating relationships.

When more than two variables are analyzed Johansen cointegration test is recommended and thus this is the reason that it is most commonly used. Compared to the previous test, Johansen test is able to test cointegrating relationships between several non-stationary time series data. Our analysis will focus on this test. It should be underlined that this test should be performed on the level form of the variables and not on their differences. There are two types of this test; the trace test and the maximum eigenvalue test.

As far as the trace tests is concerned, it evaluates the number of linear combinations in a time series data, i.e. S, to be equal to the value S_0 and the hypothesis for the the value S to be greater than S_0. The hypotheses of Johansen cointegration test are defined as bellow:

$$H_0 : S = S_0 , H_\alpha : S > S_0.$$

We usually set S_0 equal to zero when applying the trace test. Then, if the null hypothesis is rejected, we can understand that there exists a cointegration relationship in the sample. In other words, the null hypothesis should be rejected in order to confirm the presence of at least one cointegration relationship in the system.

Continuing, maximun eigenvalue test is similar to Johansen test, although the alternative hypothesis is stated differently. It should be reminded that an eigenvalue is defined as a non zero vector, which, when a linear transformation is applied to it, it changes by a scalar factor. Let us now define the two hypothesis:

$$H_0 : S = S_0 \text{ (the same as previously)}, H_\alpha : S = S_0 + 1.$$

In the scenario when S_0 is equal to zero and the null hypothesis is rejected, then there is only one possible outcome of the series to produce a stationary process. On the other hand, when S_0 = m-1 and we reject the null hypothesis, then there are m possible linear combinations.

It is possible the two types of test statistics do not be in agreement and may show different results.

Chapter 4

Empirical analysis

This chapter includes the basic part of this dissertation. The steps followed in order to reach conclusions about the relationship between a group of certain macroeconomic variables and the rates of NPLs were theoretically described in the previous chapter. The above steps as well as data used for this purpose are analyzed in this chapter through Eviews software. The results of each test are further explained below.

4.1 Data

The data used for this analysis have been taken from the official website of Bank of Greece (https://www.bankofgreece.gr/), OECD statistics website (https://stats.oecd.org/), as well as Thomson Reuters Datastream. In particular, as far as the rates of the three categories of NPLs are concerned, they have been exported from the website of Bank of Greece. Inflation and unemployment data have been taken from OECD statistics website and finally, the rest data that are related to Economic Sentiment Index and Exchange rate was collected from Thomson Reuters Datastream database (which is a global financial and macroeconomic data platform). Quarterly data have been used for the present analysis. The data selected refer to Greece for the time period from the last quarter of 2002 (2002 Q4) till the first quarter of 2019 (2019 Q1), since it is the largest time period that the data for all time series needed is available. This is therefore the time period on which the analysis of this dissertation is based. Thus, a total of 66 observations for each time series have been used in the procedure of this investigation.

All time series used in the present analysis is seasonally adjusted. Statisticians and Economists usually use the method of seasonal-adjustment to reveal trends in data. In simple terms, one can think of monthly data. These are influenced by a variety of issues such as the number of days and the number of weekends in a month, or the timing of holidays and seasonal activity. These influences make it more complicated to examine underlying changes in a given time series. In other words, seasonal adjustment is a statistical technique that tries to measure and remove the influences of predictable seasonal patterns. This is reason seasonal adjustment is commonly used.

Moreover, it should be pointed out that all variables are expressed as percentage points so they can be compared to each other. To be more specific, the variable of Inflation (meaning the rate of inflation) has been calculated based on the following formulation of Consumer Price Index (CPI): $R_{inf} = \frac{CPI_T - CPI_{T-1}}{CPI_{T-1}}$

Inflation, therefore, results from the rate of change in the consumer price index over a certain period of time. The variable of Economic Sentiment Index has been also calculated using a similar formula, like the above one. Exchange rate as well as Unemployment rate remains unchanged. As far as the time series of loans, they are also expressed in rates and in specific, each loan is considered by its rates of default.

As it is already mentioned, it is assumed that rates of default are related to macroeconomic and business conditions. So, time series are used in our investigation, as they can capture both current and future directions in the economy and the business environment as well.

To recap, the macroeconomic data used consist of the Unemployment rate, change of Inflation rate and Economic Sentiment Index, and Exchange rate. We used the aforementioned variables in our research in order to show how these factors influence the rate of default. Regarding the data on Non Performing Loans, these consist of three different types which are business, mortgage and consumer loans and are also expressed in percentage points.

4.2 Econometric methodology

In this analysis we will investigate the econometric methodology in order to show the relationship between macroeconomic factors and rates of default for consumer, business and mortgage loans. First of all, tests for stationarity in our variables are performed via unit root tests. Continuing, Vector Autoregressive model is used to determine the effect of Unemployment, Inflation, Economic Sentiment Index and Exchange rate (Euro to US) on each category of non-performing loans (business, consumer and mortgage) and reach several conclusions about the effect of the above macroeconomic factors on loans' default. Our analysis will be carried out through the Eviews software package. Each section that follows analyses the above steps.

4.3 Stationarity

4.3.1 Test for Stationarity

As already mentioned in the previous chapter, stationarity condition is essential for the further analysis of our dataset. A short recap is provided in order to remind the purpose of testing for this. With the term of stationarity we mean that the statistical characteristics of a process creating a time series do not alter over time. All variables to be included in the VAR are required to be stationary. Regarding the investigation of whether a time series is stationary or not, unit root tests are carried out and more specifically Augmented Dickey-Fuller (ADF) test is implied.

4.3.2 Unit root test (Augmented Dickey–Fuller test)

Having all variables plotted, one should have a sense that most of these series is not stationary. By visual inspection we can comprehend that most of our variables are not stationary, since means and variances do not seem to be constant over time. However, this needs to be checked through unit root tests. Hence, as previously stated, in order to find the order of integration, meaning the least number of differences needed to acquire a stationary series, all variables are subjected to Augmented Dickey–Fuller (ADF) tests via Eviews software. According to this test, P-value should be less than 0.05, at 5% level of significance (for this study's purposes), in order to reject the null hypothesis of unit root, or alternatively to make sure that the series is stationary. The null hypothesis of ADf test is that the time series being studied is not stationary. Thus, when p-value is more than 5%, we accept the null and therefore there is no stationarity in our variables. It is necessary to determine whether we should be working with levels of the variables or first (or even second) differences of the data. Thus, if the data is not stationary, we have to modify it so that it becomes stationary. A brief explanation of the presented results in the section of the Appendix is following.

Considering the output tables shown from our analysis, the only variable that is stationary in levels is the rate of default of business loans (Figure 4.3.1). So, this will remain in levels for the extension of the present research since it is already stationary and does not need any transformation. On the other hand, the macroeconomic variable of Inflation (Figures 4.3.11-4.3.12) should be transformed to first differences given the fact that it seems to be I(1). This is also true for Economic Sentiment index and exchange rate (Figures 4.3.14, 4.3.16). It can be easily noticed that applying the initial ADF test to these three variables results in probability values greater than 0.05 (Figures 4.3.13, 4.3.15), so one fails to reject the null hypothesis, whereas in the next ADF test this does not happen (and we lead to rejection of the null). As far as the rest of our variables, which include time series of unemployment, rate of default of consumer and rate of default of mortgage loans, they are I(2) which means that they require differencing twice to induce stationarity (Figures 4.3.10, 4.3.4, 4.3.7). In other words, we will continue our analysis taking second differences for these variables, since we failed to accept the alternative hypothesis twice. It is worth mentioning that studying the results of all these tests, it is easily noticed ADF statistic is a negative number. What is more interesting is that the more negative it is (which means smaller real value of number), the more robust the rejection of the hypothesis is that there exists a unit root at five percent level of confidence.

Plots of the time series that will be included in the Vector Autoregressive model are provided in the Appendix in order to identify stationary versus non stationary data. Graphs of the rates of default for each loan category before and after the transformations are presented in the Appendix (Figures 4.1, 4.2, 4.3). Moreover, Figure 4.4 presents the macroeconomic variables after the appropriate transformations so as to become stationary. The horizontal axis of all plots indicates the time period from the last quarter of 2002 (2002Q4) till the first one of 2019 (2019Q1).

4.4 Overview of VAR Model

Initially, a Vector Autoregressive model (VAR) with all the endogenous variables is performed. But why it is said to be autoregressive? The answer is that it is considered as an Autoregressive model due to the fact that each variable is examined as a function its own past values and the lagged values of the other endogenous variables. Otherwise stated, the predictors are nothing but the lags of the time series. To begin with, we apply a VAR model. The equations of our vector model that we are interested in are these that contain the rates of default for loans on the left hand side of the equation and the past values of them as well as all the other variables (including their lagged values) on the right hand side. In fact these equations are three since we are interested in the response of three loan categories, which are business, consumer and mortgage loans. However, before proceeding to the implementation of the VAR model, it is necessary to check some additional conditions. The next step of our analysis presents how to check the right number of lags that should be applied in the model. We should therefore estimate the proper number 'p', which refers to the length of time lag in a VAR(p) model.

4.5 Serial Correlation Test

A common concern in time series studies is the presence of serial correlation – also known as autocorrelation. In this part of our analysis, a test for serial correlation is conducted. We can think of autocorrelation as a systematic relationship between the residuals measured at different points in time. The disturbance term u_t includes the effect of all the variables that cannot be included in the model. However, this effect can often refer to future time periods and not to the present. This is the well known issue of serial correlation. As previously stated, the presence of serial correlation can induce a variety of problems. Autocorrelation in residuals is a common issue in most econometric models. This is the reason this specific test is applied before we proceed to further investigation.

It has already been mentioned that the Breusch-Godfrey (BG) test has some benefits compared to some other methods. Thereby, it will be implemented to check for the presence of serial correlation in our model. This test was applied using again Eviews software. The table output is illustrated in Figure 4.5.1 in the Appendix. Based on this figure, the first test (first row) will check if the first order autocorrelation is significant for our model. It should be reminded that the null hypothesis of Breusch-Godfrey test states that there is no autocorrelation, so we perform tests until we reject the null. More specifically, we can see that the probability value of the first test is certainly small (0.0045), so we reject the null hypothesis for pretty much all the levels of significance. We have determined the level of significance to be 5% for our investigation. It can be easily concluded from the tests that a VAR model with one or two lags is not appropriate for this investigation regarding Breusch-Godfrey test. The output table indicates that the residuals are correlated with first two lagged terms. This violates the assumption of serial independence of residuals. As shown from the results, p-value is larger than 0.05 for the case of three lags (since it is 0.6826). So we have to re-specify our model and increase the number of lags in order to tackle this problem. The overall test for the number of the optimal lag length to be included in our model is estimated in the following section.

4.6 Lag Length Estimation

Before we proceed to the interpretation of our Vector Autoregressive model we need to determine the appropriate number of lags. In this section, we will establish whether the fact that we used two lags in the VAR is sufficient to capture all the dynamics in the changes of our variables. In general, lag structures identify the time delay of the response to independent variables. Assuming either unnecessarily large or small enough lag length can lead us to fallacious results. It is therefore, useful to have certain procedures or criteria for choosing the adequate lag order.

In order to choose the optimal number of lags for VAR model, different information criteria are employed, such as Akaike, Schwarz and Hannan-Quinn. It is important to highlight that the inclusion of lagged values helps to avoid inference and estimations problems. As we will observe in the next part of this analysis, when VAR estimates are printed there are some more regression outputs. The regression results at the bottom part of the regression table are related to the estimates of the VAR system itself. The values of Akaike and Schwarz information criteria are also printed in this table and refer to the VAR model. Considering the size of our sample - that is rather small - the most appropriate criterion for our dataset is Akaike (AIC). Focusing on a statistical approach, these criteria help us decide how many lags should be included in our autoregressive model. Given the fact that the test for serial correlation has already been performed, a constraint for using no more than three lags is applied to the software, regarding the number of lags to include in the VAR model. AIC indicates that the proper lag order is three (Figure 4.6.1). However, if we had not taken into account the results obtained from serial correlation test, a certainly larger number of optimal lags would have been emerged. As it is obvious, including many lags leads to loss of our observations and many other difficulties, analyzed in the former chapter. The presence of the asterisk sign (*) in Figure 4.6.1 indicates the optimal lag length for each criterion. As we can observe the lag order varies among the different criteria applied. We will focus on the results of AIC, which indicate that the required number of lags for our model is three. So, we will continue our analysis using a VAR model with three lags, denoted as VAR(3), of our data. In other words, three lags are sufficient to make our results reliable, since past values affect the present values in the model.

4.7 VAR model

Since we have done all the necessary tests, we can now proceed to the implementation of the VAR model. To summarize the previous sections, all variables included in the model - to be more specific the transformations of them - are stationary, the errors are uncorrelated and 3 lagged values of each time series are sufficient for our VAR model. The variables included in our model are the rates of defaults for each loan category and four unique macroeconomic factors. We will give more emphasis to the impact of the former time series to the latter variables, since we are interested in the effects of macroeconomic variables to the default ratio. It should be reminded that all time series are treated as endogenous and thus we can study the results of each variable to the others.

The notation for business, consumer and mortgage loans is ‘BUS’, ‘DCONS’ and ‘DMORT’ respectively. As far as the macroeconomic factors, it is easily understood that ‘DUN’ denotes the unemployment rate, ‘DINF’ the inflation, ‘DESI’ is the Economic Sentiment Index and ‘DEXC’ is the Exchange rate. In fact, the first character ‘D’ denotes that these variables have been transformed to first or second differences. The purpose that our data have been differentiated is to avoid non stationarity. Figure 4.7.2-a (or alternatively 4.7.2-b) presents the roots of the characteristic polynomial. It is shown that all of them lie inside the unit circle which means that the system is stationary.

The VAR(3) estimates are illustrated in Figure 4.7.1. Each column in the table corresponds to an equation in the VAR, and each row corresponds to a regressor in the equation. It should be noted that the regressors are grouped by variable, and as we can observe all the lags of the first variable, here RD of business loans, are followed by all of the lags for the second variable, consumer loans, and so on. We have twentyone lag coefficients and one constant for each equation. For each right-hand side variable, EViews prints the estimated coefficient, its standard error, and the *t*-statistic. For example, the coefficient for BUS(-1) in the DCONS equation (meaning in the equation where the LHS variable is consumer loans) is -0.188310, the standard error is 0.14650 (presented in the parenthesis), and the corresponding *t*-statistic is -1.28537. It is apparent that it is difficult to interpret the statistical significance of these coefficients as well as the impact of one variable to the other in this complex figure. This will be conducted in the next section.

4. 8 Impulse Response Functions

Continuing on Eviews, now we will move on the interpretation of this complicated VAR system. The main tool to do this is to use Impulse Response Functions (IRFs). Considering that all variables in a VAR model depend on each other, individual coefficient estimates only provide limited information about the reaction of the system to a shock. In order to get a better picture of the model's dynamic behaviour, impulse responses are used. Impulse-response analysis is used to analyze the dynamic response of an economic variable of interest to shocks in the other economic variables. Impulse responses are best represented in graphs showing the responses of the VAR endogenous variables in time. The results of IRFs are divided into four categories for the sake of clarity and understanding. Each macroeconomic time series is presented as the 'impulse' variable and the responses of each loan category are examined.

Figure (4.6.5) in the appendix represent impulse responses. With a first glance, we observe that impulse response functions are characteristically different depending on which variable represents the 'impulse' and which is the 'response'. It is important to mention that the ordering of the time series has an effect on the impulse responses. Nevertheless, theory does not suggest an obvious ordering of the variables. After applying IRFs several times it was observed that the order of the time series does not have a significant impact on the results. This means that setting the variables with a particular order leads to IRfs having similar graph illustrations with other orders. In other words, in our case the change of ordering of variables did not cause an important change in the results. So, some sensitivity analysis should be undertaken. In our analysis we followed the procedure according to which we set first (in the Cholesky ordering) the variables we would like to interpret, and last those variables we expect to be less considerable in predicting the other variables.

Due to the large number of possible combinations between ‘independent’ and ‘dependent’ variables, we will focus on a subset of IRF diagrams (the most significant). The variables of interest are explained thoroughly next. The horizontal axis of all the diagrams below represents the number of quarters (i.e. periods) after the unexpected shock. The time horizon considered in the present analysis is two and a half years (10 quarters). As we have already mentioned we have managed to "get rid" of short-term autocorrelation in our VAR model but not of the long-term one. For this reason, we chose the above time horizon beyond the quarter that the shock was appeared. The blue line indicates the impulse response function, whereas the red lines are simply the 95% confidence intervals. So the IRF must always lie within this confidence interval. The interpretation of IRF graphs is next.

- **Unemployment rate** – How do RD of loans respond to a shock in this variable?

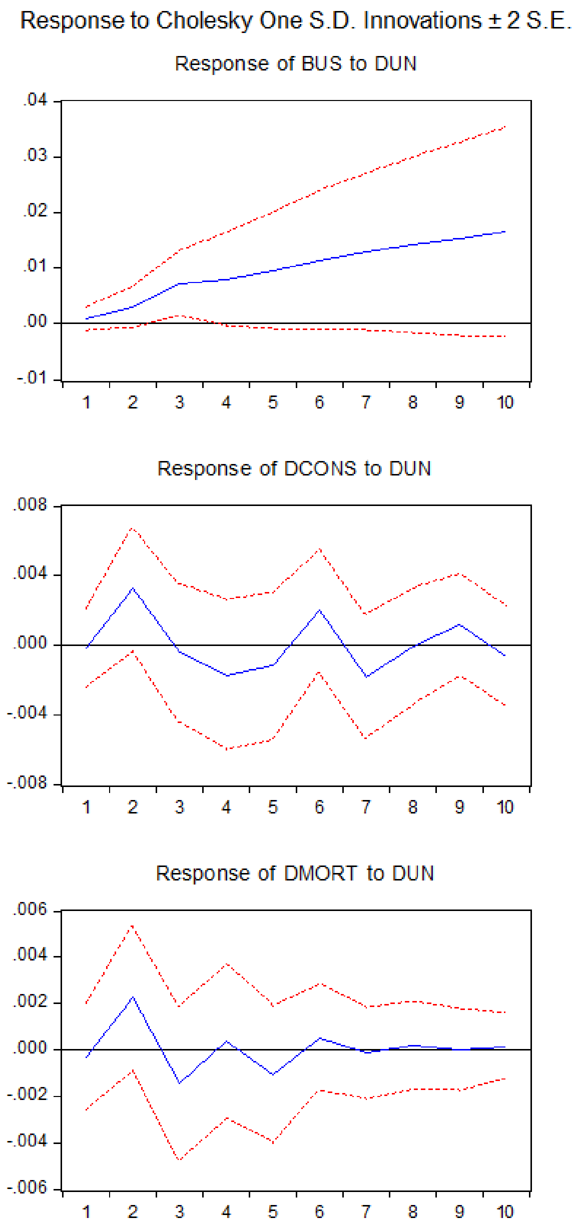

Figure 4.6.1: This figure presents the results of IRFs when the ‘impulse’ variable is the unemployment rate and the rates of default of the three loan categories discribe the responses.

Figure 4.6.1 presents the results of IRFs and more specifically it shows the response of the rates of default to a shock in the variable of unemployment rate. Looking at this graph we can understand that there is a positive reaction of business loans (actually we mean the rate of default of these loans) from the first period. To be more specific, we can observe that from the first period upward business loans gradually increase. This result is consistent with the fact that a rise in the unemployment rate can play a major role in business loans.

Continuing to the second graph of Figure 4.6.1 the response of consumer loans to a shock to the variable of unemployment is depicted. It starts first by causing the deviation between the short-run equilibrium values of the rates of consumer loans to increase after an unanticipated increase in unemployment. This positive response takes place for the first and a half period, and then there is a sharp decline to consumer loans until the third period when the blue line hits the steady state value from where it remains in the negative region till the 5th quarter. We can see an increase from the 4th period till the 6th one. Then, it decreases again and after that it increases till the 9th quarter. Finally, a small decline occurs. To sum up, consumer loans do not seem to have a 'clear' response to the variable of unemployment, instead fluctuations are obvious in the graph.

The third graph illustrates the reaction of rate of default regarding mortgages when a shock in unemployment takes place. A one standard deviation shock to unemployment initially increases mortgages reaching the peak of them in the 2nd period. This positive response sharply declines until the 3rd period. Then, there are some fluctuations on the response until the 7th quarter, when the shock seems to be absorbed, since it subsides to zero towards the end of the time horizon that is illustrated.

- **Inflation rate** - How do RD of loans respond to a shock in this variable?

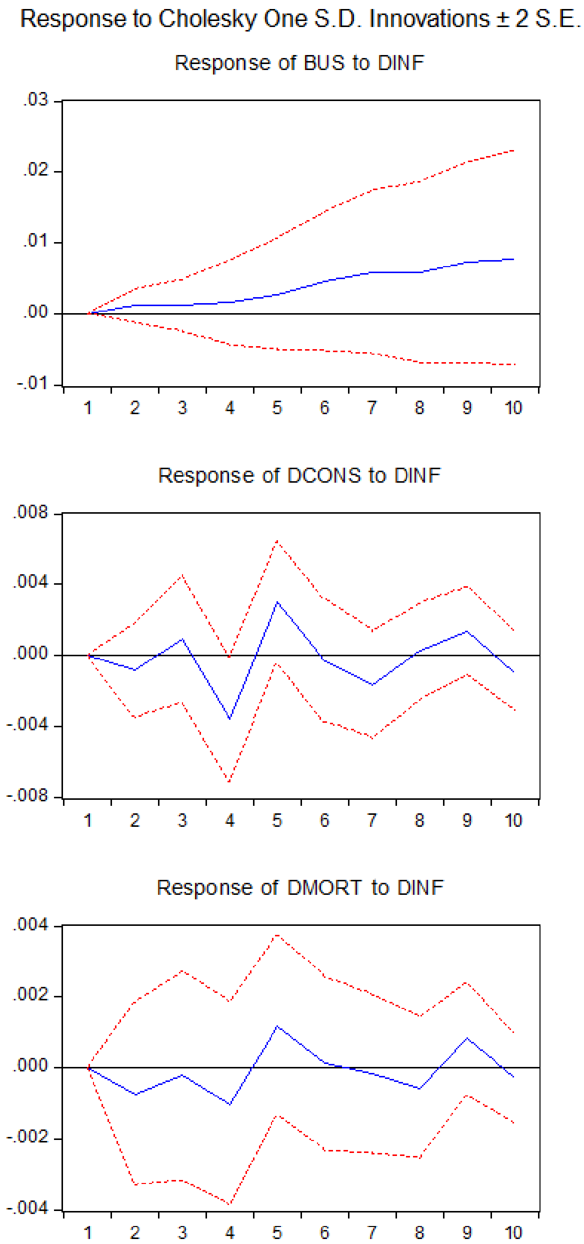

Figure 4.6.2: This figure presents the results of IRFs when the 'impulse' variable is the inflation rate and the rates of default for the three loan categories stand for the responses.

Moving to the next figure (Figure 4.6.2), we will interpret the reaction of each loan category to an unexpected shock in the variable of inflation. The first graph illustrates that the response of business loans gradually increases from the first period till the end of the time horizon depicted in the graph. We can observe that a shock to inflation will have a positive impact on business loans, meaning an increase in the inflation rate will lead to higher rate of default for this loan type. This is also reasonable and consistent with the economic theory.

As far as the response of consumer loans is concerned, there is not a general conclusion. To be more specific, the second graph of Figure 4.6.2 shows that there is a decline for the first two periods and the impact is negative. Then, a small increase takes place before a decrease till the end of the first year (4^{th} quarter). It is worth mentioning that a sharp increase occurs during the 5^{th} period when the blue line reaches its maximum value. Thereafter, we can see that the impact of the shock will cause consumer loans to decrease up to the 7^{th} quarter and remain in the negative region till the 8^{th} quarter. Then it tends to increase before it finally declines in the last period.

The last graph of this Figure indicates the shock does not seem to be absorbed by mortgage loans in the near future and it takes a long time for it to start converging. In particular, the effect is initially negative until the 4^{th} quarter when it hits its steady state value. Beyond the 5^{th} quarter, the results are pretty much the same with consumer loans, since mortgages rise above its steady state value and remain in the positive region for the next period. After that, a decline and then an increase take place again leading to a decrease two and a half years (10^{th} period) after the appearance of the shock. In other words, shocks to inflation will have asymmetric impacts on the rate of default of mortgage loans. However, one can result the shock seems to be absorbed after two and a half years (Figure 4.6.5).

- **Economic Sentiment indicator** - How do RD of loans respond to a shock in this variable?

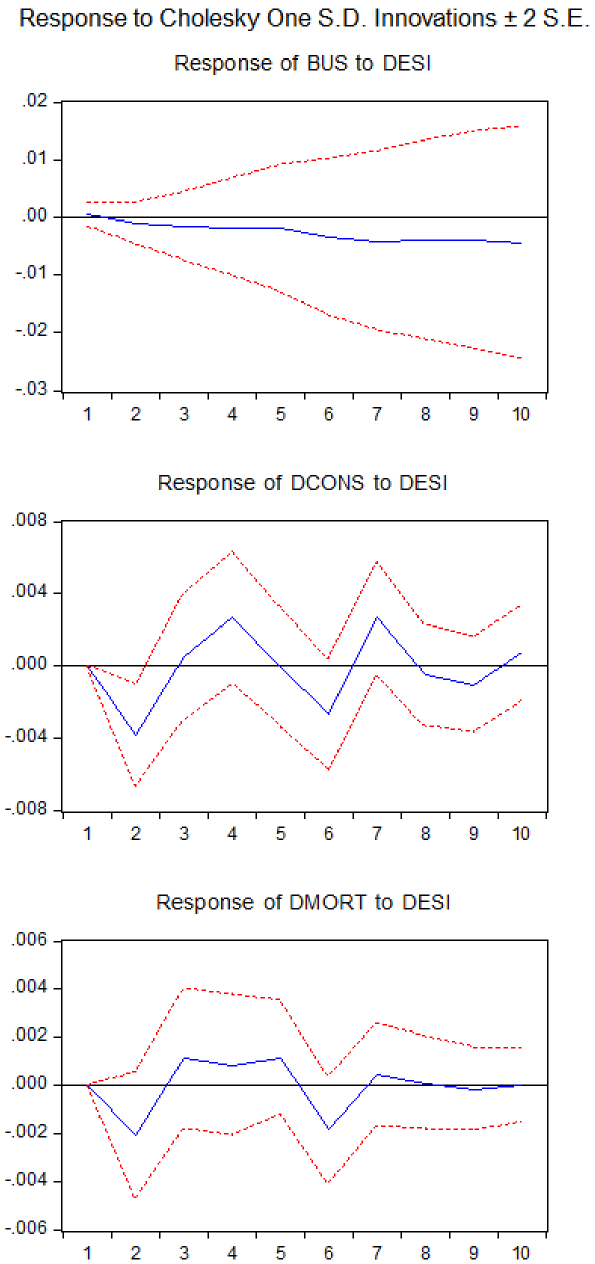

Figure 4.6.3: This figure presents the results of IRFs when the 'impulse' variable is the economic sentiment index and the rates of default of the three loan categories describe the responses.

Based on Figure 4.6.3, the one standard deviation shock in the variable of economic sentiment index (ESI) causes a slight decrease to the variable of business loans. Moreover, the business loans have a negative effect in this case, which is still persistent till the last period. This effect can be explained as a rise to ESI results in a decline in the rate of default of business loans. This is logical if we take into account the confidence indicators that ESI consists of (such as industrial confidence indicator in 40%).

As far as the reaction of consumer loans to the aforementioned shock, we can observe that there is not a specific reaction of the rate of default of this loan category. More specifically, a negative decreasing impact is illustrated during the first two periods and according to the graph the effect of consumer loans gets its minimum value in the 2nd quarter. After that there is an increase and then a decline up to the 6th period (the response is positive from the 3rd up to the 5th quarter). Afterwards, the response acts in a similar way and seems not to have such large changes after the last periods (Figure 4.6.5).

Last but not least, the response of mortgage loans to an unexpected shock in ESI goes down in the near future. Then, it continues by causing the deviation between the short-run equilibrium values of mortgages to increase after an unanticipated increase in ESI. However, the response of mortgages is less persistent over the next two periods (3rd – 5th quarter). After that a decline and an increase follow successively. Finally, the effect of such a shock is less persistent over and after the 7th time period, which means that consumer loans are then less affected by this shock.

- **Exchange Rate (EUR to USD)** - How do RD of loans respond to a shock in this variable?

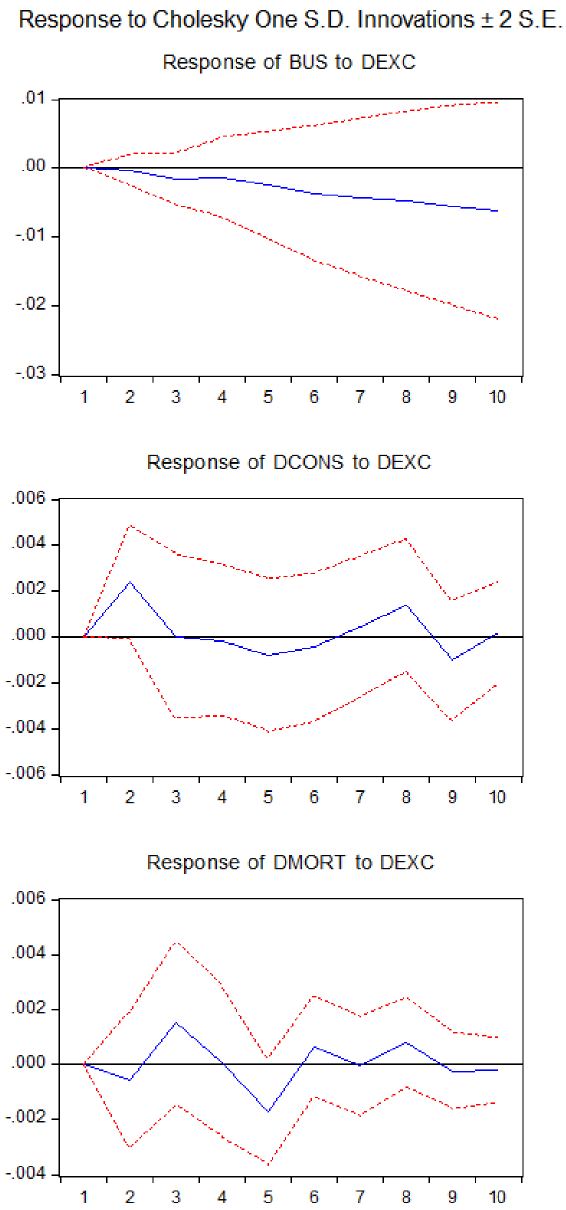

Figure 4.6.4: This figure presents the results of IRFs when the ‘impulse’ variable is the exchange rate (EUR to USD) and the responses consist of the rates of default of the three loan categories.

Figure 4.6.4 examines the response of loans’ default to a shock in the exchange rate. Let us remind to the reader that we have chosen the EUR to USD exchange rate for the present analysis. Firstly, business loans are determined. A negative impact of the shock is obvious on this loan category. In specific, business loans are not contemporaneously affected by a shock to the exchange rate. However, the response gradually declines from the 2^{nd} till the last period. The decay in the plot illustrates that, as time passes, the effects of a shock in the exchange rate today lead to a reduction to the rate of default. This is reasonable since an increase in the exchange rate between EUR and US dollar indicates EUR becomes stronger and thus we can conclude that businesses can probably repay their loans. Also we see that the blue line lies below zero, which means that a shock to the exchange rate will have a negative impact on the rate of default of business loans.

The second graph illustrates the response of consumer loans to a shock like the above one. We can see that the blue line initially increases but in continue there is the ‘opposite’ impact. The initial positive response declines until the 3^{rd} quarter when it hits the horizontal axis (its steady state value) from where it remains in the negative region up to the end of the 6^{th} period, howbeit with increasing tendencies. It becomes positive in the 7^{th} quarter. Actually, there is an increase in the 5^{th} till the 8^{th} period and then it gradually converges to zero, which signifies that the shock is starting to be absorbed over the next 10 quarter of its emergence.

The last diagram of Figure 4.6.4 presents the corresponding results regarding mortgage loans. It is apparent that the graph is characterized by fluctuations in the response. Nevertheless, a more distinct picture is shown from the 9^{th} quarter upwards since the blue line converges to the horizontal axis (Figure 4.6.5). We can therefore explain this feature of the blue line assuming that the shock is absorbed by mortgages about 9 periods after its onset.

4.9 Variance Decomposition

As it is previously stated, Variance Decomposition (VD) is another way to interpret our VAR model. Variance decomposition analysis enables us to estimate the future change of each endogenous variable due to the simultaneous variation of other endogenous variables. Essentially, VD helps us to determine the proportion of forecast error variance in one variable explained by innovations in itself and the other variables. It should be mentioned that even though the ordering of the variables is important, the results did not significantly change with different ordering for the variables in the model. The results of this technique are usually represented in table form. The forecast time horizon, which is shown in the first column of each table, is 10 periods meaning 10 quarters (same as IRFs). The following columns show the percentage of the predicted variation from each shock.

Figure 4.7 depicts three tables which correspond to each loan category. To be more specific, the first table indicates how much (in percentage points) a shock to one variable impacts the (variance of the) forecast error of the rate of default of business loans. It should be noted that own series shocks usually explain most of the error variance, although the shock will also affect the other variables in the system.

It is easily observed that in the second period the percentage of variance explained by own shock for business loans declines to about 93.5 percent and continues falling until it ends with an average of around 82 percent at the end of the 10th quarter. Moreover, in the second period the fraction of business loans (in fact we mean their rate of default) forecast error variance attributable to variations in the variable of unemployment is about 2%. Then this rate increases from the third quarter and by the end of the 10th period, the contribution averages around 10.2 percent. Lower percentages are shown for the other variables of the system.

Continuing to the next table we can observe the corresponding results for the rate of default of consumer loans. In the first row it is shown that the percentage of variance explained by its own shock is approximately 91.5 percent, while the rest 8.5 percent is explained by business loans. Let us now move to the 5th period, where the variance of inflation accounts for about 8.8 percent and it is the next higher percentage after the variable itself. It then increases reaching an average value of 7.9 percent at the end of the last quarter. At this time, the interpretation of the variability in the forecast error of consumer loans (in their equation), when a shock to the variable of economic sentiment index takes place, is about 11.3 percent. However, own series shocks explain most of the error variance in the equation of consumer loans since it is about 48 percent.

Moving to the last table, we will interpret the results regarding the equation of mortgages. In particular, an own series shock will again explain the most of the error variance (since it is about 85.5%) in the first period. The rest is explained by shocks in the other loan categories. Continuing to the next periods, one can observe that a shock in the variables of consumer loans, unemployment and economic sentiment index can explain in a significant degree the variance of the error in the equation of mortgages. It is also worth mentioning that at the end of the 10^{th} quarter the fraction of mortgage loans forecast error variance attributable to variations in consumer loans is approximately 17.7% and about 6.3% is explained by economic sentiment index.

4.10 Testing Causation using Granger's Causality Test

At the last step of this thesis, A Granger causality test is performed. The structures of the causal short-run relationships between our variables were also analyzed through the Granger causality approach. The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another. According to Granger, if a variable X_1 "Granger-causes" (or "G-causes") a variable X_2, then past values of X_1 should contain information that helps predict X_2 above and beyond the information contained in past values of X_2 solely. As usual in statistical significance tests, if the probability value is less than any $1\text{-}\alpha$ level of significance (5% for our analysis), then the null hypothesis will be rejected at this level. The null hypothesis in this test is stated as there is no Granger-causality between time series. A necessary assumption before applying the test is that the data is covariance stationary. We have already transformed our data so we can continue. The outcome tables of Granger-causality tests are presented in the Appendix (Figure 4.8.1, Figure 4.8.2). We proceed to further analysis only for the relationships that generated significant and interpretable results regarding pair-wise Granger causality tests between our variables.

One can result from Figure 4.8.1 that the variable of unemployment Granger-causes not only the variable of business loans, but also consumer loans, since the p- value in both cases is less than 0.05 (0.0041 and 0.0172 respectively). So, past values of the unemployment are useful for predicting future values of the rate of default of the above loan categories. One approach to explain this is that the increase of unemployment the last decades in our country may have a significant effect on the present repayment of business as well as consumer loans, and thus including these macroeconomic time series in the equations of these loans can improve the explanation of the latter. However, is easily observed that the majority of p-values are greater than 5% or even 10%. This is not a major issue keeping in mind the previous analysis.

Apart from that, Figure 4.8.1 shows that time series concerning economic sentiment index G-causes the rate of default of consumer loans or alternatively stated ESI helps in the prediction of consumer loans (p-value=0.0077). However, this short run relationship is not bidirectional, since consumer loans do not G-cause the variable of ESI (p-value=0.9078). Actually, all the aforementioned relationships were found to be unidirectional, which means that even if a variable 'A' G-causes a variable 'B', the opposite relationship does not occur. It is worth mentioning that for a higher significance level, for example 10%, we can lead to different conclusions. For instance, we can see that there is a two-way relationship between rate of default of consumer and the respective rate of business loans as far as Granger causality is concerned. Thus, past values of the former may play a significant role on the latter and vice versa.

Figure 4.8.2 presents the results of the VAR Granger causality tests. Similar results can be exported, however we can see that the joint tests are also shown in the tables. More specifically, one can observe that for the variable business loans it is shown that the other variables jointly Granger cause its rate of default at 5% significance level (the p-value for the joint test is 0.0204 < 0.05). Similar results are obtained for consumer loans (p-value is approximately zero) as well as exchange rate (p-value=0.0161). This can indicate the existence of Granger-causalities, despite the fact that only one (or even more) variable was found to have 'direct' relationship with the above-mentioned variables. It was also found that Granger causality exists between macroeconomic variables, nevertheless it is beyond the scope the present analysis.

Chapter 5

Conclusion and Recommendations

This chapter provides a summary of the findings of this study. Moreover, the implications of these results are presented. Apart from that, some recommendations are proposed for future research and analysis at the end of this chapter.

NPLs are an acute issue for Greek banking sector. In this paper we investigate whether a group of four specific macroeconomic factors determine NPLs, using data from the last quarter of 2002 until the first quarter of 2019. Our econometric analysis is based on Vector Autoregressive model and more specifically, on Impulse Response Functions which can interpret the results of a VAR model. The results of this analysis lead to a number of interesting conclusions since it was shown that macroeconomic conditions play a major role in the rate of default of loans in the Greek banking system. In particular, changes in the unemployment rate constitute a determinant in explaining the rate of default of business loans. In addition, the repayment of business loans was also found to depend on changes in the other independent variables. The other categories of loans - including consumer and mortgages - do not appear to have ‘clear’ influences from changes in the unemployment and the other macroeconomic determinants. This can be attributed to the fact that other factors (such as bank-specific determinants) are also significant and may have a considerable impact on these loans categories. However, it is shown that both consumer and mortgage loans can almost absorb a shock to any of the other variables, meaning inflation, economic sentiment index and exchange rate (EUR to USD) after a given time period.

Based on the findings above, banks' profitability regarding lending is not only affected by bank-specific variables. Thus, banks should also take into account the current macroeconomic conditions in order to improve their credit risk management. The present results can be very useful in understanding sources of loan defaulting from a macroeconomic standpoint.

In terms of future research, the current study can be extended in many ways. One of them could be adding more than one country in the investigation. In addition, a cointegration analysis could be performed so as to test the long-term impacts of these macroeconomic factors using a larger and more satisfying dataset or even adding more variables.

Appendix

A.1 Augmented Dickey-Fuller test – Output tables

Null Hypothesis: BUS has a unit root
Exogenous: Constant
Lag Length: 4 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.456199	0.0126
Test critical values:	1% level	-3.542097	
	5% level	-2.910019	
	10% level	-2.592645	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.1: Augmented Dickey-Fuller test for Rate of Default of Business Loans. It is proven that this variable is stationary and no transformation needed.

Null Hypothesis: CONS has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.640288	0.4562
Test critical values:	1% level	-3.540198	
	5% level	-2.909206	
	10% level	-2.592215	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.2: Augmented Dickey-Fuller test for the Rate of Default of Consumer Loans.

Null Hypothesis: D(CONS) has a unit root
Exogenous: Constant
Lag Length: 8 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.987860	0.2912
Test critical values:	1% level	-3.552666	
	5% level	-2.914517	
	10% level	-2.595033	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.3: Augmented Dickey-Fuller test for the first differences of the Rate of Default of Consumer Loans.

Null Hypothesis: D(CONS,2) has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.924320	0.0000
Test critical values:	1% level	-3.540198	
	5% level	-2.909206	
	10% level	-2.592215	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.4: Augmented Dickey-Fuller test for the second differences of the Rate of Default of Consumer Loans. It is shown that the variable of consumer loans is stationary only after taking second differences.

Null Hypothesis: MORT has a unit root
Exogenous: Constant
Lag Length: 7 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.932129	0.3156
Test critical values:	1% level	-3.548208	
	5% level	-2.912631	
	10% level	-2.594027	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.5: Augmented Dickey-Fuller test for Rate of Default of Mortgage Loans.

Null Hypothesis: D(MORT) has a unit root
Exogenous: Constant
Lag Length: 5 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.506657	0.5234
Test critical values:	1% level	-3.546099	
	5% level	-2.911730	
	10% level	-2.593551	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.6: Augmented Dickey-Fuller test for first differences for the Rate of Default of Mortgage Loans.

Null Hypothesis: D(MORT,2) has a unit root
Exogenous: Constant
Lag Length: 4 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.649441	0.0000
Test critical values:	1% level	-3.546099	
	5% level	-2.911730	
	10% level	-2.593551	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.7: Augmented Dickey-Fuller test for second differences of the Rate of Default of Mortgage Loans. The variable is stationary after this transformation.

Null Hypothesis: UNEMP has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.192351	0.2111
Test critical values:	1% level	-3.540198	
	5% level	-2.909206	
	10% level	-2.592215	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.8: Augmented Dickey-Fuller test for the variable of Unemployment rate.

Null Hypothesis: D(UNEMP) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.380722	0.1511
Test critical values:	1% level	-3.536587	
	5% level	-2.907660	
	10% level	-2.591396	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.9: Augmented Dickey-Fuller test for first differences of Unemployment rate.

Null Hypothesis: D(UNEMP,2) has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.235278	0.0013
Test critical values:	1% level	-3.544063	
	5% level	-2.910860	
	10% level	-2.593090	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.10: Augmented Dickey-Fuller test for second differences of Unemployment rate.

Null Hypothesis: INFL has a unit root
Exogenous: Constant
Lag Length: 5 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.290959	0.6284
Test critical values:	1% level	-3.544063	
	5% level	-2.910860	
	10% level	-2.593090	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.11: Augmented Dickey-Fuller test for the variable of Inflation rate.

Null Hypothesis: D(INFL) has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.847006	0.0000
Test critical values:	1% level	-3.542097	
	5% level	-2.910019	
	10% level	-2.592645	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.12: Augmented Dickey-Fuller test for first differences of Inflation rate.

Null Hypothesis: ESI has a unit root
Exogenous: Constant
Lag Length: 5 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.269350	0.0208
Test critical values:	1% level	-3.544063	
	5% level	-2.910860	
	10% level	-2.593090	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.13: Augmented Dickey-Fuller test for the change of Economic Sentiment Index.

Null Hypothesis: D(ESI) has a unit root
Exogenous: Constant
Lag Length: 10 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.322026	0.0011
Test critical values:	1% level	-3.557472	
	5% level	-2.916566	
	10% level	-2.596116	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.14: Augmented Dickey-Fuller test for the first differences of change of Economic Sentiment Index.

Null Hypothesis: EXCH has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.355847	0.1582
Test critical values:	1% level	-3.534868	
	5% level	-2.906923	
	10% level	-2.591006	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.15: Augmented Dickey-Fuller test for the variable of Exchange rate.

Null Hypothesis: D(EXCH) has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic - based on AIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.666111	0.0000
Test critical values:	1% level	-3.542097	
	5% level	-2.910019	
	10% level	-2.592645	

*MacKinnon (1996) one-sided p-values.

Figure 4.3.16: Augmented Dickey-Fuller test for the first differences of Exchange rate.

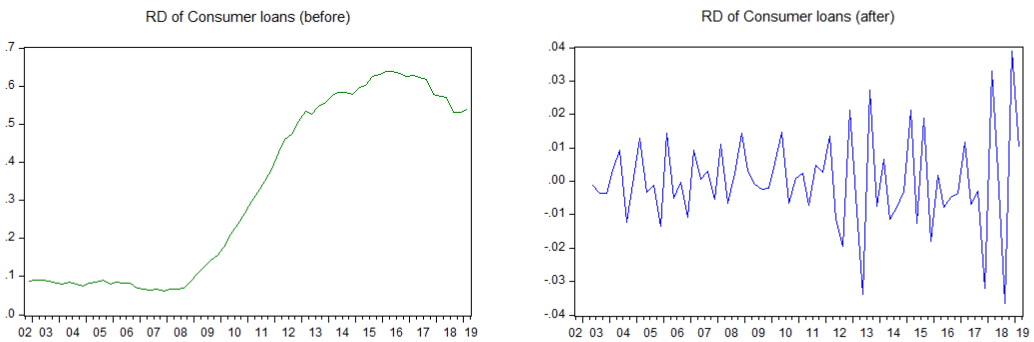

Figure 4.1: The rate of default (RD) of consumer loans is illustrated in this figure. The graph on the left hand side presents the RD before any transformation (non stationary), while the right hand side graph presents the corresponding rate after taking second differences of the variable (stationary).

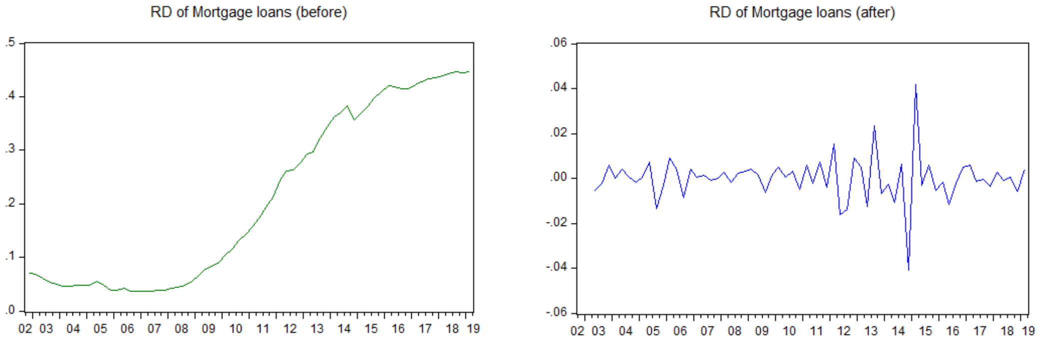

Figure 4.2: The rate of default (RD) of mortgage loans is illustrated in this figure. The graph on the left hand side presents the RD before any transformation (non stationary), while the right hand side graph presents the corresponding rate after taking second differences of the variable (stationary).

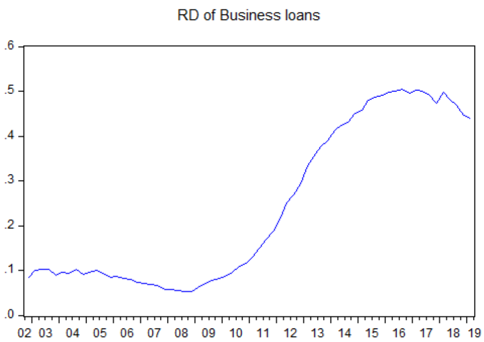

Figure 4.3: The rate of default (RD) of business loans is illustrated in this figure. This variable was found to be stationary without any transformation.

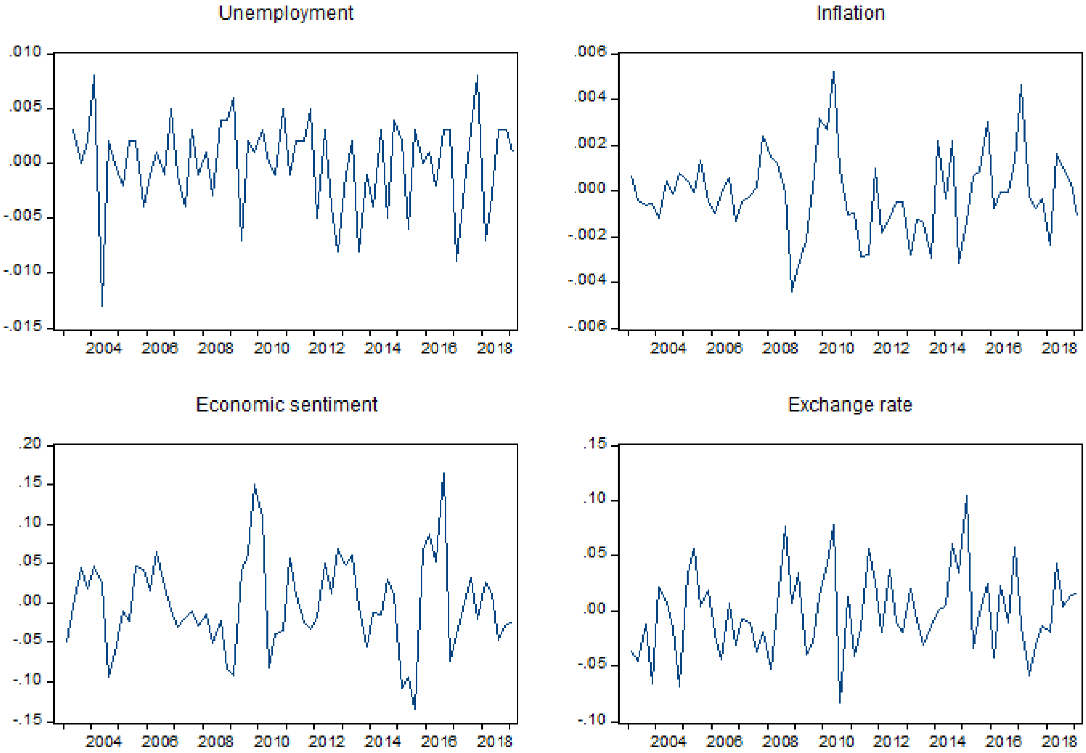

Figure 4.4: This figure presents the macroeconomic variables after their final transformations in order to become stationary (unemployment rate in 2nd differences, inflation rate, economic sentiment index and exchange rate in 1st differences).

A.2 Output Tables

VAR Residual Serial Correlation LM Tes
Null Hypothesis: no serial correlation at
Date: 02/02/20 Time: 21:04
Sample: 2002Q4 2019Q1
Included observations: 62

Lags	LM-Stat	Prob
1	78.75199	0.0045
2	80.53741	0.0030
3	43.81934	0.6826

Probs from chi-square with 49 df.

Figure 4.5.1: This figure presents the output of Breusch-Godfrey test for serial correlation of the error terms.

VAR Lag Order Selection Criteria
Endogenous variables: BUS CONS ESI EXC INF MORT UNEMP
Exogenous variables: C
Date: 02/03/20 Time: 20:29
Sample: 2002Q4 2019Q1
Included observations: 63

Lag	LogL	LR	FPE	AIC	SC	HQ
0	899.4406	NA	1.17e-21	-28.33145	-28.09332	-28.23779
1	1498.595	1046.142	3.07e-29	-45.79666	-43.89165*	-45.04741*
2	1563.100	98.29442	1.99e-29	-46.28890	-42.71701	-44.88406
3	1622.845	77.76327*	1.64e-29*	-46.63001*	-41.39124	-44.56958

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Figure 4.6: This figure presents the results of the optimal lag length (denoted with an asterisk) using Akaike, Schwarz and Hannan-Quinn information criteria.

Vector Autoregression Estimates
Date: 02/11/20 Time: 19:25
Sample (adjusted): 2004Q1 2019Q1
Included observations: 61 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	BUS	DCONS	DMORT	DINF	DUN	DESI	DEXC
BUS(-1)	1.420199	-0.188310	-0.064253	-0.014885	-0.097615	1.335241	-0.641310
	(0.13940)	(0.14650)	(0.14918)	(0.02880)	(0.06752)	(0.91124)	(0.56529)
	[10.1876]	[-1.28537]	[-0.43071]	[-0.51690]	[-1.44566]	[1.46531]	[-1.13448]
BUS(-2)	0.157417	0.163660	0.112329	0.022483	0.101867	-2.292038	1.775303
	(0.26322)	(0.27662)	(0.28168)	(0.05437)	(0.12749)	(1.72057)	(1.06737)
	[0.59804]	[0.59163]	[0.39879]	[0.41350]	[0.79899]	[-1.33214]	[1.66325]
BUS(-3)	-0.592403	0.007489	-0.055285	-0.005110	-0.004115	0.934223	-1.106008
	(0.15163)	(0.15935)	(0.16226)	(0.03132)	(0.07344)	(0.99115)	(0.61487)
	[-3.90689]	[0.04700]	[-0.34071]	[-0.16314]	[-0.05603]	[0.94257]	[-1.79878]
DCONS(-1)	-0.316402	-0.894400	0.085139	0.081778	0.120421	-1.561806	1.522981
	(0.15973)	(0.16786)	(0.17093)	(0.03299)	(0.07737)	(1.04409)	(0.64771)
	[-1.98086]	[-5.32820]	[0.49810]	[2.47855]	[1.55649]	[-1.49586]	[2.35134]
DCONS(-2)	-0.391630	-0.888334	-0.090746	0.098555	0.137966	-0.760997	0.826054
	(0.19858)	(0.20869)	(0.21250)	(0.04102)	(0.09618)	(1.29802)	(0.80524)
	[-1.97218]	[-4.25677]	[-0.42704]	[2.40267]	[1.43440]	[-0.58627]	[1.02585]
DCONS(-3)	-0.411416	-0.302036	-0.178569	0.056359	0.143394	-1.602114	0.998261
	(0.18717)	(0.19670)	(0.20030)	(0.03866)	(0.09066)	(1.22347)	(0.75899)
	[-2.19805]	[-1.53550]	[-0.89152]	[1.45768]	[1.58168]	[-1.30948]	[1.31525]
DMORT(-1)	0.101803	0.009566	-0.736122	-0.023245	0.027422	0.098813	-1.211700
	(0.17099)	(0.17970)	(0.18298)	(0.03532)	(0.08282)	(1.11769)	(0.69337)
	[0.59538]	[0.05324]	[-4.02297]	[-0.65812]	[0.33110]	[0.08841]	[-1.74755]
DMORT(-2)	-0.127846	0.227410	-0.301470	-0.019730	0.027033	-0.216848	-1.027811
	(0.21253)	(0.22335)	(0.22744)	(0.04390)	(0.10294)	(1.38925)	(0.86183)
	[-0.60153]	[1.01816]	[-1.32551]	[-0.44941]	[0.26260]	[-0.15609]	[-1.19259]
DMORT(-3)	-0.142896	-0.091058	-0.144313	-0.028791	-0.021154	1.421533	-1.435550
	(0.18833)	(0.19792)	(0.20154)	(0.03890)	(0.09122)	(1.23105)	(0.76369)
	[-0.75875]	[-0.46008]	[-0.71606]	[-0.74009]	[-0.23191]	[1.15474]	[-1.87976]
DINF(-1)	0.969113	-0.642925	-0.285932	0.274679	-0.636774	10.00228	-8.505472
	(0.87127)	(0.91563)	(0.93237)	(0.17997)	(0.42201)	(5.69516)	(3.53303)
	[1.11230]	[-0.70217]	[-0.30667]	[1.52622]	[-1.50890]	[1.75628]	[-2.40741]
DINF(-2)	-0.001144	1.957708	0.112935	0.133838	0.418090	-7.520072	12.82169
	(0.86066)	(0.90448)	(0.92101)	(0.17778)	(0.41687)	(5.62577)	(3.48999)
	[-0.00133]	[2.16447]	[0.12262]	[0.75283]	[1.00293]	[-1.33672]	[3.67385]
DINF(-3)	0.098996	-2.559274	-1.163297	-0.445573	0.388225	1.708764	-9.560318
	(0.77274)	(0.81208)	(0.82692)	(0.15962)	(0.37428)	(5.05107)	(3.13347)
	[0.12811]	[-3.15151]	[-1.40678]	[-2.79146]	[1.03725]	[0.33830]	[-3.05103]
DUN(-1)	0.541395	0.492658	0.528718	0.093314	-0.457851	0.818220	1.316393
	(0.33533)	(0.35240)	(0.35884)	(0.06927)	(0.16242)	(2.19192)	(1.35977)
	[1.61451]	[1.39800]	[1.47340]	[1.34716]	[-2.81891]	[0.37329]	[0.96810]
DUN(-2)	1.087132	0.760302	0.192386	0.025978	-0.309558	0.501426	1.648801
	(0.35891)	(0.37719)	(0.38408)	(0.07414)	(0.17384)	(2.34607)	(1.45540)
	[3.02896]	[2.01572]	[0.50090]	[0.35040]	[-1.78067]	[0.21373]	[1.13288]
DUN(-3)	0.246078	0.088570	0.015817	-0.034208	-0.092417	-0.465346	-0.942303
	(0.34459)	(0.36213)	(0.36875)	(0.07118)	(0.16691)	(2.25245)	(1.39732)
	[0.71412]	[0.24458]	[0.04289]	[-0.48058]	[-0.55371]	[-0.20660]	[-0.67436]
DESI(-1)	-0.036032	-0.054909	-0.041061	-0.002076	0.014409	0.350078	0.118959
	(0.02455)	(0.02580)	(0.02627)	(0.00507)	(0.01189)	(0.16047)	(0.09955)
	[-1.46774]	[-2.12832]	[-1.56302]	[-0.40935]	[1.21182]	[2.18161]	[1.19499]
DESI(-2)	-0.012159	-0.045366	0.011468	0.012987	-0.002623	-0.003876	-0.088044
	(0.02638)	(0.02772)	(0.02823)	(0.00545)	(0.01278)	(0.17243)	(0.10697)
	[-0.46093]	[-1.63650]	[0.40625]	[2.38349]	[-0.20533]	[-0.02248]	[-0.82310]
DESI(-3)	-0.015426	0.037809	0.025018	0.001917	-0.005057	-0.288942	0.039160
	(0.02664)	(0.02799)	(0.02851)	(0.00550)	(0.01290)	(0.17412)	(0.10802)
	[-0.57908]	[1.35060]	[0.87763]	[0.34841]	[-0.39190]	[-1.65940]	[0.36253]
DEXC(-1)	-0.012206	0.081857	-0.020175	-0.006091	0.005687	-0.310144	0.152526
	(0.04012)	(0.04216)	(0.04293)	(0.00829)	(0.01943)	(0.26222)	(0.16267)
	[-0.30429]	[1.94171]	[-0.46997]	[-0.73502]	[0.29271]	[-1.18278]	[0.93765]
DEXC(-2)	-0.015884	0.035134	0.015586	-0.014235	0.008048	-0.224232	-0.274051
	(0.04207)	(0.04421)	(0.04502)	(0.00869)	(0.02038)	(0.27497)	(0.17058)
	[-0.37760]	[0.79475]	[0.34624]	[-1.63827]	[0.39500]	[-0.81548]	[-1.60659]
DEXC(-3)	0.025497	0.011479	0.005231	0.001279	-0.029368	0.619585	-0.191419
	(0.03502)	(0.03680)	(0.03747)	(0.00723)	(0.01696)	(0.22890)	(0.14200)
	[0.72812]	[0.31192]	[0.13959]	[0.17680]	[-1.73146]	[2.70682]	[-1.34804]
C	0.003441	0.004777	0.001749	-0.000605	0.000478	0.003838	-0.007925
	(0.00212)	(0.00222)	(0.00227)	(0.00044)	(0.00103)	(0.01384)	(0.00858)
	[1.62540]	[2.14707]	[0.77225]	[-1.38385]	[0.46651]	[0.27739]	[-0.92318]
R-squared	0.998628	0.750140	0.498284	0.445081	0.412345	0.456962	0.467917
Adj. R-squared	0.997889	0.615600	0.228130	0.146278	0.095915	0.164556	0.181411
Sum sq. resids	0.002700	0.002981	0.003091	0.000115	0.000633	0.115345	0.044390
S.E. equation	0.008320	0.008743	0.008903	0.001719	0.004030	0.054384	0.033737
F-statistic	1351.322	5.575588	1.844441	1.489548	1.303115	1.562767	1.633182
Log likelihood	219.2235	216.1943	215.0895	315.4292	263.4438	104.7010	133.8261
Akaike AIC	-6.466345	-6.367025	-6.330803	-9.620628	-7.916191	-2.711509	-3.666429
Schwarz SC	-5.705046	-5.605726	-5.569504	-8.859329	-7.154892	-1.950210	-2.905130
Mean dependent	0.251009	0.000266	9.00E-05	-0.000102	-6.56E-05	-0.000567	0.001603
S.D. dependent	0.181062	0.014102	0.010134	0.001860	0.004238	0.059499	0.037289

Determinant resid covariance (dof adj.)	2.37E-29
Determinant resid covariance	1.04E-30
Log likelihood	1499.908
Akaike information criterion	-44.12812
Schwarz criterion	-38.79903

Figure 4.7.1: The VAR estimates are represented in this figure. All variables included in the model are stationary and there are three lags of each variable.

Roots of Characteristic Polynomial
Endogenous variables: BUS DCONS DMORT DINF D
Exogenous variables: C
Lag specification: 1 3
Date: 02/11/20 Time: 19:40

Root	Modulus
0.963665 + 0.070948i	0.966273
0.963665 - 0.070948i	0.966273
-0.277147 + 0.895936i	0.937823
-0.277147 - 0.895936i	0.937823
0.554065 + 0.583061i	0.804331
0.554065 - 0.583061i	0.804331
0.309597 + 0.729473i	0.792453
0.309597 - 0.729473i	0.792453
-0.740901 - 0.119531i	0.750482
-0.740901 + 0.119531i	0.750482
-0.631751 + 0.402779i	0.749226
-0.631751 - 0.402779i	0.749226
0.560197 - 0.374689i	0.673953
0.560197 + 0.374689i	0.673953
-0.329389 + 0.516881i	0.612914
-0.329389 - 0.516881i	0.612914
-0.483650	0.483650
0.097629 + 0.362896i	0.375799
0.097629 - 0.362896i	0.375799
-0.329333	0.329333
-0.089837	0.089837

No root lies outside the unit circle.
VAR satisfies the stability condition.

Figure 4.7.2-a: Roots of the characteristic polynomial are represented in this figure (modulus is the absolute value of our roots, since there may be complex roots). It is apparent that all the eigenvalues of our VAR(3) model are smaller than unity and thus our system is stationary.

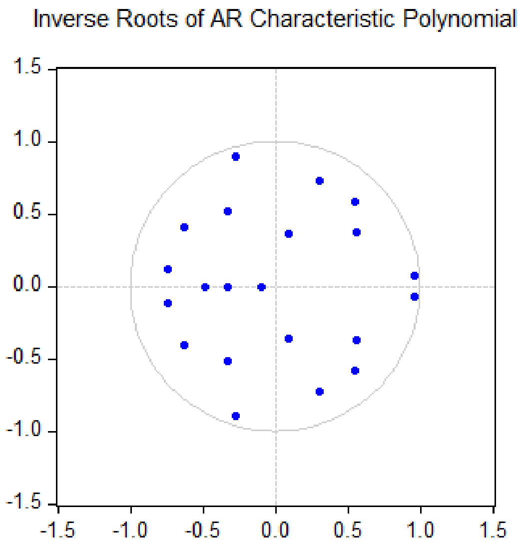

Figure 4.7.2-b: This figure represents the roots of the characteristic polynomial in a graph. All roots lie inside the unit circle and thus our system is stationary.

Variance Decomposition of BUS:								
Period	S.E.	BUS	DCONS	DMORT	DINF	DUN	DESI	DEXC
1	0.008320	100.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.013760	93.56125	1.629293	0.788078	0.253900	2.105585	1.596605	0.065290
3	0.022422	88.38048	2.184617	0.338065	0.097046	6.652464	1.829291	0.518037
4	0.030843	87.31734	3.119340	0.185511	0.057497	6.977460	1.866673	0.476179
5	0.040203	87.46105	2.545758	0.110560	0.114808	7.248853	1.835131	0.683841
6	0.049633	85.67960	2.521594	0.076177	0.328176	8.100954	2.282412	1.011091
7	0.059261	84.32238	2.211878	0.070277	0.568158	8.889195	2.692868	1.245241
8	0.069048	83.97207	1.857439	0.051810	0.654014	9.299989	2.758566	1.406108
9	0.078672	83.38144	1.668281	0.040000	0.821960	9.740207	2.773910	1.574205
10	0.088120	82.61469	1.604956	0.032347	0.947436	10.20615	2.824687	1.769731

Variance Decomposition of DCONS:								
Period	S.E.	BUS	DCONS	DMORT	DINF	DUN	DESI	DEXC
1	0.008743	8.500174	91.49983	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.013103	10.62840	71.80885	0.106416	1.428386	5.196365	7.593823	3.237759
3	0.013375	10.54623	71.70316	0.365266	1.881484	4.987214	7.409191	3.107457
4	0.015205	12.38646	57.06487	6.029131	5.514598	7.304052	9.283696	2.417190
5	0.016091	13.98616	52.38855	7.505097	8.804484	6.550844	8.343568	2.421298
6	0.016670	13.85105	50.34413	7.454344	8.411617	7.454848	10.14977	2.334240
7	0.017632	13.45142	49.88979	6.889167	7.989055	8.161190	11.47020	2.149181
8	0.017765	13.34715	49.67475	6.922474	7.887759	8.041581	11.39346	2.732821
9	0.018154	14.22986	48.27866	6.901550	7.874131	8.524212	11.25163	2.939958
10	0.018283	14.22839	48.08414	6.863764	7.949221	8.672610	11.28987	2.911999

Variance Decomposition of DMORT:								
Period	S.E.	BUS	DCONS	DMORT	DINF	DUN	DESI	DEXC
1	0.008903	0.234024	14.28708	85.47890	0.000000	0.000000	0.000000	0.000000
2	0.010924	0.213613	12.34761	79.99486	1.184861	2.882391	3.093685	0.282977
3	0.011298	0.455513	11.63118	76.68804	1.115574	4.239229	3.807936	2.062525
4	0.011496	0.693624	12.01327	74.95374	1.939962	4.094590	4.307523	1.997291
5	0.011891	0.700891	13.17295	70.06529	3.172816	4.171301	4.742630	3.974127
6	0.012368	1.510976	15.75250	65.22389	2.932929	4.060750	6.582913	3.936040
7	0.012562	1.535164	18.03466	63.29424	2.858530	3.945615	6.513474	3.818315
8	0.012656	1.700531	17.78970	63.01737	3.028623	3.887506	6.427406	4.148864
9	0.012744	2.306045	17.72145	62.22695	3.361952	3.890142	6.366268	4.127187
10	0.012752	2.318294	17.70543	62.15683	3.419822	3.887754	6.360034	4.151834

Cholesky Ordering: BUS DCONS DMORT DINF DUN DESI DEXC

Figure 4.7: This figure presents the results of Variance Decomposition regarding the rate of default of business, consumer and mortgage loans.

Pairwise Granger Causality Tests
Date: 02/23/20   Time: 17:06
Sample: 2002Q4 2019Q1
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
DCONS does not Granger Cause BUS	61	3.40206	0.0241
BUS does not Granger Cause DCONS		2.63031	0.0593
DMORT does not Granger Cause BUS	61	2.33465	0.0841
BUS does not Granger Cause DMORT		0.44135	0.7244
DINF does not Granger Cause BUS	62	0.29627	0.8279
BUS does not Granger Cause DINF		0.30820	0.8193
DUN does not Granger Cause BUS	61	4.95452	0.0041
BUS does not Granger Cause DUN		1.70609	0.1766
DESI does not Granger Cause BUS	62	0.58306	0.6286
BUS does not Granger Cause DESI		0.21588	0.8850
DEXC does not Granger Cause BUS	62	0.38298	0.7657
BUS does not Granger Cause DEXC		0.45850	0.7124
DMORT does not Granger Cause DCONS	61	1.52148	0.2194
DCONS does not Granger Cause DMORT		0.97567	0.4111
DINF does not Granger Cause DCONS	61	1.29913	0.2842
DCONS does not Granger Cause DINF		0.59840	0.6188
DUN does not Granger Cause DCONS	61	3.69308	0.0172
DCONS does not Granger Cause DUN		1.35356	0.2668
DESI does not Granger Cause DCONS	61	4.39713	0.0077
DCONS does not Granger Cause DESI		0.18258	0.9078
DEXC does not Granger Cause DCONS	61	2.18010	0.1010
DCONS does not Granger Cause DEXC		0.79351	0.5028
DINF does not Granger Cause DMORT	61	0.68695	0.5639
DMORT does not Granger Cause DINF		0.11026	0.9537
DUN does not Granger Cause DMORT	61	1.10163	0.3565
DMORT does not Granger Cause DUN		0.16065	0.9223
DESI does not Granger Cause DMORT	61	2.15491	0.1040
DMORT does not Granger Cause DESI		0.70661	0.5523
DEXC does not Granger Cause DMORT	61	0.20090	0.8953
DMORT does not Granger Cause DEXC		1.83636	0.1515
DUN does not Granger Cause DINF	61	1.31044	0.2805
DINF does not Granger Cause DUN		1.50006	0.2249
DESI does not Granger Cause DINF	62	2.22857	0.0951
DINF does not Granger Cause DESI		0.53886	0.6577
DEXC does not Granger Cause DINF	62	1.53087	0.2168
DINF does not Granger Cause DEXC		5.20592	0.0031
DESI does not Granger Cause DUN	61	0.86088	0.4671
DUN does not Granger Cause DESI		0.15704	0.9247
DEXC does not Granger Cause DUN	61	0.80828	0.4948
DUN does not Granger Cause DEXC		0.20779	0.8906
DEXC does not Granger Cause DESI	62	3.89328	0.0136
DESI does not Granger Cause DEXC		1.55734	0.2101

Figure 4.8.1: This figure presents pair-wise Granger causality tests between the variables included in our model.

VAR Granger Causality/Block Exogeneity Wald Tests
Date: 02/23/20 Time: 16:12
Sample: 2002Q4 2019Q1
Included observations: 61

Dependent variable: BUS

Excluded	Chi-sq	df	Prob.
DCONS	5.467220	3	0.1406
DMORT	2.227093	3	0.5266
DINF	1.629977	3	0.6526
DUN	9.798904	3	0.0204
DESI	3.714408	3	0.2940
DEXC	0.944728	3	0.8146
All	32.26983	18	0.0204

Dependent variable: DCONS

Excluded	Chi-sq	df	Prob.
BUS	8.110500	3	0.0438
DMORT	3.745955	3	0.2902
DINF	10.71744	3	0.0134
DUN	4.952022	3	0.1753
DESI	11.16183	3	0.0109
DEXC	4.541926	3	0.2086
All	61.84209	18	0.0000

Dependent variable: DMORT

Excluded	Chi-sq	df	Prob.
BUS	1.074498	3	0.7832
DCONS	3.058390	3	0.3827
DINF	2.418432	3	0.4902
DUN	2.226412	3	0.5268
DESI	4.003007	3	0.2611
DEXC	0.340912	3	0.9522
All	16.07434	18	0.5874

Dependent variable: DINF

Excluded	Chi-sq	df	Prob.
BUS	2.929345	3	0.4026
DCONS	7.061349	3	0.0700
DMORT	0.832621	3	0.8416
DUN	2.337188	3	0.5054
DESI	8.121046	3	0.0436
DEXC	3.625902	3	0.3048
All	19.62420	18	0.3544

Dependent variable: DUN

Excluded	Chi-sq	df	Prob.
BUS	3.304204	3	0.3471
DCONS	3.025662	3	0.3877
DMORT	0.339769	3	0.9524
DINF	5.348297	3	0.1480
DESI	1.855993	3	0.6028
DEXC	3.716641	3	0.2937
All	20.05790	18	0.3296

Dependent variable: DESI

Excluded	Chi-sq	df	Prob.
BUS	2.309155	3	0.5108
DCONS	3.804472	3	0.2834
DMORT	2.646348	3	0.4494
DINF	3.513059	3	0.3191
DUN	0.246640	3	0.9697
DEXC	11.17609	3	0.0108
All	18.14852	18	0.4459

Dependent variable: DEXC

Excluded	Chi-sq	df	Prob.
BUS	4.027088	3	0.2586
DCONS	6.992364	3	0.0721
DMORT	5.522460	3	0.1373
DINF	16.24006	3	0.0010
DUN	2.949254	3	0.3995
DESI	1.650994	3	0.6479
All	33.13462	18	0.0161

Figurem4.8.2: This figure presents the VAR Granger Causality / Block Exogeneity tests.

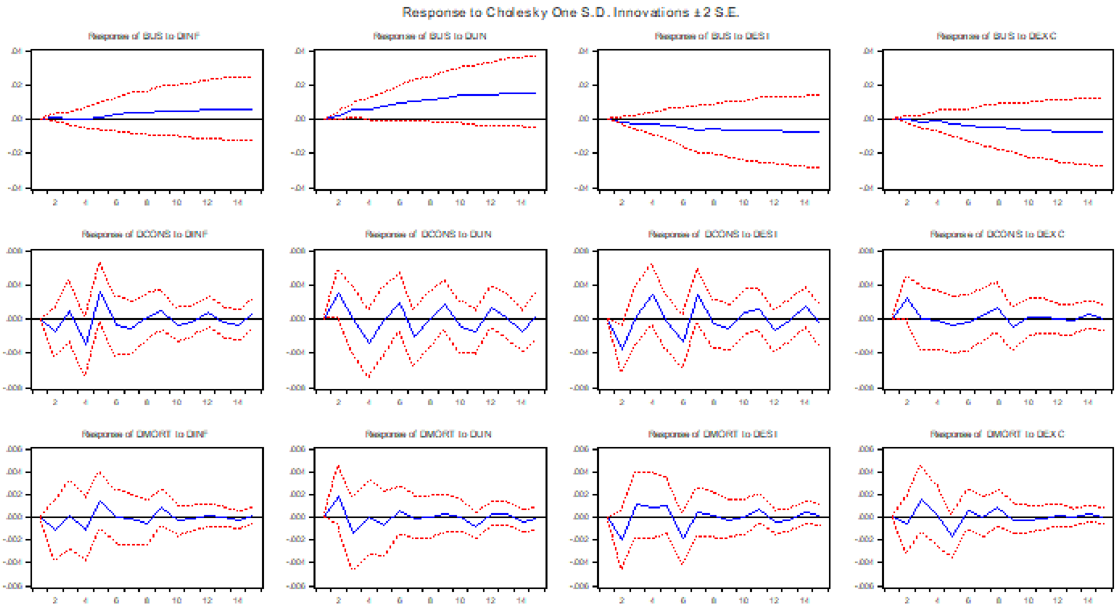

Figure 4.6.5: This figure presents the output graphs from Impulse Response Functions (IRFs). The macroeconomic factors describe the impulses while rates of default stand for responses.

References

Greek

- Σαπουντζόγλου Γ., Πεντότης Χ., (2017), *“Τραπεζική Οικονομική”*, Εκδόσεις Μπένου, Αθήνα
- Δημέλη Σ., (2010), *“Μακροοικονομικά Μεγέθη και Ανάπτυξη της Ελληνικής Οικονομίας”*, Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών, Αθήνα
- Δημέλη Σ., (2013), *“Σύγχρονες Μέθοδοι Ανάλυσης Χρονολογικών Σειρών*”, Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών, Αθήνα
- Τζαβαλής Η., (2008), *“Οικονομετρία”*, Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών, Αθήνα

Foreign

- Charalambakis E., Dendramis Y., Tzavalis E., (2017), *“On the determinants of NPLs: Lessons from Greece”*, Bank of Greece, Athens
- Dendramis Y., Tzavalis E., Adractas G., (2017), *“Credit risk modelling under recessionary and financially distressed conditions”*, Journal of Banking and Finance
- Dendramis Y., Tzavalis E., Varthalitis P., & Athanasiou E., (2019), *“Predicting Default Risk under Asymmetric Binary Link Functions”*
- Anastasiou D., Bragoudakis Z., Malandrakis I., (2019), *“Non-performing loans, governance indicators and systemic liquidity risk: evidence from Greece”*, Bank of Greece, Athens
- Louzis D., Vouldis A., Metaxas V., (2010), *“Macroeconomic and bank-specific determinants of non-performing loans in Greece: a comparative study of mortgage, business and consumer loan portfolios”’*, Bank of Greece, Athens
- Karadima M., Louri H., (2019), *“Non-performing loans in the euro area: does market power matter?”*, Bank of Greece, Athens
- Anastasiou D., Louri H., Tsionas M., (2016), *“Non-performing loans in the euro area: are core-periphery banking markets fragmented?”*, Bank of Greece, Athens
- Georgoutsos D., Migiakis P., (2012), *“Heterogeneity of the determinants of euro-area sovereign bond spreads; what does it tell us about financial stability?”*, Bank of Greece, Athens
- Athanasoglou P., Brissimis S., Delis M., (2006), *“Bank-specific, industry-specific and macroeconomic determinants of bank profitability*”, Journal of International Financial Markets, Institutions & Money, Athens
- Brooks C., (2019), *“Introductory Econometrics for Finance”*, Cambridge university press
- James G., Witten D., Hastie T., Tibshirani R., (2017), *“An introduction to Statistical Learning”*, Springer New York Heidelberg Dordrecht London
- Hull J., (2012), *“Risk Management and Financial Institutions”*, John Wiley & Sons
- Lutkepohl H., (2005), “*New Introduction to Multiple Time Series Analysis”*, Springer Science & Business Media

- Asmah E., (2013), *"Sources of Real Exchange Rate Fluctuations in Ghana"*, American Journal of Economics
- Xindong Zhao, Guoqi Qian, (2014), *"On Multivariate Time Series Model Selection Involving Many Candidate VAR Models"*, European Journal of Pure and Applied Mathematics
- Augustin P., Chernov M., Song D., (2018), *"Sovereign Credit Risk and Exchange Rates: Evidence from CDs Quanto Spreads"*, National Bureau of Economic Research
- Schonbucher P. (2000), *"Factor Models for Portfolio Credit Risk"*, University of Bonn
- Lopez J., Saidenberg M., (2000), *"Evaluating credit risk models"*, Journal of Banking & Finance, USA
- Giannopoulos V., (2018), *"What Causes Non-Performing Loans? The Case of Greece Using Primary Accounting Data"*, Open Journal of Accounting
- Nyathi K., Ndlovu S., Moyo S, Nyathi T., (2014), *"Optimisation of the Linear Probability Model for Credit Risk Management"*, International Journal of Computer and Information Technology
- Hang Bai, (2016), *"Unemployment and Credit Risk"*, School of Business, University of Connecticut